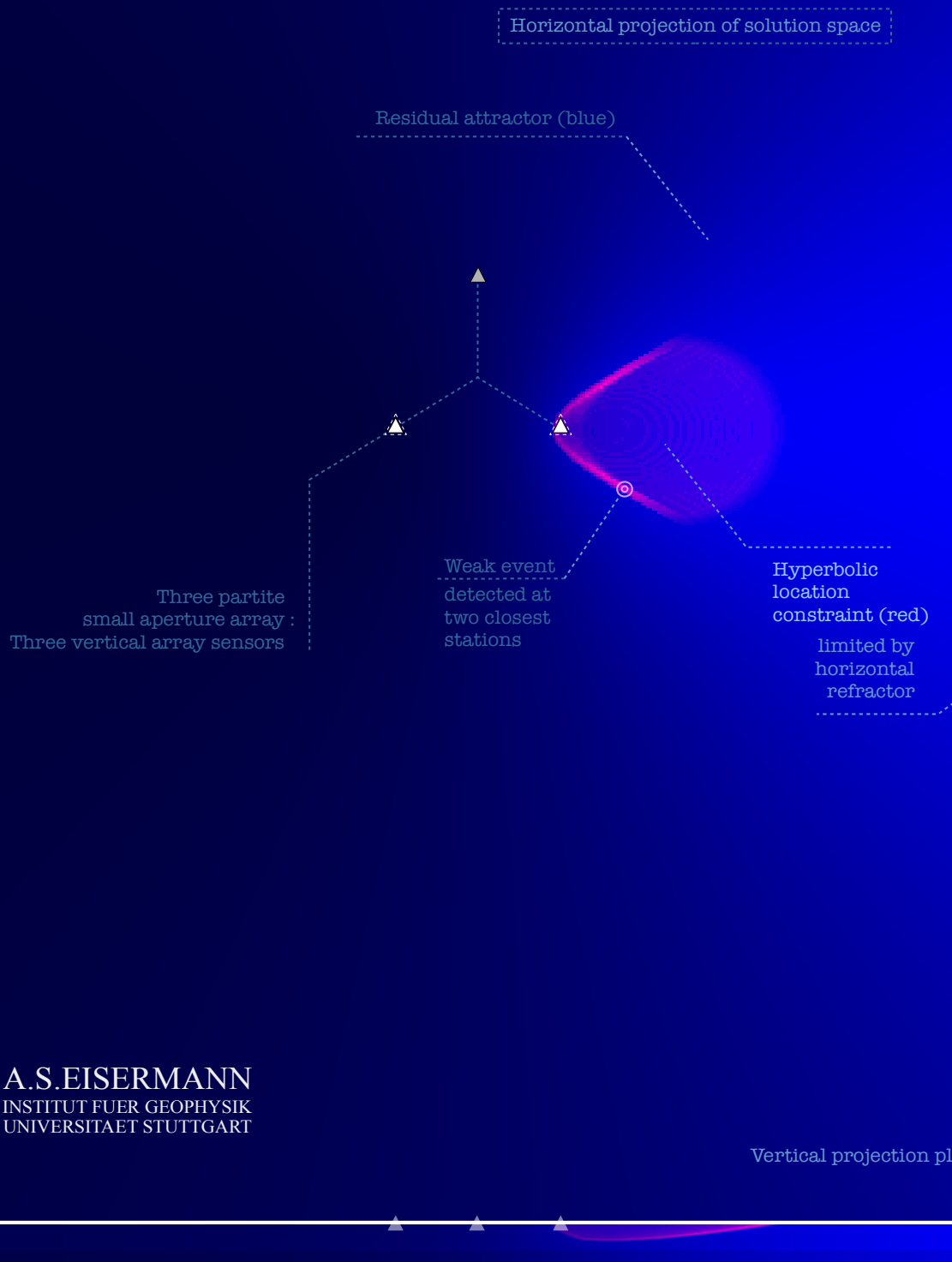
EARTHQUAKE LOCATION *by* DISTINCT CONSTRAINTS *for* SPARSE AND DOUBTFUL DATA



A.S.EISERMANN
INSTITUT FUER GEOPHYSIK
UNIVERSITAET STUTTGART

EARTHQUAKE LOCATION *by* DISTINCT CONSTRAINTS *for* SPARSE AND DOUBTFUL DATA

Von der Fakultät für Bau- und Umweltingenieurwissenschaften der Universität Stuttgart zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung.

Vorgelegt von
Andreas Samuel Eisermann
aus Berlin.

Hauptberichter: Prof. Dr. rer. nat. Manfred Joswig
1.Mitberichter: Prof. Dr. Edi Kissling, ETH Zuerich
2.Mitberichter: Prof. Dr. David Steinberg, Tel Aviv University

Tag der mündlichen Prüfung: 03. Mai 2017

Institut für Geophysik der Universität Stuttgart
2018

הללויה הודו ליהוה
כי-טוב כי לעולמ חסדו
ps. 106:1

To my family.

Table of Contents

Abstract

Earthquakes affected mankind since the days of old, claiming more human casualties than any other nature catastrophe. The determination of the hypocenter location displays one of the key subjects of seismology. Today's standard approaches provide for trustable location estimates - given a large amount of stations with high quality data and proper assumptions about the sub-surface velocity structure. The same approaches fail, however, when this amount of stations and quality of data is not given, e.g. in the mapping of seismically active zones using low magnitude events: Here, only few stations detect the signals that sometimes barely exceed the noise level. Seismic phases appear hence unclear, rendering the information of arrival times doubtful. Another example are real-time location schemes (e.g. in Earthquake Early Warning Systems): Here, events need to be evaluated and located within fractions of seconds without knowledge of the complete waveform, and data available only from the first few stations that already detected.

The objective of this thesis lies in a methodological development that provides for more accurate single event locations in the context of sparse and doubtful data. The less data is available the more the location estimate is determined and affected by the individual datum, its uncertainties and errors. The method of choice must therefore be outlier-resistant (e.g. ignore false picks) and incorporate all parameter uncertainties. When data is few, solutions may further be ambiguous (not due to errors in the input parameters, being exact solutions to a set of even ideal arrival times), meaningly: Multiple, significantly separated location candidates may exist. Also, models are usually only rough and simplified representations of the subsurface structure and will often not explain the observed data well enough. Today's standard approaches often disregard the corresponding uncertainties and, hence, often displace the hypocenter significantly to the true location - outside of the assumed error margins. Mislocations in earthquake early warning or forensic seismology may have far-reaching implications for society and on the political level. This thesis provides therefore a novel location methodology that incorporates the important uncertainties, naturally disregards outliers and thereby leads to robust hypocenter estimates.

The work presented builds on the concept of (graphical) jackknifing, which contrary to most of today's standard approaches doesn't attempt to minimize the error in the over-determined system directly, but decomposes the system first into exactly- or even under-determined subsystems. This results in distinct location constraints that are based on arrival time differences between two- or three phases, only. Each constraint identifies a subset of space as possible hypocenter region. The combination of multiple constraints consecutively constrains the final hypocenter region. Since a single constraint relies on a minimum amount of phase onsets, only, solution discrepancies can be traced back to the individual phase data, which allows the data base (e.g. outliers) to be re-evaluated. The global solution is finally recomposed based on the sub-solutions deemed trustworthy, which provides robust and outlier resistant solutions.

This concept is built on, supporting for three dimensional station layouts, complex velocity models and a volumetric computation, which render this approach suitable for a wide class of modern applications. A real-time methodology that regards uncertainties in phase picks, phase types and model parameters provides for robust and accurate locations when data is uncertain and sparse. New constraints are introduced, which allow to resolve ambiguities and provide for faster hypocenter and magnitude estimates in Earthquake Early Warning. A new direct search scheme is developed that integrates constraint probabilities over grid cells, which ensures the identification of sharp hypocenter regions independent of the grid's resolution, satisfying the demand for a complete search. The improvement in location quality is demonstrated using several examples ranging from gas-field low-magnitude event monitoring, forensic seismology to examples of real-time locations in Earthquake Early Warning.

Zusammenfassung

Erdbeben fordern die meisten menschlichen Opfer unter allen Naturkatastrophen und beschäftigen die Menschheit seit alters her. Die Bestimmung des Hypozentrums ist eine der fundamentalen Themen und Aufgaben der Seismologie. Heutige Standardverfahren bieten eine hohe Ortungsgenauigkeit, insofern Daten hoher Qualität genügend vieler Stationen vorliegen und die Untergrundstruktur hinreichend bekannt ist. Die Verfahren versagen aber, insofern diese Bedingungen nicht erfüllt sind. So z.B. in der Kartographierung seismisch aktiver Zonen unter Nutzung schwacher seismischer Ereignisse: Nur wenige Stationen registrieren deren Signale, die sich unter Umständen nur schwach von dem Untergrundrauschen abheben, und erlauben eine nur unsichere Wahl der Phaseneinsätze. Ein weiteres Beispiel liefert die Echtzeitortung (z.B. in Erdbeben-Frühwarnsystemen): Dort müssen Ereignisse innerhalb von Bruchteilen von Sekunden bewertet und geortet werden, ohne Kenntnis des vollständigen Signals, und basierend auf nur den wenigen Stationen, die das Beben bereits registrierten.

Diese Doktorarbeit befasst sich mit der Entwicklung einer Methode, die akkuratere Ortungen von Einzelereignissen im Kontext weniger und ungenauer Daten bereitstellt. Umso weniger Daten verfügbar sind, umso stärker ist die Lösung abhängig von einzelnen Phaseneinsätzen, ihren Fehlern und Unsicherheiten. Das Verfahren muss daher die gegebenen Unsicherheiten berücksichtigen und resistent gegenüber Ausreißern (z.B. falsch identifizierten Phaseneinsätzen) in den Daten sein. Bei sehr wenigen verfügbaren Stationen können Lösungen des Weiteren ihre Eindeutigkeit verlieren (nicht aufgrund von Fehlern in den Eingabeparametern, vielmehr existieren nun mehrere mathematisch exakte Lösungen selbst bei idealen Phaseneinsätzen). Dies heißt, dass mehrere, klar getrennte Regionen als mögliches Hypozentrum in Frage kommen können. Die verwendeten Untergrundmodelle sind für gewöhnlich nur grobe Annäherungen an die bestehende Untergrundstruktur. Die Anwendung von Standardverfahren, die diese Unsicherheiten oft ignorieren, kann zu Fehlortungen führen, welche die erwarteten Fehlergrenzen überschreiten. Dies ist insbesondere kritisch für Frühwarnsysteme oder die forensische Seismologie, da Fehlortungen dort signifikante Auswirkungen für die Gesellschaft oder auf der politischen Ebene haben können. Diese Arbeit stellt eine neue Methodik vor, welche die gegebenen Unsicherheiten berücksichtigt, inhärent Ausreißer vernachlässigt und dadurch zu robusteren Hypozentren führt.

Die entwickelte Methodik basiert auf dem Konzept des (graphischen) “Jackkifing's”. Entgegen den meisten Standardverfahren wird hier nicht versucht, den Fehler des überbestimmten Systems direkt zu minimieren. Stattdessen wird das System zuerst in exakt- und unterbestimmte Subsysteme zerlegt. Daraus ergeben sich verschiedene Ortsbeschränkungen, die ihrerseits nur auf der Einsatzzeitdifferenz zwischen zwei- oder drei Phasen beruhen. Jede individuelle Ortseinschränkung identifiziert eine Teilmenge des Raumes als möglichen Ursprung des Erdbebens. Durch Kombination mehrerer Einschränkungen wird schließlich das Gebiet eingegrenzt, welches das Hypozentrum in seiner Unsicherheit beschreibt. Da die Ortseinschränkungen jeweils nur auf wenigen Phaseneinsätzen beruhen, können Unstimmigkeiten in der Lösung auf einzelne Phaseneinsätze zurückgeführt werden, und die Datengrundlage (insbesondere Ausreißer) revidiert werden. Die globale Lösung wird schließlich aus den bewerteten und korrigierten Sublösungen rekombiniert, wodurch stabile Ausreißer-resistente Lösungen erreicht werden können.

Auf diesem Konzept beruhend, wurde die Methode um drei-dimensionale Stationsverteilungen, komplexe Geschwindigkeitsmodelle und eine räumliche Berechnung erweitert, welche die Methode für viele moderne Anwendungen öffnet. Echzeitfähige Algorithmen, die Unsicherheiten in Phaseneinsatzwahl, Phasenklassifizierung und Modelparametern berücksichtigen, erlauben robuste und akkurate Ortungen, selbst wenn die Datengrundlage karg und unsicher ist. Neue Ortseinschränkungen werden vorgestellt, die Mehrdeutigkeiten in der Lösung auflösen und eine schnellere Bestimmung des Hypozentrums und der Erdbebenmagnitude in Frühwarnsystemen erlauben. Eine neue Rastersuche, welche die Wahrscheinlichkeit der Erfüllung der Einschränkungsbedingungen über Gitterzellen integriert, sichert die Identifizierung kompaktester Minima unabhängig von der Rasterauflösung und erfüllt somit die Bedingung nach Vollständigkeit der Lösung. Die verbesserte Ortungsqualität wird an Hand mehrerer Beispiele aus den Bereichen der Gasfeld-Überwachung, der forensischen Seismologie und der Echtzeitortungen in Frühwarnsystemen aufgezeigt.

I **Introduction**

I.1 **Overview**

The phenomenon “Earthquake” equally affected and fascinated mankind since the days of old. The earthquake location problem, which can be described as the search for the source location from which all released energy radiates, displays one of the early and key subjects of seismology. Today, this field does not only pertain to classical earthquake location, but is applied for various kinds of source processes, usually several magnitudes of energy weaker: Minute fracture processes, explosions and implosions, impact events, structure collapses, events that often would only be noted as noise bursts. In most of today’s seismological applications enough seismic stations exist to deliver a sufficient amount of seismogram and arrival time data to satisfy the commonly used models with redundant data. Yet, in other cases the amount of stations and redundancy and quality of data is not given. One such example is the analysis of M_L -3 shear-quakes in soft shale landslides using the methodology of Nanoseismic Monitoring (Joswig, 2008) to identify destabilization processes. Here the event is often only detected by very few stations and signals close to noise level often impede a clear identification of phase onsets. In the field of forensic seismology, searching for M_L -2 aftershocks in “Comprehensive Test Ban Treaty On-Site Inspections” (CTBT-OSI), only few stations actually see the low SNR (Signal-to-Noise-Ratio) signals. In the real time location framework of Earthquake Early Warning (EEW), the event needs to be located and evaluated within fractions of seconds without knowledge of the full signal and detections only on a minimum amount of stations. In other cases, only few stations can be deployed as in the exotic case of the Rosetta space mission's Acoustic Surface Sounding Experiment (CASSE) on the Philae lander (which just landed on comet 67P/Churyumov-Gerasimenko), deriving information from the three sensors in its legs.

The less data is available the more the location estimate depends on the individual datum and the more it is affected by its uncertainties and errors. Mislocations in applications of early warning or forensic seismology may have far-reaching implications for society and on the political level. The location analysis therefore requires robust methods that allow to resolve the correlation between individual data and global solution and reveal all physically possible solutions.

The work presented in this thesis builds on the concept of graphical jackknifing (Joswig, 2006), which complementary to the widely used purely numerical, iterative localization methods offers insight into the location process in a way that is crucial for the analysis of sparse and doubtful data.

Contrasting most of today's standard approaches, which attempt to directly minimize the error in over-determined systems (e.g. sum of L2-normed residuals), here the (over-determined) system is first decomposed into exactly- and even under-determined subsystems. This results in distinct location constraints that are based on arrival time differences between arriving phases. These are individually evaluated, which provides the possibility to trace back discrepancies in the solutions to individual phase data, allowing to re-evaluate and possibly correct the datum or underlying information base. The global solution is finally recomposed based on the solutions of the trustworthy subsystems, which allows to obtain stable and outlier resistant solutions.

Even when principal phases (e.g Pg- or S-phases) can be identified and picked sufficiently accurate - an even larger impact on the source location is exerted by the choice of the velocity model. Often, the velocity structure is not well known and only roughly approximated through 1D models. Given only few data, the effect of mis-chosen models can be significant.

The work presented in this thesis therefore focussed on two key points. As the event interpretation relies on a high location quality, one aspect focusses on ways to increase the accuracy of single event locations through methodological advancements, to support complex velocity models and three-dimensional station distributions in the context of distinct constraints. As this yields complex constraint behavior, new approaches will be introduced to gain insight to the solution space. The second aspect relates to the proper treatment of the corresponding location uncertainties which may be significant in the context of sparse and doubtful data, to identify all mathematically possible locations and those most probable. Husen and Hardebeck (2010) point out that the uncertainties in earthquake location are dominated by three factors, which are 1) measurement errors in the arrival times, 2) modelling errors in the computed travel times and 3) the general non-linearity of the location problem. If implemented in location codes, measurement errors are usually assumed to be Gauss distributed, yielding Gaussian error distributions for the location. As this prerequisite is not given for doubtful data with unclean phases in noisy condition nor for the scenario of few stations, a general probabilistic form will be derived (A large amount of stations with small errors, on the other hand, will always yield Gauss-distributed probabilities for the location error, independent of the form of the probability distribution of the parameter errors (Central-Limit-Theorem)). Model errors -if at all implemented- are based on simplified functions, usually also Gauss-distributed and treated in a combined manner with measurement errors (Lomax, 2009), but not representing the true uncertainties derived from the velocity model or travel time curves. Pavlis (1986) shows that this leads to biased solutions and beside Lomax (2011), Husen and Hardebeck (2010) conclude that “all

location methods suffer from an inappropriate handling of velocity model errors". Yet, the proper treatment of these uncertainties is crucial for the analysis of sparse data. A novel approach will be presented that allows to account for these uncertainties within the framework of distinct constraints. Additional focus will be laid on extending the concept of distinct constraints to the estimates of small-aperture arrays, which are an important component in the data collection, offering important location constraints, in particular back-azimuth information. Chapter III.1 will evaluate the precision in location that the various constraints can achieve, but it will further shed light on the important aspect of location ambiguities – ambiguities that are not caused by errors in the input parameters (which would lead to a scatter of minima in e.g. L1-residua maps) but are exact solutions to a set of (*even ideal)* arrival times of any given phase, e.g. the P-phase arrival recorded at several stations. Since these ambiguities appear only for special geometrical configurations when many stations are given, they are hardly observed - but with less than five stations, these ambiguities arise commonly and may cause a significant misinterpretation of the source. The method of distinct constraints is formulated to properly identify these ambiguities and provide constraints that allow the identification of the true solution, to yield accurate hypocenter locations. This is of fundamental importance for earthquake early warning systems. Due to a real time capable implementation, distinct constraints is well suited for the application of real time locations in earthquake early warning, which chapter IV is dedicated to.

The circumstance of few available data, weak or non-standard events defines the context of the location problem laid out in this thesis. One such example is shown in the following figure (I.1:1), displaying surface waves on seismic traces recorded at several stations at the US East Coast.

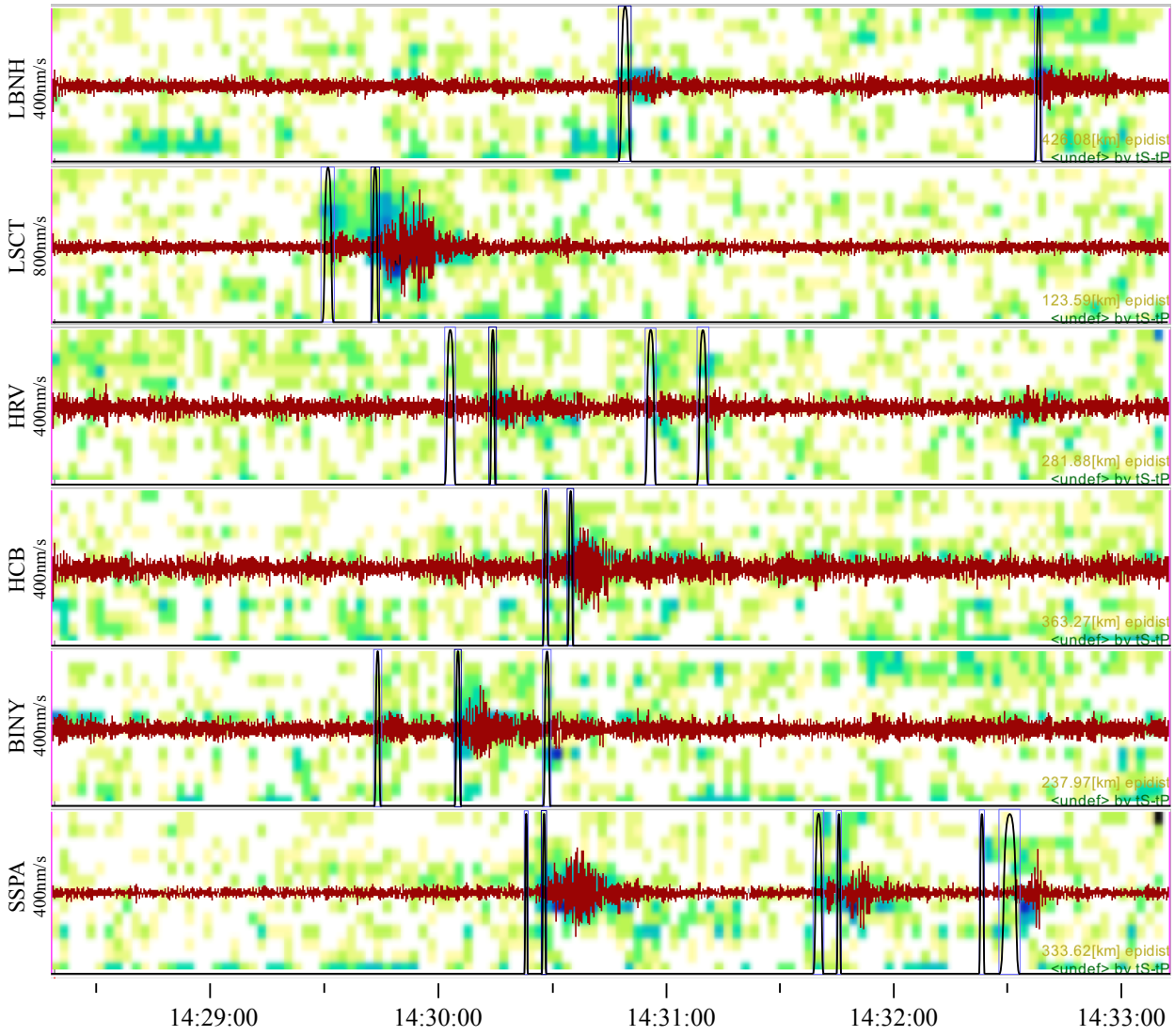


Fig. I.1:1: Registration of low SNR surface wave signals on the vertical BHZ component of six broad band stations. The blue curve marks all possible phase onsets. The unfiltered seismograms (red) are underlaid by sonograms (i.e. noise-adapted spectrograms, displaying the signal energies per frequency band over time) of fig.I.1:2. While the seismogram at HRV doesn't allow a correlation with the waveform of the closest station LSCT (and impedes an automatic detection or phase identification), the sonogram (fig.I.1:2) clearly shows the same signature. Station SSPA in far distance shows several signatures similar to LSCT, two of them are due to local events.

The signals in this seismic record originate from the catastrophic collapse of the Twin Towers, on Sept. 11th 2001. Solely based on this record it would be difficult to locate their origin. Several vague phase onset candidates render the association to one single event difficult. Weak signatures at distant stations (e.g. HRV, $SNR \approx 1$) can only be recognized using a noise- adapted spectral analysis (Sonograms) (Joswig, 2008) (fig. I.1:2). Arrival time picking uncertainties are therefore large, as picking is not based on a phase identification in raw seismograms but by similarity in the spectral signature.

The existence of additional spectral energy signatures, possibly originating from local sources, leads to an ambiguity in picking that impedes the use of standard techniques as Root-Mean-Square (RMS) residua sum based approaches.

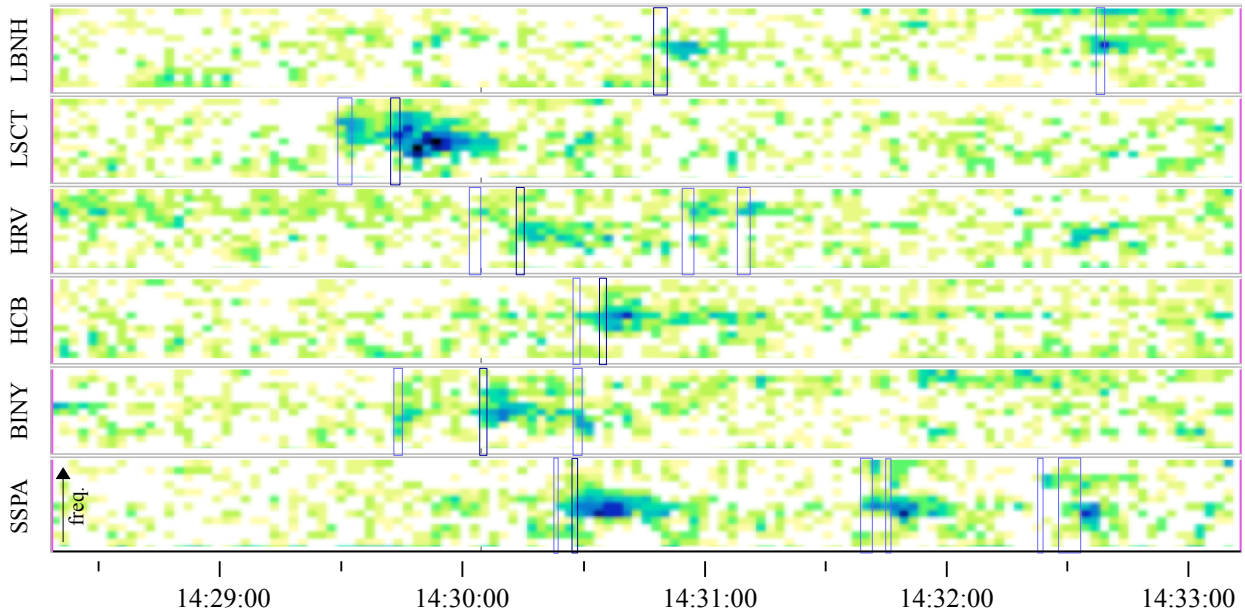


Fig. I.1:2: Sonogram traces of fig. I.1:1. The strong signature of the closest station LSCT can still be identified on HRV. On SSPA several signatures could match the one of LSCT. The blue boxes mark all possible signature onsets. The color scale spans from white (no energy above noise level in given frequency band) over green and blue to black (indicating significant energies above noise level in the respective frequency band).

The following chapters present the development of a novel location methodology that provides for accurate locations when uncertainties are large and information is few: Ranging from weak and ambiguous signals (as in fig. I.1:2) in microseismic monitoring, to large-magnitude events that need to be located within fractions of seconds on the basis of a few stations in earthquake early warning. The approach presented here further develops solutions found in the most advanced location techniques available today and blends them with key components that are, however, found much earlier, in the location schemes being introduced when modern seismology was just born.

I.2 Earthquake Location: “State of the Art” and “Back to the Roots”

Given a sufficient amount of good data, various advanced location methods allow to obtain detailed information on the source location, its extent, as even the whole space-time history of the rupture process. Given many stations and multiple events, joint hypocenter relocation schemes as Master Event (Deichmann and Garcia-Fernandez, 1992) or Double Difference (Waldhausen, 2001) may be applied. This greatly improves relative locations, often resolving sharp structural information in the form of tight event clusters. It, however, does not provide more accurate absolute locations than obtained by single-event location techniques (Menke and Schaff, 2004) except in cases where ground truth information is available. Given a sufficiently large amount of stations (>100), not only the hypocenter but the full space-time history of a single rupture process can be determined by back propagation, back projection and Time Reversal techniques (Larmat, 2006). The signals being recorded at transducer-arrays are here time-reversed and retransmitted through the model-medium. The wave travels back and refocusses at the source location due to the time invariance of the wave equation. Passive Seismic (Duncan, 2005) allows to locate events with signals below noise level using automated beam steered waveform stacking. Here the emitted energy is calculated at every location of interest by stacking waveforms, each being time shifted according to the travel time from trial location to station before stacking. However, also this method requires data of between 40-100 stations.

When this amount of data is not available, single-event location falls back to feature (seismic phase) based approaches. The focus of this work lies on the evaluation of local to regional events using a phase based analysis, incorporating the information of direct and refracted P and S waves and small-aperture-array beam forming within a jackknifing concept. For phase based single event location predominately two inversion classes exist today (Husen and Hardebeck, 2010), both attempting the minimization of residual sums. The first is based on iterative searches, in which the forward problem is linearized and the solution found by step wise improvement of a initial estimate. Matrices of partial derivatives representing the local gradient of the penalty function at given test vectors in solution space are inverted to obtain improved estimates. In the second class, direct, global searches are performed. The solution space is sampled using regular, irregular or nested grids which allows a mapping of the solution quality and finds the global and all local minima, given that they lie in the sampled interval. Linearized inversions only sample the solution space at few test locations and are therefore incomplete. They depend on an initial test estimate which may lead to

convergence into local minima. As they rely on partial derivatives they may become unstable in heterogeneous environments with first order discontinuities and may misrepresent the true solution for poorly-constrained problems. Outliers in the data strongly affect the solution (Husen et al., 2003). Yet, linearizing iterative schemes are very efficient and of low computational cost, making them generally the method of choice. Software packages like HYPO71 (Lee et al., 1975), HYPOINVERSE (Klein, 1978), HYPOCENTER (Lienert et al., 1986), HYPOELLIPSE (Lahr, 1989) and HYPOSAT (Schweitzer, 1997) are well known representatives of this class. In the following the underlying iterative scheme and its error measures shall be outlined.

I.2.1 **Linearized Inversion Schemes**

The fundamental concept of many iterative location procedures used today has its roots in the scheme developed by Geiger (1910), which initially inverted in two spatial dimensions as deep hypocenters were ruled out at the beginning of the 20th century. The objects of inversion are hypocenter location $(x, y, z)^t$ and origin time t_0. The two are combined in the model vector

$$\boldsymbol{m}=(x, y, z, t_0) \tag{I.2.1-1}$$

which starting from an initial guess location is improved stepwise until convergence is met. The data given are the observed arrival times t_i^O picked from a seismogram recorded at location $\boldsymbol{x_i}$ of station i

$$\boldsymbol{x_i}=(x_i, y_i, z_i)\,. \tag{I.2.1-2}$$

Arrival times can be expressed as the sum of origin time t_0 and the wave's travel time Tt between source $\boldsymbol{x}$ and i-th station $\boldsymbol{x}_i$

$$t_i^O=t_0+Tt(\boldsymbol{x}, \boldsymbol{x_i})\quad. \tag{I.2.1-3}$$

Given a model representing the existing velocity structure, a proposed hypocenter and origin time, the synthetic arrival time t_i^M can be predicted by a forward calculation of the travel time

$$t_i^M = F_i(\boldsymbol{m}) \,. \tag{I.2.1-4}$$

Here, F_i is the operator projecting the four dimensional space-time model vector $\boldsymbol{m}$ into the temporal one dimensional solution space. F describes all calculations of wave propagation between source and i-th station through the velocity model. These calculations are performed either semi-analytically, but usually numerically using ray-tracers or wavefront solvers, or practically by interpolating tables of recorded travel-times. Eq. (I.2.1-4) can be written vectorially as

$$\boldsymbol{t}^M = F(\boldsymbol{m}) \quad . \tag{I.2.1-5}$$

The residual $\boldsymbol{r}_i$ defines the discrepancy between forward calculated (modeled) and observed onset times

$$\boldsymbol{r}_i = t_i^O - t_i^M \,. \tag{I.2.1-6}$$

Now, a model vector $\boldsymbol{m}^*$ is sought which minimizes the sum over all (absolute) values of residuals (L1-norm). For this vector, calculated (predicted) and observed onset times become similar

$$R^{(L1)} = \sum_i |r_i| = \sum_i \left| t_i^O - t_i^M \right| \stackrel{!}{=} min \quad . \tag{I.2.1-7}$$

This L1-norm is a robust measure for the discrepancy between predicted (modeled) and observed data, as it is insensitive to outliers. However, it is not differentiable at the point of vanishing residuals which complicates the numerical minimization. The commonly used L2-norm

$$R^{(L2)} = \sum_i r_i^2 = \sum_i \left(t_i^O - t_i^M \right)^2 \stackrel{!}{=} min \,, \tag{I.2.1-8}$$

on the other hand is once continuously differentiable and assigns comparably higher penalties to larger residuals. While this is reasonable for data with smaller measurement errors (as it is more

important to bring larger discrepancies into agreement, than small ones for which the requirements are nearly satisfied), the least-squares norm causes misguiding outliers to dominate the solution. To dampen these effects, mixed L1/L2-norms may be applied, e.g. the Huber-norm (Guitton and Symes, 2003) uses a L2 norm for smaller residuals, while using an L1 norm when they grow larger. To compensate the impact of large residuals, weights are often attributed in the root-mean-squared residuals (RMS) minimization (eq. I.2.1-9).

$$RMS = \frac{\sum_{1}^{n} w_i r_i^2}{\sum_{1}^{n} w_i} \tag{I.2.1-9}$$

Residuals may be weighed directly based on the quality of the corresponding phase pick. The weight w_i of station i is also generally set in relation to the distance as closer sources tend to have clearer phase onsets and uncertainties in the local velocity model show less effect for shorter distances. A standard form of distance weighting is given in (Havskov et al., 2011)

$$w_i^d \propto \begin{cases} x_{far}, & x_{far} < \Delta_i \\ \dfrac{x_{far} - \Delta_i}{x_{far} - x_{near}}, & x_{near} < \Delta_i < x_{far} \\ x_{near}, & \Delta_i < x_{near} \end{cases}, \tag{I.2.1-10}$$

with Δ_i being the epicentral distance, x_{near} the distance yielding full weight and x_{far} the distance corresponding to reduced weight. x_{near} is suggested to be in the order of magnitude of network diameter and x_{far} being about twice x_{near}. For more than six available phases, HYPOELLIPSE (Lahr, 1999) makes use of truncation or boxcar weighting which reduces the weight, and fully neglects residuals larger then a multiple of the standard deviation, respectively.

The inversion scheme solves the inverse problem iteratively. Starting with an initial guess value $\boldsymbol{m}^0$, the hypocenter is approached by searching for a better $\boldsymbol{m}$ in its vicinity.

$$\boldsymbol{m} = \boldsymbol{m}^0 + \Delta \boldsymbol{m} \tag{I.2.1-11}$$

$\Delta\mathbf{m}$ is assumed to be small, so that the relationship between modeled arrival time t^M and $\mathbf{m}$ may be linearized neglecting higher order Taylor-terms

$$t_i^M(\boldsymbol{m}) = t_i^M(\boldsymbol{m}^0) + \sum_j \left.\frac{\partial t_i^M}{\partial m_j}\right|_{m^0} \Delta m_j + \ldots \quad . \tag{I.2.1-12}$$

A "better" model vector $\boldsymbol{m}$ is characterized by lower residuals $r(\boldsymbol{m})$ compared to $r(\boldsymbol{m}^0)$.

$$\begin{aligned} r_i(\boldsymbol{m}) &= t^O - t_i^M(\boldsymbol{m}) \\ &= \underbrace{t^O - t_i^M(\boldsymbol{m}^0)}_{r_i(\boldsymbol{m}^0)} + \sum_j \left.\frac{\partial t_i^M}{\partial m_j}\right|_{m^0} \Delta m_j \end{aligned} \tag{I.2.1-13}$$

$r(\boldsymbol{m})$ can be minimized by requiring

$$r_i(\boldsymbol{m}^0) = \sum_j \left.\frac{\partial t_i^M}{\partial m_j}\right|_{m^0} \Delta m_j \, . \tag{I.2.1-14}$$

This, in vectorial notation, takes the form of

$$\boldsymbol{r}(\boldsymbol{m}^0) = \underline{\underline{G}}\, \Delta \boldsymbol{m} \tag{I.2.1-15}$$

with $\underline{\underline{G}}$ being the Jacobian matrix of partial derivatives

$$G_{ij} = \frac{\partial t_i^M}{\partial m_j} \, . \tag{I.2.1-16}$$

As $\boldsymbol{m}$ is four dimensional, $\underline{\underline{G}}$ becomes a square matrix in case of four given onsets. In this case,

$\Delta \boldsymbol{m}$ may be obtained by direct multiplication of $\boldsymbol{r}$ with the inverse of $\underline{\underline{G}}$, $\underline{\underline{G}}^{-1}$. Using eq. (I.2.1-11) an improved model vector is given. The process is repeated until $\boldsymbol{m}$ converges. If more than four stations are given, the dimensionality of $\underline{\underline{G}}$ differs from the one of $\boldsymbol{m}$. In this case $\Delta \boldsymbol{m}$ can not be obtained directly, since no inverse $\underline{\underline{G}}^{-1}$ exists ($\underline{\underline{G}}^{-1}\underline{\underline{G}} \stackrel{!}{=} 1$). The overdetermined equation system is solved by minimizing a global misfit function χ^2. Every onset t_i^O is bound to errors, expressed by the standard deviation σ_i by which it is weighted in χ^2. In analogy to eq. (I.2.1-7) it is written as

$$\chi^2 = \sum_i \frac{1}{\sigma_i^2}\left(r_i - \sum_j G_{ij}\Delta m_j\right)^2 \quad . \tag{I.2.1-17}$$

Requiring the partial derivative $\partial\chi^2/\partial\Delta m_k$ to vanish leads to a system of linear equations that yields the set of $\Delta \boldsymbol{m_i}$ for which χ^2 is minimized. Since all $\boldsymbol{m}_i$ are independent ($\partial\Delta m_j/\partial\Delta m_k = 0$) eq. (I.2.1-17) can be reformulated to

$$\frac{\partial\chi^2}{\partial\Delta m_k} = 0 = 2\sum_i \frac{1}{\sigma_i^2}\left(r_i - \sum_j G_{ij}\Delta m_j\right)G_{ik} \quad . \tag{I.2.1-18}$$

Splitting up the summation terms

$$\sum_i \frac{1}{\sigma_i^2} r_i G_{ik} = \sum_i \frac{1}{\sigma_i^2}\left(\sum_j G_{ij}\Delta m_j\right)G_{ik} \, , \tag{I.2.1-19}$$

and assuming similar standard deviations $\sigma_i \approx \sigma \, , \forall i$ yields

$$\stackrel{\sigma_i=\sigma}{\rightarrow} \sum_i r_i G_{ik} = \sum_i \left(\sum_j G_{ij}\Delta m_j\right)G_{ik} \quad , \tag{I.2.1-20}$$

or

$$\underline{\underline{G}}^T \boldsymbol{r} = \underline{\underline{G}}^T \underline{\underline{G}}\,\Delta \boldsymbol{m} \tag{I.2.1-21}$$

in vector notation. It can be solved for $\Delta \boldsymbol{m}$:

$$\Delta \boldsymbol{m} = \left(\underline{\underline{G}}^T \underline{\underline{G}}\right)^{-1} \underline{\underline{G}}^T \boldsymbol{r} = \underline{\underline{G}}^{-g} \boldsymbol{r} \,. \qquad \text{(I.2.1-22)}$$

Due to its similarity with eq. (I.2.1-15), $\underline{\underline{G}}^{-g}$ is also called the generalized Inverse of $\underline{\underline{G}}$, allowing to solve for $\Delta \boldsymbol{m}$ following a least squares formalism. $\underline{\underline{G}}^{-g}$ is equivalent to $\underline{\underline{G}}^{-1}$ for dim(**r**)=dim(**m**)=4. Successive application of eq.(I.2.1-11) and (I.2.1-22) leads to an steadily improved $\boldsymbol{m}^*$. The corresponding onset times being the solution of the forward problem, are calculated using eq. (I.2.1-5). A good choice of the start vector $\boldsymbol{m}^0$ is important as a $\boldsymbol{m}^0$ distant to the true hypocenter may lead to convergence in local minima instead of the global one. An initial start vector may be obtained using an origin time extracted from Wadati-diagrams and an location obtained from hyperbolic location (chapter I.2.4). For a stable iteration the operator F has to be continuously differentiable. An inhomogeneous medium e.g. containing layer boundaries is therefore problematic and may lead to convergence in local minima, e.g. in vicinity of the layer boundary.

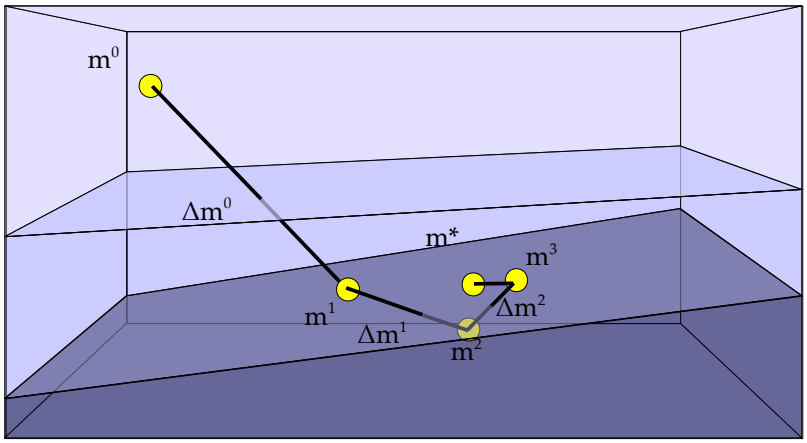


Fig. I.2.1:1: Iterative inversion in a 3-layer medium with dipping layer boundaries, showing the spatial evolution of the model vector. The solution m is found by step wise improvement, starting from the estimate m^0. Not shown is the evolution of the origin time, which -in well behaved models- converges with the spatial solution.*

The quality of the solution is indicated by the behavior of the RMS in the vicinity of $\boldsymbol{m}^*$, which could be directly observed in a global, direct search approach: A RMS, growing quickly with distance from the minimum, corresponds to a well-constrained and well-localized solution. A slowly growing RMS indicates that small deviations in the onset times would lead to large spatial

dislocations of the estimated hypocenter: The true solution could consequently be found in a larger area around $\boldsymbol{m}^*$. This can be quantified through the variance-covariance error ellipsoid, which extrapolates the information given in the variance-covariation found at the point of $\boldsymbol{m}^*$. The variance-covariance matrix $\boldsymbol{\sigma}_X^2$ is defined as

$$\boldsymbol{\sigma}_X^2 = \left(\underline{\underline{G}}^T \underline{\underline{\sigma}}^{-1} \underline{\underline{G}}\right)^{-1} \tag{I.2.1-23}$$

with $\underline{\underline{\sigma}}$ being the matrix of onset covariances while the second term expresses the covariance of the partial derivatives laid out in eq. (I.2.1-16). The diagonal elements of $\boldsymbol{\sigma}_X^2$: σ_{xx}^2, σ_{yy}^2, σ_{zz}^2 and σ_{tt}^2 are the variances of hypocenter location (x,y,z) and origin time (t^0), while off-diagonal elements display the relationship e.g. trade-offs between the four model quantities (Bormann, 2012). A singular value decomposition (SVD) or principle component analysis (PCA) of the submatrix containing the spatial components yields the ellipsoid's principal axes. The resulting eigenvectors and -values orient the ellipsoid in space. With $\boldsymbol{\sigma}_X^2$ being a variance matrix, the square root of diagonal elements corresponds to the standard deviation of the hypocentral parameter. This 1 σ ellipse corresponds to a 68% confidence interval of the hypocenter falling into this volume (Chew, 1966). Scaling the variance-covariance matrix to correspond to two or three standard deviations enlarges the ellipsoid to express a 95% resp. 99.7% confidence region. Yet, as the ellipsoid is extrapolated from the single point $\boldsymbol{m}^*$ this error measure looses its applicability in the presence of inhomogeneities, in which case the hypocenter might be located well outside the confidence region. Pavlis (1986) showed for a simple example that due to the disregarding of model uncertainties none of the evaluated events true hypocenters was located inside the corresponding 95% confidence ellipsoids. Also, when the system is ill-conditioned, e.g. poorly constrained solutions due to few unfavorably places stations, the ellipsoid is an inadequate representation of uncertainties (Husen and Hardebeck, 2010). Nonetheless, the main axes of the confidence-ellipsoid are used as the standard error measure in many of todays earthquake catalogs.

A mayor drawback of this inversion methodology lies in the fact, that the solution space remains largely unexplored, except for the few points that are evaluated during the iteration or in the final analysis around the found minimum. In cases with only a few data and expected misidentifications of phase onsets, several minima will exist and need to be evaluated. It is important to reveal these

model vectors with similar RMS (competing side minima), as their scatter in solution space indicates the true solution uncertainty. The identification of an outlier datum is difficult using linearizing RMS based approaches in which the effect of individual parameter changes are only observable in an integral manner. Due to the underlying least-squares formalism data outliers have a large impact on the solution and small deviations in observed onset times may result in large hypocenter dislocations (Lin&Sanford, 2001). The application of the more robust L1 norm leads to instabilities in the iteration process, due to misleading gradient estimates near the minima. Linearizing schemes generally face the danger to iterate into local minima and may get "trapped" at velocity contrasts such as layer boundaries. Many of these disadvantages can be resolved using the computationally more expensive approach of direct, global searches often referred to as "Grid searches", which do not rely on gradient computations.

I.2.2 **Direct Global Search Schemes**

To obtain "complete" solutions, i.e. identify all local and the global minima, a "direct global search" attempts to evaluate all possible model vectors contained in a certain interval, mapping the solution space rather than sparsely sampling it. The problem is inverted by testing the whole solution space via forward calculations for the model vector that explains the observed data best. For each test model vector the discrepancy between observed and calculated arrival times is computed to find the one yielding the lowest discrepancy. As the whole model space is sampled, this methodology provides overview over the solution space, identifying the global as well as local minima. As no partial derivatives are required for a direct search, other discrepancy measures e.g. the L1-norm may be used, enhancing the outlier resistance of this scheme, thereby increasing the robustness of the solution. For the same reason, heterogeneous velocity models with high velocity contrasts may directly be implemented and analyzed. The term "global" denotes a completeness in covering the model space, which is the harder aspect to realize, as in fact also this method only *samples* the solution space, although orders of magnitude higher resolved than the linearizing inversion. With a four dimensional model space in the simplest case it is sampled into a four dimensional regular grid. The maximum possible resolution is limited by the available computational power and memory. As for too low resolutions the true vector might not be found, adaptively refined nested grids often find application e.g the oct-tree algorithm implemented in NonLinLoc which is about 100 times faster

than a standard grid search that is of higher resolution, initially (Lomax, 2009). NonLinLoc represents one of the commonly used grid search approaches and will serve as example for a direct search approach. Here, a initially coarse grid is recursively refined. At the center of each cell, misfit and probability measures are computed. The grid cell with the largest probability to contain the source is subdivided into 8 cells, for which misfit and probabilities are again calculated. The process is repeated until termination criteria are met.

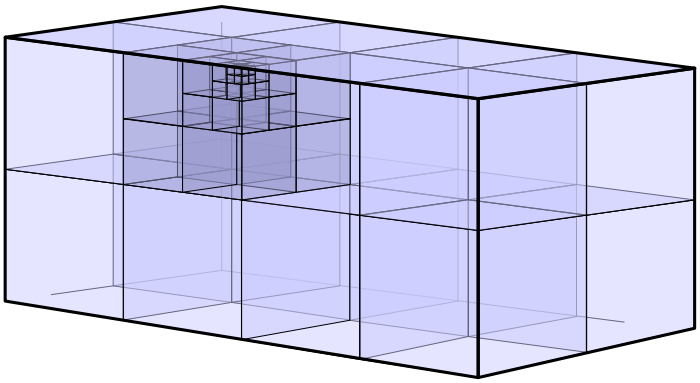


Fig. I.2.2:1: Recursively nested grid. Oct-tree grid.

NonLinLoc follows the probabilistic inversion approach of Tarantola and Valette (1982) to derive a posterior probability density function (PDF), identifying the optimal source location as the point of maximum likelihood. As the oct-tree's recursive subdivision follows the maximum of the PDF, the relative probability P_i that a source is located in any given cell i is given by

$$P_i = V_i L(\boldsymbol{x}_i)\,, \tag{I.2.2-1}$$

with V_i being the cell volume and $\boldsymbol{x}_i$ denoting the coordinate vector at the cell center. $L(\boldsymbol{x})$ Is the likelihood function pertaining to the spatial part of the model vector $\boldsymbol{m}$. The general likelihood $L(\boldsymbol{m})$ is derived from a marginal posterior PDF under the assumption of Gauss distributed observed data and negligible uncertainties in the forward problem $F(\boldsymbol{m})$ to

$$L(\boldsymbol{m}) = \exp\left\{-\frac{1}{2}\left[\boldsymbol{d_0} - F(\boldsymbol{m})\right]^T \boldsymbol{C_d}^{-1}\left[\boldsymbol{d_0} - F(\boldsymbol{m})\right]\right\} \tag{I.2.2-2}$$

in analogy to eq. (I.2.1-5). Here, $\boldsymbol{d}_0$ and $\boldsymbol{C}_d$ describe the mean and covariance matrix of the gauss-distributed observed data. Moser et al. (1992) showed that for these assumptions a maximum likelihood origin time for a hypocenter at location $\boldsymbol{x}$ can be obtained analytically as

$$t_0^{ml}(\boldsymbol{x}) = \frac{\sum_i \sum_j w_{ij} \cdot [t_i - h_i(\boldsymbol{x})]}{\sum_i \sum_j w_{ij}} \tag{I.2.2-3}$$

with w_{ij} being weight components standing in relation to the covariance matrices of computed travel- and observed arrival times, both assumed to have Gaussian uncertainties. t_i and $h_i(\boldsymbol{x})$ denote the mean corrected arrival- and modeled travel times, respectively. Using this origin time estimate, the problem's dimensionality can be reduced to a purely spatial three-dimensional analysis. The least squares probability function $L_{LS}(\boldsymbol{x})$ can now be derived from eq. (I.2.2-2) to

$$L_{LS}(\boldsymbol{x}) = \exp\left\{ -1/2 \sum_i \frac{\left[t_i^O - t_i^M(\boldsymbol{x}) \right]^2}{\sigma_i^2} \right\} \tag{I.2.2-4}$$

with t_i^O being the observed arrival time and t_i^M being the modeled arrival time obtained from modeled travel times Tt^M minus the maximum likelihood origin time $t_0^{ml}(\boldsymbol{x})$. σ_i^2 expresses the combined standard deviation for arrival and modeled travel times. Using the gaussian assumption, the search for the maximum likelihood location is equivalent to a least squares problem (Tarantola and Valette, 1982) and hence sensitive to outliers. An alternative measure was therefore derived from the Equal Differential Time (EDT) approach (Milne, 1886), which relies on the arrival time difference of the same phase at station pairs (Its details will be described in chapter II.1.2.2). In Tarantola and Valette's formulation the likelihood $L_{EDT}(\boldsymbol{x})$ is given by

$$L_{EDT}(\boldsymbol{x}) = \left[\sum_{a,b} \frac{1}{\sqrt{\sigma_a^2 + \sigma_b^2}} \cdot \exp\left(- \frac{\left\{ \left[t_a^O - t_b^O \right] - \left[Tt_a^M(\boldsymbol{x}) - Tt_b^M(\boldsymbol{x}) \right] \right\}^2}{\sigma_a^2 + \sigma_b^2} \right) \right]^N . \tag{I.2.2-5}$$

As the sum over arrival time pairs from station a and b lies here outside the exponential function,

the PDF based on L_{EDT} is more robust in presence of outliers than one based on L2 least-squares, in case of large discrepancies even of the L1-norm (see chapter II.1.1, fig. II.1.1:2). The exponent N denotes the number of observations. As with $L_{LS}(\boldsymbol{x})$, the maximum of the PDF identifies the "optimal" location while the PDF itself indicates its uncertainties. The general formulation of Tarantola and Valette (1982) provides for a complete, probabilistic solution including picking and model uncertainties. However, to obtain analytical solutions for earthquake location, Gauss-distributed uncertainties were applied. While normally distributed and unbiased errors are often assumed, the true distributions are asymmetric and more complex. Picking is biased as picks tend to be late (wherefore no Gauss-shaped distribution around the true arrival time is given). Using Lomnitz' words: "Generations of station seismologists faced the thankless chore of trying to pick an emerging signal in the presence of noise, yet the fact that this operation is inherently biased was ignored" (Lomnitz, 2006). One sided errors do not yield Gauss-like distributions. Also model uncertainties are more complex and can yield asymmetric distributions already for simple 1D-models (chapter II.3.5). Lomax (2011) notes that "estimating these travel-time errors is difficult and often not attempted". Simple estimates of travel-time errors are used instead. For the $L_{EDT}(\boldsymbol{x})$ formulation, the error can optionally be estimated as a fraction of the travel-time. This describes the error well for a homogeneous half space, but is insufficient for more complex models. Lomax concludes "Because it is difficult or impossible to obtain, a more complete estimate of the travel-time errors (or, equivalently, a robust estimate of the errors in the velocity model) is not used. This is a serious limitation of this and most location algorithms, particularly for the study of absolute event locations". Chapter II.3.5 will establish practical estimates for the travel-time error caused by model uncertainties and will introduce a methodology generally working in non-gaussian analysis. The oct-tree importance-sampling of the solution space is fast and efficient, identifying the global maximum even in multi-modal probability density functions. However, it may not identify narrow, local maxima (see chapter II.4, fig. II.4:1 highlights the importance of this point). This scheme is further not intended or able to display the correlation between an individual datum and the solution, which is important for the analysis of doubtful or ambiguous data.

Compared to the residual sum based measure $L_{LS}(\boldsymbol{x})$, the likelihood based on EDTs proves itself more stable in presence of outliers (e.g. fig.II.1.1:1, chapter II.1.1). The underlying concept of EDTs Milne described already in 1886 as a consequence of the *method of circles* which came to be known as the location based on "S-P circles". Chapter II.1 will show that both methods are members of a

set of methods that can be derived applying the concept of jackknifing, yielding location estimates that are robust in presence of outliers. We will look at these two early key methods to understand how the earthquake location may be determined in a robust way, without the use of residual sum estimates.

I.2.3 **Location based on S-P Information**

The *method of circles* mentioned by Milne (1886) locates the earthquake by drawing circles around the stations, searching for a common intersection. The radius is determined by the time difference of P-phase onset and origin time. As, initially, the origin time is unknown, this method is a *method of trial*. To compensate for the unknown, first, circles of a minimum radius were drawn around the stations. If no common intersection was given, the radius was enlarged proportionally among the circles and the circles were redrawn. At the right radius, most circles would intersect in a small region. This approach works for the assumption of that time, that "hypocenters are shallow", but also since this was used on regional and global scale where depth is negligible. However, this approach can be transformed into an explicit one as the origin time may be obtained using e.g. Wadati-diagrams. For the class of velocity models for which P- and S-wave travel on the same paths, the difference in travel time of S- and P-wave scales linearly and proportional to the distance between station and epicenter due to the linear relationship between P- and S-wave velocity (eq. I.2.3-1)

$$\frac{v_p}{v_s} = \sqrt{\frac{2\mu+\lambda}{\mu}} = \sqrt{1+\frac{1}{1-2\nu}} > 1 \qquad \begin{array}{l} \lambda : 1^{st} \text{ Lamé parameter} \\ \mu : 2^{nd} \text{ Lamé parameter} \\ \nu : \text{Poisson's ratio} \end{array} . \qquad \text{(I.2.3-1)}$$

For such models this ratio remains constant over the travelled path. But if this ratio is known to fluctuate significantly in the analyzed region, the epicentral distance can only be found by computing the travel paths and times for P- and S-phase independently and solving eq. I.2.3-3. For the following derivation we will assume a constant ratio.

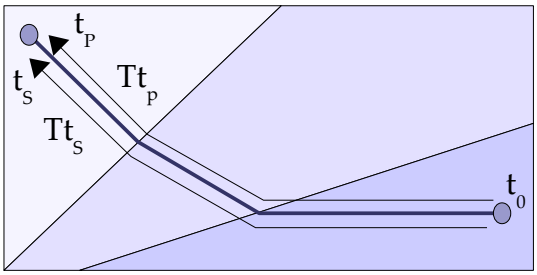


Fig. I.2.3:1: For constant v_p / v_s-ratios the ray paths of S and P wave are similar and both travel times are linearly proportional to each other. Blue shading correlates with the seismic velocity)

Eliminating the origin time from eq. (I.2.1-3), one may write for two given phases, S and P

$$t_p - tt_p(\boldsymbol{x}, \boldsymbol{x}_{station}) = t_0 = t_s - tt_s(\boldsymbol{x}, \boldsymbol{x}_{station}) \quad , \tag{I.2.3-2}$$

yielding

$$t_s - t_p = tt_s - tt_p \, . \tag{I.2.3-3}$$

Using

$$\frac{tt_s}{tt_p} = \frac{v_p}{v_s} \, , \tag{I.2.3-4}$$

eq. (I.2.3-3) becomes

$$t_s - t_p = \frac{v_p}{v_s} tt_p - tt_p = tt_p\left(\frac{v_p}{v_s} - 1\right) \tag{I.2.3-5}$$

and hence the travel time can be obtained from the difference of S- and P- arrival time

$$tt_p = \frac{t_s - t_p}{(v_p / v_s) - 1} \, . \tag{I.2.3-6}$$

For a homogeneous velocity model, this corresponds to a radially constant hypocentral distance d

$$d = v_p \cdot tt_p = \frac{v_p \cdot (t_s - t_p)}{(v_p / v_s) - 1} \,. \qquad \text{(I.2.3-7)}$$

Similar to the concept of trilateration, in which a navigator obtains his own location using the intersection of three circles around fix points, the epicentral distances obtained from the arrival time difference of S- and P-phase can be used to span out several spheres around the given stations. Each sphere corresponds to all possible hypocenters pertaining to the information given at one station. The global solution must comply with all given information, and is ideally determined by the one intersection of all spheres, found in the correct focal depth. Historically, the focal depth was neglected, which allowed the reduction from spheres (3D) to circles (2D). Since an ideal intersection was usually not given due to picking and model errors, a representative epicenter estimate was commonly chosen by crossing the chords, which connect the intersection points of circle pairs (fig. I.2.3:2).

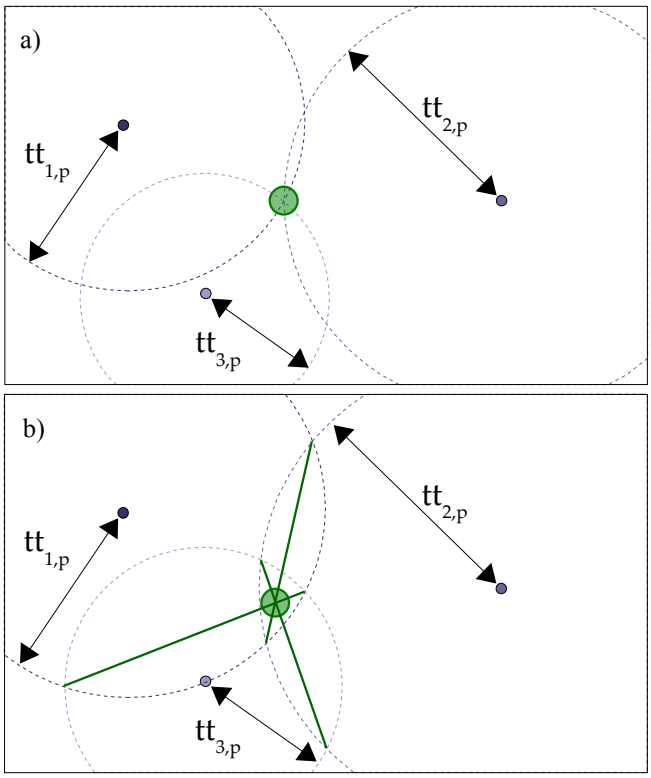


Fig. I.2.3:2: Trilateration. a) Location determined by the intersection of several circles. The solution is marked by the common intersection (green). b) If no common intersection exists, it is estimated by the crossing of chords.

I.2.4 **Location based on P^1-P^2 Information**

Milne's *method of hyperbolas*, which is generally known today by the name of *hyperbolic location* or *Time-Difference-Of-Arrival (TDOA) location*, can be understood as a derivate of the trial *method of circles* for which the origin time was unknown – now applied to station pairs. Let two stations be given for which the onset of the same phase has been recorded. Let the wave first arrive at station 1 at time t_1^p and Δt seconds later at station 2,

$$t_2^p = t_1^p + \Delta t \ . \tag{I.2.4-1}$$

The travel time tt_i from hypocenter to station i is given by

$$t\,t_i = t_i^p - t^0 \ . \tag{I.2.4-2}$$

As the origin time is unknown, the circle's radius is unknown. In the method of hyperbolas (like in the *method of circles)* the origin time is varied, but here synchronized for station pairs. A trial origin time equal to the arrival time at station 1 yields a travel time of $t\,t_1 = 0\,s$ for the first station, and a travel time of $tt_2 = \Delta t$ for station 2. These travel times can be interpreted as proxy for hypocentral distance radii around the stations, allowing to construct circles as done using S-P information. If the distance between both stations is larger than the radius corresponding to Δt, no intersection between both circles is achieved: No agreement exists between both given distance information. Continuing to test with earlier origin times causes both circles to grow until they eventually intersect in one common point. With the trial origin time further decreasing, both circles continue to grow and the subsequent intersections of both circles build up the geometrical shape of a hyperbola (fig. I.2.4:1).

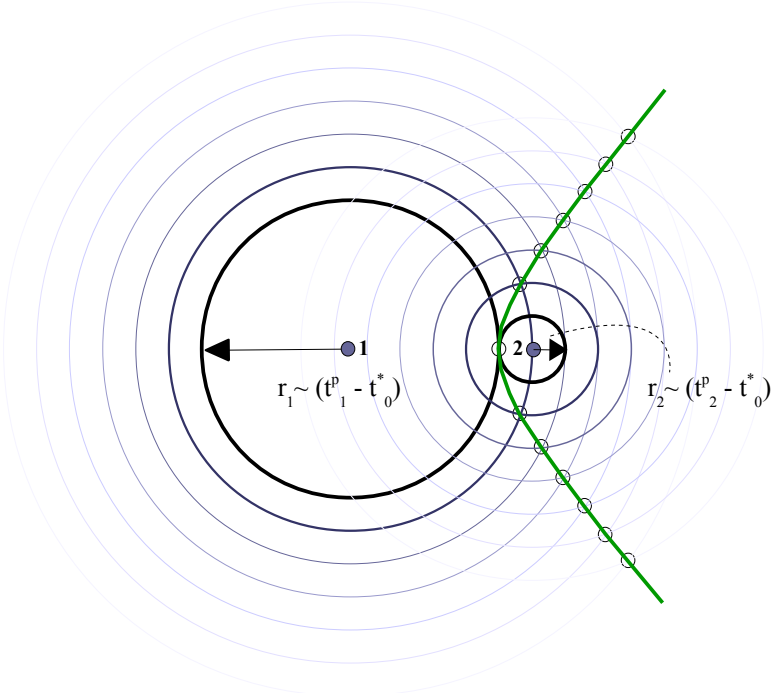


Fig. I.2.4:1: Hyperbola (green) as intersection of two growing circles. The earlier the trial origin time t^*_0 *, the larger the radius.*

The hyperbola displays therefore all possible epicenter locations for all possible origin times, based on the time difference between two stations. The first common point at which both growing circles touch, corresponds to the hyperbola's apex and the origin time decreases with distance to it. The origin time may therefore be interpreted as the parameter of a parametric hyperbolic curve.

At $\Delta t = 0$s the hyperbola takes the form of a line, perpendicular to the axis connecting both stations and located centered. The hyperbola's curvature increases with Δt: The larger the difference between both observed phases, the closer the hyperbola is located toward the first station, bending around it. The origin time difference is limited by physical constraints: There is no solution for a Δt larger than Δt_{crit} which corresponds to the the travel time that the phase needs to travel from the first to the second station tt_{12} (fig. I.2.4:2).

$$\Delta t_{crit} = tt_{12} \qquad \text{(I.2.4-3)}$$

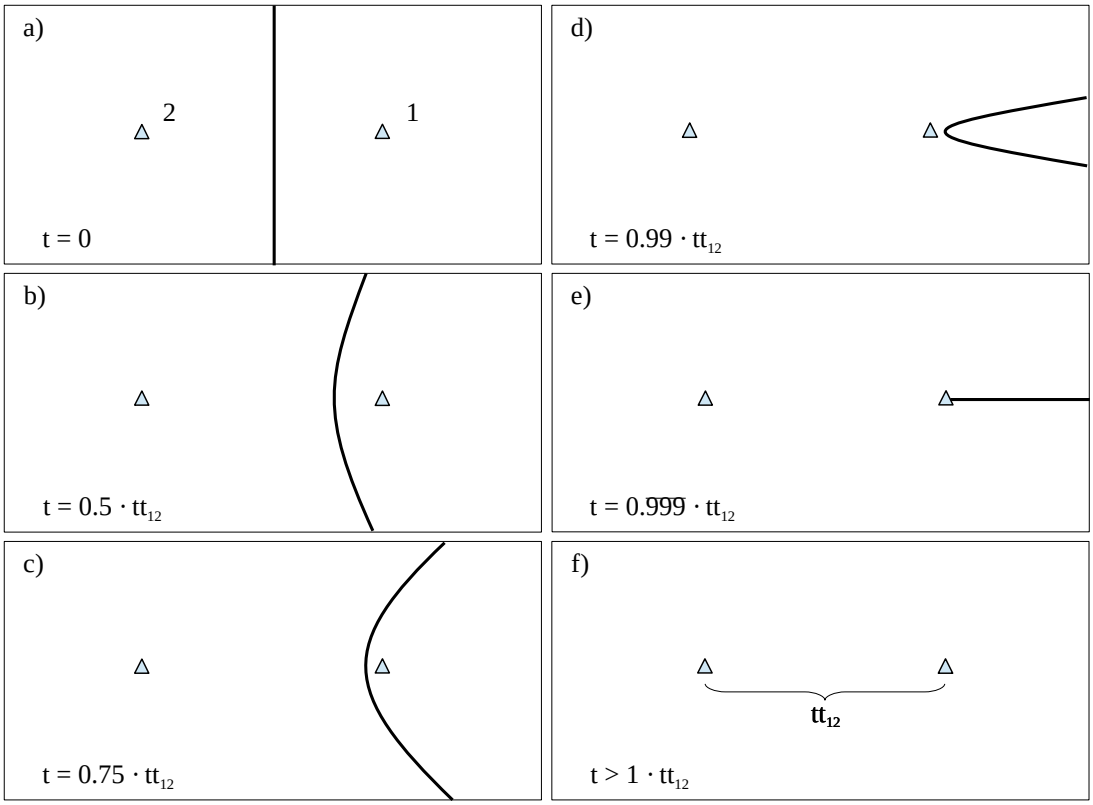


Fig. I.2.4:2: Hyperbolic curves for different arrival time differences in a homogeneous model. In a) both stations detect at the same time. b-d) The earlier station 1 detects, the stronger the hyperbola is bent towards the station. e) When station 1 detects the phase tt_{12} seconds earlier than station 2, the hyperbola collapses into a line. tt_{12} being the time a signal travels from one station to the other. For every point on the line, the travel time difference equals tt_{12} (see text). For larger arrival time differences than tt_{12} the hyperbola ceases to exist (f).

This may be understood, considering that the most extreme Δt are achieved for source locations lying on the connecting axis of both stations: $|\Delta t|$ is minimal for a source centered between both stations (in a laterally homogeneous model). For a source collocated with the station of first detection, holds $\Delta t = tt_{12}$. For a source on the same axis outside of the network, let tt_{s1} define the travel time between source and first station and tt_{s2} the travel time between source and second station. For the latter, the wave must travel past station 1, continuing to station 2. Therefore,

$$tt_{s2} \approx tt_{s1} + tt_{12} \tag{I.2.4-4}$$

and

$$\begin{aligned}\Delta t &= t_2 - t_1 \\ &= (tt_{s2} + t_0) - (tt_{s1} + t_0) \\ &= tt_{s2} - tt_{s1} \\ &\approx tt_{s1} + tt_{12} - tt_{s1} \\ &= tt_{12}\end{aligned} \qquad \text{(I.2.4-5)}$$

The maximum possible arrival time difference for any source location outside of the network is given by $\Delta t = \Delta t_{crit}$. Approaching Δt_{crit}, the hyperbola's apex approaches the station of first detection while its opening angle (the angle between both asymptotes) goes toward zero, the "hyperbola collapses into a line". Larger onset time difference are physically impossible for the given velocity model. This defines the existence criterion for hyperbolas

$$tt_{12} < \Delta t_{crit}\,. \qquad \text{(I.2.4-6)}$$

Larger onset time difference indicate discrepancies in the data. This fact often allows to discard ambiguous phases for which this criterion is violated. On the other hand, if phases at different stations can clearly be identified to be the same (e.g. by their pattern and/or spectral signature), but the existence criterion is violated, an indication for model-discrepancies is given.

When more than two stations are given, the *hyperbolic location method* may be applied. The following two (space-)dimensional example (i.e. known focal depth) illustrates the concept of hyperbolic location with three given stations (fig. I.2.4:3).

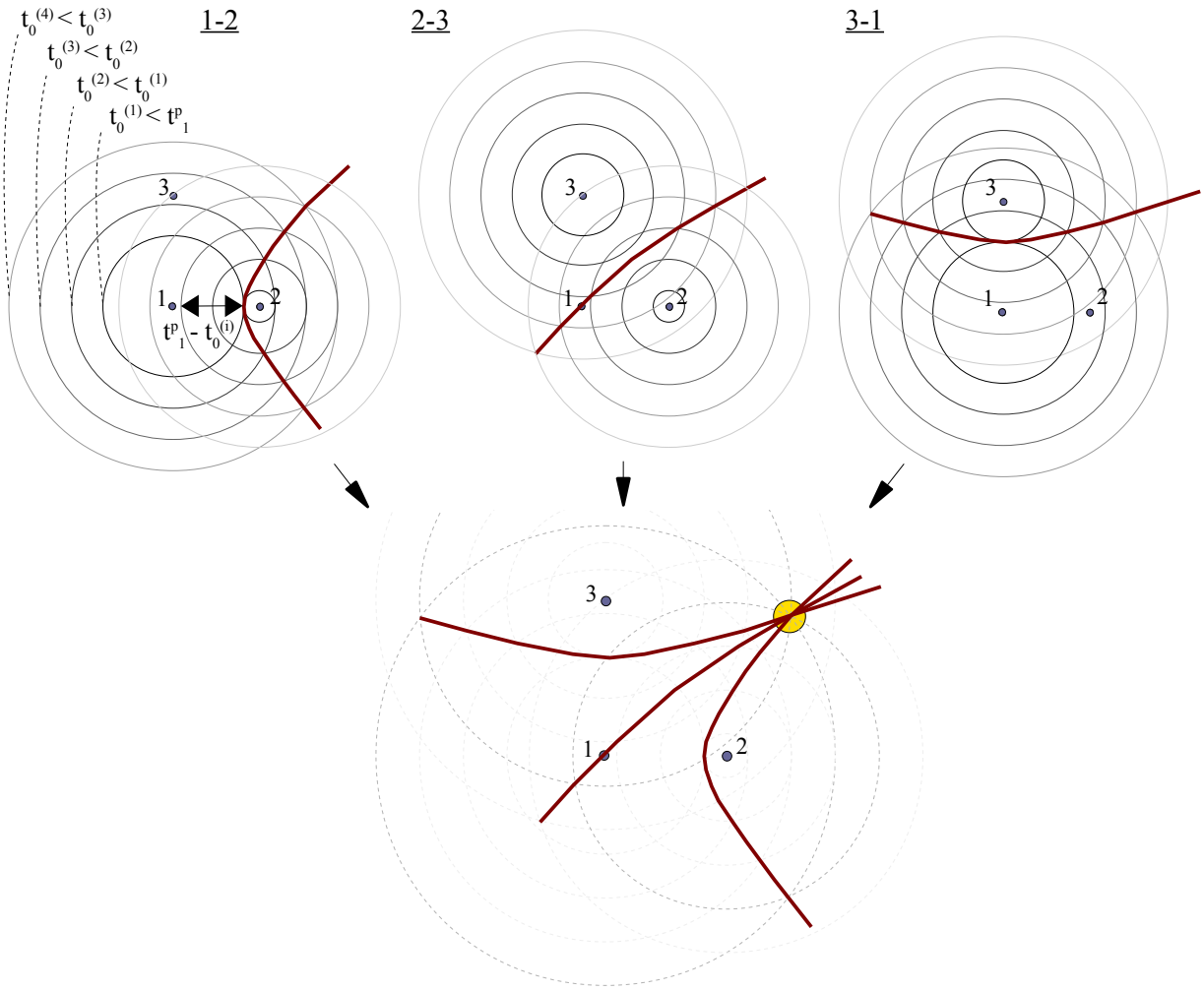


Fig. I.2.4:3: Hyperbolic location. Between pairs of stations a hyperbola is constructed by varying trial origin times (each circle corresponds to a constant origin time). With three given stations three pairs of station can be formed, yielding three hyperbolas (top). These three intersect in a common point (bottom), marking the epicenter (yellow circle). This point bears the same origin time (intersection of three dashed circles of same origin time). Chapter 208 will show that for certain sets of arrival times also multiple intersection points may exist, leaving the solution ambiguous (fig. I.2.4:5).

The permutation of stations 1,2 and 3 into pairs yields three hyperbolas which represent a coupled system of equations (station 1&2, 2&3 and 3&1) being exactly determined. Every hyperbola represents all possible epicenters (and origin times) based on the subset of information it relies on. The solution pertaining to the full set of information must also fulfill the condition of every hyperbolic constraints, individually, i.e. will correspond to one of the solution of each subset. The solution is therefore identified as the intersection point of the three hyperbolas. This point (all constraints meet in the same location) bears also the same origin time on all hyperbolas and will be called the “triple-point” of the triple-group that consists of the three involved hyperbolas. If the arrival time reading of a phase is altered, the solution will be spatially shifted: The triple-point will

move along the third hyperbola which is independent of that reading and fixed in space (fig. I.2.4:4). This allows the analyst to estimate the geometric impact that a time shift (e.g. onset correction) has on the spatial solution. For large arrival time-differences, secondary solutions (multiple "triple-points") may appear in a triple-group. This is due to the fact that though the underlying equation system is exactly determined (3 unknowns, 3 constraints), it is not linear (fig. I.2.4:5). It requires additional data to to resolve this ambiguity (chapter III.2).

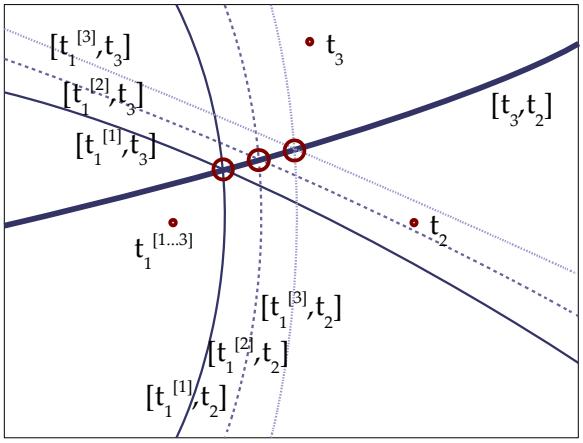


Fig. I.2.4:4: Three different phase readings are evaluated at station 1 (red point), each one corresponding to another set of lines (solid line, dashed line, thin solid line) The solution (red circle) moves on the third hyperbola [2,3] that does not depend on the altered phase.

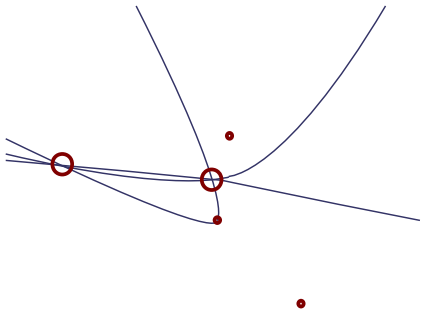


Fig. I.2.4:5: Three station network (red points) in a 2D location scenario with a degenerated solution: Three detected P-onsets constrain three hyperbolae which intersect in two locations.

With additional stations more hyperbolas are generated and the system becomes overdetermined. The number of hyperboloids is given by

$$H = \frac{1}{2} \cdot \left(n^2 - n \right) \quad , \tag{I.2.4-7}$$

with n being the amount of stations that detected the phase, given that the condition of I.2.4-3 is met for all. The number of triple-groups is then given by

$$T = \frac{1}{6} \left(n^3 - 3n^2 + 2\mathrm{n} \right) \quad . \tag{I.2.4-8}$$

The actual number of triple-points per -group may actually be higher due to degeneration: Already in a homogeneous model it can be twice as high as T , in more complex models even higher. In an error-free scenario all triple-points intersect in one single point, marking the global solution.

Errors in the data cause the triple-points to scatter (fig. I.2.4:6). As each triple-point intersection corresponds to the agreement of multiple location constraints, their spread correlates with the solution quality. The larger the spread, the higher is the uncertainty in location based on the given information. The lower the spread, the higher is the line density in the triple-point region, expressing the agreement of the majority of subsystems. (Note: A low spread (i.e. the agreement of subsystems) does not necessarily guarantee accuracy. A compact triple-point cluster (with low spread) might actually be biased by a locally inaccurate model. Adding a station situated in a region for which the model may be more accurate (and hence provides better location estimates), will therefore increase the system's accuracy - but also enlarge the spread. Spread indicates therefore the quality of the solution in means of indicating the extent of discrepancies in the given information. The analyst, however, still needs to decide which data is most trustworthy.)

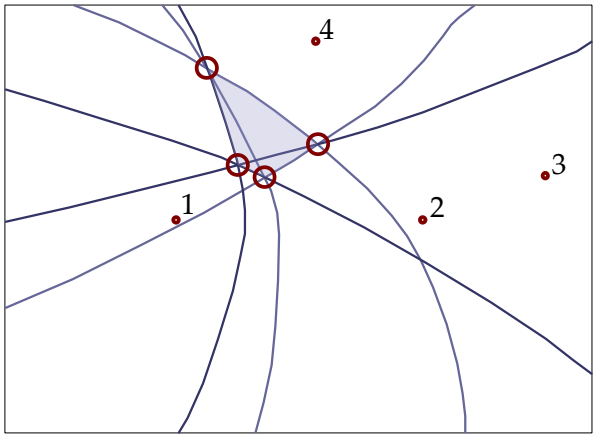


Fig. I.2.4:6: Overdetermined System and triple-point spread. Four stations yield 6 hyperbolas and 4 triple-points clustered in a region marking the global solution.

In the following chapters we will use the fact that the same information base that is usually evaluated in residual sum analysis to obtain a best fit solution, can also be decomposed into small subgroups, with a member size just big enough to still give constraints for the location. In this context we will again find the S-P- and hyperbolic constraints playing a key role, and will introduce additional constraints. Depending on the situation, certain constraints gain special importance and are applied and are given weight according to the given problem. The method developed will generalize the constraints to the context of complex 3D velocity models and introduce uncertainty measures to form a solid base to analyze sparse and low SNR data, allowing to track the effect of an input datum to the global solution while preserving the aspects of solution completeness (i.e. to reveal ambiguous solutions) and a probabilistic analysis as described in chapter I.2.2.

II **Distinct Constraints**

The application of location schemes based on residual sums has drawbacks for the analysis of uncertain data that need to be overcome: Due to the summation, such schemes are sensitive to outliers. This point is crucial as outliers (e.g. mispicks) are expected due to unclear phase identifications. The summation further prohibits to perceive the impact that an individual datum exerts on the solution, e.g. how it independently constrains the solution and which effect parameter changes (e.g. modification of the arrival time) will cause. This insight is important as it supports to find the true onset when ambiguous phase onsets exists. Furthermore, parameter uncertainties cannot easily be translated into uncertainty regions enclosing the hypocenter: The solution is given as a point with minimum residuum (linearized inversion) or as residual map (grid searches), but no threshold exists that would determine which residual is too large to still yield a valid solution, and hence no robust region is defined that encloses the hypocenter. At best confidence intervals can be derived based on the variance in solution. They do, however, not guarantee location accuracy (Pavlis, 1986). This quality, however, is important whenever the location is the basis for far reaching decisions. Earthquake early warning serves as good example. There, reliable boundaries for the location can be translated into lower (and upper) bounds of magnitude estimates, that directly trigger the warning for the society (fig. II.4:1 in chapter II.4 displays a recent example of an Earthquake-Early-Warning module using a RMS grid search scheme that dislocated the hypocenter by over 200km). The spatial extent of this region is directly dependent on the parameter uncertainties. For analytical solvability, these uncertainties are often taken as Gauss-distributed. This, however, does not reflect the actual uncertainties, in neither picking nor model. This is one of the reasons, why obtained confidence intervals may be inaccurate. Lomnitz (2006) summarizes the loss of information due to the summation process as that “the problem is compounded by the fact that we locate an object in space by using measurements in the time domain. By reducing the problem to the time domain we blind ourselves to the fact that the residuals are actually vectors. A 1-sec residual is not the same thing (i.e., should not have the same effect on the solution) as a residual of 1 sec to the north.” Lomnitz continues “In theory there is no obvious way of translating a least-square algorithm in the time domain to the space domain. Should we minimize the vectorial sum of the distance residuals? This is not the correct answer either.” He leaves this problem unsolved, but points to the only direction that in his opinion may offer a remedy: “Beno Gutenberg used a graphical location method that was routinely utilized in Pasadena for years, particularly by

Charles F. Richter in the classical joint research work *Seismicity of the Earth* (Gutenberg and Richter, 1954). " Although context and application differ, we will find that "graphical" location indeed allows to overcome the shortcomings mentioned above, and that it yields a methodology well suited for the context of uncertain data. "Graphical" should be understood in the sense, that only the spatial component of **m** is inverted for, not referring to whether the solution space is actually visually inspected by an analyst or wether it is numerically evaluated. The hypocenter problem will be solved independent from the origin time, which is estimated via forward modelling after the location is determined. This is realized by jackknifing the dataset into smallest units of parameter information still providing information on the location. These distinct units, individually and collectively constraining the model vector **m** consecutively to smaller regions, are therefore called *distinct constraints* and are the subject of the presented work. Their derivation shall be outlined in the following chapter.

II.1 **Elementary Jackknifing and Phase Difference Groups**

Jackknifing is a resampling technique often applied for variance or bias estimation (Quenouille, 1949). Solution robustness is evaluated by subsequent analysis of varying subsets of the available data. Disregarding parts of the data during the analysis allows to understand the impact of the excluded data on the solution, whether it establishes the global trend or diverts from it, and provides the means to construct a global solution based on the most coherent subsets.

Let $\hat{\theta}$ be an estimator of the parameter θ regarding the full sample η consisting of n elements ($\eta_i, i \in [1,n]$). Let $\hat{\theta}_{-E}$ represent the estimate regarding a subset η_{-E} of the full sample, consisting of all data but e sample elements (e being set according to the maximum expected number of faulty data). The Jackknife estimate of θ is obtained by averaging the estimates $\hat{\theta}_{-E}$ for all possible subsets η_{-E} that can be formed by combination of the n sample elements into e-member groups. The sample is hence distributed into g groups of $h = n - e$ elements, with

$$g = \frac{n!}{(n-e)!\,e!} = \frac{n!}{(n-h)!\,h!} = \binom{n}{h} , \qquad \text{(II.1-1)}$$

each group excluding a different subset of e elements.

The common form of Jackknifing excludes single observations (Miller, 1974). In this case applies

$e=1$, $h=n-1$ and $g=n$. For this case let $\hat{\theta}_{-i}$ denote the estimate regarding the subset η_{-i}, consisting of all data but element i. The mean jackknife estimate $\hat{\theta}_J$ pertaining to the whole data is now obtained by averaging the estimates $\hat{\theta}_{-i}$ of all η_{-i},

$$\hat{\theta}_J = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i}\,. \tag{II.1-2}$$

Commonly, a jackknife estimate $\tilde{\theta}_J$ is applied that reduces the methodological bias

$$\tilde{\theta}_J = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i \;=\; n\hat{\theta}-(n-1)\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i}\,, \tag{II.1-3}$$

using the pseudo-value

$$\tilde{\theta}_i = n\hat{\theta}-(n-1)\hat{\theta}_{-i}\,,\quad (i=1,\,\ldots,\,n)\,. \tag{II.1-4}$$

A measure of the impact D_i that the excluded observation η_i exerts on the solution (Rabinowitz and Steinberg, 1988) is given by

$$D_i = (\hat{\theta}_{-i}-\hat{\theta})^T\, cov(\hat{\theta})\,(\hat{\theta}_{-i}-\hat{\theta})/\,dim(\hat{\theta})\,. \tag{II.1-5}$$

This basic concept, laid out for a single excluded observation, holds for every group size (Quenouille, 1956). While excluding one datum reveals whether the related information is (a) redundant, stabilizing the trend or (b) significantly changing the solution (which in well-conditioned problems allows to identify outlier candidates), it does not reveal how the datum itself constructs the location (i.e. constrains the hypocenter location to a subset of space): We might exclude a station from the location procedure after we located an event by e.g. the L1-minimum of residua and relocate to see how the hypocenter location changes – we will often still not be able to understand why the solution changed the way it did. This is due to the impact, which the data of all other stations exert on the solution. When we now expect faulty data among the few data provided, this becomes problematic as both, $\hat{\theta}$ and $\hat{\theta}_{-i}$ are based on data, that themselves are not error-free and the estimate D_i may still be biased by the error of the remaining data. This can be demonstrated with a simple scenario: Let e.g. seven stations be given with five providing good onsets (holding only minor errors) that point out a certain hypocenter. Let the dataset further be contaminated by two erroneous onsets (outliers, holding large errors), that both have a similar effect

on the solution. In such a case, excluding one datum will identify neither of the two as outlier, as the second outlier, still contained in the dataset, biases the sub-solution towards the other. The only way to identify the two outliers contrasting the five well-agreeing stations would be given in repeating the jackknife with sub-groups excluding two members (group size $h=n-2$). Since the number of bad data is initially unknown, this process could be repeated with decreasing sub-group size h down to the minimum number required by the estimator. The analysis, however, would be highly dimensional, since the jackknife would have to be repeated $n-4$ times. It would further be computationally expensive, as the inversion has to be repeated g times with each repetition. Additionally, the amount of sub groups g is initially quickly growing with decreasing sub-group size h, following the binomial coefficients (eq. II.1-1). The inversion would therefore be performed orders of magnitude more often than in a standard location procedure and the result would be complex, multidimensional, and difficult to interpret.

Let's assume the same scenario of seven stations, but containing only one outlier and let the jackknife be performed only once, but right away excluding two members ($h=n-2$). Most sub-groups (all those excluding the outlier) will point to the same hypocenter. Few sub-groups (those including the outlier) will not follow the trend. Looking at two of the discrepant sub-solutions will identify the outlier as common datum in both groups. This simple example illustrates the fact that the dimensionality of the solution can be reduced by directly using smaller sub-group sizes, maintaining the ability to identify and isolate bad data, as long as $h<n-o$ (o being the number of outliers/bad data). Following this thought to the end, it can be inferred that all outliers can be identified by only using the smallest sub-group size possible. To dissect the information to the lowest level, we could therefore attempt to apply Jackknifing in the strongest form possible, by reducing the group size to the minimum number required by the estimator. The standard estimator for earthquake location (operator of inversion) relies on 4 data (since $dim(\boldsymbol{m})=4$). The global, overdetermined system would therefore be decomposed into exactly determinable subsystems of 4 members (h=4), which would be individually solved. The global solution could then be recomposed based on the solutions of the subsystems, whereby stable, outlier resistant solutions could be obtained. Fig. II.1:1 shows an example for a two-(space)-dimensional problem with five given stations. In this example, the depth is no unknown, and the dimension of the problem and the size of the exactly determinable subgroups is hence reduced to three. All data, except one outlier (the P-onset of station 5) are chosen to be ideal (error-free). The five stations (panel A) are permutated into 10 three-member groups (surrounding panels).

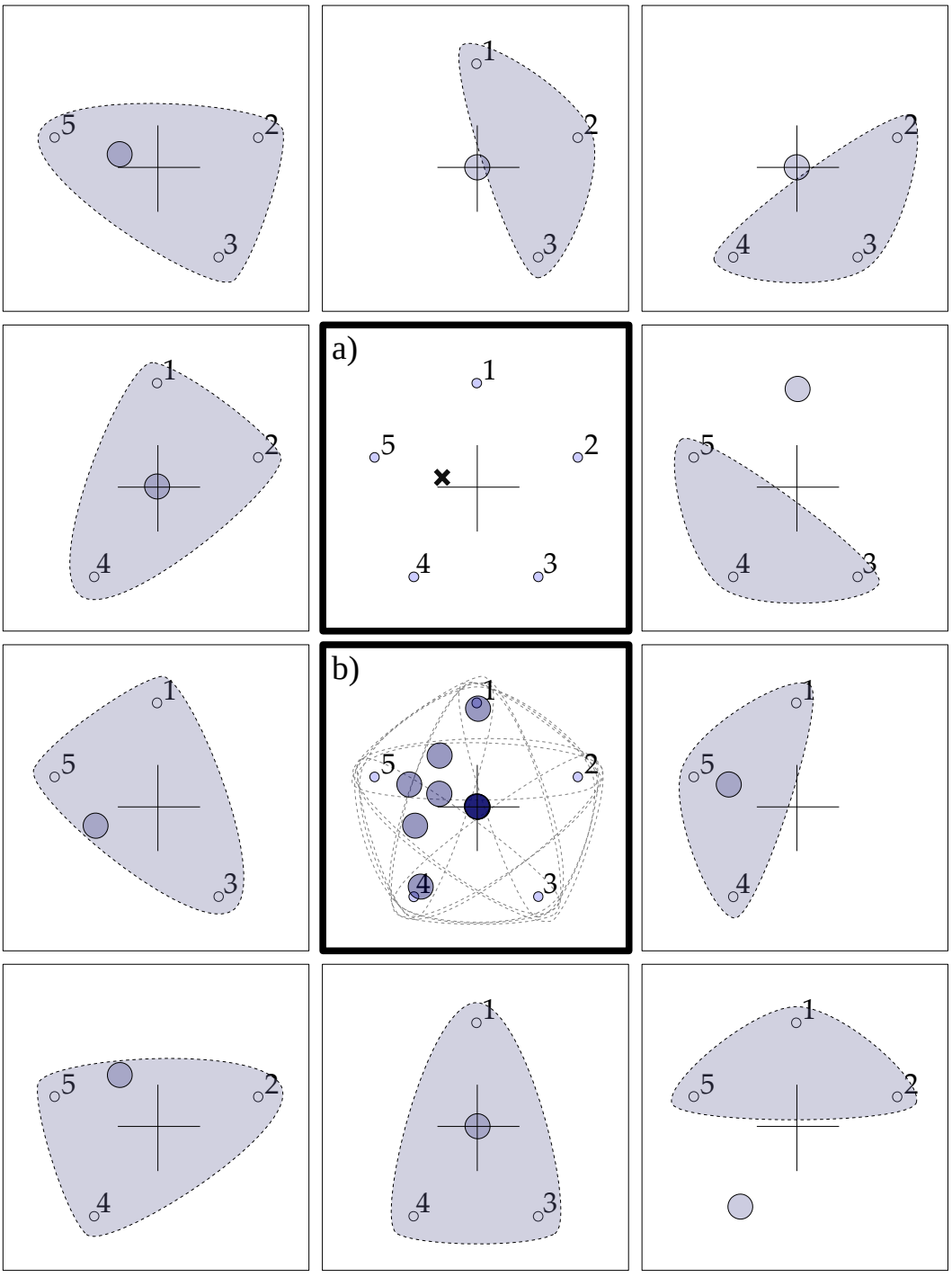


Fig. II.1:1: Principle of Location Jackknifing, displayed for a two dimensional location scenario with five given stations. The central Panel A shows the five stations, the true location (cross-hair) and the solution (small cross) that minimizes the L2-residua sum. The global system of five stations 1,2,3,4 and station 5 of faulty data is decomposed into ten exactly determined subsystems of three stations (surrounding panels), which are independently solved (solid circles). The solution of the groups not relying on station 5 correspond with the true location, all other sub-solutions are scattered. The global solution for the whole system is composed based on the superposition of sub-solutions (Panel b). The scatter of sub-solutions indicates the solution quality, a large scatter indicating a lower quality. Although the large scatter indicates the existence of a significant discrepancy in the data, the maximum still identifies the true location.

The superposition of the 10 subgroup solutions shows a maximum at the true location, as several groups do not rely on the faulty data of station 5. The solutions of all other groups relying on the faulty data are scattered. This large scatter indicates the existence of a significant discrepancy in the data. The maximum identifies the true location, nonetheless.

The amount of combinations and subgroups is given by the binomial coefficient, respectively in the factorials of n and h according to the principle of combination without order and repetition according to eq. II.1-1. The amount of combination groups grows therefore quickly with the amount of stations, e.g. 10 stations already form 210 subsystem groups for h=4. It should therefore be noted that such an elementary decomposition is -computationally speaking- only practical for smaller data sets.

When only few data are given, a subgroup of four stations is still too large to display the correlation between data and solution on an elementary level. For a better breakdown of information, it is desirable to reduce the size of the data subgroups to the smallest unit possible. This can be done by using underdetermining estimators. The smallest unit providing explicit location information (the "quantum of information") are pairs of onsets. We will therefore reduce the sub-group size down to only two members (h=2), yielding underdetermined systems. This will reduce the complexity of the problem, but will also ease the computational load as the amount of sub-groups with two members (h=2) equals the one with h=n-2 of the earlier example, being significantly smaller than the amount of sub-groups with four members (h=4), i.e. for 10 detecting stations, two-member groups (h=2) result in 45 sub-groups compared to 210 for four-member groups (h=4).

Rather than averaging the sub-group solutions as in eq. II.1-3 (fig. II.1:1), we will superpose the solutions $\tilde{\theta}_i$ in solution space. This serves two purposes. For any regression the jackknifing solution (eq. II.1-3) will be less sensitive to outliers than L1- or L2-normed RMS schemes as the large amount of groups not dependent on the outlier significantly outweighs the impact of the few groups dependent on it. By superposing the sub-results and using a center-of-gravity (cog) approach over the zone(s) of highest congruity (rather than averaging over all sub-group solutions), outliers are additionally weighed less as they appear far off the cluster of sub-solutions. The second advantage lies in the ability to directly observe the effect of parameter changes on the solution scatter. It will become perceivable which datum constraints the solution most, i.e. input data for which small perturbation cause little effect on the solution. Vice versa, an input datum for which a small perturbation results in a large change in the corresponding spacial sub solution, constrains the

global solution only poorly and may be weighed less. Summarizing, two important characteristics should be highlighted: 1) The two-member sub-groups rely on the minimal possible subset of data. 2) Any given datum (e.g. phase-pick) will only affect the sub-groups depending on it. An Analysis based on these two-member groups therefore allows to reveal the impact that smallest data units exert on the solution. Further, it allows to identify outliers even for small datasets as non-conform sub-solutions can be linked back to the few data its group relies on, allowing to identify the discrepant information. An example is given in fig. II.1.1:1.

II.1.1 **2-Phase Groups**

The earthquake location problem is a four dimensional problem (x, y, z, t^0). However, using sub-groups of two members and a pair wise analysis between all observed phases, this dimensionality can be reduced within the inversion by choosing a suitable evaluation operator. According to eq. (I.2.1-3), two observed phase onsets t^A and t^B stand in relation to origin time and location according to

$$t^A = t^0 + Tt(\boldsymbol{x}, \boldsymbol{x}_A) \qquad \text{(II.1.1-1)}$$

and

$$t^B = t^0 + Tt(\boldsymbol{x}, \boldsymbol{x}_B) \quad . \qquad \text{(II.1.1-2)}$$

Forming the onset time difference Δt eliminates the origin time from the equations and reduces the problems dimensionality to three dimensions, yielding a solely spatial inversion problem (which bypasses the need for origin time estimates within the inversion process, chapter I.2.2).

$$\Delta t = t^B - t^A = Tt(\boldsymbol{x}, \boldsymbol{x}_B) - Tt(\boldsymbol{x}, \boldsymbol{x}_A) \qquad \text{(II.1.1-3)}$$

Analyzing the difference of the two members (i.e the arrival time difference), the use of the jackknife in this strong form not only dissects the data set to the lowest level, it also has technical implications: It allows to find the solution (i.e. the source location, and from the source location the origin time) using a three dimensional direct search solely based on travel time differences, by comparing such with the observed arrival time differences.

A sub group of two members can be formed by choosing two onsets of different phases at one station (class I constraint) or using phase onsets that are recorded at two stations (class II constraints). For the latter, the phase type is usually the same for both phases (e.g. P) but may also differ. A well known phase onset pair for one station is given in S- and P-phase, but the difference between direct- and refracted P-phase is another useful choice. The first corresponds to the classical *method of circles* explained in Milne (1886) (chapter II.1.2.1). For two stations, the difference between P-phases is the dominant choice, which corresponds to the classical *method of hyperbolas* (Milne, 1886) also known as *hyperbolic location* (chapter II.1.2.2). The same is applicable for S-phases which increases the stability of the solution due to the lower S-wave propagation velocity, or surface waves which are dispersive but propagate even slower. However, the lower precision in onset arrival time picking often compensates that effect. For local earthquakes P-onsets are therefore the phase of choice.

Each group of constraints displays the consistency of the involved phase data: A group of P^1-P^2 constraints will reveal the consistency between all P-phases, e.g. whether two phases P^1 and P^2 agree with the solution that is indicated by the other stations' P-phases. A group of S-P constraints will validate the consistency of the S- and P-phases at individual stations. If e.g. a certain S-P constraint appears far off the point of highest constraint density (i.e. the region pointed out by most S-P constraints), it indicates that either the P-phase- or the S-phase picking at that station is inaccurate (if the respective velocity models are trustworthy). In combination with P-P constraints it can be determined wether the P- or S-phase caused the discrepancy: If e.g. the corresponding P-P constraints (those, that depend on that station's P-onset) show a high congruity with the other P-P constraints (i.e. agree on a certain hypocenter region), only the S-onset or the local S-velocity could be inaccurate, identifying this station's S-information as problematic (The same would have been found by analyzing the scatter of S^1-S^2 constraints). This concept of constraint consistency can be extended to any phase combination to identify outlier onsets. It can further be used to evaluate the assumptions that were made for underlying parameters which the constraints depend on: For example, the consistency of constraints which are based on the difference of direct and refracted P-phase onsets P-P_{refr} can be used to evaluate the particular model assumption of a refractor's depth or velocity contrast. Similar to the P^1-P^2 Equal Differential Time (EDT) constraints, S^1-P^2 EDT's that use different phases at different stations (and appear as Pseudo-Circular shapes for a given depth) may be used to evaluate whether the S-phase at one station yields consistent location information with the P-phase at another station (Ziv, personal communication, 2015). Using the concept of

constraint consistency, problematic data can be excluded and a trustworthy set of information be constructed, on which the location can be based.

The two independent subgroup types of S-P and P^2-P^1 provide an optimal representation of the information given in P- and S phases, due to their specific constraining characteristics, which will be laid out in chapter III.1. This information can be supplemented by a group defined through the double difference between phase onsets of three stations (e.g. $[t^{P,2}\text{-}t^{P,1}] - [t^{P,3}\text{-}t^{P,1}]$). This combination of phase differences in conjunction with a plane wave assumption allows to cancel out the local velocity dependency and yields the spacial information of back azimuth (BA), the local direction of wave propagation, and the apparent velocity with which the wavefront propagates in between the stations. This constraint is especially useful for laterally homogeneous structures where the local velocity might not be well known, but can be assumed to be constant over the inter-station distance. This double difference hence forms the basis for the jackknife-analysis of small-aperture arrays (chapter II.1.2.3) where the latter condition is usually met. With knowledge of the local velocity for the analyzed phase, the apparent velocity may be translated into an incidence angle, so that the 3dim. propagation-vector can be obtained – a quantity which may be derived independently from the three dimensional particle motion of a three component sensor (chapter IV.1). This shows that the computation of the propagation vector can be obtained in two independent ways based on different information, but always requiring three components, either three collocated orthogonal components or three distributed parallel ones. In principle, both methods can be used to obtain the BA information and are laid out in chapter II.1.2.3 and IV.1.

The three groups, S-P, P^2-P^1 and BA, define the set of basic subsystems, applied in the work presented, to constrain the solution in the space domain (Joswig, 2005). These three constraint classes will be referred to as S-P, P^2-P^1 and BA constraints. Each constraint type has its own constraining characteristics (chapter III.1) which give it complementary advantages in different situations (e.g. locations inside- vs. outside the network, in near vs. far distance or concerning depth constraining). As each sub group relies on a minimal set of data, it allows to trace back outlier sub-solutions directly to the datum they rely on, offering the analysts a deeper understanding of the data impact on the global solution. Since 2-phase groups are the result of a dataset which is jackknifed to the lowest possible level, a pair-wise analysis of phase-onsets will be most robust to outliers in the dataset. Schweitzer, who applied a pair wise phase analysis for HYPOSAT (Schweitzer, 2001) in the methodological context of a linearizing inversion using a modified $\boldsymbol{G}$ matrix, observed that in all test cases with erroneous data, an inversion using travel-time differences yielded better solutions to

the true source than least-square approaches. The class II P-P constraints, which appear as hyperbolic bands in a two space-dimensional problem (e.g. fixed hypocentral depth), shall be used in the following paragraph to demonstrate this solution stability. Although the following chapters will focus for the formal derivation on the classes of P-P and S-P constraints as the predominant examples of constraints being based on one phase type at two stations and constraints with two phase types at one station, respectively, the derivation is valid for any other phase-type combination.

Let four stations with P onsets be given, generating six P^i-P^j constraints, with $i, j \in \{1,2,3,4\}$, $i \neq j$. Recalling hyperbolic location in which the solution was identified by the intersection of the hyperbolas (Chapter I.2.4), we will now likewise look for the intersection of the P-P constraints (regions of constraint agreement) to identify the epicenter. Of those six constraints, one will be based on the onsets of station 1 and 2, a second one based on station 2 and 3 and a third one on the phase onsets of station 3 and 4. (The other three constraints we will disregard for a moment.) For all time differences consistent with the velocity model, the first and third hyperbolic constraint must always intersect the second, since both constraints share one onset with the second constraint. The first and third hyperbolic constraint, however, are fully independent and may not intersect/agree, indicating an inconsistency between the phase onsets of station 1 and 4. Adding the remaining three constraints, those based on consistent information will yield intersections, clustered in a certain region, while those based on inconsistent information will produce spreading outlier solutions (see e.g. fig. II.1.1:1b). A pairwise comparison between the given phases of all stations therefore weighs the consistency of information. An explicit example is shown in fig. II.1.1:1. Panel a) displays hyperbolic constraints, generated from well picked P phases of five stations with proper velocity model assumptions for a surface source in proximity to a small aperture array (ALNB) of the DeadSeaNet array of arrays. Compared to the hyperbolic location in chapter I.2.4 the constraints here bear a certain “thickness”. This is due to applied uncertainties in the input parameters which cause an uncertainty in the solution space (see Chapter II.3). All constraints intersect in a common region. Panel c) shows the corresponding L2-residual sum least-squares representation, with a well defined minimum. This changes when the algorithm misidentifies one station's P-onset (panel b and d) and triggers 0.02s too early. The least squares analysis yields a large minimum residuum, with a wide spatial extent, strongly displaced from the true location (d). The constraint based approach proves itself more stable: Although the total value of constraint agreement (intersections) decreases, the location based on hyperbolic constraints remains stable at the original location.

The more constraint types are used, the more is the consistency of the information base being evaluated - revealing possible discrepancies.

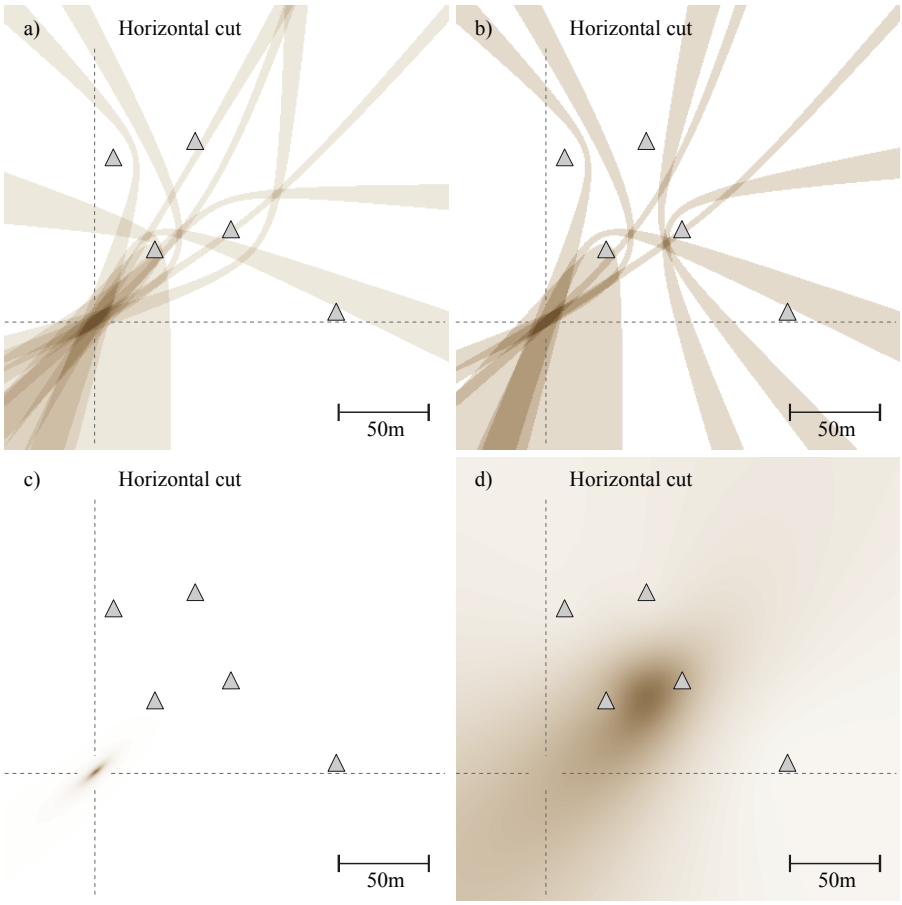


Fig. II.1.1:1: Hyperbolic constraints (a) and L2 least-squares (c) for good information in a local event scenario: The constraints yield a clear maximum, and the least-squares residuum a clear minimum at the source. With one station mis-picking, the least-squares minimum indicates a different location (d) while the hyperbolic location remains with its maximum at the true location (b).

Beside the stability of the solution given by the consistency measure of subgroups, outlier data can directly be identified by constraints that don't follow the main trend: Identifying the two hyperbolic constraints that disagree with the majority of other constraints (fig. II.1.1:2a, marked blue), constraint P^D-P^A and P^D-P^C, identifies the common P phase of station D as inconsistent information. Fig. II.1.1:2a) highlights all constraints which depend on station's D P-phase. One may notice that actually no hyperbolic constraints are visible between station D&B (P^D-P^B), and D&E (P^D-P^E). These onset time difference do not yield information consistent with the chosen velocity model. The common datum in both constraints is again onset D and indicates again, that onset D is erroneous. The drop of the total amount of constraints under the theoretical value of eq. II.1-1 (or

specifically I.2.4-7 for the hyperbolic sub system) indicates therefore wrongly chosen phase onsets or inconsistencies with the chosen velocity model.

When we compare the results of the hyperbolic constraints with the performance of a L1-residuum (which is significantly more robust than the L2-norm), we find that also the global L1-minimum identifies the true location, yet, less distinct (fig. II.1.1:2): A secondary minimum (similar to the global one of the L2-norm) exists, which renders the solution poorer in contrast than the maximum given by the hyperbolic constraints. It is interesting to note that several of the hyperbolic shapes can be identified in the shading of the L1 map in vicinity of the minima (compare fig. II.1.1:2 (a) and (b)). However, their (Jackknifing) character, which allows to directly spot the outlier, is not given.

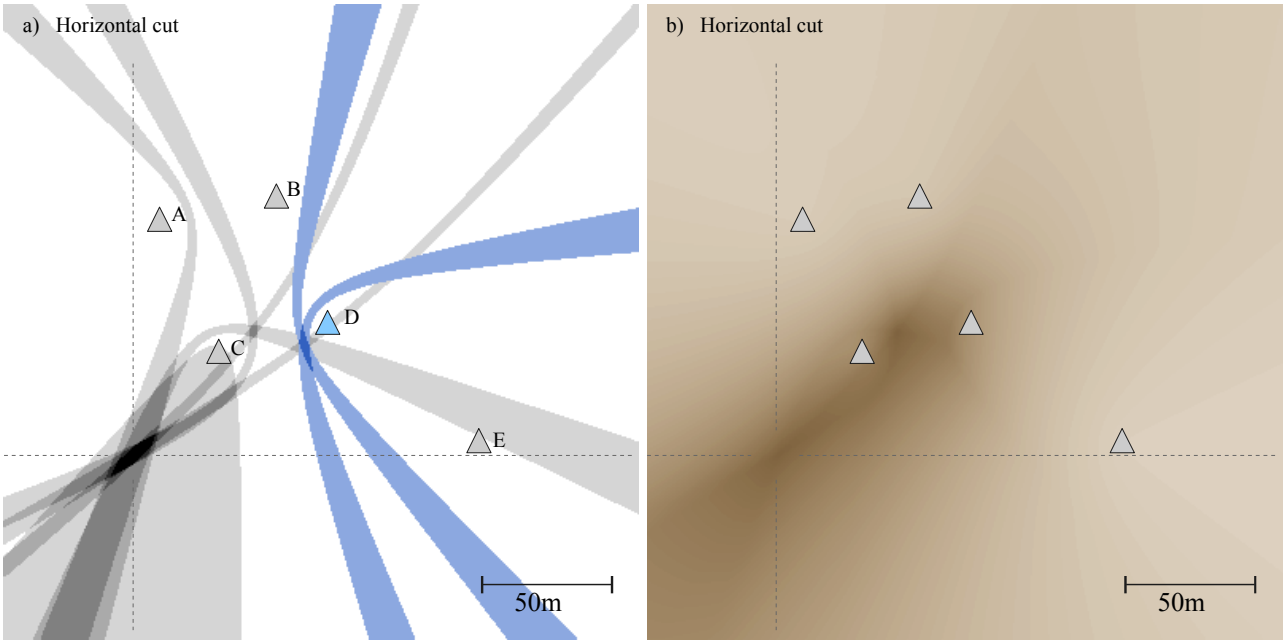


Fig. II.1.1:2: a) Identifying the constraints that do not follow the trend (two blue hyperbolic constraints), identifies the inconsistent phase information (phase onset at station D). Panel (b) displays the least-squares residuum using the L1 norm for comparison: Being significantly more robust than the L2 norm, the global minimum remains at the true location. However, a comparable minima exists, which is similar to the global minimum of the L2 norm (fig. II.1.1:1d). Comparing (a) and (b) shows that the true location is significantly richer in contrast (i.e. better identified) with the hyperbolic constraints. It is interesting to note that -looking at the shading of the L1 map- in vicinity of the minima clearly several of the hyperbolic curves can be identified. Yet, one of their advantages, to directly spot the source of discrepancy (i.e. outlier), is not possible in the residua map.

After understanding how the constraint based approach adds robustness to the location, we will investigate how (i.e. to what spatial regions) the individual constraints actually constrain the hypocenter location.

II.1.2 **Constraining of Location**

When few data are given, the location may at times be determined based on few or even single constraints, only. In the latter case, consistency of information can obviously not be evaluated to gain robustness. In such a case it is vitally important to regard uncertainties in the input parameters to obtain useful solutions. The formal and quantitative derivation will follow in chapter II.3, and for now we will settle with the fact that uncertainties add a certain “thickness” to the mathematical surfaces that represent the constraint. This single constraint scenario allows us to study how the constraint, being parameterized in the time domain, actually constrains the solution in space. For a single constraint within the four dimensional location problem, the equation-system can be described as undetermined polynomial system, which if consistent, i.e. if it has a solution, usually has infinitely many solutions. With a four dimensional problem and one given constraint the solution is determined to be three dimensional. The found spatial solution consists of all points that hold the same time difference defined by the two-phase group. Although the solutions are degenerate for a single constraint, they nonetheless give us information about the location which in combination with a-priori knowledge may allow us to successfully identify events. A typical example is an aftershock sequence, where weak aftershocks might only be registered at few or single stations: A back-azimuth similar to the one of the main-shock indicates that this event may indeed be an aftershock. The same holds for a S-P constraint, defining a fixed hypocentral distance, which if consistent with the main-shock, also provides this indication. Both constraints in combination at a single 3comp. station provide a well defined location region that can link the event to the main-shock source region, allowing a trustworthy interpretation. In the following three sub-chapters we will study how individual S-P, P-P and BA constraints define the possible hypocenter region, before we combine them in chapter II.1.3.

II.1.2.1 **Single Station S-P (Sphere) Constraint**

S-P constraints are anchored to one spatial coordinate (the station's location). The time difference between the arrivals (of the two phases traveling with different velocities) increases monotonically with distance. For homogeneous models, the constraint consequently appears as spherical surface,

on which the hypocenter is bound to lie on. For surface stations the solution can naturally only be located beneath and appears for that reason as half sphere (fig. II.1.2.1:1).

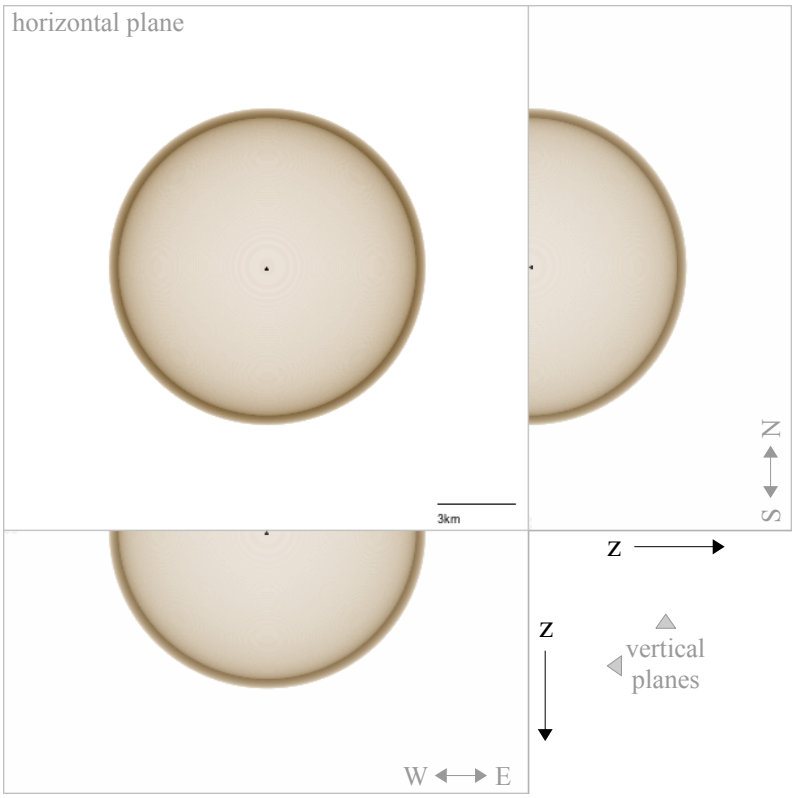


Fig. II.1.2.1:1: Top and side views of all S-P hypocenter solutions for a surface station (black point in the center). The set of solutions forms a half sphere. The top left panel displays the half sphere from top, the bottom and right panels show it from two sides (East-West and North-South direction). Because the body is displayed as transparent object, the color intensity (darkness) scales with the amount of solutions in the direction of sight. The plot may also be interpreted as a stack of all solutions in the direction of sight. While the horizontal panel displays the lateral constraint geometry, the side panels reveal the constraint over depth.

The location of an event based on several S-P constraints hence boils down to the search of the intersections of spherical shells. The classical *method of circles* treated the focal depth as negligible compared to the epicentral distances, and reduced the spheres to two-dimensional circles with radii corresponding to the S-P interval. But since these radii pertain to spheres ("the circle's radius varies over depth"), the classical *method of circles* was bound to errors whenever the focal depth was not negligible, e.g. for local events.

The single S-P constraint is limited to a certain region. Its radius provides a lower bound for the focal depth and an upper bound for the epicentral distance. When multiple S-P constraints are given,

the lower bound for the focal depth is given by the radius of the smallest sphere constructed in the ensemble of S-P constraints, as the intersection point of all spheres can only lie in the space above. The S-P constraint also fixes the origin time: all locations on the S-P-sphere hold the same origin time. For non-homogenous velocity models, these spheres will show deformations causing a complex relationship between radius and depth. Figure II.1.2.1:2 shows examples for surface and sub-surface stations in layered velocity models of increasing velocity. The varying velocities and the inverse effect of refraction cause a complex surface that yields ambiguities in the inversion process: Equal epicentral distances (“circles” of same radius) are found in several depths. This can cause multiple, ambiguous, equally possible earthquake locations that require additional data to discriminate wrong solutions, reducing the uncertainty.

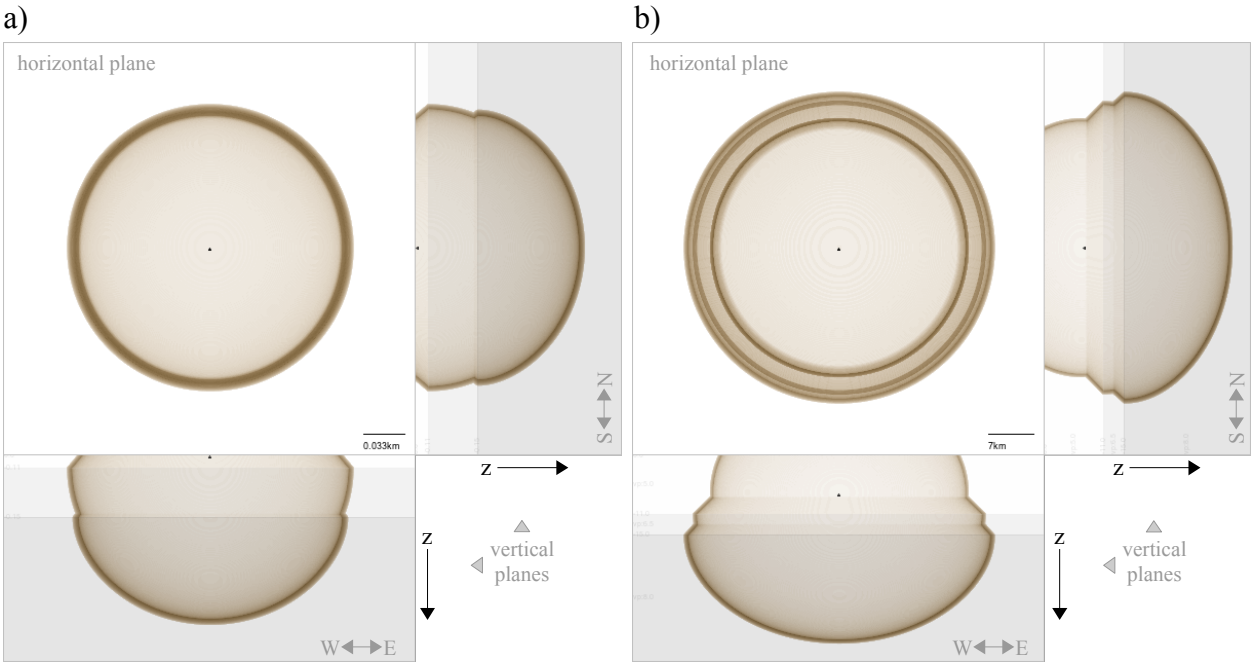


Fig. II.1.2.1:2: Examples of S-P half spheres in 1D-models of monotonically increasing velocity: a) surface station b) subsurface station. The station location is marked by the black point.

S-P constraints are of the few that are limited to a certain region and are therefore of special importance. On the downside, they hold a higher location uncertainty due to the lower picking accuracy of the S-phase. In the following we will look at the unlimited but more accurate P-P constraints which are of special importance for real-time location as they are derived from the first arriving phase.

II.1.2.2 **Two Station P-P (Hyperboloid) Constraint**

For S-P spheres, which depend on one spatial reference coordinate only, three dimensional station distributions added little complexity to the problem. P^2-P^1 constraints, depending on two spatial reference points (two stations' locations), show a more complex behavior. The P-P constraint can be thought of being constructed by the intersections of two growing spheres (fig. II.1.2.2:1): Assuming a origin time t_0^*, the time interval between t_0^* and the following P-onset provides time for the seismic signal to travel. Likewise S-P constraints, a fixed travel time can be translated into a spherical constraint around each station. The region that fulfills the demands of both constraints (the intersection of both spheres) resembles a circle. The earlier the trial origin time t_0^*, the larger the two sphere's radii, and the larger the circular region of agreement. The set of circles obtained over all trial origin times constructs a hyperbolic surface. P^2-P^1 constraints therefore generally appear as hyperbolic surfaces (fig. II.1.2.2:1c) on which the hypocenter is bound to lie. This is a fundamental difference to S-P constraints which limit the solution in time and space: P-P constraints in itself are infinite in time and space, i.e the source-receiver distance and origin time intervals are only one-sided (there exists an upper limit for the origin time and a lower limit for the hypocentral distance). Amplitude decay and SNR may be used to set an upper limit for the distance.

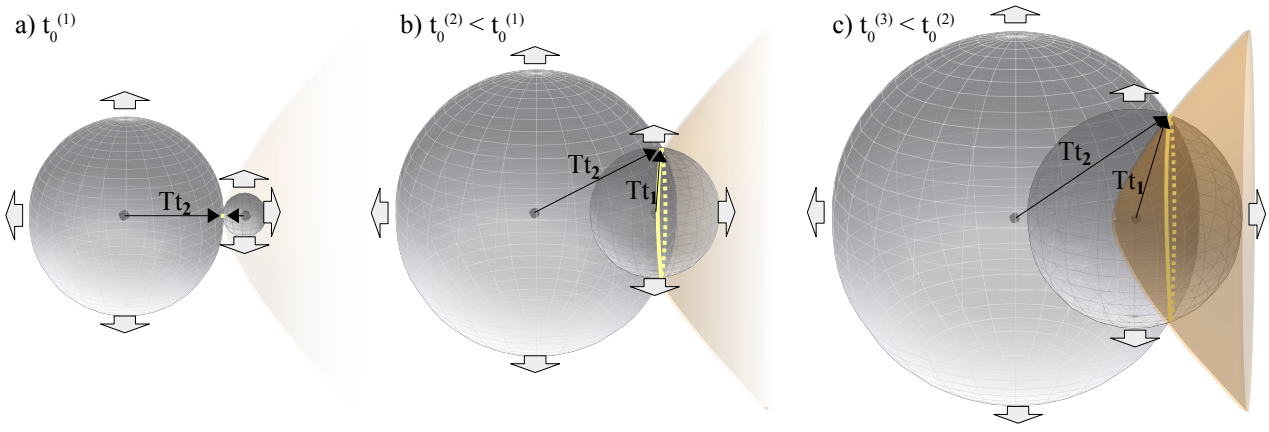


Fig. II.1.2.2:1: The hyperboloid as intersection of trial t_0 spheres. Shifting the trial origin time t_0 to earlier times, the time difference between trial t_0 and observed arrival time increases (a-c). This time difference corresponds to a travel-time (Tt) which similar to S-P constraints constrains the location to be found on a spherical surface, growing in size with decreasing t_0. Taking two stations into account (located at the centers of the spheres), two spheres are formed. b) The possible locations fulfilling both constraints lie on the circle of intersection (yellow). c) Over all trial origin times the set of constructed intersection circles forms the hyperboloid (brown).

The method *of hyperbolas* was historically solely applied for surface stations which yielded a set of hyperbolas in the analyzed horizontal cross section. Today, hyperbolic constraints will have to be analyzed in the context of three dimensional station distributions and events in non-negligible depths as given in the analysis of borehole-stations, networks on strong topography or mine-monitoring. In these cases the hyperboloids will not only be oriented horizontally (like in the classic method) but may be rotated into any space direction.

In interactive analysis and visual inspection, the solution space is often scanned, leaving one search parameter fixed to reduce the problem's complexity. Commonly, the focal depth is fixed to an expected value and the corresponding epicenter is evaluated. Based on the results the focal depth is then altered and the process repeated until a best fit solution is found. This corresponds to analyzing the three-dimensional solution space in horizontal cross-sections. With the hyperbolic constraints being able to be oriented in any space direction (fig. II.1.2.2:2 a-c), horizontal cross section will no longer only lead to hyperbolic curves, but also to pseudo-ellipses and circles. This, on one hand, renders the interpretation challenging, as P-P constraints may appear in the shape of S-P constraints. It does, on the other hand, hold valuable practical implications, since pseudo-ellipses and circles – contrary to the hyperbola – are finite: Vertical near-surface station layouts consisting of only two stations would allow to constrain the hypocenter to a limited region: The arrival time difference between the two stations constructs a vertically oriented hyperboloid, growing in epicentral extent with depth (flipped version of fig. II.1.2.2:2c). It reaches its maximum extent at the lowest seismogenic depth which bounds the hyperboloid (as no deeper solutions are possible). The deeper the event, the narrower the hyperboloid and the smaller the maximum possible epicentral distance. Early warning networks using this particular layout (e.g. in Japan) could therefore use this explicit and latency independent region as first real-time location estimate, avoiding the otherwise often applied common – but error-prone – practice of fixing it to the location of the first triggered station. Combined with back-azimuth information, this would allow a location based on a single site that is similarly precise as the intersection of S-P sphere and back-azimuth beam. Yet, solely working with P-phases, it avoids the time-loss waiting for the S-phase arrival and the complications of its real-time detection. Future early warning systems would optimally consist of three-dimensional arrays.

With being finite, pseudo-ellipses and circles only hold a limited band of origin time values (fig. II.1.2.2:2). While the interval for hyperbolas is unrestricted regarding earlier origin times $\left]-\infty, t_0^{cmax}\right]$ (fig. II.1.2.2:2a), it is limited for the pseudo-elliptical cuts $[t_0^{min}, t_0^{cmax}]$ which occur with tilted hyperboloids (fig. II.1.2.2:2b). t_0^{cmax} denotes the cut's maximum value of origin time,

being equal to- or earlier than the latest physically possible origin time t_0^{max} at the tip of the hyperboloid, depending on the cut position. For borehole analysis with two vertically aligned stations, the hyperboloid will stand upright and the horizontal cross section will display a circle (fig. II.1.2.2:2c). In this case, all possible hypocenter locations even hold the same origin time t_0^c.

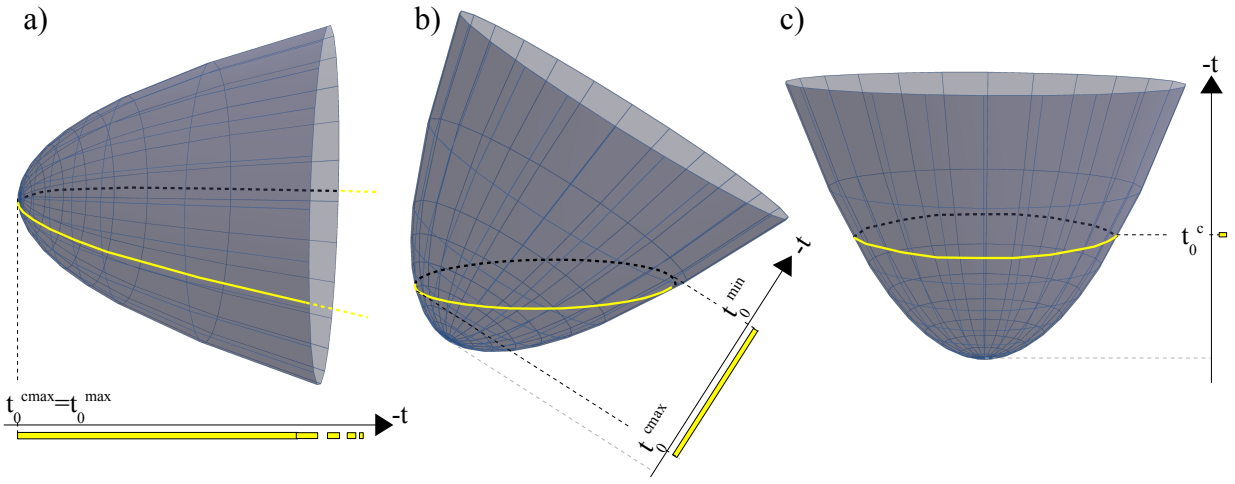


Fig. II.1.2.2:2: Horizontal cuts of freely orientated hyperboloids. As the origin time decreases on the hyperboloid with distance from it's apex, the different intersection curves occupy different intervals. a) Hyperbola: unlimited interval, b) Ellipse : limited interval c) Circle: single origin time value

More complex velocity models lead to deformations in the hyperboloid body and cause a more complex distribution of possible hypocenter locations. Where refracted phases caused discontinuities in classical hyperbolic location with planar station distributions (fig. II.1.2.2:4a), they may even cause multiple distinct curve runs for three dimensional station distributions.

The variation of appearance of P-P hyperboloids for different three-dimensional station distributions and velocity models shall be portrayed in fig. II.1.2.2:3 to II.1.2.2:8. In the first example we will look at the classic case of two surface stations over a homogeneous half space. With this layout the earthquake location is constrained to lie on a horizontally oriented hyperboloid (fig. II.1.2.2:3). The stations are aligned on the East-West axis, wherefore Vertical North-South cross-sections yield spherical intersection curves. Exchanging the homogeneous model for a 1D model of multiple layers of increasing velocity causes refractions in the wave field which lead to discontinuities in the hyperboloid's body (fig. II.1.2.2:4).

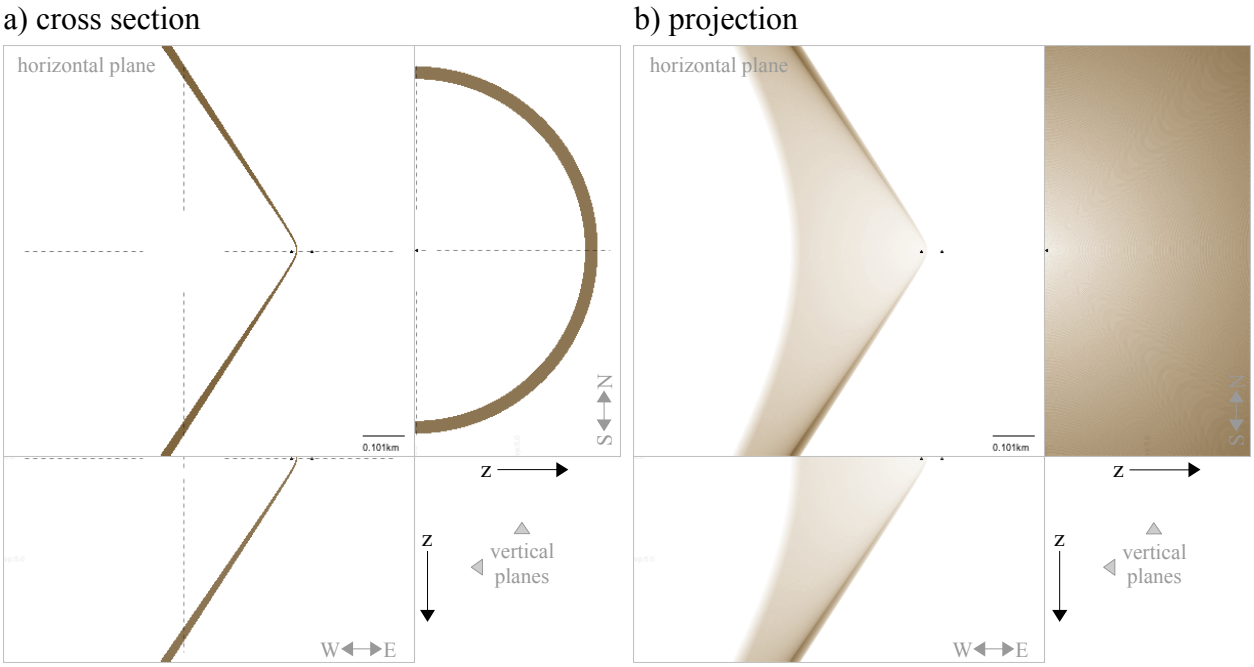


Fig. II.1.2.2:3: a) cross-sections and b) projections in three space dimensions of a hyperboloid formed by the arrival time difference between two surface stations (black points) applying a homogeneous velocity model. In the cross-section view (a) the space is intersected by three slices (horizontal, E-W and N-S), showing the constrained regions. The position of a slice is indicated by the dashed lines in the other two respective slices. The projection-view (b) provides an X-ray like view on (or more accurate "through") the constraint and space from three different sides, conveying a volumetric visualization of the constraint's body. The cross-sections (a) on the other side provide to view the exact constrained region intersected at a given coordinate, e.g. the top panel here showing the constrained region on surface.

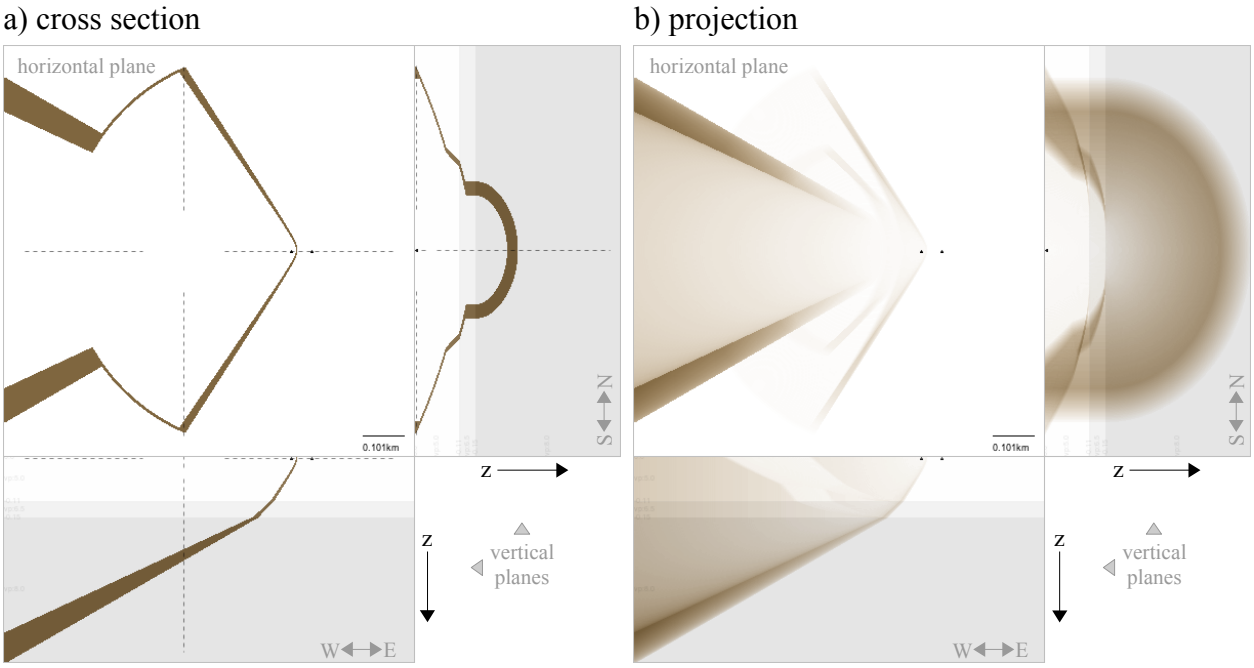


Fig. II.1.2.2:4: a) cross-sections and b) projections in three space dimensions of a hyperboloid formed by the arrival time difference between two surface stations (black points) applying a layered model. Refractions cause sections of different opening angle corresponding to the slope of the different travel-time-curve segments. The "thickening" of the constrained region with distance is an inherent location characteristic, showing that any uncertainty in parameters will affect the spatial solution the more, the further we distance ourselves from the recording stations.

For a minutely larger measured onset time difference, the infinite hyperboloid collapses, limiting the possible location to a small volume (fig. II.1.2.2:5).

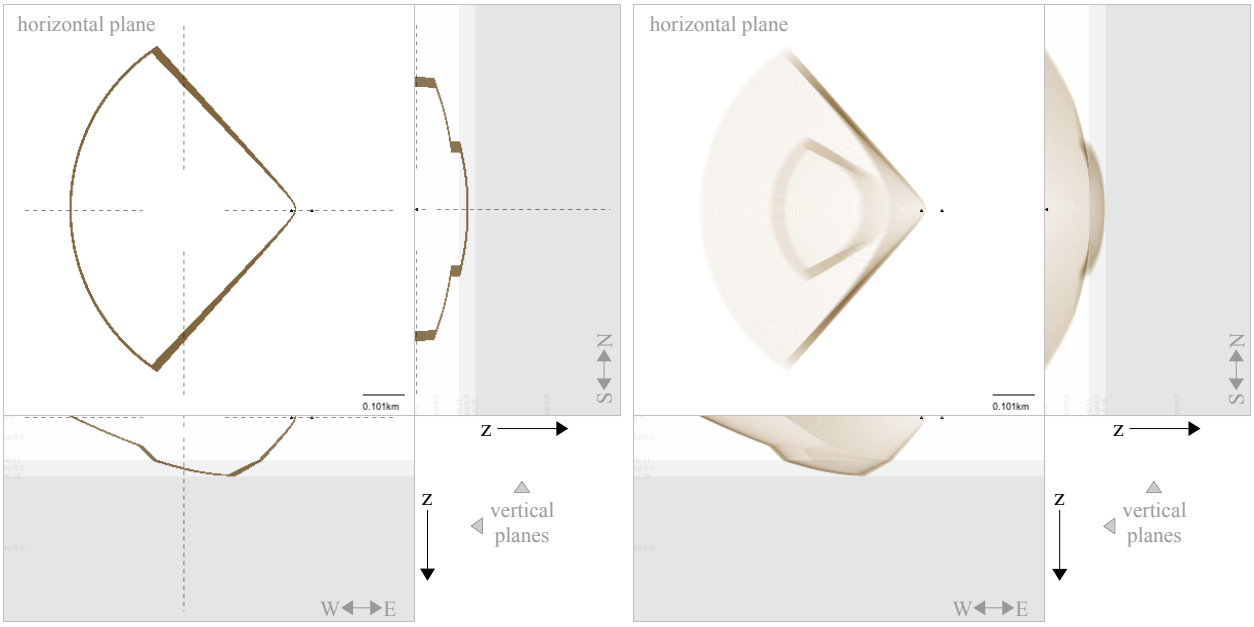


Fig. II.1.2.2:5: Example of a hyperboloid formed by the arrival time difference between two surface stations (black circles) applying a layered model. The inverse effect of total reflection limits the usually infinite body to a small region.

This example demonstrates the complexity of the location problem given already by a single hyperbolic constraint and shows the impact that small parameter changes may have on the overall solution. A small change in the measurement of arrival times makes the difference between the compact, limited hyperboloid segment (II.1.2.2:5) and one of infinite extent (II.1.2.2:4). It is therefore crucial to regard uncertainties in the input parameters properly to interpret the location and its precision accurately. This example also emphasizes the importance of the ability to illuminate the solution space and its uncertainties when information is sparse or questionable, to weigh the given information properly in the overall location procedure. Had, for instance, other constraints been given that pointed out a solution further west, the analyst (or algorithm) would have pulled the best fit solution eastwards from there, averaging between the solution cluster in the west and the limited hyperbolic constraint in the east. Yet, had the example of fig. II.1.2.2:5 been analyzed including (even small) model uncertainties, the hyperbolic constraint would have been expanded, similar to the one of fig. II.1.2.2:4; The analyst wouldn't have mistakenly overrated the constraint, pulling the location eastward, as he would have realized that a small model variation is sufficient to bring all constraints into agreement at the western cluster.

After looking at this simple example for a standard network layout of horizontal stations, let us now look at another complication we will face when dealing with three-dimensional station distributions, as in mine monitoring. Let us therefore start with the simple model of a homogeneous half space, but shift one of the two stations subsurface. This causes the hyperboloids to become tilted. Consequently, the solutions for a assumed hypocentral depth (a horizontal cross-section of the solution space) will no longer appear as unlimited hyperbolic curves but as closed pseudo-elliptical constraints (fig. II.1.2.2:6).

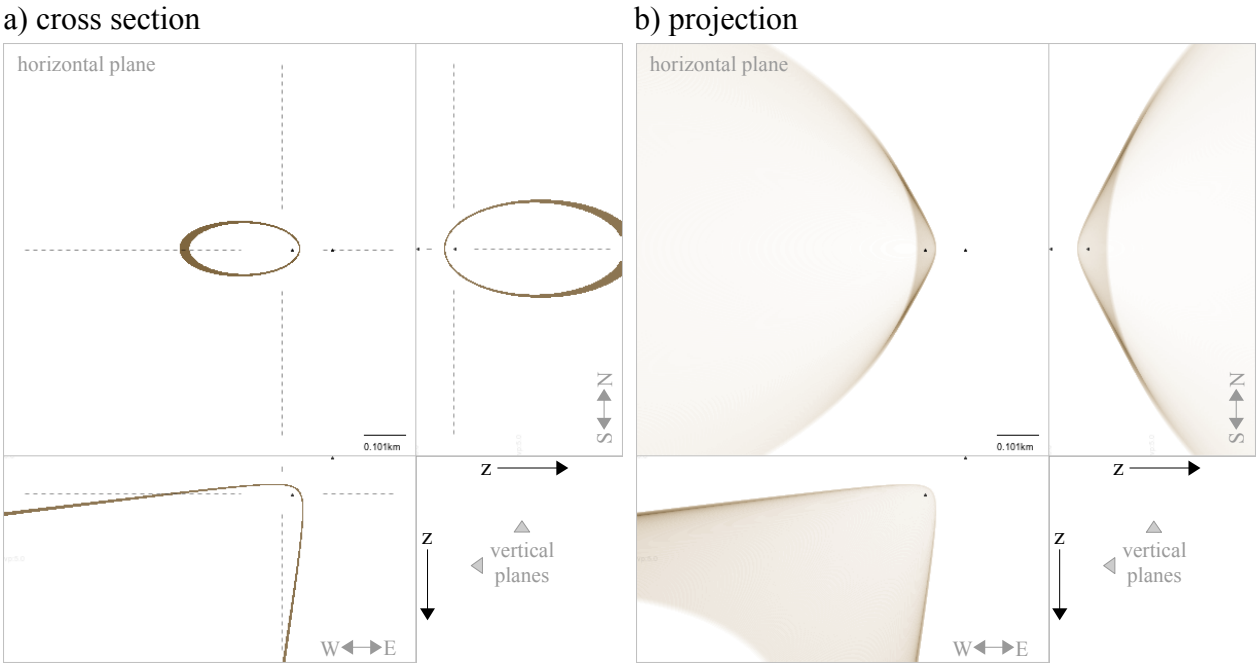


Fig. II.1.2.2:6: a) cross-sections and b) projections in three space dimensions of a hyperboloid formed by the arrival time difference between two stations (black points) lying in different depths, applying a homogeneous model. The cross-sections show pseudo-elliptical intersection curves under a certain depth.

Such pseudo-ellipses can also be observed with surface stations, given a significant topography (e.g. mountain ranges as given in North California or Graben systems as in the Dead Sea plain of the Rift Valley). When several constraints are given, these closed curves cause multiple intersection points (regions of constraint agreement), yielding ambiguous epicenter solutions. Where three surface stations (constructing three hyperbolas) usually yield one solution for a given depth, three stations with one being located subsurface (constructing one hyperbola and two pseudo-ellipses) usually yield two solutions. This ambiguity appears already for an inversion with fixed focal depth. Such a limited inversion is not uncommon, as automatic location schemes for e.g. catalog generation initially fix the focal depth to an average depth, being based on the catalog. Due to the high

uncertainty in depth resolution, these initial estimates often remain unchanged. (Catalogs therefore tend to show a large amount of events on this same depth, corresponding to the initial estimate. This creates a positive-feed-back for the depth estimate of future events, resulting in a biased dataset.)
While these tilted hyperboloids can be a source of ambiguity on the one hand, they can limit the hypocenter region on the other hand with very little data: In regions where the seismogenic zone is known not to exceed a certain depth (e.g. 10km, South California), already a single hyperbolic constraint as in fig. II.1.2.2:6 may constrain the possible location to a compact region and origin time window.
Unlike the last example, in practice we rarely deal with homogeneous half-space models. Adding a deeper layer of higher velocity to the model of fig.II.1.2.2:6 causes refractions in the wave field. This now leads to two seemingly independent distinct curves in the horizontal cross-sections for certain focal depths.

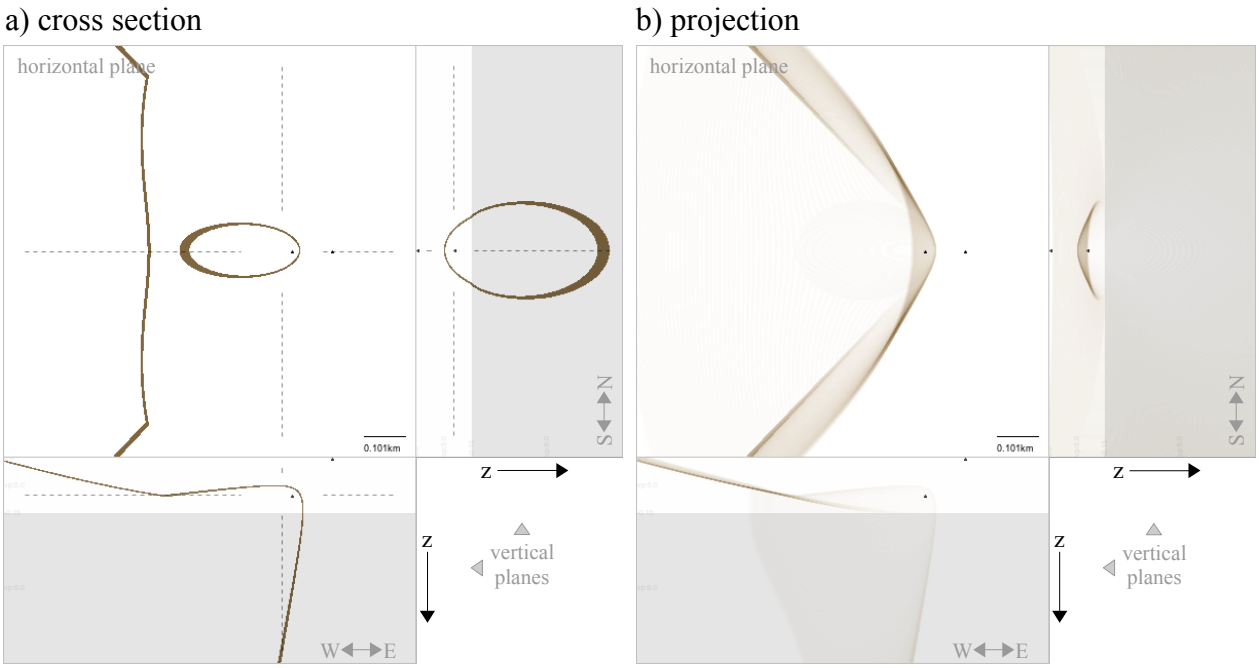


Fig. II.1.2.2:7: a) cross-sections and b) projections in three space dimensions of a hyperboloid formed by the arrival time difference between two stations (black points) lying in different depths, applying a layer-over-half space model. The inverted refractions cause distinct curve runs in the horizontal cross-section in a small band of focal depths (which the horizontal cross-section is showing).

This counter-intuitive appearance of the constraint(s) makes the interpretation challenging as single constraints may yield multiple, seemingly independent curves. For interactive location this effect emphasizes the need for multiple orthogonal cross-sections as they complete the picture, revealing the connection between the appearingly separated curves. However, as these are computationally

expensive and unstable using existing semi-analytical iterative solvers (Eisermann, 2008; Joswig, 2005), the computation has to be approached in new ways. Multiple distinct curve-runs will also arise with vertically-aligned stations as given in a chain of sensors in bore-holes in presence of deeper layers of higher velocity. Fig. II.1.2.2:8 shows such an example. Here the event is detected first by the deeper station, causing the hyperboloid to be pointing upwards, growing in size with depth. Horizontal intersection curves (solutions for a given focal depth) will in this case appear circular. The solution that pertains to the onset caused by the head wave, however, bends the hyperboloid up again, starting at the take-over distance from direct to refracted wave-front. This lets the hyperbolic constraint appear in the shape of a “Sombrero”, touching the surface (fig. II.1.2.2:8).

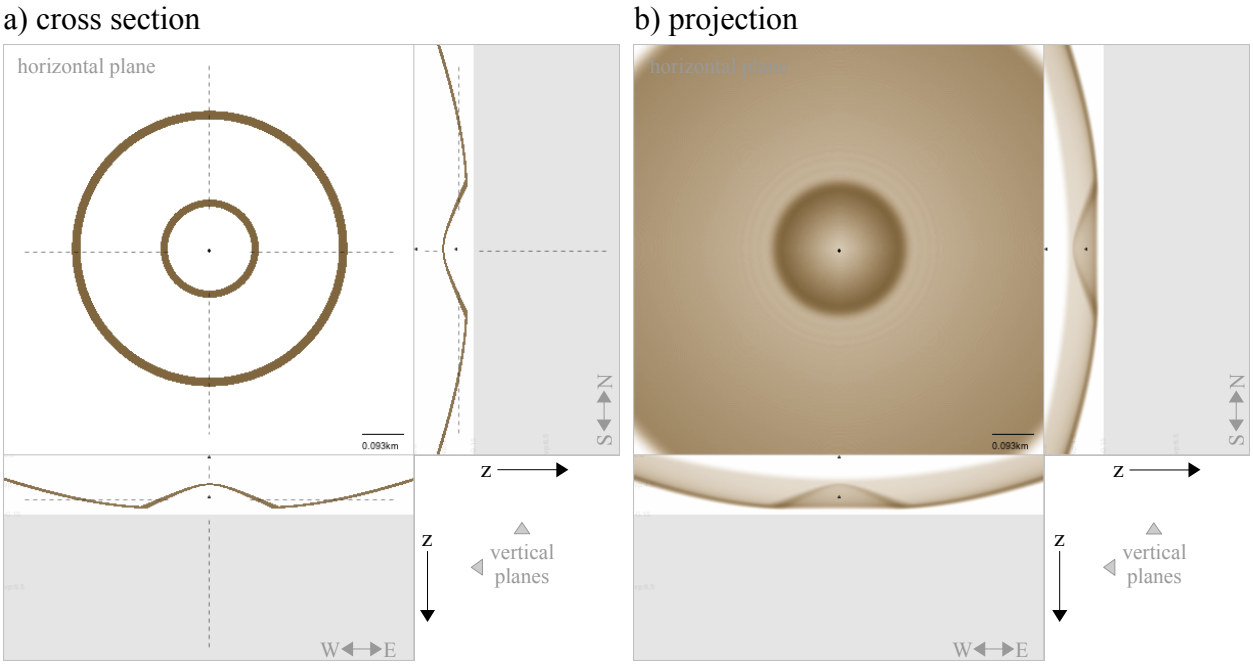


Fig. II.1.2.2:8: a) cross-sections and b) projections in three space dimensions of a hyperboloid formed by the arrival time difference between two stations (black points) being positioned vertically aligned, applying a layer-over-half space model. The horizontal hyperboloid cut resembles circular shape. The inverted refractions cause two distinct curve runs for the horizontal cross-section of a single constraint.

The otherwise infinite hyperboloid is now of limited extent and causes a second distinct circle appearing in the horizontal cross-section for certain focal depths. Each circle corresponds to a single origin time, similar to S-P circles. Yet here, these “circles” are connected - and as the outer shrinks with depth, so the inner grows, until both unite just above the refractor and cease to exist for lower depths. To properly interpret the constraints in the solution space, it is important to be able to distinguish between the constraint types (S-P, P-P, ...), to weigh their impact accordingly. In numerical schemes this knowledge is always given. For visual inspection however, where

constraints used to be distinguishable by shape in classic scenarios, this quality is no longer given (the curves in fig. II.1.2.2:8 might easily falsely be interpreted as S-P "circles"). It hence becomes necessary to apply means as e.g. color-coding for the different constraint-classes. All these examples show that the solution space needs to be visualized in a way that allows to perceive the constraint's three-dimensional extent, to decipher the complexity given in the location problem. A behavior as given in these last two examples is counter-intuitive and makes the search for the hypocenter challenging when several stations are given. This is especially true in the context of ambiguous phase arrivals. Such a case shall be given in this final example: Here, two surface stations are added to the station distribution of fig. (II.1.2.2:8), posing an exactly determined problem (four stations). A grid search generated residual plot, however, reveals two distinct minima for the observed arrival times (II.1.2.2:9a). The importance of comprehensive location schemes (e.g. grid schemes) becomes obvious, as simple iterative schemes would have converged into either one of the two minima, leaving the other solution unevaluated. A residual plot (or scatter plot like in NonLinLoc) reveals both solutions, but does not allow to understand their coming about nor how to interpret them. The representation of the same information using distinct constraints reveals the connection between both solutions in the vertical cross sections.

a) residual plot b) phase encoded constraint plot

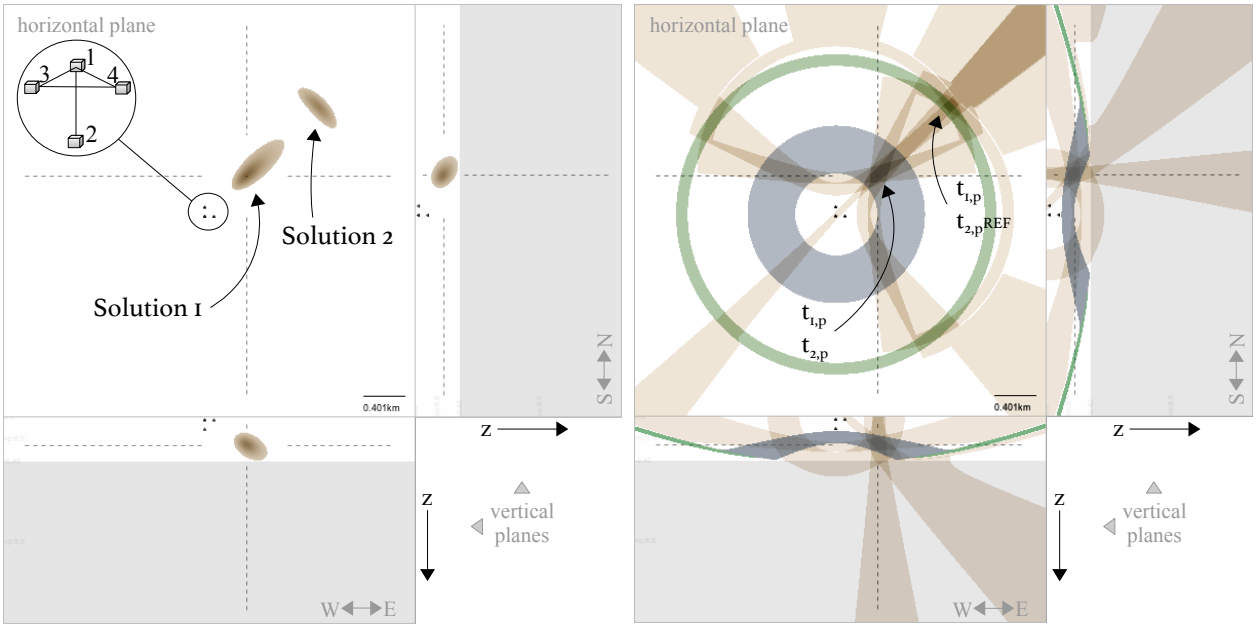


Fig. II.1.2.2:9: Example of three surface stations with a fourth subsurface station. a) The residual plot shows two distinct solutions. b) Using phase encoded distinct constraints displayed on vertical cross-sections (green indicating an involved refracted phase), reveals the connection of both solutions: The P-P constraint between station 1 and 2 (indicated in grey&green) shows that solution (1) (lying in the grey circle) corresponds to the case of $t_{1,p}$ being the onset of a P-phase that travelled on a direct wave path and solution (2) corresponding to the case that the P-phase is of a refracted wave (green circle).

The up-bending ("Sombrero") hyperboloid of fig. (II.1.2.2:8), being formed by the two vertical stations 1 and 2 and the refractor, explains the ambiguity. Encoding it according to the involved phases (fig. II.1.2.2:9b) reveals that this constraint is the result of two different travel paths. The downward branch, in which one solution is found, corresponds to the direct path while the up-bending branch, in which the other solution is found, is constructed by the refracted path of the P-wave. The distant solution (2) would therefore mark the true hypocenter if the first arrival at station 2 was a P-phase of a refracted (head) wave, while solution 1 would mark the right location, if the P-phase is of a direct path. Given this information the analyst may re-evaluate the P-onset at the second station towards indication for the type of onset (direct or refracted) and possibly resolve this ambiguity. This example shows the strength of graphical Jackknifing in resolving data to minimum groups, which allows to trace back the influence of individual phase onsets on the global solution. Although a constraint based location procedure can be well performed automatically in standard situations, for examples like these, the human mind is still indispensable as guiding instance to find the plausible solution.
After analyzing how individual S-P and P-P constraints constrain the location, we will complete this chapter with the analysis of the important array back-azimuth constraints.

II.1.2.3 **Tripartite Array Back Azimuth (Beam) Constraint**

When the inter-station distance is small in relation to the epicentral distance but comparable or larger than the dominant wavelength of the recorded signal, array techniques may be applied to obtain location constraints which are independent of the local seismic velocity. Further they provide slowness estimates and inner-array amplitude ratios, that offer advantages in the event classification. Nanoseismic Monitoring (Joswig, 2008) uses networks of small aperture arrays. Here, a small array consists of a central three component sensor, combined with a small scale tripartite array of vertical sensors with an average aperture of 100m (Joswig, 2008) in its minimal configuration. This forms a four-partite small-aperture array identified as SNS (Seismic Navigation System) (Wust-Bloch & Joswig, 2006) (fig.II.1.2.3:1).

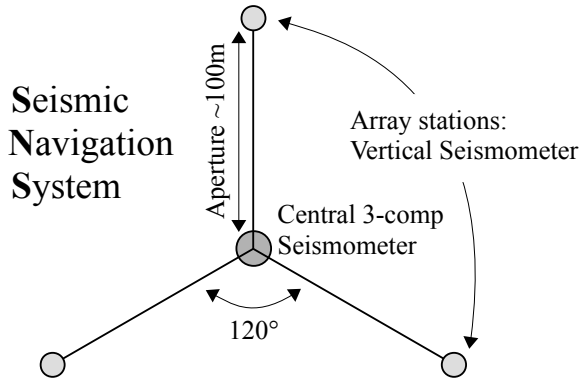


Fig. II.1.2.3:1: Tripartite small array configuration.

For distant sources (for which the plane wave assumption is fulfilled and the stations still lie within the correlation radius of the signals) the sparse seismic array provides the velocity model independent back azimuth and apparent slowness estimates. In dense arrays with multiple stations, the Direction Of Arrival (DOA) and corresponding slowness estimate is usually obtained using an *f-k* analysis (Capon, 1969), which in the time domain corresponds to a two-dimensional grid search over azimuth and slowness values in various filter-bands: Claiming a certain slowness and propagation direction of an incoming plane-wave-front determines delay times given by the time that the signal travels to later stations after passing the first. Shifting the recorded seismograms according to these predicted time delays leads to signals that are in-phase for the correct combination of azimuth and slowness value. The summation of the n signals yields constructive interference that amplifies the signal amplitude by the factor n. This process is called "beam-forming" as it amplifies the signals coming from a certain direction ("beam") defined by back-azimuth and slowness. Integration over the squared amplitudes of the summation trace yields the total energy that is recorded at the array for the given beam. A (2dim.) mapping over ranges of azimuth and slowness values identifies the strongest beam – the horizontal direction and slowness of the most prominent signal in the used frequency band.

As the local noise is uncorrelated, its amplitude is increased by $\sqrt{n}$. The SNR of the array trace is therefore improved by this factor $\sqrt{n}$ (Green et al., 1965). The increased SNR allows to improve the phase identification of weak or masked seismic signals. Due to the significantly lower detection threshold compared to single stations, the total amount of necessary stations and infrastructure can be significantly reduced by using SNS networks (e.g. DeadSeaNet, Wust-Bloch et. al., 2016). The slowness information allows a characterization of the propagation type, e.g. the identification of

acoustic signals or surface waves, since it provides a lower bound for the seismic wave-speed (e.g. a P-phase cannot appear with slownesses significantly larger than the inverse of the local P-velocity. A slowness larger than this value can therefore only pertain to a shear-/surface- or acoustic phase). Small-aperture arrays are therefore powerful instruments for the analysis of non-standard events. Since a delay-and-sum beamforming may not deliver good estimates in presence of high noise conditions, usually frequency domain based approaches are applied (Capon, J., 1969; Kvaerna and Doornbos, 1986). Nonetheless, when Kvaerna (1992) and Kvaerna & Ringdal (1992) evaluated the *f-k* analysis for sparse arrays, they found that the low amount of stations may not sample the wave field sufficiently, resulting in highly ambiguous results. Also, at least one of the traces usually provided a higher SNR than a well-weighed array-trace of a four-partite SNS, which at best could yield a 6dB ($= 20 \log_{10} \sqrt{4}$) improvement (Joswig, 2008). While these approaches may find application for extended small arrays of more stations (e.g. a central 3-comp stations and up to 9 vertical stations on 12-channel data-loggers), they are not suitable for sparse arrays. Such extended arrays with a sufficiently high number of well distributed stations break the symmetry axes and periodicities that lead to ambiguous results with sparse arrays. It is interesting to point out that the f-k analysis computes the energy on the stacked trace (which would correspond to squaring the mean corrected samples in the time domain) and therefore carries methodologically characteristics of RMS based schemes. As concerning location with sparse data in general, we will not apply summation schemes but will analyze the data using the Jackknife framework. Where the beam-forming approach corresponds to a grid search in the time domain in which the seismic traces are shifted according to a predefined set of slowness-vectors (naturally satisfying the plain wave front condition) and stacked, we will perform the analysis on the signal's side, in deriving the time offsets between stations using the cross-correlation of their given seismic signals. One of the benefits, starting from this side, is given in less ambiguous results: In short windows, signals are often dominated by a certain frequency. Shifts, in multiples of the dominant wave-length will lead to constructive interference in the stacked trace and yield side-maxima in the slowness map, which can be of the same order of magnitude as the main maximum. By cross-correlating a window consisting of pre-signal noise and the initial signal, these side-maxima are weakened. The initial signal is further expected to yield the best back-azimuth estimates as it consists of the wave packages that travelled on the fastest path, laterally generally pointing towards the epicenter. Another benefit is, that the array can be generalized to behave like a network for local signals while acting as array for distant events, as no prerequisite of a plane wave assumption is taken up to the cross-correlation

process. Rather than inverting all obtained arrival time differences in one error minimizing scheme, the array will be dissected in Jackknife manner into minimal sub-groups (sub-arrays) that yield exact solutions for back-azimuth and slowness estimate. Afterwards, the set of generated sub-solution will be used to construct the full array's back-azimuth and slowness estimates. The spread of sub-solutions provides for a quality measure, and outlier sub-solutions (e.g. obtained through inaccurate arrival time differences, due to low signal similarity or velocity heterogeneities) are naturally weighed less, improving the overall accuracy. The equation for the back azimuth beam, which computes the direction of the event's incoming signal, is derived assuming incoming plane wave fronts and requires three stations for an exact solution (fig. II.1.2.3:2)

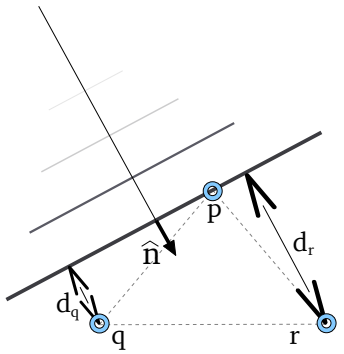


Fig. II.1.2.3:2: Tripartite array and incoming wave front. Plane wave assumption.

For the distances between the wave front arriving at station p, and station q and r applies:

$$\begin{aligned} d_r &= c\cdot(t_r - t_p) = \hat{n}\cdot(\vec{r}-\vec{p}) \\ d_q &= c\cdot(t_q - t_p) = \hat{n}\cdot(\vec{q}-\vec{p}) \end{aligned} \quad , \qquad \text{(II.1.2.3-1)}$$

with $\vec{p}, \vec{q}, \vec{r}$ being the station locations, and t_p, t_q, t_r the corresponding onset times.

From this, the common form of the azimuth vector can be calculated as

$$\vec{n} = \begin{pmatrix} (r_y - p_y)\cdot(t_q - t_p) - (q_y - p_y)\cdot(t_r - t_p) \\ (q_x - p_x)\cdot(t_r - t_p) - (r_x - p_x)\cdot(t_q - t_p) \end{pmatrix} \qquad \text{(II.1.2.3-2)}$$

and the slowness vector $\boldsymbol{S}$ to

$$\boldsymbol{S} = \frac{1}{(q_y - p_y)\cdot(r_x - p_x) - (r_y - p_y)\cdot(q_x - p_x)}\ \vec{n} \qquad \text{(II.1.2.3-3)}$$

This common form is valid when all traces are cross-correlated to one common reference trace. Here holds the equality $(t_p - t_q) + (t_q - t_r) + (t_r - t_p) = \Delta_{qp} + \Delta_{rq} + \Delta_{pr} = 0$. However, when the similarity of signals varies over the array, a station will show a higher correlation factor to a neighboring station rather than to a distant reference trace. More accurate time differences $\Delta_{\{i\}\{j\}}$ can therefore be found by a pair-wise signal cross-correlation of neighboring stations i and j (Cansi, 1995). However, here the time differences are usually not consistent $\Delta_{qp} + \Delta_{rq} + \Delta_{pr} \neq 0$. For such a case we can derive a more robust form from eq. II.1.2.3-1, yielding

$$\hat{n} = (\gamma^2 + 1)^{-1/2} \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \qquad \text{(II.1.2.3-4)}$$

with

$$\gamma = -\frac{r_y\cdot(t_p - t_q) + q_y\cdot(t_r - t_p) + p_y\cdot(t_q - t_r)}{r_x\cdot(t_p - t_q) + q_x\cdot(t_r - t_p) + p_x\cdot(t_q - t_r)}\ . \qquad \text{(II.1.2.3-5)}$$

The slowness vector $\boldsymbol{S}$ is hereby given as

$$\boldsymbol{S} = \frac{\hat{n}}{3}\left(\frac{t_q - t_r}{(\vec{q} - \vec{r})\cdot\hat{n}} + \frac{t_p - t_q}{(\vec{p} - \vec{q})\cdot\hat{n}} + \frac{t_r - t_p}{(\vec{r} - \vec{p})\cdot\hat{n}}\right). \qquad \text{(II.1.2.3-6)}$$

Following the jackknifing methodology of decomposition, computing sub-solutions and recomposition, the array stations are permuted into sub array groups of three stations. For a SNS of four stations this would yield four groups which each would yield one exact sub-solution (fig. II.1.2.3:3). The spread of sub-solutions indicates the solution quality. The most plausible e.g. averaged solution is chosen as the representative global back-azimuth value, while the sub-solution spread serves as parameter uncertainty.

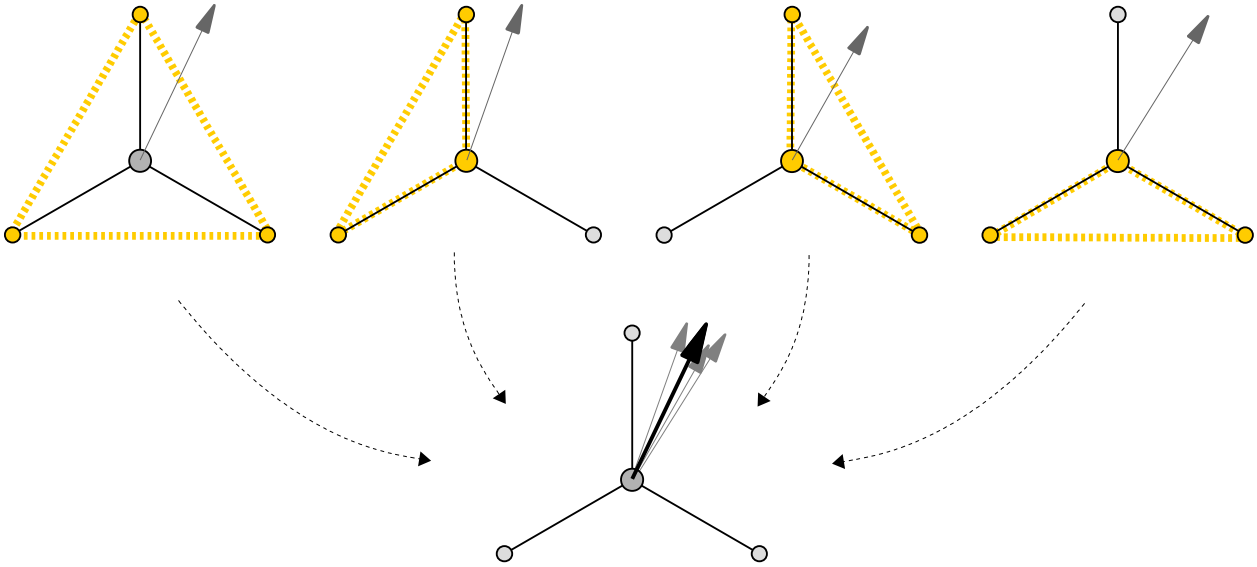


Fig. II.1.2.3:3: Beam forming via jackknifing in a four-partite array. Decomposition of the four stations into sub arrays of three stations, each yielding one solution (gray arrow starting at the center site), and recomposition of the sub solution to obtain the global solution (bold black arrow).

The uncertainty of the back-azimuth is further a function of the correlation factor of the correlated array traces. Consequently, it is implicitly also impacted by the sampling rate, especially for steep incidence angles where the energy arrives at all traces with minimal time delay (although subsample precision can be achieved with frequency based correlation methods). These uncertainties (see chapter II.3.6) cause the back-azimuth direction to fan-out. As the back-azimuth is also independent of depth, its spatial location constraint resembles a (wedge-shaped) disc segment (fig. II.1.2.3:4).

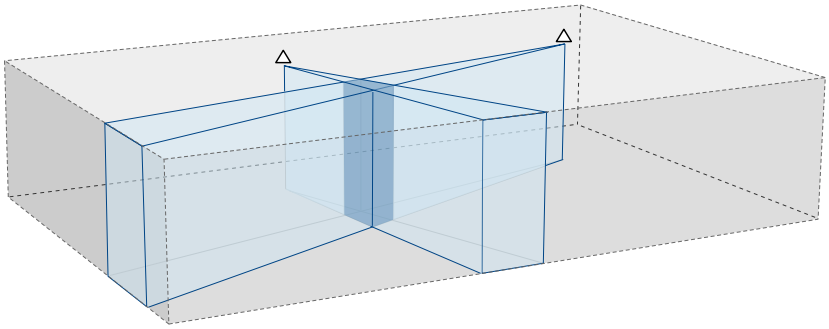


Fig. II.1.2.3:4: Two surface arrays with back-azimuth information. Due to a uncertainty in the azimuth estimate, the constraint fans out, forming a beam. Since the back-azimuth is depth independent, its shape is resembled by a disc segment. The intersection region (of mutual location agreement) of both back-azimuth constraints (highlighted) defines the epicenter, but yields no depth constraints.

As the velocity structure is commonly not well known, it is often approximated by 1D-layered models. Since back azimuth information is independent of the quantity of this (laterally homogeneous) velocity, it is not being affected by velocity assumptions which affect all other constraints. It therefore provides a robust constraint, when the structure can be assumed to be laterally homogeneous, but the velocity is not well known (i.e. near-surface events in sedimentary beds). Generally, the back-azimuth beam could be constructed -starting at the station in the direction of the measured back-azimuth- by following the steepest gradient of the wave field that a point source, collocated with the station, would generate. For a laterally homogeneous model the gradient shows a radial pattern and consequently the back-azimuth information appears as a beam, increasing its width linearly with distance. This is true until the beam enters a region of lateral inhomogeneity (e.g. a dipping layer boundary), where it will be “refracted” out of the initial direction. Its shape loses its linear form and its computation requires gradient back-tracing algorithms. Although this “redirection” of the beam can computationally be accounted for, the detailed information about existing velocity inhomogeneities are usually not given. In the work presented, the back-azimuth beam will therefore only be applied within the context of laterally homogeneous models, and hence be computed as linear beam.

Back-azimuth information can be obtained at a single point derived from the amplitude ratios of 3-component seismometer data (Chapter IV.1). The use of arrays, however, often provides more robust estimates as they “sample” the local wave-front, deriving the back-azimuth from the more stable arrival time differences. But to analyze the local wave-front sweeping over the array, it requires an aperture large enough to capture the signal's dominant wave-length (*e.g. the slow rise time in signals of large magnitude teleseismic events may cause a less precise cross-correlation, which causes higher relative errors for smaller apertures and arrival time differences*), yet small enough to guaranty constant wave velocity within the array and wave-form similarity between the array traces to support signal cross-correlation. Small arrays with multiple stations in varying apertures fulfill both criteria for a broad band of signals. At the same time, the chosen aperture requires a minimum epicentral distance, since the analysis requires a planar wave front to interpret the measured time offsets. Chapter III.1.1 will evaluate the methodological error for nearby sources for which this assumption is violated. This allows to define distance-dependent uncertainty measures to account for this effect. As particle-motion back-azimuth analysis is not bound to any aperture-related restriction, it can be used as complementary approach to obtain back-azimuth information for larger magnitude events which bear lower dominant frequencies and rise-times in

the onset. Here again the configuration of the small-aperture array with a central three-component sensor is favorable. Back azimuth constraints are typically used in conjunction with S-P constraints to yield robust locations by combining distance and direction information. A real-time application using Not-Yet-Arrived S-phases will be laid out in chapter IV.2.

II.1.3 **Combining Constraints**

An individual constraint, whether BA or any 2-phase group like P-P and S-P, constrains the location to a certain sub-set of the full space. According to the constraint's uncertainty, this region may be larger or smaller, but is in most cases not completely bounded. Among the three dominant groups, only S-P constraints will always result in a limited region. The location can only be constrained further by additional data, yielding additional constraints: As a solution should fulfill the demands of all constraints (i.e. in the time-domain yield the observed arrival-time differences, in the space-domain lie in constrained space), the possible solution interval is reduced when constraints (holding different parameter values) are added. If all parameters were error-free (i.e. exact onsets, a perfectly matching model) constraints were thin surfaces and the location would be found in the one intersection point of all constraints, the point for which all information is consistent and constraints agree. Naturally, all parameters are bound to errors. These would cause the curves to scatter, an exact intersection would not exist and a best fit hypocenter would need to be chosen instead, based on the region of highest local constraint density. However, if the uncertainties in the input parameters can properly be estimated and mapped to space, constraints will gain a certain thickness and the solution will be bound to the region in which the constraints overlap (fig. II.1.3:1). This overlap region can be interpreted as the location uncertainty, being the set of all possible hypocenters agreeing with the data and its uncertainties. However, since a region corresponding to a 100% consistency of information is not robust to outliers or insufficiently sized uncertainties, a more robust bounding region, defined by a 70% agreement of constraints, will represent the location uncertainty (detailed in chapter II.3.2).

Each sub-group class (i.e. P-P, S-P or BA) may be used individually to constrain the earthquake location (although BA information only epicentrally). However, a joint analysis that includes multiple classes, evaluates the consistency of independent information and hence provides more robust locations. A simple example is found in the problem of time drift in data-loggers, that occurs

when e.g. GPS synchronization fails. When one station's data is un-synchronized, P-P constraints and their intersections will scatter. However, since the constraint classes are based on different amounts of stations (e.g. S-P on one station, P-P on two stations,...), time drift errors will only cause inconsistencies in the ensemble of P-P constraints, while leaving S-P constraints unaffected. Likewise particle-motion-based back-azimuth relying on one station, as also array-based back-azimuth with all array traces having the same time base (being recorded on one data-logger) will be unaffected by these. Even for data with totally unsynchronized time stamps over the network, these constraints allow proper location (as each is based on data running on the same clock).

While for cases where the phase type association is clear and information is consistent, automatic algorithms can be applied to scan the solution space for the most plausible solution, manual analysis may become necessary when information is significantly inconsistent (e.g. a large spread between the solutions obtained by the individual classes or several outliers within one class). For an interactive location, the solution space needs to be represented in a way that maximizes insight for the analyst. To analyze multi-dimensional problems, the dimensionality is usually reduced by splitting off one or more dimensions. Hereby, a parameter (typically the hypocentral depth) is kept fixed during the analysis and the analysis is subsequently repeated with different quantities of the parameter until a best fit is found. For a fixed hypocentral depth, this would result in a 2(-space)-dimensional evaluation of the location constraints. To visualize the three-dimensional space, a common approach lies in displaying intersecting slices, each keeping one spatial coordinate fixed. Following this approach, for planar surface networks the horizontal slices reveal Milne's well known S-P “circles” or P^2-P^1 “hyperbolas” (fig. II.1.3:1, A-C). Vertical slices display the plausible hypocentral interval.

Fig. II.1.3:1 displays a location scenario given several constraints, using the constraint classes individually (a-c) and combined (d). While BA constraints (c) may provide precise epicenter locations, they do not constrain the focal depth. P-P constraints (a) constrain the hypocenter well for local sources, but yield elongated, even open regions for distant sources, resembling BA beams. The combined use of all constraint classes provides for precise location, robust location boundaries and a quality measure based on the amount of fulfilled constraints (chapter II.3.2).

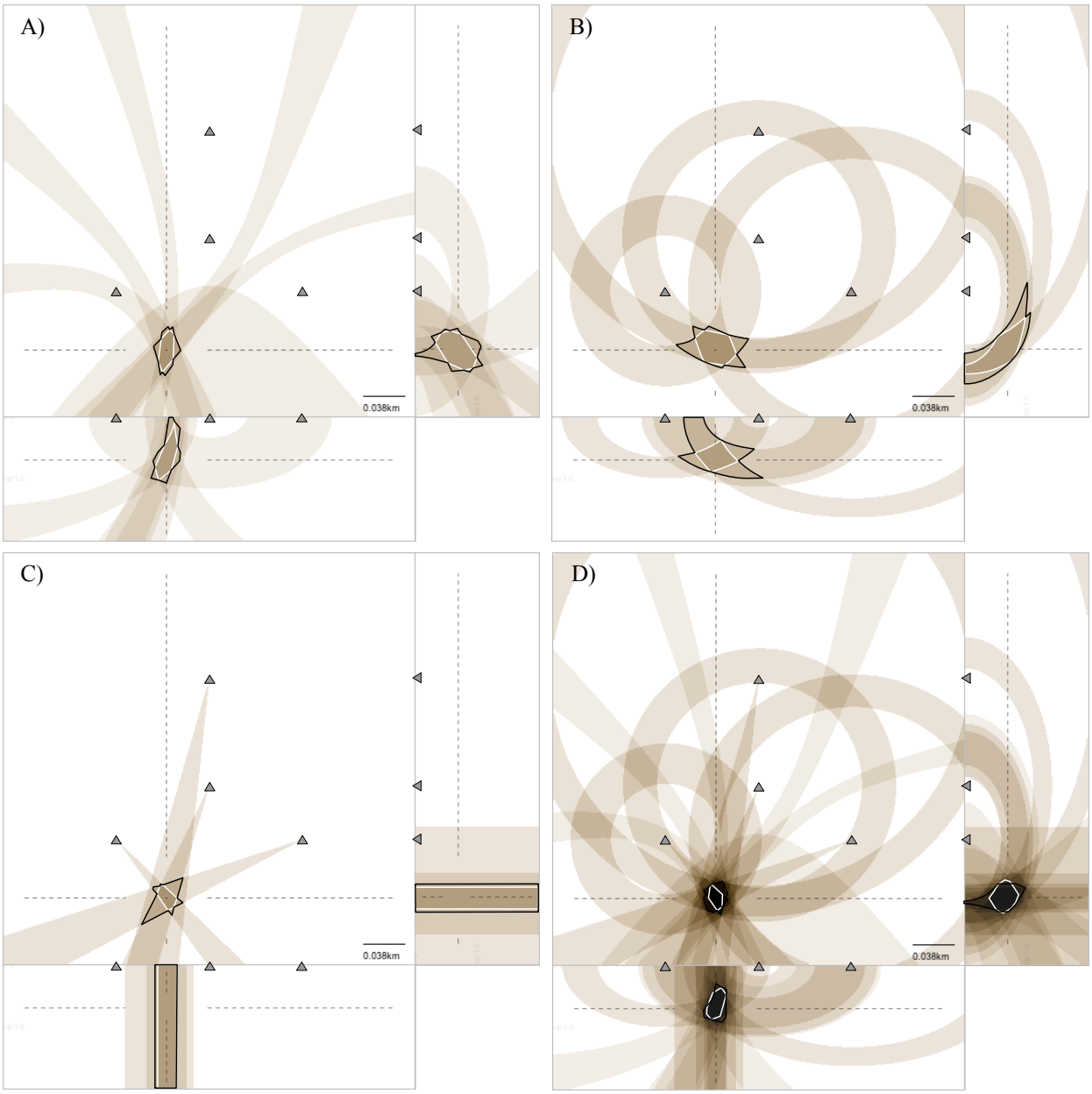


Fig. II.1.3:1: Jackknifing scheme of distinct constraints. A) Hyperbolic location B) S-P trilateration C) Back azimuth beams. A-D) The superposition of the individual constraints allows to identify the hypocenter as region of strongest constraint intersection/overlay (white contour) (i.e. highest constraint agreement), marking the global solution. The hypocenter region is defined as the region that corresponds to a 70% constraint agreement (black contour), yielding robustness in presence of outliers. D) shows the combined use of all constraint classes, which are in agreement in this synthetic example.

Since the constraints are based on arrival time differences, the origin time is not explicitly solved for during the inversion. After the hypocenter region is constrained, the origin time interval is determined by forward calculation of the travel times between hypocenter and stations.

II.2 Real-Time Direct Search Methodology

This chapter outlines the methodological innovations developed in this work. Graphical Jackknifing using location constraints given in S-P-, P^2-P^1 and BA information was developed as interactive method for Nanoseismic Monitoring (Joswig, 2008). It enables the analyst to gain insight into the solution space and instantaneously observe the effect of parameter changes visually to find the most plausible location: This instantaneous response enables the analyst to test ambiguous phases for plausibility and to change model parameters, evaluating the stability of solution. Inconsistencies can be traced back to minimal groups of data, allowing to re-evaluate the corresponding phases and applied models. Using this concept of elementary constraints, outlier resistant location estimates may be obtained.

However, the requirement of a real-time applicability (to provide for a responsive visualization) was a limiting factor for the complexity of the problem description and its analysis. A constraint calculation dependent on live travel time calculations limited the complexity of the velocity model to layered models of monotonic increase in velocity with depth. As the constraints were calculated as exact mathematical bodies resp. intersections thereof (Eisermann, 2008), the semi-analytical algorithms tend to fail in case of the complex intersection curves that may occur for a three dimensional distribution of stations. This complexity also limited the analysis to an evaluation on horizontal cross sections only, not allowing for vertical cross sections revealing the constraint behavior over depth. Detailed picking and model uncertainties had to remain unaccounted for. An idea of the uncertainties would instead be gained through interactive modification of the input-parameters (e.g. phase onsets or the model parameters) with a real-time response, which would reveal their impact on the solution and individual constraints. The solution quality was (qualitatively) indicated by the spread of constraint curves, but could not yet be quantified.

The application of Nanoseismic Monitoring in environments of strong topography and a high local variance in the velocity field raised the need for the treatment of three-dimensional station distributions and heterogeneous (three-dimensional) velocity models. The political importance and concern for correct locations in e.g. CTBT (“Comprehensive Test Ban Treaty”) applications or the interpretation of sources in gas-fields called for the correct treatment of picking and model uncertainties to obtain robust hypocentral regions that would provide reliable location bounds. The same is of high importance for earthquake early warning systems, which derive information as

magnitude and expected arrival times from the early and uncertain estimate to issue public warnings.

To support all given requirements, yet uphold the real time applicability to provide for real-time algorithms as used in Earthquake Early Warning Systems on the one hand and for analyst guided interactive Visualizations on the other, posed the technological challenge of this work. It was solved by reformulating the methodology into a direct, global search scheme (chapter II.4): In a direct search scheme the algorithms for constraint calculation simplify and become extremely robust, allowing to map complex constraint geometries as occur with three-dimensional station distributions or heterogeneous velocity models. The constraint condition based on arrival time differences $[t^B - t^A]$ is in principle evaluated by equalities in the form of

$$[Tt^B_{xyz} - Tt^A_{xyz}] = t^B - t^A \qquad \text{(II.2-1)}$$

at each grid node, where Tt^A_{xyz} denotes the travel time between corresponding station and grid cell $[x, y, z]$ of phase A. As there is no difference in complexity for evaluating this condition on horizontal or vertical cross-sections, the display of the constraint behavior over depth may in this manner directly be obtained. A picking uncertainty (u) (chapter II.3.1) can directly be evaluated using inequalities in form of

$$\left|[Tt^B_{xyz} - Tt^A_{xyz}] - [t^B - t^A]\right| < u\,. \qquad \text{(II.2-2)}$$

As travel times do not change for a fixed station layout and velocity model, they may be pre-cached at each grid point with the need to be recomputed only when velocity model or spatial parameters of the grid are changed (fig. II.2:1). The decoupling of travel time and constraint calculation allows for a real-time analysis using velocity models of any complexity as e.g. inverse layering or three-dimensional heterogeneous models with cavities as needed in e.g. mine-monitoring.
More important than picking uncertainties are the uncertainties in the applied model(s). Velocity model uncertainties will therefore be introduced by using two travel time grids per station, holding the values of the minimal and maximal possible travel times (chapter II.3.5).

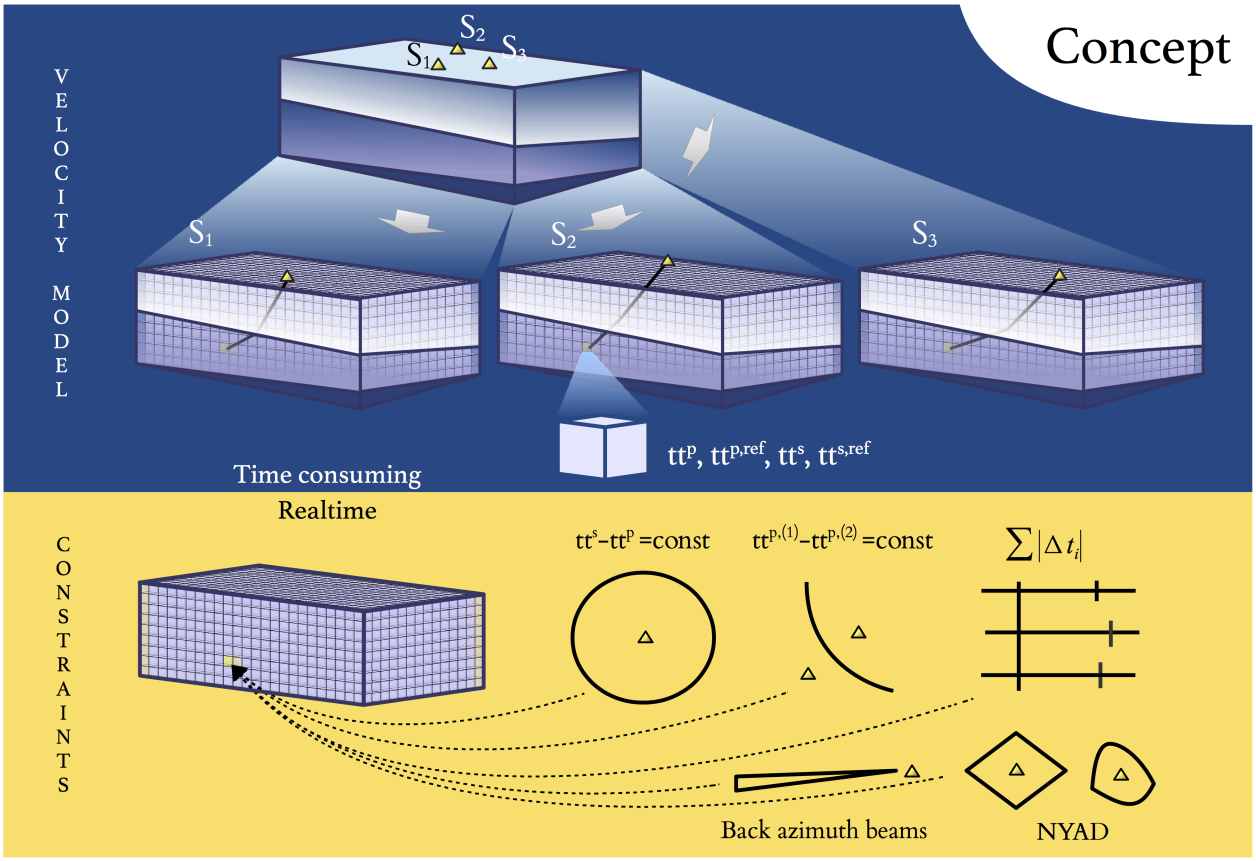


Fig. II.2:1: Conceptual sketch for the location scheme. The inversion is split into a rendering and analysis process. First, all required travel times (e.g. direct and refracted P- and S-phases tt^{p},$tt^{p,ref}$,tt^{s},$tt^{s,ref}$) are computed between all grid points and stations. Each station "sees" its own travel time field. Depending on the velocity model this step can take between fractions of- up to several seconds (Using model uncertainties of 3D Grids the rendering process may take up to minutes). In the analysis, all given information as S-P, P^1-P^2, back azimuth beams, Not-Yet-Arrived-Data (being introduced in chapter) and residual maps can be calculated in real time, being evaluated at all grid cells at given cross sections.

Taking account of both, picking and model uncertainties, constraints will occupy "volumes" (i.e. 3dim. regions) in space rather than being represented by (thin) surfaces. Multiple constraints which due to their errors did not intersect in a common region, will now intersect in congruent regions of mutual constraint agreement due to appropriate uncertainty measures, defining the hypocenter region (chapter II.3.2). Congruent regions, corresponding to a region of 70% constraint agreement, are robust in presence of outliers (e.g. wrongly chosen ambiguous phases), but may grow large in presence of model uncertainties. A probabilistic analysis for non-gaussian error distributions will be introduced (chapter II.3.3) to find the most probable location within these large regions and to quantify the trustworthiness of location.

An optimal visualization of the solution space is important for the interactive feedback system, in which the analyst re-evaluates questionable phase onsets after understanding their behavior in the solution space, evaluating their impact on the global solution. The real-time computation of constraints provides the analyst with the ability to directly evaluate the effect of parameter changes and explore the solution space in the search for the most plausible location. The solution space will for this be visualized using three cross-sections, projections or 3D rendered graphics. To differentiate between the constraint classes, the constraint groups are displayed in different color channels, that superposed add up to white. The intensity/brightness of each channel scales with the amount of overlapping constraints in that class. This supports the psychophysical perception, as regions of higher constraint agreement appear in higher intensity. The likelihood of a solution therefore scales with the perceived brightness. The region of global solution is featured by high brightness and a white color if all constraint groups agree (fig. II.2:2), leading the analyst to the most plausible solution.

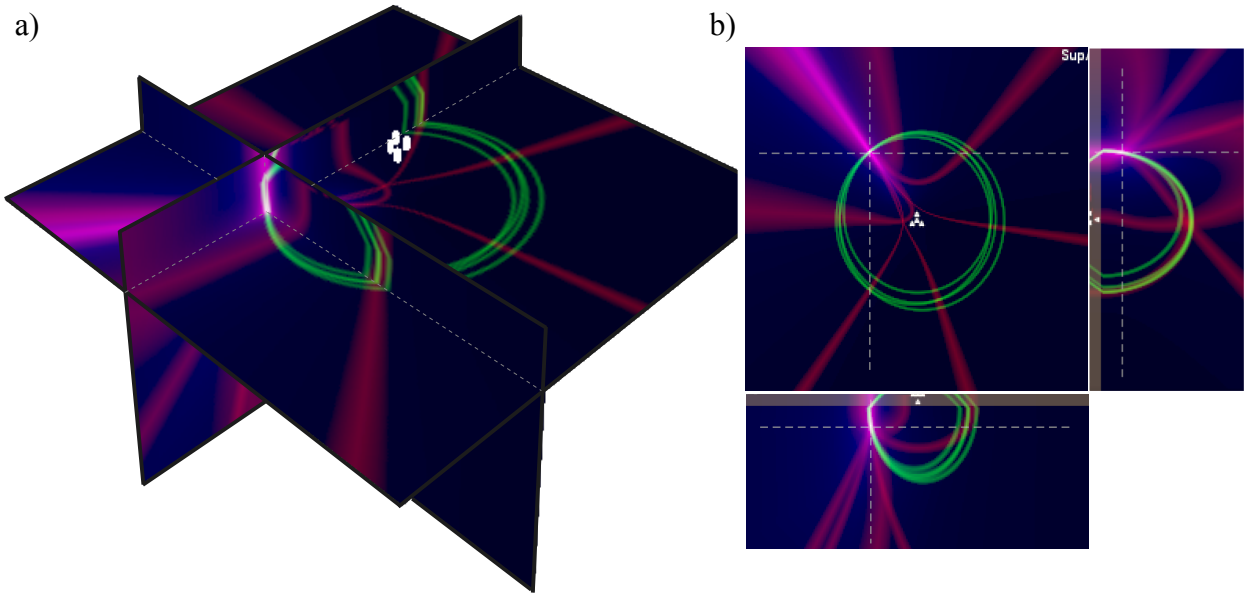


Fig. II.2:2: Display of constraints on cross-sections in a) 3D- and b) planar visualization, consisting of horizontal, E-W and N-S cross-section. Probabilistic representation of P-P (red) and S-P (green) information (see chapter II.3.3). Here, additionally the L2-residua sum is displayed on the blue channel. Colors superimpose to white at points of mutual agreement, while the brightness of a given color generally correlates with the agreement of information within the corresponding constraint class.

When information is sparse, the low number of constraints will appear as distinct objects, each having a high impact on the solution. For a large amount of information resulting in a high number of constraints, the individual constraint will comparably fade weaker. Since the normalized density

of constraints at a given point scales with the point's likelihood to be a solution, the graph will eventually resemble a topological likelihood function (fig. II.2:3b). The distinct visibility of the individual constraint is therefore also a measure of its impact on the global solution.

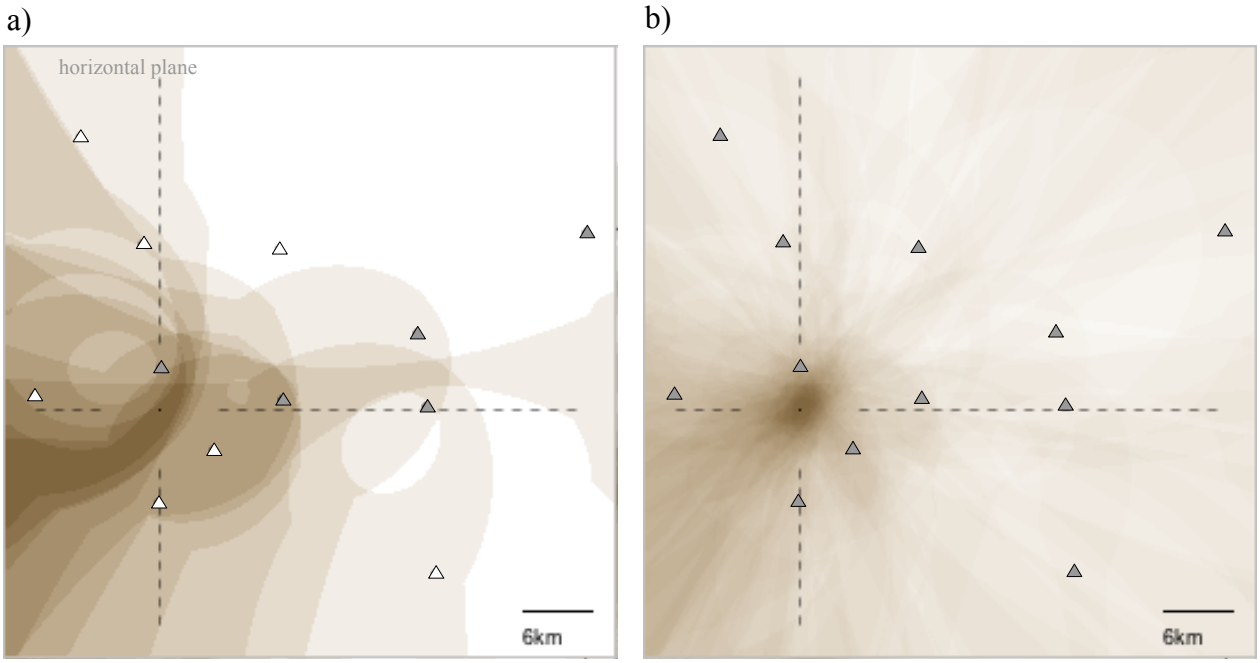


Fig. II.2:3: Superposition of a) 10 P-P constraints generated from the P arrival times of 5 stations (grey triangles) and b) 66 constraints generated from the P arrival times of all 12 stations. While the constraints still appear as individual objects in (a), they are indistinguishable, resembling a likelihood distribution in (b). Example taken from chapter V.1.

Earlier chapters stressed the importance of global and complete solutions, referring to an algorithm's ability to avoid converging into local minima, but to find the global extremum, even revealing the set of best fit locations if multiple ambiguous solutions exist. The requirement for global solutions calls for coverage on two levels. On the large scale, the grid must cover the complete space containing all relevant solutions. Due to secondary solutions outside the network having amplitude ratios and arrival time differences similar to a solution within the network, grids of large spatial extent are needed. Secondly, as grid searches only sample the search domain at certain trial positions, they may overlook sharp solutions that lie in between grid points. On the lower end of the scale the distance between grid points should therefore be very short. Neither of both demands can technically be met, especially in the context of a real-time analysis. To provide truly complete solutions, the grid search will be modified on both ends of the scale. On the lower end, we will integrate constraint probabilities over grid cells rather than evaluating constraint at grid points only (chapter II.4.1). This renders it impossible to miss sharp minima or maxima. On the

upper end of the scale the problem is met by a variable grid resolution (chapter II.4.2). As the spatial extent of constraint uncertainties grows with the distance from corresponding stations, the grid resolution may be reduced with distance without loss of precision. Larger space intervals can therefore be analyzed with equal computational power, providing for truly complete solutions. The following chapters will shed light on all these subjects.

For robust locations, the proper description and mapping of uncertainties plays a key role to cover all associated errors. We will therefore begin with the formulation of constraints in the presence of uncertainties.

II.3 **Constraint Uncertainties**

The less data are given, the stronger is their individual impact on the solution and the stronger their errors will affect the solution. Analyzing sparse data as if it was error-free will produce results showing a spurious precision. Given an exactly determined system (e.g. four surface stations with P-triggers, a scenario common in realtime location), constraints are bound to agree perfectly in at least one location (i.e. in one point). This result of highest precision will most likely, due to inaccuracies in the velocity model and possibly unprecisely picked phase-arrivals, be inaccurate. Especially with the ill-conditioned problem for the hypocentral depth, the focal depth will often be non-representative of the true one. Trusting this spurious precision would lead to wrong interpretations. We face the same problem if we have less data (underdetermined systems), only here the solutions are additionally degenerated. With multiple given constraints in overdetermined systems on the other hand, we would usually notice a discrepancy- as no zone of complete constraint agreement could be found. However, also in this case it would be difficult to define a bounding region for the hypocenter. The key component for accurate results and an estimate of their precision is therefore the estimation of the input parameters' uncertainties and their mapping to the solution space. Doing so will reveal the fact that the solution might be found in a much broader space than the solution, that was based on an error-free assumption, did indicate. In the following chapters we will address the question how these parameter uncertainties are mapped into (uncertainties in) solution space. In the work presented, two types of uncertainties are regarded: those of picking and those of the model. Hereby, the picking uncertainty consists of a) the uncertainty in arrival time estimation and b) uncertainties that are due to phase type determination and ambiguous phase-arrivals. Mapping the uncertainties will have two major effects: 1) Individual constraints will gain a certain “thickness”, in such a way that a forward calculation of the travel times from any location inside the constrained region will always yield arrival times that correspond with the two measured phase-arrivals within their assigned uncertainties. The constraint corresponds therefore to the set of solutions (locations) that all possible perturbations of the input parameters within their uncertainties could have produced. As such, it portrays already during the inversion process the result of a sensitivity analysis that otherwise would only have been performed after inversion. This allows to find trustworthy locations even when only few data is given, e.g. in exactly determined or even underdetermined problems where no redundancy exists to evaluate the

solution quality. 2) With multiple constraints, overlap regions of constraint agreement will be found. Such a region provides for robust boundaries for the hypocenter and signifies the hypocenter uncertainty. To account for outliers, the hypocenter region will be defined as the region in which at least 70% of constraints overlap i.e. agree (chapter II.3.2). The degree of agreement (congruity) serves as measure for the solution quality. This measure allows for robust automatic constraint based location algorithms, e.g. for real-time location. To obtain best-fit estimates, chapter II.3.3 will introduce a probabilistic approach for the analysis of constraints and their uncertainties. The focus of this work lies, however, not in finding a best fit solution, but to find a solution interval that is robust and accurate: Accuracy is weighed higher than a high precision, that may be spurious. As such, the guiding philosophy is to only constrain the solution as much, as the data in its uncertainties truly provides: Initially (applying no constraints), the possible solution interval is given in the full set of points that make up the sub-surface space. From here, only those points will be eliminated that can not be brought into agreement with the data in its uncertainties. Solution regions may therefore be less precise (i.e. larger), but also more robust in truly containing the hypocenter.

It was mentioned before that uncertainties cause the constraint to acquire a certain “thickness”. However, the relationship between the uncertainties in parameter space (time) and those in solution space (space) is not trivial, as fig. II.3:1 reveals. Here a two dimensional example is displayed, in which two stations detect three surface events. For each event a P-P constraint can be formed according to the measured arrival time difference between the two stations. Here, each event location yields a different arrival time difference, but the picking uncertainty is the same for all detections. The constraint's “width” (i.e. the uncertainty mapped into solution space) varies strongly between the three examples, contrary to intuition, as all hold the same parameter uncertainty. For the constraint based on the largest arrival time difference, the “width” is also strongly asymmetric around the “exact” (error-free) constraint. Most significant though is the fact, that the constrained region significantly differs between a constraint incorporating uncertainties vs. one neglecting them.

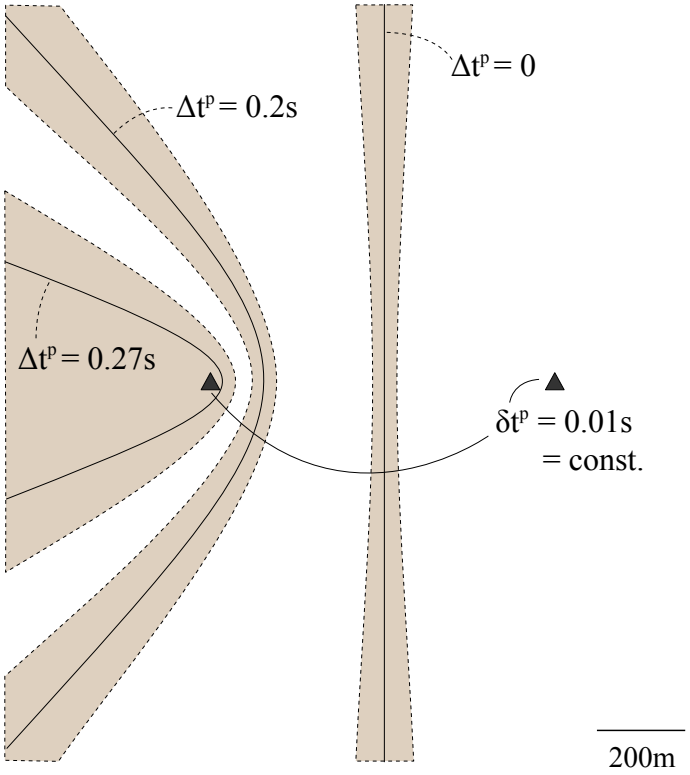


Fig. II.3:1: Three hyperbolic constraints with the same picking uncertainty of δt^p=0.01s yield a very different spacial precision, depending on the arrival time difference Δt^p. Especially in the left example of largest arrival time difference, the epicenter may lie in a much larger region than the exact constraint (solid curve) indicates. Arrival time perturbations alter the location the stronger, the larger the arrival time difference is.

This simple example shows that neglecting uncertainties may cause a fundamental misinterpretation in event location and that their proper consideration is crucial when data are few. The following chapters will establish different types of uncertainties. Beginning with non-Gauss-shaped picking uncertainties, which are important for the evaluation of unclear and ambiguous phases, the velocity model will be assumed to be perfectly known. After introducing a probabilistic formalism, this concept will be expanded to a multi-pick evaluation, formulated for a direct search approach. Finally, velocity model uncertainties will be added, completing the evaluation of uncertainties considered in this work.

II.3.1 **Picking Uncertainties**

The determination of phase onsets, whether manually or automatically picked, is bound to uncertainties. This uncertainty is caused by the fact, that the point at which the signal arises from

the surrounding noise or signal, can only be determined with limited accuracy. For dispersive surface waves the signal's different wave lengths arrive at different times. The apparent onset therefore varies with the chosen frequency band (also due to the filter's phase shift which remains often unaccounted for). Also, the corner frequency of a fracture decreases as the fracture grows with time, causing lower frequencies to arrive later. The data-logger time drift that occurs when GPS time synchronization delays due to reception problems, causes additional uncertainty in the absolute picking time. In the analysis of doubtful data with signals close to noise level, onsets might additionally be masked and multiple possible onset candidates might exist. Here, the incorporation of error and uncertainty analysis is crucial for proper location, as the absence would feign an unrealistic high constraint accuracy and may cause misleading interpretations of results and subsequent mislocation. Especially for large arrival time differences, small deviations of phase onsets have a large impact on the constraint location and "width" (fig. II.3:1) - a counter-intuitive behavior (increasing impact with decreasing relative error) that is linked to the existence of a maximum value (chapter I.2.4) that growing arrival-time differences are approaching. For the formal derivation, errors will be first constructed as symmetrical uncertainty around the arrival time, similar to physical parameters with absolute errors. Within the later concept of probabilistic constraints the uncertainty will then be generalized to a (weighted) arrival time interval, for which only the interval boundaries and weights are used as information parameterizing the constraint. This allows to properly address the fact, that picking errors are usually one-sided (late) and the uncertainties therefore asymmetrical. Such an approach will mark point(s) in space according to their likelihood to correspond with the hypocenter location. For now, however, we will not focus on discriminating the point(s) most likely, but rather illuminate the whole region of physically possible solutions. While this will disregard the level of likelihood varying over this region, it will allow to define a solution region in space, which can be trusted to contain the hypocenter. For this, the phase onset time t will now be expressed as measured physical parameter $\bar{t}$ with associated error δt

$$t = \bar{t} \pm \delta t \,, \tag{II.3.1-1}$$

the error expressing a strict bound for the parameter value. The true value of t will therefore lie between the minimum and maximum value

$$t_{min} = \bar{t} - \delta t \text{ and } t_{max} = \bar{t} + \delta t \tag{II.3.1-2}$$

and the maximum discrepancy between any possible t and $\bar{t}$ is given by δt. As all location constraints rely on more than one onset, the collective error associated with the constraint has to be evaluated. When deriving values from uncertain parameters, the uncertainties propagate to the computed values. The rules of error propagation for fundamental mathematical operations on uncorrelated variables are

$$\delta(x+y) = \delta x + \delta y \tag{II.3.1-3}$$

$$\delta(x-y) = \delta x + \delta y \tag{II.3.1-4}$$

$$\delta(x \cdot y) = \bar{x} \cdot \bar{y} \left(\frac{\delta x}{x} + \frac{\delta y}{y}\right) \tag{II.3.1-5}$$

$$\delta(x/y) = \frac{\bar{x}}{\bar{y}} \left(\frac{\delta x}{x} + \frac{\delta y}{y}\right). \tag{II.3.1-6}$$

Most fundamental location constraints rely on the difference of two onsets, as e.g. P-P-hyperboloids which are based on the difference of the P onsets at two stations:

$$T^{pp} = \Delta t^{p} = t_2^{p} - t_1^{p}, \tag{II.3.1-7}$$

defining the constraint parameter T. According to eq. II.3.1-4, the overall constraint uncertainty is given by the sum of both onset's uncertainties, which can be confirmed evaluating the extreme values of the possible parameter interval. Introducing uncertainties, (II.3.1-7) can be written using eq. II.3.1-1 as

$$\bar{T}^{pp} \pm \delta T^{pp} = \left(\bar{t}_2^{p} \pm \delta t_2^{p}\right) - \left(\bar{t}_1^{p} \pm \delta t_1^{p}\right) \tag{II.3.1-8}$$

Sorting the terms to obtain the extreme values, the minimum and maximum value of the interval of possible values for $T^{pp} \in [T_{min}^{pp}, T_{max}^{pp}]$ become

$$T_{max}^{pp} = \bar{T}^{pp} + \delta T^{pp} = \bar{t}_2^{p} - \bar{t}_1^{p} + \left(\delta t_2^{p} + \delta t_1^{p}\right) \tag{II.3.1-9}$$

$$T_{min}^{pp} = \bar{T}^{pp} - \delta T^{pp} = \bar{t}_2^{p} - \bar{t}_1^{p} - \left(\delta t_2^{p} + \delta t_1^{p}\right) \tag{II.3.1-10}$$

The collective constraint uncertainty δT^{p} is given in agreement with eq. II.3.1-4 by

$$\delta T^{pp} = \left(\delta t_2^{p} + \delta t_1^{p}\right), \tag{II.3.1-11}$$

defining the general error for constraints based on the time difference of two uncertain onsets *A* and *B* as

$$\delta T_{AB} = (\delta t_A + \delta t_B) . \tag{II.3.1-12}$$

The following example (Fig. II.3.1:1) for two simultaneously detected onsets, shows the validity of this relation (II.3.1-11) also for overlapping or coinciding uncertainty intervals:

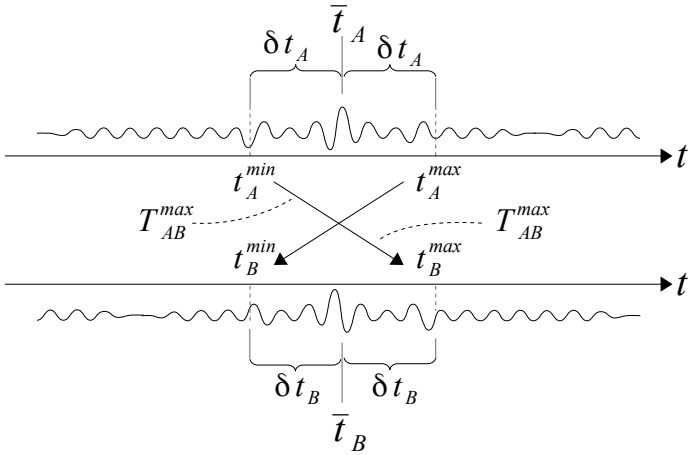


Fig. II.3.1:1: Overlapping uncertainty intervals. Even when intervals are overlapping, the minimum and maximum arrival time differences follow eq. II.3.1-11.

In this case both onsets were detected at the same time, thus $\bar{T}_{AB}$ equals zero.

$$\bar{T}_{AB} = \bar{t}_B - \bar{t}_A = 0 \tag{II.3.1-13}$$

Nonetheless, the interval of possible T_{AB} extends from $-(\delta t_A + \delta t_B)$ to $(\delta t_A + \delta t_B)$ according to equations (II.3.1-9) to (II.3.1-11)

$$T_{AB}^{max} = \underbrace{\bar{t}_B + \delta t_B}_{t_B^{max}} - \underbrace{(\bar{t}_A - \delta t_A)}_{t_A^{min}} = \delta t_A + \delta t_B = \delta T_{AB} \tag{II.3.1-14}$$

$$T_{AB}^{min} = \underbrace{\bar{t}_B - \delta t_B}_{t_B^{min}} - \underbrace{(\bar{t}_A + \delta t_A)}_{t_A^{max}} = -(\delta t_A + \delta t_B) = -\delta T_{AB} \tag{II.3.1-15}$$

Each $T \epsilon [T_{min}, T_{max}]$ corresponds to an exact (uncertainty free) constraint surface $S(T)$. The set of constraints, that can be constructed from all possible $T \epsilon [T_{min}, T_{max}]$ forms a region R in solution space which - if the whole interval corresponds to physically sound solutions - is bounded by the two constraint surfaces $S(T_{min})$ and $S(T_{max})$, enclosing the mean constraint surface relating to $\bar{T}$. This is demonstrated in figure II.3.1:2 for a hyperbolic constraint, being based on two stations observing a P-onset at the same time. Each $T \epsilon [T_{min}, T_{max}]$ corresponds to a hyperbola in 2D (a hyperboloid in full space), the central hyperbola ($\bar{T}_{AB}=0$) resembling a line (plane). The set of all hyperbolas constrains the hyperbolic region R, which with $\bar{T}_{AB}=0$ is in this case symmetric biconcave.

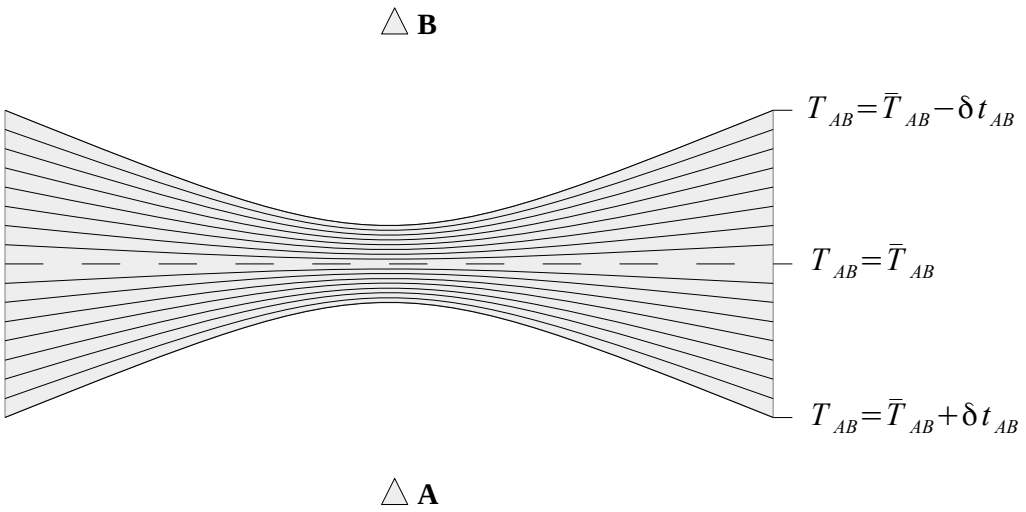


Fig. II.3.1:2: Constraint region constructed by set a set of constraint surfaces for the example of hyperbolic constraints. $\bar{T}_{AB}=0$

It is, however, not a given that the full parameter interval $[T_{min}, T_{max}]$ corresponds to physically possible solutions. In case of P-P constraints, parts of the interval may correspond to (absolute) time differences too large to be consistent (i.e. physically possible) with the given velocity model and station distance (See chapter III.2). In case of S-P constraints, an observed arrival time difference smaller than the associated picking errors would lead to overlapping uncertainty intervals. Part of the S-P interval would therefore include negative values, which would express that the S-phase had arrived earlier than the P-phase. Although this is physically impossible, the formalism is still valid and does not affect the results since the constraints are analyzed in the space domain. The spacial constraint is not affected by unphysical arrival-time differences due to the fact that a grid search comparing arrival time differences with travel time differences will only map the physically

possible part of the parameter interval: The negative S-P arrival time difference, for example, will be compared to the difference of the times that P- and S-phase travel from the station to any evaluated point in space. Since the S-phase will always require a longer time to travel to any point in space than the corresponding P-phase, only positive travel time differences will be found. The negative part of the interval will not "find" a corresponding spatial solution. Hence, these pseudo solutions will naturally be ignored and the interval effectively reduced by the velocity model to an interval of consistent information (i.e. a consistency between observables and model: Model uncertainties hence need to be addressed, so that observations being critical but consistent within the vagueness of the model, are not excluded). In the solution space, a reduced interval corresponds to a constraint appearing with a reduced "width". If a T interval does not contain any valid arrival time difference information (e.g. due to mis-picked phases), no space is being constrained: the constraint is not being "displayed". This is an indication that either a wrong phase was picked or that the applied model does not sufficiently represent the actual velocity structure.
After laying the ground-works of constraint uncertainties, in viewing how timing errors affect the constrained hypocenter region, we will investigate how such constraints may be analyzed.

II.3.2 **Constraints Congruity, Feasible Regions and Representative Best Fit Locations**

Given several stations and phase onsets, various constraints are generated. A global solution would be found in the common intersection of all constraints, as this point fulfills all constraints' demands (fig. II.3.2:1). This constraint based location approach can mathematically be defined as a flexible constraint satisfaction problem (SCP) in the framework of optimization where all constraints are given in equalities or inequalities (Dubois et al., 1996). In standard SCP, the problem is given in finding the set of points that fulfills all given constraints. A set of points that meets all the constraints' conditions is called a feasible region and defines a first set of solutions which at later stages may be refined e.g. using likelihood measures. If the constraints are mutually contradictory, no feasible region can be found and the solution corresponds to the null set. If we disregard uncertainties, location constraints correspond to equalities and hardly ever yield a feasible region in SCP. Hence, the estimation of location and location uncertainty is left to qualitative measures. The last chapter showed that "exact" location constraints appear as "thin" surfaces in space (curves in 2D). When uncertainties are accounted for, the constraints can mathematically be described as inequalities, spanning out regions (fig. II.3.2:1b).

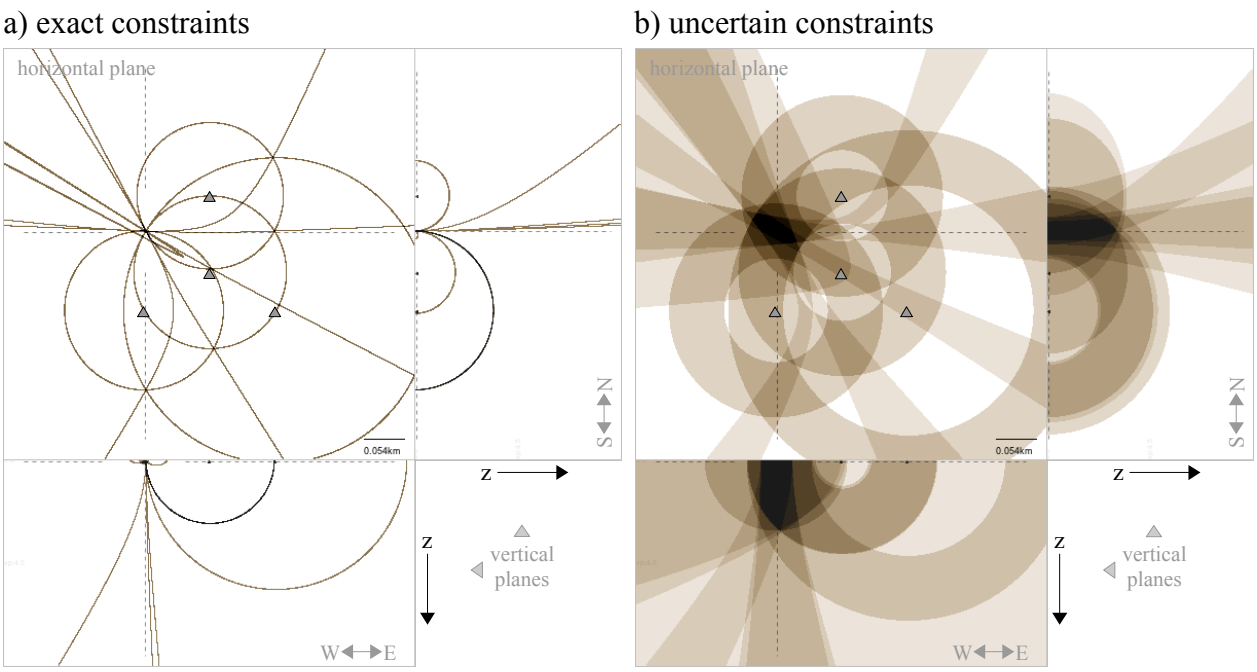


Fig. II.3.2:1: Comparison of constraints a) disregarding and b) regarding uncertainties. The four spheroids and four hyperboloids constrain a congruent region (intersection of all constraints, black), identifying all points possible for location corresponding to the given uncertainties. The cluster of sensors has an aperture of 100m. A homogenous model with 4.5km/s and uniform picking uncertainties resembling 1 sample for P picks and 2 samples for S-picks at a sample rate of 500Hz are being used. While the congruent region in panel (b) allows to quantify the precision of the solution, (a) this is difficult for panel (a).

Where deviations from the true onset would lead to a spread of curves and intersection points in the classical approach with "exact" constraints, here the intersection remains a congruent region that is robust concerning picking errors within the assigned uncertainties. This is demonstrated in figure II.3.2:2: Where exact constraints scatter for perturbated arrival times, uncertain constraints still form congruent regions containing the event location. Also, attempting to define location uncertainties qualitatively from the set of exact constraints will lead to very different results than the computed congruent regions indicate! (The interested reader may try this experiment with fig. II.3.2:2, covering the right panel.) To visualize the impact of smallest parameter uncertainties, here an uncertainty of only one sample (δt^p=0.002s, @500Hz sample rate) has been applied. For this small scale scenario, this already causes large spatial uncertainties. Leonard (2000) reports an average of ±0.067 sec for the difference in onset time picking by different analysts for teleseismic P phases, with a standard deviation of 0.15 sec. The actual picking and solution uncertainty is therefore expected to be significantly larger.

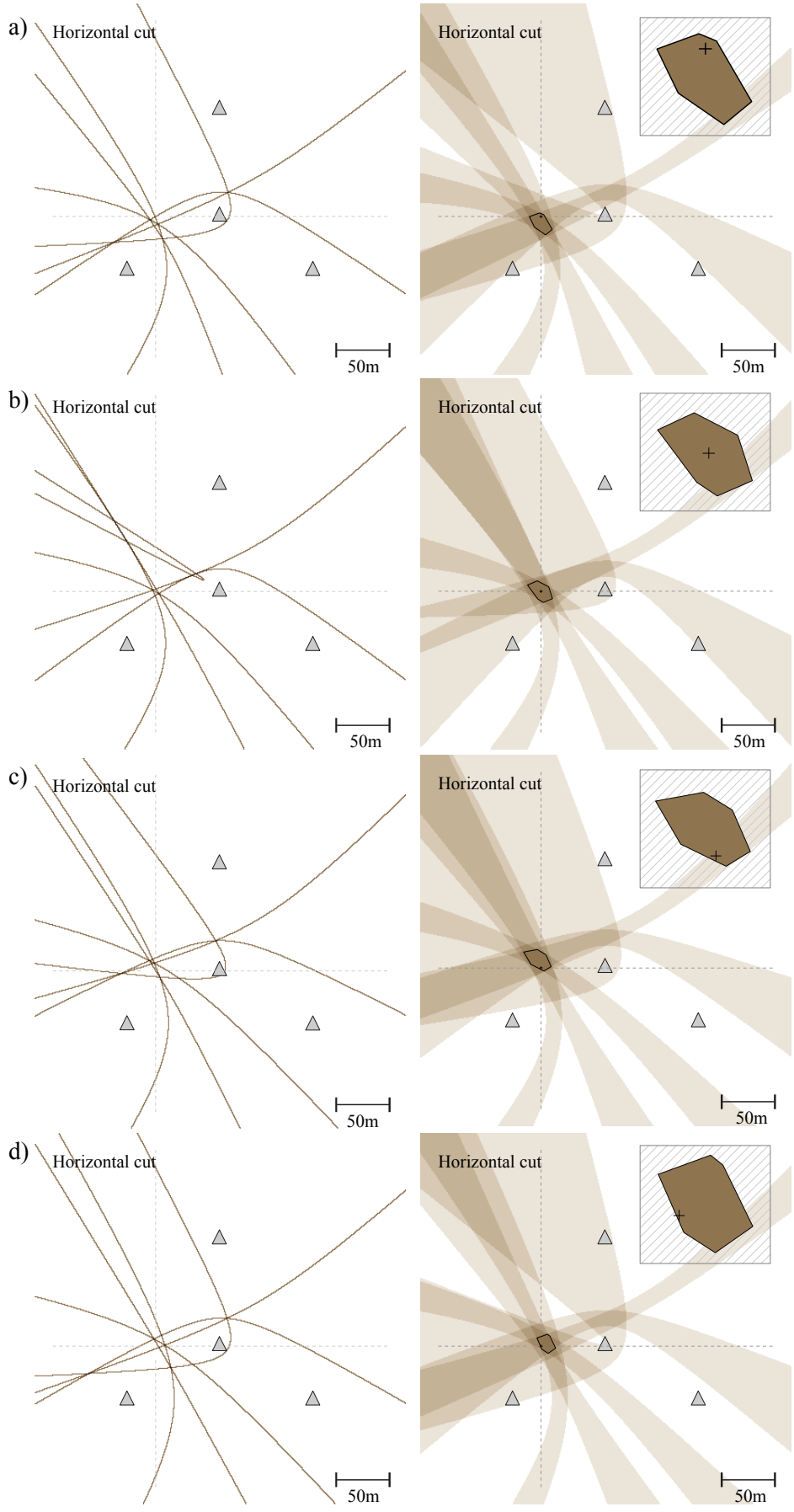


Fig. II.3.2:2: Exact constraints (left) vs. uncertain constraints.

Panels a-d) show a set of hyperbolic constraints for the same event with the onset picks only varying up to one sample (500Hz data, $\delta t^p \leq 0.002$s. *This time interval can be viewed as the minimum uncertainty concerning the signal due to discretization. On the left, exact constraints (neglecting uncertainties) are used, showing the spread of curves that arises from this minimum uncertainty. The right side, displaying uncertain constraints with a uniform uncertainty of 0.002s, reveals the corresponding uncertainty in space, which for some constraints is quite large. Note, that the spatial uncertainty ('width') is strongly varying between the constraints, although all hold the same the (time-based) parameter uncertainty.*

While exact curves scatter and require an averaging procedure to identify the rough location of the epicenter, constraints regarding uncertainties intersect in feasible regions of congruent information. The hatched area (top right), shows the congruent region magnified: The epicenter (black plus) is always enclosed.

Accounting for uncertainties therefore serves two purposes. First, it adds confidence that the event truly lies in the computed region. Secondly, it displays the state of knowledge concerning the precision of the result.

Accounting for uncertainties can serve several purposes. First and most, it adds confidence that the event may be found in the region having been computed. Secondly, using an expected representative uncertainty, the expected precision of future location estimates (based on the given network geometry) can be mapped over the region of interest, to evaluate whether the given network suffices to reach the required resolution or whether it would be beneficial to strategically deploy additional stations (chapter III.1).

The congruent zone displays all possible locations for the uncertain parameters, e.g. the set of locations that would be obtained by all combinations of exact picks possible, within assigned uncertainty bounds. If all uncertainties are properly addressed, the congruent zone will form a maximum region, in which the true location should lie. This maximum boundary is of special importance, e.g. for earthquake early warning location estimates, allowing to derive minimum and maximum hypocentral distances and magnitude estimates; or in CTBT applications, to yield the confidence that the possible source location area does not extend over geographic boundaries of political entities. Again, as timing (picking) errors represent only a part of the total error, it is important to incorporate all significant uncertainty measures (e.g. sensor orientation for BA constraints), but especially model uncertainties (Chapter II.3.5). Now, applying a “hard” SCP for earthquake location, whereby “hard” indicates that all constraints need to be fulfilled, is not robust: Outlier picks, for instance, will generate constraints appearing outside the zone of highest congruity and reduce the maximum congruity of the system. As not all “hard” constraints are satisfied, no “feasible” region exists and no solution would be given. Rather than using “hard” constraints that have to be fulfilled for a solution to exist, the SCP construct of “flexible” constraints needs to be applied. Flexible constraints do not necessarily have to be satisfied. The quality of the systems solution is rather valued by the number of satisfied constraints. This corresponds to the theory of coherence, in which the degree of a claim's trustworthiness is assessed by evaluating the consistency of a set of related statements. An event's location would accordingly be trusted to be found within that region that is supported consistently by most constraints. Where all constraints (κ) agree (i.e. overlap or intersect), the congruity c is defined as $c=1$

$$\left[\bigcap_{i=0}^{n_\kappa} \kappa_i\right] \neq \emptyset \;\rightarrow\; c=1 \qquad \text{(II.3.2-1)}$$

with n_κ being the amount of given constraints. In regions of lower constraints agreement, c drops in value, indicating a less robust or trustworthy solution. To compensate for mis-picks and the fact that the true uncertainties might not always have been applied, a solution will be declared "feasible" if at least 70% of constraints agree. A region is therefore becoming part of the hypocenter region if it holds a congruity $c \geq 0.7$ (fig. II.3.2:3). This particular choice of threshold causes the feasible region to be identical to a region of 100% constraint-agreement for underdetermined systems of hyperbolic constraints (3 or less P-phase detections), which lack the redundancy to support outlier-robust solutions. [With three stations, three hyperbolic constraints are generated: A region where only two out of the three constraints agree holds a congruity of $c = 2/3 = 0.\bar{6} < 0.7$ and therefore doesn't qualify as "feasible" – the feasible region is in this case identical to the region of 100% constraint agreement: the region on which all three constraints agree. This avoids an overestimation of the feasible region since zones, in which only two out of three constraints agree, may be extensive. Given four different P-phases, also regions where five out of the six P-P constraints agree will become part of the feasible region, since $c = 5/6 = 0.8\bar{3} \geq 0.7$).] This particular threshold acts therefore as natural "switch", adding robustness to the feasible region when sufficient data is available, at the same time avoiding its overestimation when data is few.

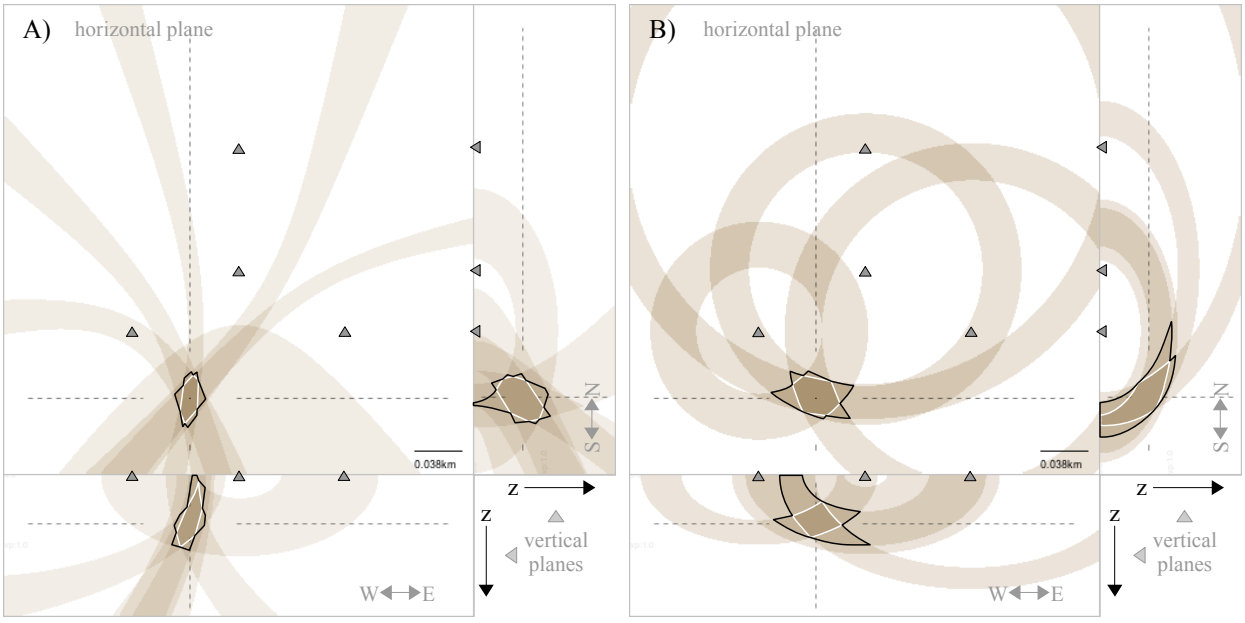


Fig. II.3.2:3: The superposition of individual constraints allows to identify the hypocenter as region of strongest constraint agreement/congruity (white contour), marking the global solution. The hypocenter region is defined as the region that corresponds to a 70% constraint agreement (black contour), yielding robust hypocenter regions in presence of outliers. Panel A) shows hyperbolic constraints, panel B) S-P constraints.

Using the concept of distinct constraints and their uncertainties, the normalized congruity can be defined in a general form as

$$c_{x,y,z} = \frac{1}{n_\kappa} \sum_{i=1}^{n_\kappa} H(V_i, U_i) \qquad \text{with} \qquad H(v,u) = \begin{cases} 1, & |v| < u \\ 0, & else \end{cases} \tag{II.3.2-2}$$

with $c_{x,y,z}$ denoting the congruity at coordinate $(x,y,z)^T$. V_i denotes the constraint deviation and U_i the constraint uncertainty. The constraint deviation V_i ("violation of the constraint's demand") denotes the difference between observed and modeled constraint parameters: If the modeled parameters confirm the observation, the constraint's demand is fulfilled; if the modeled parameters deviate from the observation, the constraint is violates. H is a boxcar function that increases the congruency by one increment if the deviation is smaller than the assigned uncertainty. For e.g. a P-P hyperboloid, eq. (II.3.2-2) takes the specific form

$$c_{x,y,z} = \frac{1}{n_\kappa} \sum_{i=1}^{n_\kappa} H(T^M_{x,y,z} - T^O_i, \ \delta T) \tag{II.3.2-3}$$

with $T^M_{x,y,z}$ being the modeled constraint parameter for a given coordinate (i.e. the difference of modeled travel times between trial location and both involved stations) and T^O_i being the observed constraint parameter (i.e. the observed difference between the arrival times of both stations). The difference between both quantities defines the constraint deviation V_i. The constraint uncertainty U_i is e.g. given in the picking uncertainty δT.

The congruent ("feasible") region allows a reliable interpretation of the source location, reflecting the possible variance in location. Its shape is determined by a) the network geometry, b) the used constraint classes and c) the assigned uncertainties and represents the uncertainty in the location of the (point source) hypocenter.

Note: It was mentioned before that a network's ability to yield precise locations can be evaluated by using synthetic arrivals of test locations in conjunction with the expected uncertainties, analyzing the extent of the feasible region. A mapping of the precision (chapter III.1) will reveal

the regions, in whom precise locations can be obtained, and those where the deployment of additional sensors would be beneficial. In this context it is worth to note that the congruent region (usually obtained from the overlap of distinct constraints) may also be obtained from residua-sum approaches, as the number of hyperbolic constraints grows with $(N^s)^2$, (N^s being the number of involved stations) while the residual sum scales linearly with N^s. For automatic schemes only focussing on the congruent region (e.g. network precision mapping/optimization) the computational cost can therefore be reduced. The congruent region corresponds to the region for which an "uncertainty truncated residuum" (eq. II.3.2-4) vanishes, identifying all locations which are possible within the given uncertainty intervals. Where the residuum is defined as the difference between observed and modeled arrival time of the corresponding phase

$$R_i^P = t_i^P - t_i^{P,M} \quad , \tag{II.3.2-4}$$

a uncertainty truncated residuum will be defined as

$$\bar{R}_i^P = \mathbf{\Lambda}\left(\left|t_i^P - t_i^{P,M}\right| - \delta t^P\right) \tag{II.3.2-5}$$

with $\mathbf{\Lambda}$ being the ramp function

$$\mathbf{\Lambda}(x) = \begin{cases} x, & x \geq 0 \\ 0, & else \end{cases} , \tag{II.3.2-6}$$

which simply sets the residuum to zero, if it's within error bounds (δt^P).
The region for which the truncated RMS

$$\overline{RMS} = \sum_i \left(\bar{R}_i^P\right)^2 \tag{II.3.2-7}$$

vanishes ($\overline{RMS} = 0$) corresponds to the constraint intersection region as formed of P-P constraints (solely P-phase based location). Since only the area of $\bar{RMS} = 0$ is evaluated, the choice of the used norm is of no effect. For the congruent region obtained from the intersection of S-P and P-P constraints (i.e. the region constrained by P- and S-information, as in fig.

II.3.2:1b) the truncated residuum sum is given by

$$\overline{RMS}^{SP} = \sum_i (\bar{R}_i^P)^2 + \sum_j (\bar{R}_j^S)^2 , \qquad \text{(II.3.2-8)}$$

with i and j iterating over all stations with given P- resp. S-detection. Instead of $1/2(n^2-n)$ hyperbolas the RMS estimate requires only n iterations, however, demands an additional iterative search for the origin time. The RMS estimate of the congruent region is suitable for the analysis of denser networks.

After the feasible region is constrained, one might look for an optimal, representative (e.g. most likely) solution to represent the point-source hypocenter location. Such a representative solution can be estimated using the center of congruity $\boldsymbol{R_c}$ which is defined in analogy to the physical quantity of the center of mass $\boldsymbol{R_m}$, which is described as

$$\boldsymbol{R_m} = \frac{1}{M}\sum_{i=1}^{n} m_i \boldsymbol{r_i} \quad \text{with} \quad M = \sum_{i=1}^{n} m_i . \qquad \text{(II.3.2-9)}$$

There, M is the total mass of n particles with locations $\boldsymbol{r}_i$ and particle masses m_i.

Following $\boldsymbol{R_m}$, the center of congruity $\boldsymbol{R_c}$ can be defined as

$$\boldsymbol{R_c} = \frac{1}{C}\sum_x \sum_y \sum_z \boldsymbol{\Lambda}^{\mathbf{0.7}}(c_{x,y,z}) \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{with} \quad C = \sum_x \sum_y \sum_z \boldsymbol{\Lambda}^{\mathbf{0.7}}(c_{x,y,z}) . \qquad \text{(II.3.2-10)}$$

Here, $c_{x,y,z}$ represents the congruity value at grid point (x, y, z), representing that coordinate's weight ($0 \geq c_{x,y,z} \geq 1$), while the Ramp function $\boldsymbol{\Lambda}^{\mathbf{0.7}}$ truncates every congruity value below 0.7 to 0, making that coordinate's influence void.

$$\boldsymbol{\Lambda}^{\mathbf{0.7}}(x) = \begin{cases} x, & x \geq 0.7 \\ 0, & else \end{cases} \qquad \text{(II.3.2-11)}$$

In this way, points of low congruity are disregarded while points of highest congruity exert the highest attraction, pulling the solution into the center of highest constraint agreement.

Chapter III.2 will lay out that in certain cases one *even ideal* set of arrival times may yield multiple distributed feasible regions (with $c=1$), i.e. ambiguous solution candidates. Seeking one representative single-point location in such a case (where actually several equivalent maxima exist) is critical as it hides the existence of the other solutions. For ambiguous solutions, the center of congruity approach (as any other averaging scheme) will fail, yielding a solution that is not representative - that might even not be part of the set of feasible regions, as the averaged location might lie in between all of them. Automatic algorithms should therefore rather use all distributed feasible regions to define an extended region (i.e. a convex polyhedron enclosing all maxima) to represent the uncertainty in hypocenter location.

With an increasing amount of stations, these ambiguous solutions that fulfill all constraints perfectly ($c=1$) are less and less likely to exist. However, when assigned parameter uncertainties are too small to cover the true errors (e.g. mispicks), no congruent region might be found in overdetermined problems. Instead several, spatially separated weak solutions (holding a $c<1$) would be found, surrounding the true location. Fig. II.3.2:4 shows such a case. A center-of-congruity approach as described above, which would attempt to average over the distributed weakly congruent regions, would fail to determine the best fit solution (red plus, fig. II.3.2:4). Again, here a polyhedron formed by the distributed weak solutions could be used to define the possible hypocenter region. A best fit solution could be obtained by "artificially" increasing the assigned uncertainties until a region with sufficiently high congruency ($c\geq 0.7$) is found. For constraints with comparable uncertainties the so found congruent region generally agrees well with the minimum found by residua-sum approaches (while remaining to be more accurate in presence of very discrepant outliers as such would be weighed less). However, this approach is slow and only patching the problem. In the following we will derive a solution for this problem concerning the most predominant case for automatic and real-time locations, the P-phase based location, to provide robust estimates even when uncertainties are not properly assigned. Therefore we recall that the scattered hyperbolic constraints (fig. II.3.2:4) were obtained using underdetermining estimators in the jackknifing scheme, relying on only two data.

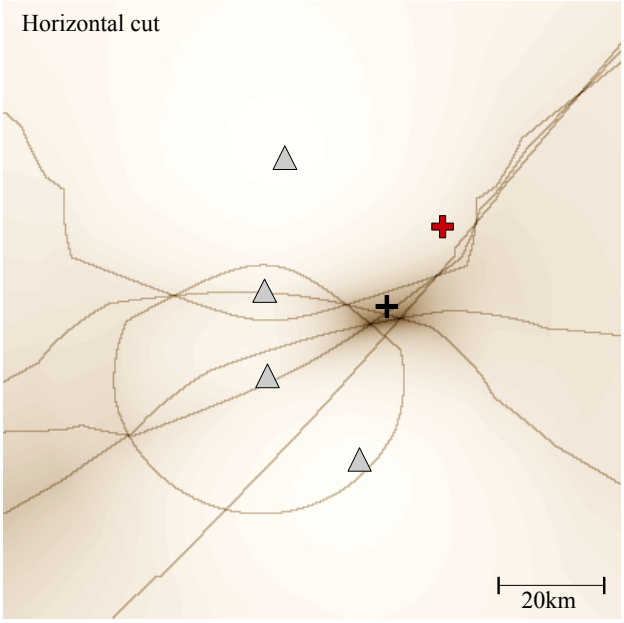


Fig. II.3.2:4: Due to too small assigned uncertainties, no zone of congruity ($c \geq 0.7$) is formed. Instead, sub-solutions of lower congruity (e.g. the intersection point of three constraints) scatter around the optimal location (black plus). Using a simple center-of-congruity approach, ambiguous sub-solutions (multiple crossing points of the same three constrains) pull the "optimal" solution (red plus) away.

If the concept of constraining fails due to discrepancies, we need to remember that it is actually the solutions of 4-member sub-groups ("four stations with P-onsets") that relate to the 4dim. earthquake location problem, yielding exactly determined sub-solutions. It is *their* scatter that describes the location uncertainty. To find an optimal solution ("explaining the data best"), it is therefore the ensemble of exactly determined sub-solutions that needs to be evaluated - not the hyperboloids themselves. Understanding their relationship to the 4-member groups, however, will allow us to define a set of rules, which will allow us to use the hyperbolic constraints to find the exact 4-member ("quad")-group solutions, to eventually construct the optimal solution.

To do so we will look at the relationship between related groups of different member size, following the hierarchy of information downwards and then upwards again. The key for understanding this walk lies in the following Lemma: *All groups that can be derived by rearranging n elements of a n-member group into (n-1)-member groups, will always intersect in the solution of the n-member group.* The decrease of the member size by one increases its solution's dimension by one: A quad-group (yielding one solution with no degree of freedom: A point) can be split into 4 triple-groups (yielding 1dim. solutions: Curves). These four curves are bound to intersect in the point of the quad-

group solution. Each triple-group itself can be split into 3 two-phase-groups (yielding 2dim. solutions: Surfaces): Here, the P-P-hyperboloids which we use to constrain the earthquake solution.
The P-P-hyperboloid, being defined by the arrival times and positions of two stations, is the lowest object in this hierarchy of information and the highest in dimension. We will now step up the hierarchy again and keep track of the stations on which the groups rely: Three hyperboloids *a,b,c* (based on stations {1,2},{2,3} and {3,1}) intersect in a single curve corresponding to the solution of the superior "parent" triple-group $T_{a,b,c}$ (being based on stations {1,2,3}). [Note: The intersection of any two of these hyperboloids forms the same curve. The third hyperboloid is not independent but redundant being based on the same parameters that determine the other two hyperboloids and is therefore bound to run through the same intersection curve]. Intersecting the four triple-group curves (being based on stations {1,2,3}, {1,2,4}, {1,3,4} & {2,3,4}), where each can be found by intersecting the corresponding "child" hyperboloids, would bring us back to the quad-group solution (being based on stations {1,2,3,4}. In total we find with the P-phases of four given stations (1,2,3,4), six hyperbolic constraints (a,b,c,d,e,f) based on stations {1,2}, {1,3}, {1,4}, {2,3}, {2,4} & {3,4}, which can be grouped into 4 triple-groups ({a,b,d}, {a,c,e}, {b,c,f}, {d,e,f} which use the information of stations {1,2,3}, {1,2,4}, {1,3,4} & {2,3,4}). The exact solution of the quad-group can therefore either be identified over the exact intersection of the 4 triple-curves or likewise the exact intersection of the 6 hyperbolic constraints. The algorithm for the optimal solution, however, has to evaluate the intersections (congruent regions) according to their belonging to certain quad-group stations. To find the optimal solution we will again use the center-of-gravity (center-of-congruity) approach, but now only consider congruent regions that contain all constraints belonging to a quad-group. Two complications need to be overcome:
The first is of a mathematical nature: There can be more than one exact solution of a quad-group. Since the travel time equation is already of second order for the most simple model (homogeneous half-space), there can be two exact locations for the same set of arrival times (chapter III.2). In more complex models this number can even be higher, e.g. in a 1D-model the number of exact solutions can possibly be as high as twice the number of existing refractors (see fig. II.3.2:5). Averaging over these ambiguous, equivalent solutions will usually divert from the best fit location. Instead, these ambiguous solutions should be interpreted as possible solution candidates. Only the one agreeing best with all data is to be evaluated in the averaging process.
The second complication is technical in nature and lies in the fact that distant solutions of inherently higher location uncertainty would additionally pull the solution towards them. This is due to their

larger spatial extent which causes them to be spread over more grid cells than precise near by solutions, which in turn causes them to exert a higher weight in eq. II.3.2-10. Rather than weighing the region of congruity, only the one point of this region that causes the least violation should therefore be evaluated. Both problems can be observed in this example (fig. II.3.2:5b).

Instead of identifying the exact sub-solutions in the three-dimensional space and averaging them, we will continue to see the focal depth as a free but fixed parameter, meaningly the inversion is being performed in two space dimensions, repeated for several depths (i.e. horizontal cross-sections). This is beneficial as the depth is inherently the least constrained solution parameter for the standard case of planar surface networks and doesn't need to be scanned as fine as the horizontal components. Now, as we temporary reduce the dimensionality of the problem, our focus will shift from the quad-group solutions to the solutions of the triple group, the triple points. The triple points correspond to the crossing point(s) of three hyperbolic constraints being based on three stations' onsets. To find the optimal solution in this 2D-problem we will apply the center-of-congruity approach to the scattered triple-group solutions. Recalling, that the scatter of sub-solutions correlates with the agreement of information (solution quality), the optimal solution for the global 3D problem is near the 2D-optimum at the depth where the triple-points are least scattered. What we therefore actually do corresponds to following each set of scattered 3D triple-group curves until they all intersect in the corresponding exact quad-group solution, which corresponds to one of the jackknifed hypocenter location estimates. In a example of four stations, four triple-point curves exist that intersect in exactly one quad-group solution. With more than four stations, several quad-group exist, each having it's own set of triple-group curves that exactly intersects in it. As this optimum is computed by the general spread of all triple-group solutions, it can deviate from the true optimum of the quad-groups if these are located in significantly different depths.
Fig. II.3.2:5a shows an example of four given stations with the inversion done in a certain focal depth. Triple-group solutions (curves in 3D) appear as exact intersection points of three hyperbolas in 2D cuts. As mentioned before, do some triple-groups not only produce a single intersection but multiple ambiguous solutions (in fig. II.3.2:5a connected with dashed lines). If we regarded all ambiguous solutions it would pull the best fit solution (fig. II.3.2:5a, red plus) from the true optimum (fig. II.3.2:5a, black plus). For the true optimum, only the one solution of the set of ambiguous solutions that is most consistent with the other data, is meaningful - the others need to be disregarded. The same holds for solutions in larger distance which due to their larger uncertainty (i.e. larger region) would impose more weight and pull the optimal solution towards them (fig.

II.3.2:5b, red plus) – also here only the one point that causes the least discrepancy within that region is weighed, and the others disregarded. These two mechanisms correspond to the general rule, that among ambiguous solutions only those are meaningful that follow the trend of the whole dataset. Additional solutions can be disregarded as non-qualifying solution candidates. However, if there is not enough data to evaluate the candidates, the ambiguity cannot be resolved and all solutions have to be kept, increasing the location uncertainty. Following these mechanisms yields the expected and true optimum (fig. II.3.2:5, black plus), which in this example is equivalent to the one found by artificially increasing the uncertainties until congruity is reached.

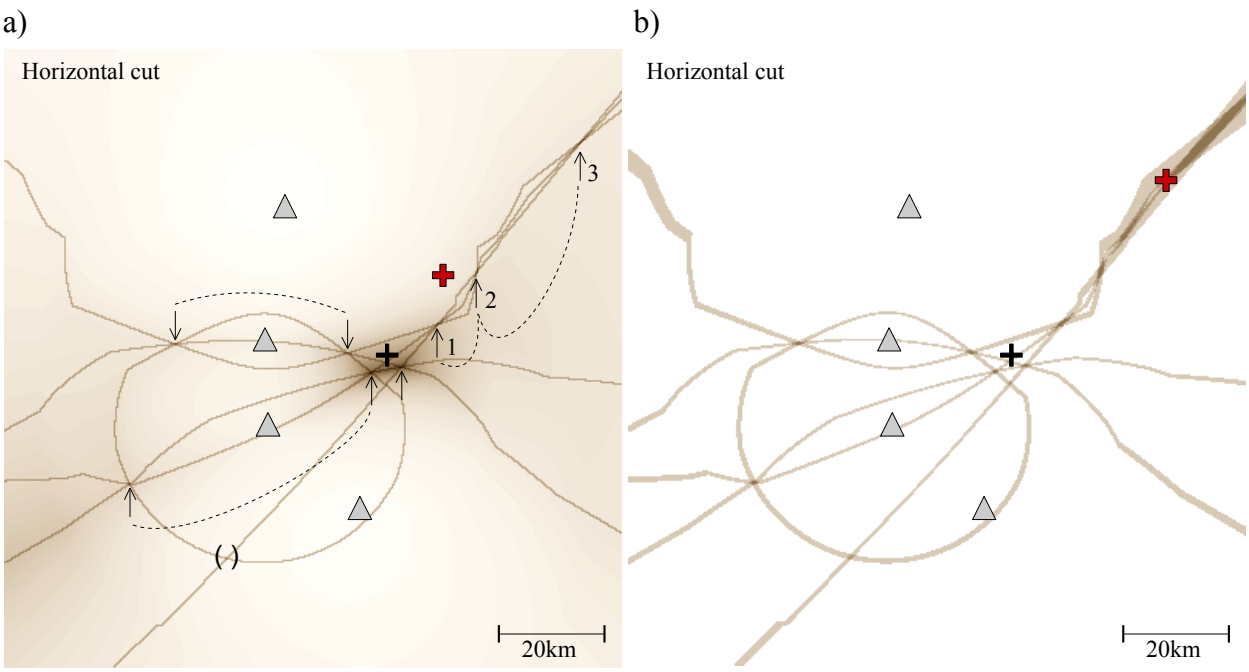


Fig. II.3.2:5: Averaging over all weak congruent solutions vs. triple point evaluation. Given is a set of hyperbolic constraints without congruent (c>0.7) region. The background shading in (a) corresponds to the L1-residua-sum, darker regions indicating a lower residua sum. The best fit solution obtained from the hyperbolic constraints should lie close to the darker region. a) The red cross marks the best fit solution obtained by averaging over all existing solutions belonging to the set of triple groups. These triple points are marked by the black arrow. In this example multiple equivalent solutions exist, and pull the solution from the expected location. Ambiguous solutions (multiple intersection points of the same three hyperbolic constraints) belonging to the same triple group are connected by dashed lines. The three triple points marked with 1-3 belong to the same triple group (this means that the onsets of three stations yield three location for the given depth), and are caused by two refractors at 10.6km and 15.5km depth increasing the near surface velocity of 3.5km/s to 5.7km/s and 6.4km/s. Points where only two hyperbolas intersect (indicated with brackets ( )) correspond to an intersection point with inconsistent time base (i.e. the constraints describe the same location but a different origin time: This means that the indicated location cannot yield an origin-time that would be consistent with the subset of the three or four involved travel- and arrival times: The hyperbola between station 1 and 2 may intersect with the hyperbola of station 2 and 3, yet the arrival times of station 1 and 3 do not yield a hyperbola running through the same point). b) For the same set of non-congruent constraints, yet with higher uncertainties assigned, the regions of sub solutions in greater distance are inherently larger than those close by. Using the averaging scheme, these extended zones attract the solution, although their location uncertainty is larger than the compact zones of the solutions close by. The optimal location ("explaining the data best") can be found by focussing on one representative triple point per triple group instead (black cross, in panel a&b).

The set of triple points $\mathbb{K}_g$ of triple group g are identified by grid cells that involve all stations of the g-th triple group and have three or more hyperbolic constraints running through them.

$$\mathbb{K}_g := \{\vec{r} \mid c^H(\vec{r}) \cdot N^H \geq 3 \ \wedge \ S_g^{TG} \subseteq s(\vec{r})\} \quad \text{and} \quad \{\vec{r} \mid \textit{all grid points}\} \tag{II.3.2-12}$$

with $c^H(\vec{r})$ being the congruity value concerning hyperbolic constraints and N^H the total number of hyperbolic constraints. The term $c^H(\vec{r}) \cdot N^H$ therefore relates to the amount of hyperbolic constraints running through the grid cell at location $\vec{r}$. S^{TG} is the 3-tuple (triple) containing all 3-combinations of stations with P-onsets S.

$$S = \{1, 2, \ldots, N^s\} \quad \text{with} \quad N^s : \text{total number of stations with P-onsets} \tag{II.3.2-13}$$

$$S^{TG} = S_3 = \binom{S}{3} \tag{II.3.2-14}$$

S_g^{TG} is therefore the *g*-th element of the tuple S^{TG}, the set of 3 stations that defines the g-th triple group. With e.g four stations S^{TG} consists of $S^{TG} = \{\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\}$.
$s(\vec{r})$ refers to the set of stations, whose information constructs the hyperbolic constraints running through the grid cell at location $\vec{r}$. G is the number of possible 3-combinations of the set S, i.e. the total number of triple groups.

$$G = \binom{N^s}{3}. \tag{II.3.2-15}$$

The best fit solution $\vec{R}$ is now the point that is closest to the representative solution $\mathbb{K}_g^{N(\vec{r})}$ of each triple group g: the solution of the set of possible solutions per triple group that causes the smallest violation.

$$\vec{R} = \underset{\vec{r}}{argmin} \sum_{g}^{G} \left| \vec{r} - \text{Ж}_g^{N(\vec{r})} \right|^2 \qquad \text{(II.3.2-16)}$$

$\text{Ж}_g^{N(\vec{r})}$ is identified as the closest point of all triple points $\vec{\text{Ж}}_g^n$ of group g to the trial location $\vec{r}$

$$N(\vec{r}) = \underset{n}{argmin} \left| \vec{r} - \vec{\text{Ж}}_g^n \right| . \qquad \text{(II.3.2-17)}$$

By focussing only on that solution of the set of possible solutions per triple group that causes the smallest violation, both problems are solved: Firstly, for multiple equivalent solutions only the one fitting all data "best" is used. Secondly, the paradox that less precise locations (larger areas) assert more weight than more precise solutions is solved by only evaluating one solution of each patch. For spheroidal S-P constraints the "best fit" solution holding the least constraint violation follows the same scheme established here for P-P constraints, but using binary groups.

When limiting the analysis to regions of constraint congruity, the theory of error and error propagation is sufficient to map the parameter uncertainties into the solution space. Averaging schemes can provide an estimate for the best fit solution. When model uncertainties are applied, constraints grow wider, forming larger congruent regions. However, as these constraints wrongly convey the same likelihood for a solution to be found over the whole constraints "width" (i.e. claiming, that the locations being constrained by parameters from the edges of the uncertainty intervals are as likely as those being obtained from the interval center), they must inherently fail when it comes to the estimation of "best fit" solutions. For such it is therefore necessary to use more advanced approaches to determine the solution most probable. Applying a probabilistic approach will fill this gap and allow to use variable weights over the picking interval and later model uncertainties. It further opens up for new applications in the field of automatic location within the context of ambiguous phase onsets, i.e. noisy data. For a complete treatment of uncertainties the problem will be examined in the context of probability theory, outlined in the next chapter.

II.3.3 **Probabilistic Constraints**

Congruent regions, as regions which comply with most or all constraint conditions within error bounds, define an (extensive) bounding region for the hypocenter location. This forms a robust measure but offers no information on the most likely hypocenter location within that region. Often, significant parts of the region may be highly unlikely to be the solution sought for, although physically possible. Choosing a probabilistic perspective will allow us to obtain the needed information and yield enhanced quality control measures, beside opening doors for advanced methods, like multi-pick evaluation (chapter II.3.4) for ambiguous arrival times.

Already on single constraint level, locations that correspond to a vanishing constraint deviation (i.e. the locations that fulfill the exact "uncertainty-free" constraint, corresponding to the "central" constraint slice relating to $\bar{T}$) should be of higher probability than locations for which the constraint deviation equals the uncertainty, being located on the constraint's outer edge. This should be true for two reasons. Firstly, the probability that the true phase is located at $t=\bar{t}$ should be higher than the chances of being located at $t=\bar{t}+\delta t$ or $t=\bar{t}-\delta t$. Secondly, even if a flat level of confidence over the whole interval of picking uncertainty was given, $T=\bar{T}$ may statistically occur more often within this time window than $T=\bar{T}+\delta T$ or $T=\bar{T}-\delta T$. To obtain the most likely location we will introduce level-of-confidence measures, that allow to express the likelihood that the phase arrival is found at a certain sample. This will yield constraint weights which vary over the deviation axis (e.g. the constraint will "fade out" at it's edges). A simple weight distribution would be given in a triangular weight function, having it's maximum at $T=\bar{T}$. As picking errors are generally assumed to be gauss distributed, Gaussians are often used as weight function (Lomax, 2009). Over the course of this chapter both functions will prove themselves to be indeed special cases of a general probabilistic form. More realistic though are functions with a shallow ramp, reaching it's maximum close to the later end of the interval followed by a quick decay to zero, since it is easier to determine the point in time that follows the phase-arrival than the point that is truly an early bound for the emerging signal. For low SNR data, onsets are less distinct (even being ambiguous) and vary in degree of confidence. It is hence necessary to allow and solve for more complex probability functions. The phases within $\bar{t}\pm\delta t$ will be weighed with their level of confidence, yielding a constraint probability density function defined in the uncertainty window, which will then be projected onto the constraint region.

Given two independent random variables X and Y, the probability $p_{XY}(x, y)$ that X takes the specific value X=x and Y the value Y=y is given by the product of their individual probabilities

$$p_{XY}(x, y) = p_X(x)\, p_Y(y)\,. \tag{II.3.3-1}$$

Let $p_A(t_A)$ define the probability that the true onset A is found at $t = t_A \pm \varepsilon$, $\varepsilon = \delta t/2$ and $p_B(t_B)$ define the probability that the true onset B is found at $t = t_B \pm \varepsilon$, $\varepsilon = \delta t/2$, with δt being the inverse of the sample rate. Let further $\pi(t)$ be the corresponding probability density at time t, e.g. given in the user defined level-of-confidence curve. The probability

$$p(t) = \int_{t-\delta t}^{t+\delta t} \pi(t)\, dt \tag{II.3.3-2}$$

is therefore sample-based: describing the probability that a phase onset is found at a certain sample. For a low SNR, waveforms vary significantly between stations wherefore manual picks may be assumed to be uncorrelated. Even when the signal is matched (i.e. cross-correlated) via it's spectral pattern in the sonogram event identification, the following phase picking in the seismograms still fulfills the requirement of independency: Due to the frequency-time uncertainty, the frequency-analysis requires significantly larger time slots and allows only a rough determination of the phase arrivals using a cross-correlation of the spectral pattern, which is afterwards refined performing a picking in the seismograms. Both variables may therefore be treated as independent. The probability for the coincidental occurrence of two such events ('event' according to probabilistic terminology, not seismological), i.e. finding phase onset A at the sample of t_A and onset B at the sample of t_B, is given according to (II.3.3-1) by

$$p_{AB}(t_A, t_B) = p_A(t_A) \cdot p_B(t_B)\,. \tag{II.3.3-3}$$

For a constraint with $\bar{T}_{AB} = \bar{t}_B - \bar{t}_A$, $p_{AB}(\bar{t}_A, \bar{t}_B)$ describes the probability of the event that the onsets are located at the midpoints $\bar{t}_A$ and $\bar{t}_B$ of the uncertainty interval. For a given probability

mass function (PMF) representing the sample-wise level-of-confidence over the uncertainty interval of the onset, the constraint parameter $T=\bar{T}_{AB}$ is not only found between the midpoints of both uncertainty intervals, but also for the event that both onsets $t_A=\bar{t}_A$ and $t_B=\bar{t}_B$ are shifted by a constant η from their interval midpoint (Fig. II.3.3:1)

$$T_{AB}=(\bar{t}_B+\eta)-(\bar{t}_A+\eta)=\bar{t}_B-\bar{t}_A=\bar{T}_{AB} \tag{II.3.3-4}$$

as long as $t_A=\bar{t}_A+\eta$ and $t_B=\bar{t}_B+\eta$ remain in their corresponding uncertainty intervals

$$\bar{t}_A-\delta t_A<t_A<\bar{t}_A+\delta t_A \tag{II.3.3-5}$$

and

$$\bar{t}_B-\delta t_B<t_B<\bar{t}_B+\delta t_B\,, \tag{II.3.3-6}$$

effectively limiting the shifting range to

$$|\eta|\leq min(\delta t_A,\delta t_B)\,. \tag{II.3.3-7}$$

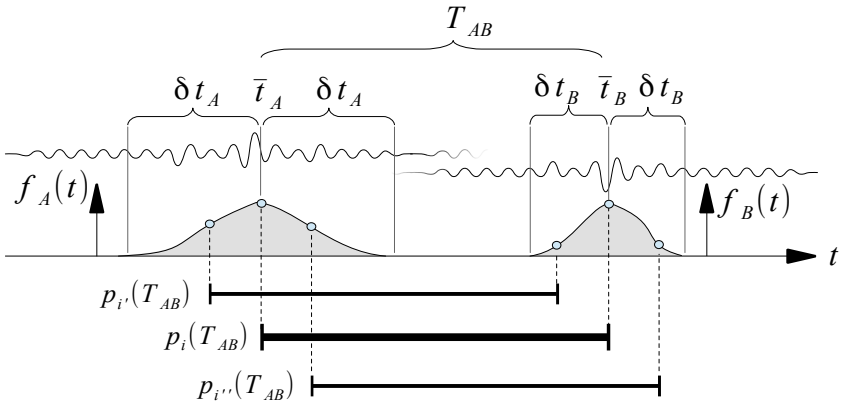


Fig. II.3.3:1: Constraint parameter probabilities in the uncertainty interval. The same arrival time difference is found multiple times between the intervals. The total probability for the arrival time difference is therefore given by the sum over all such occurrences, weighed by the (product of the) individual likelihoods in interval A and B.

The total probability corresponding to the constraint parameter $T=\bar{T}_{AB}$ is therefore given by the sum of all events

$$p(T=\bar{T}_{AB})=\sum_{\{\eta\ |\ |\eta|<min(\delta t_A,\delta t_B)\}} p_{AB}(\bar{t}_A+\eta,\bar{t}_B+\eta)\,. \tag{II.3.3-8}$$

Using (II.3.3-3) this yields

$$p(T=\bar{T}_{AB})=\sum_{\{\eta \mid |\eta|<min(\delta t_A,\delta t_B)\}} p_A(\bar{t}_A+\eta)\cdot p_B(\bar{t}_B+\eta)\ . \qquad \text{(II.3.3-9)}$$

Recalling, that a single constraint (based on uncertain parameters) can be thought of being constructed by a continuous set of constraints based on *exact* parameters, obtained from all value combinations in the uncertainty intervals (i.e. the constrained space can be constructed by a continuous set of surfaces, chapter II.3.1), this probability $p(\bar{T}_{AB})$ weighs the surface that relates to $T=\bar{T}_{AB}$. Computing the probability $p(T)$ for all other $T\epsilon[T_{min},T_{max}]$ provides the complete probability distribution over the whole constraint's "width". Mapped to space, this attributes certain likelihoods to spatial coordinates, describing the likelihood to find the hypocenter at the given coordinate. The general relation for any $T\epsilon[T_{min},T_{max}]$ can be derived from eq. (II.3.3-8) using $t_A=t_B-T_{AB}$

$$p(T_{AB})=\sum_{\{\eta \mid \substack{-\delta t_B<\eta<\delta t_B \\ \bar{t}_A-\delta t_A<\bar{t}_B+\eta-T_{AB}<\bar{t}_A+\delta t_A}\}} p_A(\bar{t}_B-T_{AB}+\eta)\cdot p_B(\bar{t}_B+\eta)\ , \qquad \text{(II.3.3-10)}$$

imposing further restrictions for the range of η: A deviation of T_{AB} from $\bar{T}_{AB}$ has the effect, that the condition for the arguments of p_A and p_B to lie within the uncertainty intervals

$$\begin{gathered}-\delta t_B<\eta<\delta t_B \\ \bar{t}_A-\delta t_A<\bar{t}_B+\eta-T_{AB}<\bar{t}_A+\delta t_A\end{gathered} \qquad \text{(II.3.3-11)}$$

is less often met. The quantity of events η is the highest for $T=\bar{T}_{AB}$ and is equal or less for all other T. Fig. (II.3.3:2) shows a discrete example for the frequency distribution of events over the interval $[T_{min},T_{max}]$.

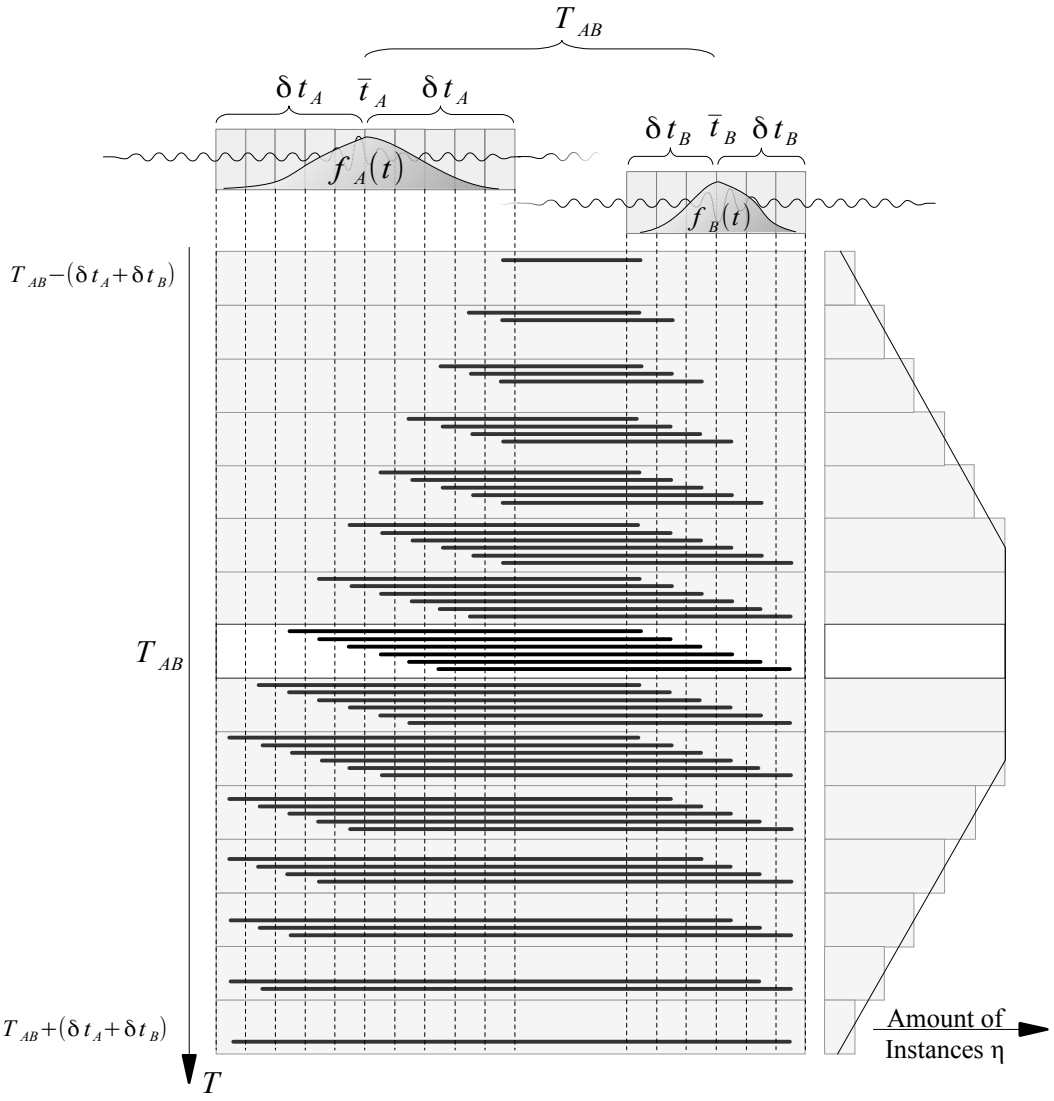


Fig. II.3.3:2: Discrete frequency distribution of events (i.e. the occurrence of a certain arrival-time difference between the intervals) in probabilistic interpretation. The continuous signal and level-of-confidence function is discretized into time bins (samples). This yields a quantized shift η *and a quantized amount of this shift to be found for the given arrival-time in the interval bounds. The arrival-time difference between the interval-midpoints (mid-samples, respectively) is statistically found most often. This causes a triangular frequency-distribution, showing a central plateau for intervals of different length.*

With constant level-of-confidence curves over two uncertainty intervals of equal length, the frequency distribution of η would cause a triangular constraint probability distribution. For dissimilar arrival-time uncertainties $\delta t_A \neq \delta t_B$, the frequency distribution of η causes a central plateau to appear in the probability distribution (see fig. II.3.3:2). In more complex distributions this effect causes the central region to be stretched out. This is generally observed with S-P constraints whose weight appears "flat" over the constraint's middle axis only fading out at it's edges (see fig.

II.3.3:5b), due to the S-phase uncertainty being at least twice as high as the uncertainty of the corresponding P-phase.

Reformulating eq. (II.3.3-10), using explicit summation boundaries that satisfy equation (II.3.3-11) allows to compute the discrete distribution p_T for the arrival-time difference $T_{AB}=t_B-t_A$, based on both finite level-of-confidence curves. These are represented by the discrete distributions p_A and p_B with lengths n_A and n_B, weighing each sample. Following the scheme displayed in fig. II.3.3:2, the vector representing the distribution p_T with length $n_T=n_A+n_B$ is given in

$$p_T[i]=\sum_{d=0}^{D(i)} p_A[a_{min}(i)+d]\cdot p_B[b_{min}(i)+d] \quad \text{with} \quad 0<i<n_A+n_B,\ i\in\mathbb{N} \tag{II.3.3-12}$$

and

$$a_{min}(i)=\begin{cases} n_A-i, & n_A\geq i \\ 0, & else \end{cases} \tag{II.3.3-13}$$

$$b_{min}(i)=b_{min}(i,a_{min})=a_{min}+i-n_A \tag{II.3.3-14}$$

$$b_{max}(i)=\begin{cases} n_B, & m_B<i \\ i, & else \end{cases} \tag{II.3.3-15}$$

$$a_{max}(i)=a_{max}(i,b_{max})=n_A-i+b_{max} \tag{II.3.3-16}$$

$$D(i)=a_{max}(i)-a_{min}(i) \tag{II.3.3-17}$$

with a_{min} being the minimal possible argument of p_A at position i, b_{max} being the maximal possible argument of p_B for given i, b_{min} the minimal possible index for given a_{min} and i, and a_{max} the maximal possible index for given b_{max} and i. The position i (i.e. the i^{th} vector component) relates to the parameter T_{AB} via $T_{AB}=\bar{T}_{AB}+(2\cdot i/n_T-1)\,\delta T_{AB}$, wherefore

$$p(T_{AB})=p_T[\,(\frac{T_{AB}-\bar{T}_{AB}}{\delta T_{AB}}+1)\cdot n_T/2\,]\,. \tag{II.3.3-18}$$

The mathematical operation behind eq. II.3.3-12 stands out, when we rewrite equation (II.3.3-10) to

$$p(T_{AB}) = \sum_{t=-\infty}^{\infty} p_A(t - T_{AB}) \cdot p_B(t) \quad \text{with} \quad t = \bar{t}_B + \eta \,. \tag{II.3.3-19}$$

The reformulation of the summation boundaries is valid, since p_A and p_B are vanishing outside their intervals (II.3.3-5) - (II.3.3-6). This form corresponds to the mathematical definition of the real valued discrete correlation

$$\{f * g(-(.))\}(k) = \sum_{i=-\infty}^{\infty} f(i) g(i-k) \,, \tag{II.3.3-20}$$

which in statistical inference describes the difference Z=F-G (here $T_{AB} = t_B - t_A$) of two discrete random variables F and G (while the convolution $f * g$ describes their sum Z=F+G, as can be easily derived, defining $G' = -G$). This operation naturally preserves the constraint error $\delta T_{AB} = (\delta t_A + \delta t_B)$ as laid out in chapter (II.3.1). The level-of-confidence curve indicates the likelihood for a phase arrival to be found at a certain sample (by definition this sample is contained in the chosen uncertainty interval). The level-of-confidence is therefore normalized over the uncertainty interval

$$\int_{t-\delta t}^{t+\delta t} \pi(\tau) \, d\tau = \sum_{\tau} p(\tau) = 1 \quad . \tag{II.3.3-21}$$

Again, the operation of correlation (of both level-of-confidence functions over their intervals) preserves this quality, yielding a normalized distribution. Mapping the likelihood distribution of T_{AB} over the constraint's width to space, ascribes these likelihoods to the covered regions, expressing how likely the constraint values them to be part of the hypocenter region (see fig. II.3.3:4). While the level-of-confidence functions are usually discrete (weighing each sample in the uncertainty interval surrounding the pick), analytical level-of-confidence functions may be used instead, which allows to solve the problem analytically and speed up calculation time. The constraint likelihood function is then calculated using the continuous form of the cross-correlation,

$$\{f * g(-(.))\}(\tau) = \int_{-\infty}^{\infty} f(t) g(t - \tau) dt \,. \tag{II.3.3-22}$$

This allows to enhance the detail (resolution) with which the weighted constraints can be mapped. For the discrete numerical code, the computed likelihood function is finally sampled back into a discrete form with probability slot lengths usually corresponding to the waveform's inverse sample rate f, weighing each sample using

$$p_{mf}(t) = \int_{t-\delta t}^{t+\delta t} p_{df}(t)\,dt \qquad \text{with} \qquad \delta t = \frac{1}{2 \cdot f} . \tag{II.3.3-23}$$

Fig. II.3.3:3 displays the constraint likelihood function for different types of the parameter's level-of-confidence functions p_A and p_B. The constraint weights in form of triangular and gauss functions having been mentioned in the beginning of the chapter are special solutions of the general probabilistic approach, which may result in complex distributions (fig. II.3.3:3d).

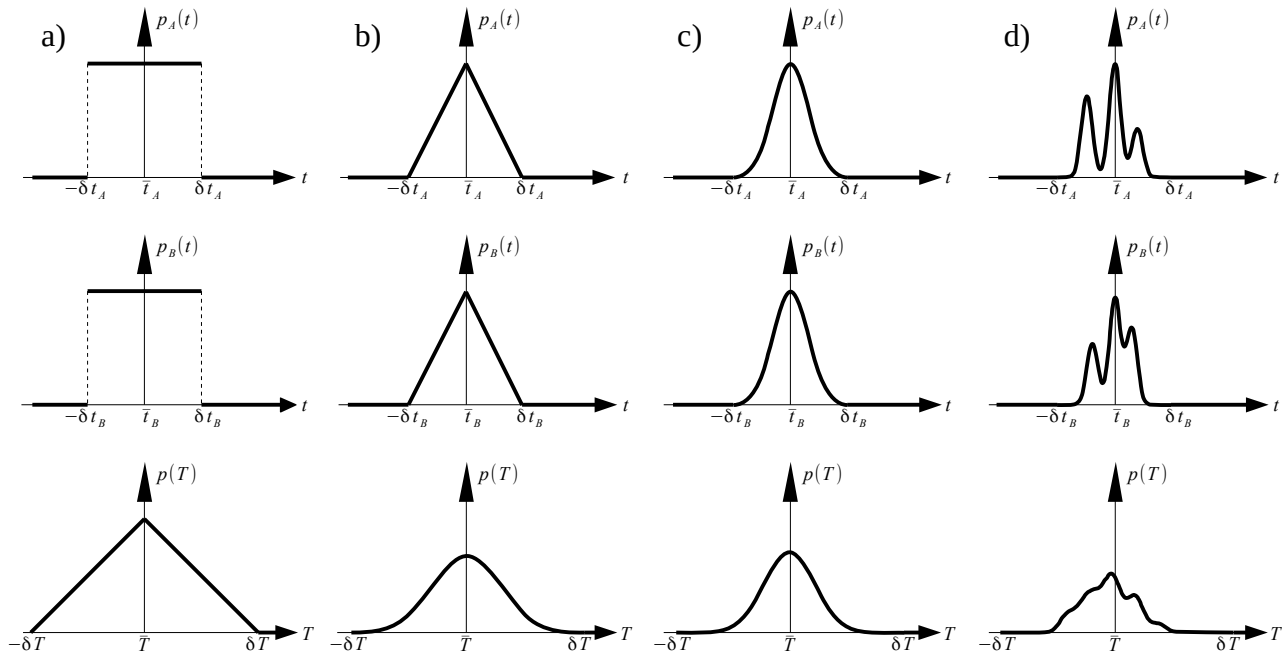


Fig. II.3.3:3: Cross-correlation of level-of-confidence curves p_A and p_B for a) constant levels of confidence yielding a triangular likelihood function b) triangular distributions yielding a Gaussian-like shape c) Gaussian distributed levels of confidence yielding a Gaussian distribution d) complex level-of-confidence distribution as given in a multi-pick-analysis (chapter II.3.4) yielding a irregular distribution. The upper two rows relate to the parameter uncertainty distributions; The bottom row displays the resulting constraint likelihood curve. The ordinate scaling varies over a)-d).

Fig. II.3.3:3(a) displays the advantage of the probabilistic analysis over the approach of chapter II.3.1, which solely mapped parameter uncertainty intervals (unweighed picking errors) into space to identify regions of congruity as physically possible solutions, but disregarded how likely they are: Such uncertainty intervals would correspond to a flat weight as in fig. II.3.3:3(a). Where the mapping of chapter II.3.1 yielded a (flat-weighed) spatial constraint (e.g. fig. II.3.3:5a), the probabilistic approach shows that even in this simple case the constraint's edge is highly unlikely to mark the hypocenter location (fig. II.3.3:5b).

A typical level-of-confidence curve for emerging signals is shown in fig. II.3.3:4, displaying negative-skewed distributions. These account for the fact that the point at which the signal arises from the noise floor cannot be determined as precise as the latter end of the uncertainty interval.

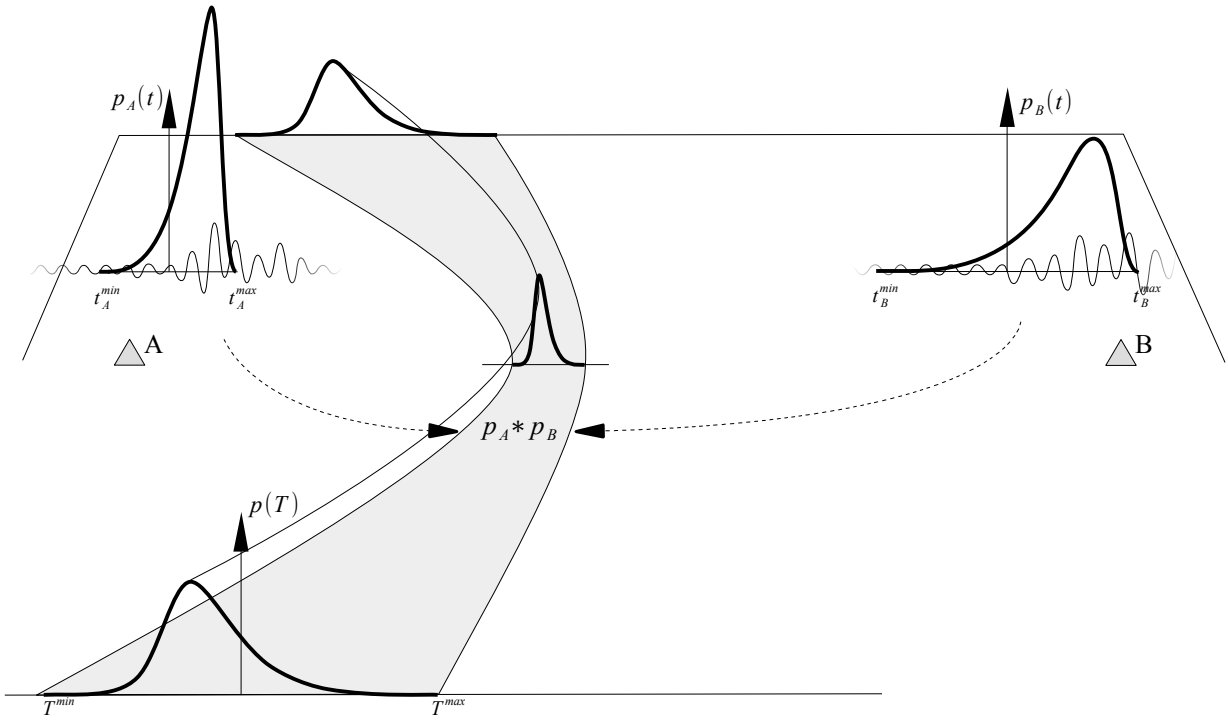


Fig. II.3.3:4: Two typical negative-skewed level-of-confidence distributions at station A and B construct a positive-skewed constraint likelihood function. In this example the picking uncertainty at distant station B is twice as large as at station A. The skewed level-of-confidence distribution is a typical shape for emergent signals, reflecting the difficulty of pinpointing the phase onset emerging from the noise. The uncertain constraint is weighed by the likelihood function over it's deviation axis, forming the probabilistic constraint.

The constraint's likelihood function (as in fig. II.3.3:4) describes the likelihood for a given location to correspond with the hypocenter location. Outside the constrained space this likelihood is zero.

With several given constraints, the global likelihood function will be defined as the sum of the individual constraint likelihoods, following the concept of congruity as used for uncertain constraints.

The example of the skewed level-of-confidence distribution (fig. II.3.3:3) shows that the parametrization used for the derivation, i.e. onset time (interval midpoint) and measurement error, now lose their physical meaning as the maximum of the level-of-confidence curve might be significantly dislocated from the interval's center. Rather than using the measurement and measurement error as input parameter representing the observables (e.g. eq.II.3.1-1), we will generalize the concept by directly using the boundaries of the level-of-confidence interval to represent the parameter uncertainties

$$t = t \pm \delta t \;\rightarrow\; t \in [t_{min}, t_{max}] \tag{II.3.3-24}$$

and likewise for the constraint parameter

$$T = \bar{T} \pm \delta T \;\rightarrow\; T \in [T_{min}, T_{max}], \tag{II.3.3-25}$$

while using their distributions over the interval to describe the corresponding likelihoods.

Fig. II.3.3:5 shows an example using a constant level-of-confidence function over the parameters' uncertainty intervals. This allows us to compare the results of two approaches, based on the same information: a) uncertain constraints, yielding “flat weights” solely highlighting the regions of physically possible solution regions and b) probabilistic constraints, representing the likelihood of such solutions. The probabilistic approach yields a triangular constraint weight having its maximum in the constraint's center, while falling to a zero-likelihood at the constraint's edge. This reveals that the solution at the constraint boundaries are unlikely to mark the hypocenter location. The likely solution interval for several given constraints is therefore a subset of the congruent region.

a) uncertain constraints b) probabilistic constraints

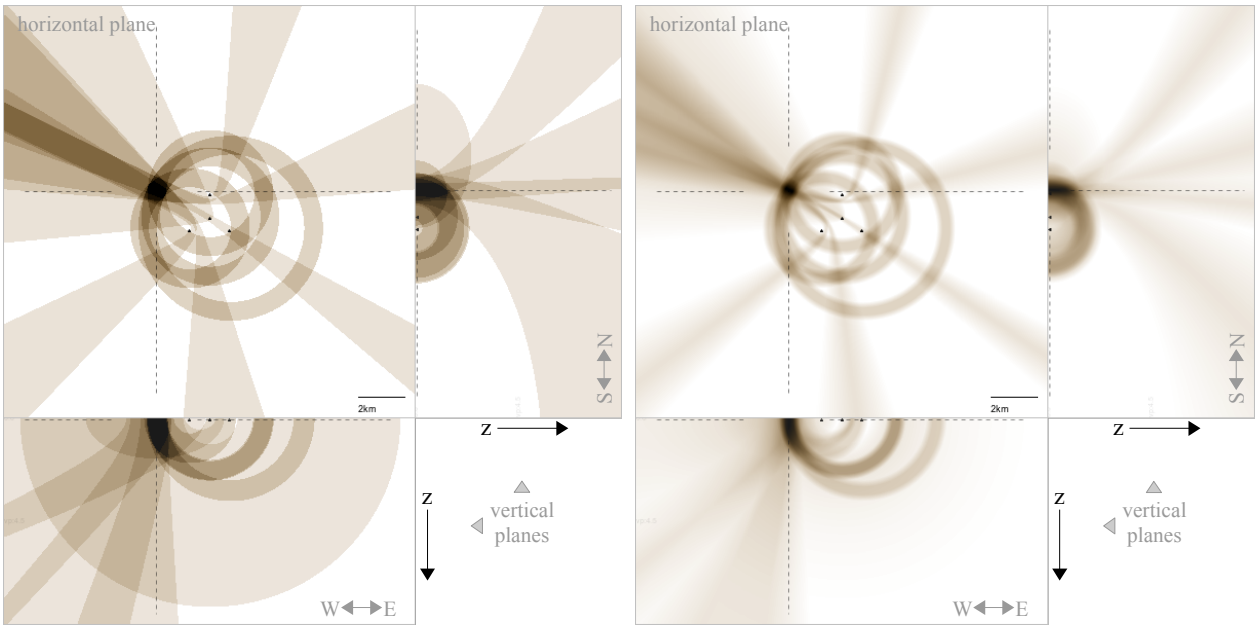


Fig. II.3.3:5: Comparison of S-P and P-P constraints with a constant level-of-confidence distribution over the uncertainty interval, using uncertain constraints (a) and probabilistic constraints (b). The point of highest probability is found within the region of congruity. While P-P constraints show a single peak in the likelihood function, S-P constraints show a wide plateau (they appear "flat") as the uncertainties for S onsets are significantly larger than the P-uncertainty (see example of fig. II.3.3:2).

Probabilistic constraints are an instrument to refine the solution obtained by uncertain constraints. The underlying concept of the level-of-confidence, however, opens doors to several advanced approaches like weak constraints and a multi-pick analysis, yielding complex multi-peak constraint weights by the use of irregular level-of-confidence functions, as laid out in the next chapter.

II.3.4 **Variable Weights and Multi Pick Evaluation**

In the last chapter we laid the theoretical foundation for the handling of timing uncertainties concerning a single phase. This phase-onset uncertainty, mainly used to account for the picking uncertainty, further allows to account for e.g. the time-drift of unsynchronized data-loggers or the data-loggers internal timestamp error, which both affect class II constraints (using the data of two stations). In this chapter we will build on last chapter's concepts to handle the uncertainty that arises when the true phase of interest cannot be distinguished from multiple ambiguous, equally-possible phases or when the type of the arrived phase is unclear. This is a common problem with low SNR data, where the true phase of interest can often not be discerned from several other possible phases in the vicinity of the true onset, often related to local noise bursts. Let fig. I.1:1 of the introduction chapter serve as example.

We will use three concepts to improve the location accuracy when phase onsets are questionable, often even yielding precise hypocenter locations, even though phases may not be properly identified. The foundational concept behind this is, that the true set of correlated phases must share a common source region: The location can therefore serve as a parameter itself, feeding back into the phase identification process. The second concept lies in the fact that the information about the hypocenter region cannot be falsified by extending the region (increasing the location uncertainty, decreasing it's precision). This follows the fundamental demand for a location analysis: To display the true state of knowledge concerning the location region, weighing accuracy higher than precision. This can be achieved by using uncertain parameters: As long as the data interval (e.g. the lower and upper time limit around the (unknown) phase) can properly be determined, the data is accurate, even though the true datum itself is not precisely known. When the true data is enclosed in the data intervals, the corresponding solution interval consequently has to be accurate, enclosing the true hypocenter. So, to obtain accurate results the inversion has to use all possible data, rather than using a (spuriously) precise phase information. This in itself will expand the location region and can therefore only increase the accuracy of a result, and never diminish it. However, as this decreases the results precision, the solution may become meaningless if its uncertainties are too large to draw useful conclusions concerning the source.

The challenge of the location problem therefore lies in reducing the full space to a meaningful

subset thereof by subsequently using additional information, without erroneously eliminating the true (but unknown) location from the solution interval (which would produce an inaccurate result with a spurious precision). Countering the second concept, which assures accuracy but lowers the precision, the third concept will increase the precision by using the fact, that not all given information will lead to physically possible solutions. Where the second concept enlarges the solution interval (i.e. the hypocenter region), the third one diminishes it.

When now *n* competing phase onsets are found at one station, they could individually be evaluated by solving the location problem *n* times, each time for a different phase candidate. Using the fact that the true set of correlated phases must share a common source region, we could evaluate which candidate yielded a solution of high congruity in conjunction with the other data. Such a solution would identify the used phases as the set of correlated phases. If multiple phase combinations yielded highly congruent solutions, then their distribution will represent the location's macro-uncertainty due to unknown phase onsets or types. However, when ambiguous phases are found at more than one station (e.g. fig. I.1:1) such an approach would require a large amount of inversions to identify the most plausible location. Instead, we can use the concept of the level-of-confidence (LOC), which provides us with a framework that allows to evaluate all possible phase combination within a single inversion. Hereby, the location constraint is formulated in a way that may decrease its precision but will guarantee its accuracy, meaningly forming an extended region that ensures to enclose the hypocenter that is constructed by the true set of correlating phases. The simplest way to obtain such a result would be to increase the onset uncertainty window until it encompassed all possible phases. So formed congruent regions would indicate the extent of the location region, but their large size would most probably prohibit to draw useful conclusions concerning the event. More detailed information can be obtained by extending the concept of confidence level of single onsets for multiple phase onsets. To do so, we will weigh all phase candidates within one LOC interval, which will describe the macro likelihood for the phase arrival (fig. II.3.4:1). Contrary to the single pick approach, where the LOC-curve $p(t)$ has *one* clear maximum, here *several* phases will be weighed using comparable weights over a larger uncertainty window $[t_{min}, t_{max}]$, encompassing all phases (fig. II.3.4:1).

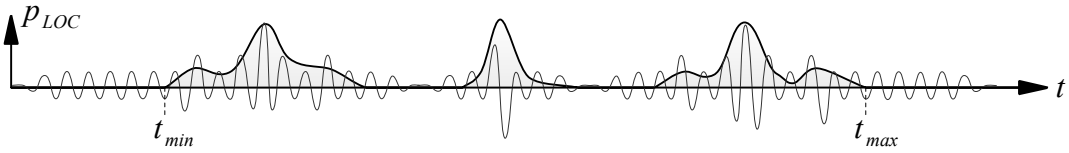


Fig. II.3.4:1: Multi pick analysis: Level of confidence defined over several ambiguous phases inside the interval $[t_{min}, t_{max}]$.

The constraint is therefore no longer observed and treated as an object relating to the time difference of two distinct observables. The constraint parameter T, expressing the arrival time difference, will be rather interpreted as a parameter of a family of surfaces $S(x, y, z, T)$ which span over the whole three-dimensional Euclidian space. If the LOC-interval would be used to describe e.g. ambiguous P-phases, $S(T)$ may be thought of the set of all possible hyperboloids that can be formed over the interval of the minimal and maximal physically possible T, occupying the whole three-dimensional space. (That $S(T)$ indeed occupies the whole space becomes obvious when the problem is reversed: Given a seismic source, a forward travel time calculation for the two involved phases would determine an arrival time difference, defining parameter T^*, which in turn causes the surface $S(x, y, z, T^*)$ to pass through that coordinate. As the seismic source could be located at any coordinate in space, S must consequently fill the whole space.) Via T, certain regions (modes) in space can now be weighted with values between 0 and 1 according to the likelihood function obtained from the cross-correlation of the involved LOC functions. With an interval $[T_{min}, T_{max}]$ (which is defined by the two time series of ambiguous phases), the set of T-modes outside this interval will be muted, shrinking the solution space to a subset of full space. Over all T within the interval, now regions in space are weighed according to the constraint's likelihood function, which is determined by the level-of-confidence being defined over the questionable phases (fig. II.3.4:2).

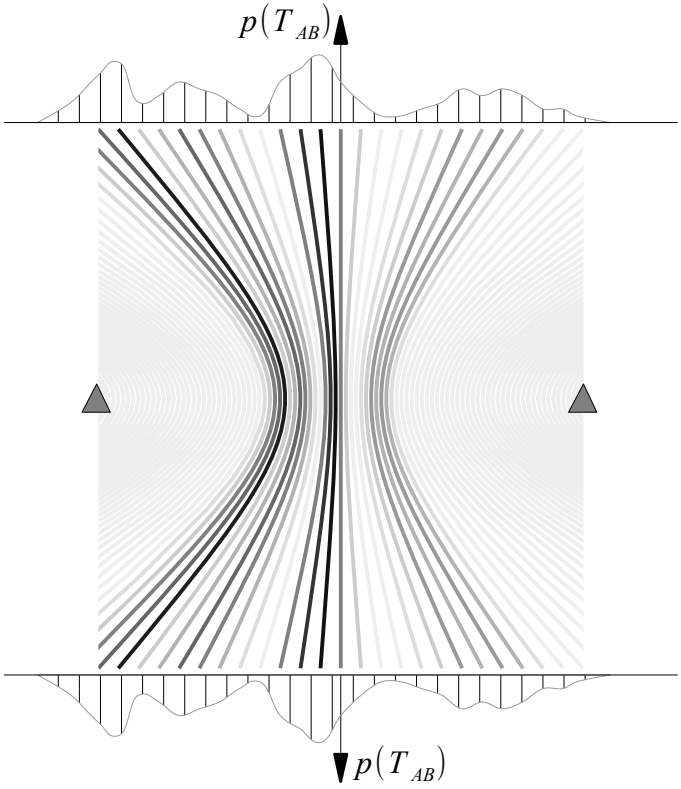


Fig. II.3.4:2: Hyperbolic T-Modes between two stations. The weight of the modes $p(T)$ is given by the two observables' LOC-curves.

The location and orientation of the T-modes depends on the geometric layout of the involved stations. With multiple LOC-intervals given, the different and differently oriented modes will superpose (as observed for classic constraints) and highlight regions of congruity. The resulting map displays the spread and clustering of maximum likelihood zones and is equivalent to the combined representation of all possible combinations of onset choices, as if evaluated via single onset picking (fig. II.3.4:3 (a)=(c)+(d)). This scheme provides therefore a complete view on all location possibilities, with the scatter of congruent regions displaying the macro-uncertainty in location due to the unknown or ambiguous phase.

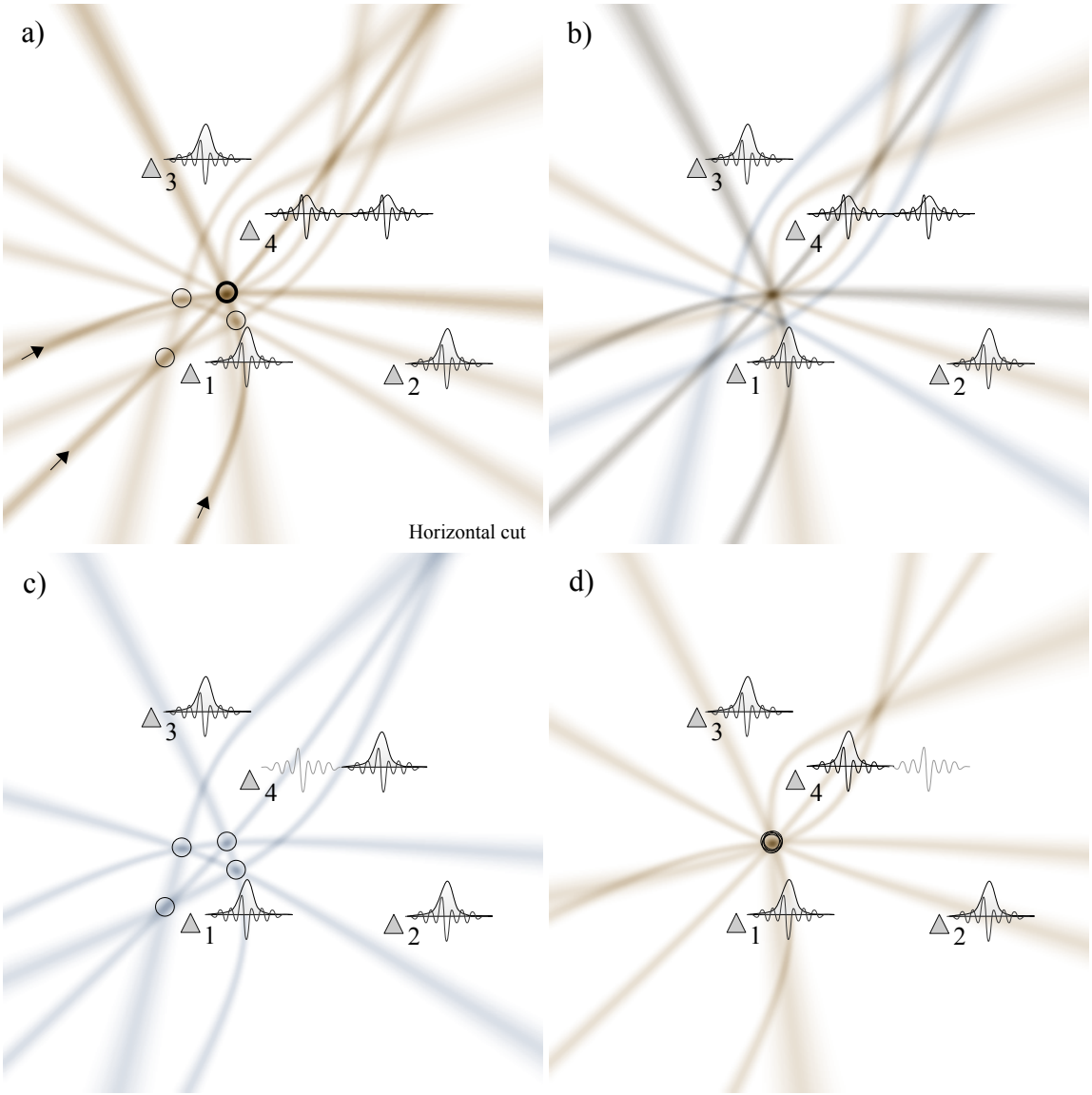


Fig. II.3.4:3: Example of four stations in a P-phase based location, with station 4 showing two ambiguous phase arrivals. a) Weighing both phases equally with a 0.5 LOC yields three weak constraints (those being based on the 0.5 LOC phases) and three strong constraints (those being based on the unique phase arrivals with a 1.0 LOC, marked by arrows). This yields 4 solutions, three being weak (being dominantly determined by weak constraints) and one showing a high congruity, marking the best fit hypocenter location. This corresponds to the superposition of the results of two individual location runs, one for each possible phase candidate: c) choosing the second candidate as the suspected phase and d) choosing the first candidate. The superposition of both individual results (panel b) corresponds to the one of a single run with two 0.5 LOC curves.

Since the true phase of interest is expected to be among the ambiguous phases contained in the LOC-interval (and can only be one of them), the LOC-curve is normalized (eq. II.3.3-21).

$$\int_{t_{min}}^{t_{max}} \pi(\tau)\, d\tau \;=\; \sum_{i}^{n} \{ \int_{t_{min,i}}^{t_{max,i}} \pi_i(\tau)\, d\tau \} \;=\; 1 \qquad \text{(II.3.4-1)}$$

with the sub-interval over each phase onset holding one n -th of the total likelihood

$$\int_{t_{min,i}}^{t_{max,i}} \pi_i(\tau)\, d\tau \;=\; 1/n \qquad \text{(II.3.4-2)}$$

with n being the amount of ambiguous phases. This normalization brings in an interesting concept, which can be visualized if we interpret the T-modes again classically as location constraints in a P-phase based location: Let e.g. four station be given of which three have weak but unique phase onsets while the fourth one shows two possible onset candidates due to local noise spikes (fig. II.3.4:3). For the first three stations the LOC-curve only spreads over one phase, which due to the normalization causes all location constraints based on these phases to obtain a 100% weight. At the fourth station the 100% probability of the onset to be found in the LOC-interval distributes over two onsets, both sub-intervals holding a 50% chance to contain the phase arrival of interest. The P-P constraints being based on station pairs [1,2], [1,3] and [2,3] will therefore each hold a 100% likelihood to contain the hypocenter location (since a cross-correlation of two normalized distribution yields a normalized distribution). The constraints based on pairs [1,4], [2,4] and [3,4] will each only yield a 50% likelihood. This expands on the concept of flexible constraints (chapter II.3.2), applying different weights to the constraints by the level of trust in their underlying data. In this example, this leads to one strong solution given in the exact intersection of the first three constraints (being based on trustable data given in unique onsets), and three weak solutions for each of the two ambiguous onset candidates. Here, each weak solution depends on two weak constraints, which each depend on one ambiguous phase. Compared to the strong solution, the weak solutions only hold a 2/3rd likelihood to contain the location that corresponds to the true phase-arrivals. Since the method determines the possible solutions for all arrival time combinations, at least one maximum corresponding to a high congruity of phase information should be found among the solutions, corresponding to the set of correlated phases.

In this example, from the -in total- seven sub-solutions, five are closely clustered while only three weak ones are scattered – the three that utilize the incorrect phase onset. A following forward travel time calculation from the likely solutions to the stations identifies the set of correlated phases by the predicted arrival times. This demonstrated that we can use the fact that the corresponding phases originate from the same source location, to identify the location even though ambiguous or questionable phases are given. Since it is the velocity model that links the corresponding phases by

their source location, this method requires an accurate model or the treatment of model uncertainties in the inversion (as will be introduced in chapter II.3.5.1).

The concept of weak&strong constraints, i.e. applying different weights to different phases (using non-normalized LOC-intervals) is also useful in standard scenarios with unique onsets, to weigh the constraints and solutions according to the trustworthiness of their underlying data. Now, however, we will need to make the concept of weighted (weak & strong) constraints void to identify the set of correlated phases: Probabilistic constraints point out the likely location but do not inform about the congruity of the solution, i.e. how many LOC-intervals support this solution. All locations which are supported by any of the ambiguous phase arrivals in one LOC interval should yield a congruity of $c=1$ as expected for non-probabilistic constraints. This can be achieved by transforming these, by cross-correlation of LOC-curves obtained multi-peak probabilistic constraints, into multi-peak flat-weight constraints (fig. II.3.4:4)

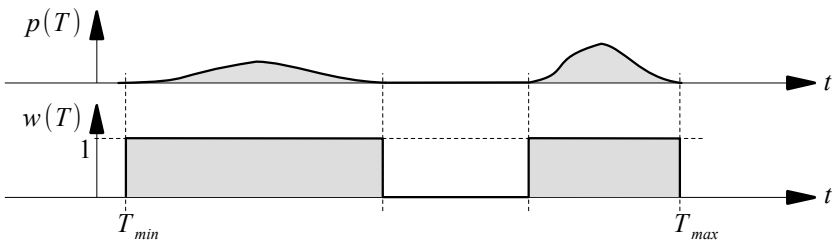


Fig. II.3.4:4: Heaviside transformation of p(T) yielding w(T). w(T) corresponds to a flat-weighed uncertainty based constraint for ambiguous onsets.

using the Heaviside transformation

$$w(T)=H[\,p(T)\,] \qquad \text{with} \qquad H(x)=\begin{cases}1, & x>0\\ 0, & else\end{cases}. \tag{II.3.4-3}$$

Applying this transformation shows that the determined hypocenter location in fig. II.3.4:3 is not only likely but holds a congruity of $c=1$, being confirmed by all LOC-intervals.

When multiple ambiguous phase arrivals are observed at several stations, there is a chance that the ambiguities are not due to uncorrelated local noise spikes but that simultaneous events were

recorded. In such a case more than one region of high congruity may be found. A forward travel time calculation from each congruent region would again identify the sets of correlated phases. However, ambiguities may also lead to (random) congruent regions that aren't caused by a common seismic source at the indicated location, but are merely coincidental (mathematically possible) solutions of uncorrelated phase arrivals and station geometries. If such location ambiguities cannot be resolved (e.g. by additional amplitude criteria evaluating the consistency between distances and the attenuated observed amplitudes), they must be accepted as macro-uncertainty due to unknown or ambiguous phases. Here, however, the third concept initially mentioned may often reduce this uncertainty and the amount of possible solutions. In chapter II.3.4 we laid out, that only physically possible solutions, i.e. arrival time differences consistent with the velocity model can be mapped. Arrival-time differences too large, will be “filtered out” by the model. This quality of “filtering” supports the identification of inconsistent phase onset combinations: Unphysical combinations of phases are often “filtered out” and the true combination of arrival phases yields often locations of higher congruity and likelihood than other phase combinations. This supports to identify the true set of correlated phases and the most probable location. The scenario of the introduction chapter (I.1) can serve as example (fig. II.3.4:5): There, six stations with multiple ambiguous weak phase arrivals are given. The low SNR does not permit a precise phase packing in the velocity waveform and several weak phases can only be identified using sonograms, a noise-adapted spectral representation of the data. These however, allow only for a rough estimation of the onset due to their low time resolution.

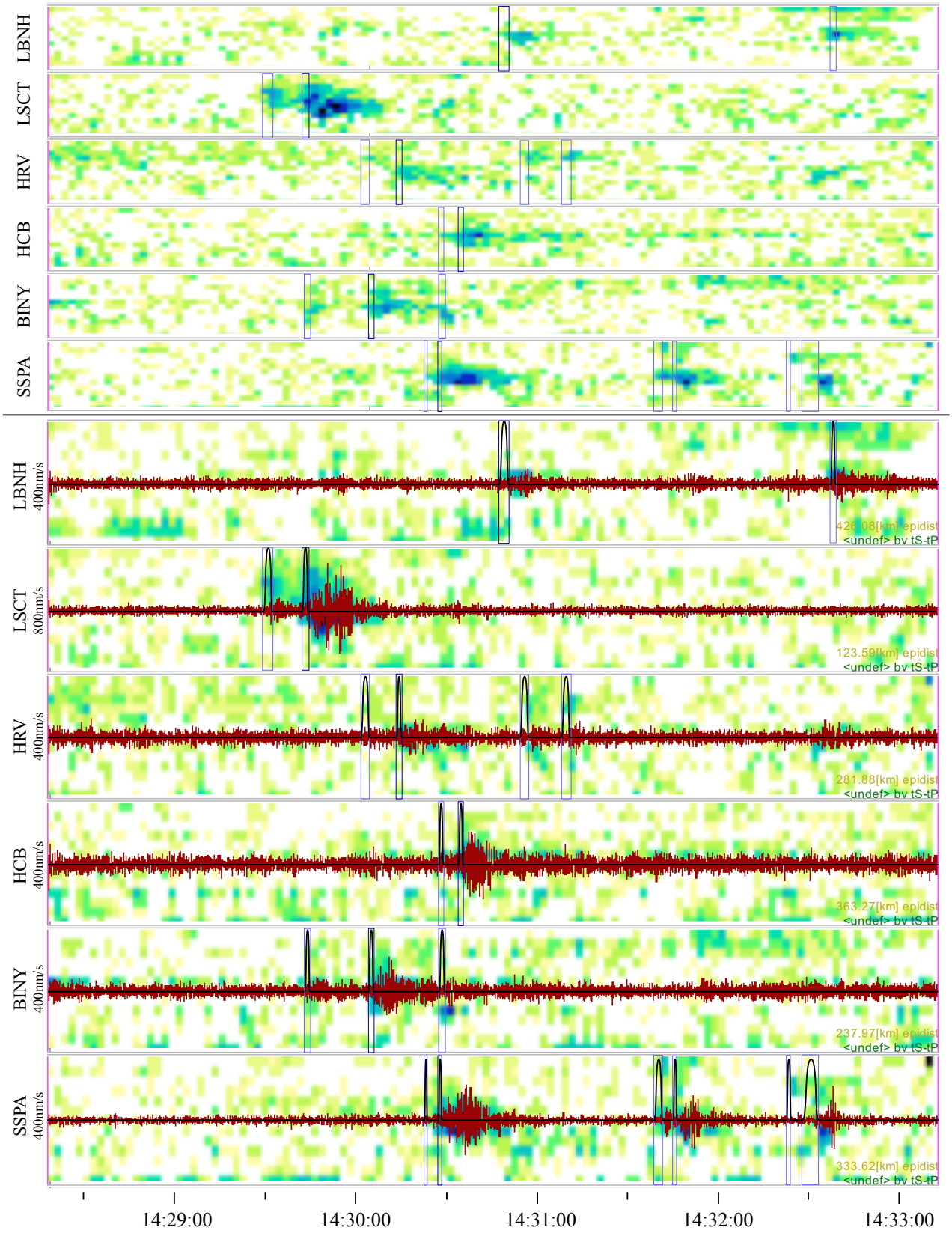


Fig. II.3.4:5: Sonogram (i.e. noise-adapted spectrogram) traces (top) and seismogram traces (bottom, red) of 6 stations showing ambiguous phase-arrivals. Blue boxes indicate onset uncertainties (top). The LOC curves (black solid lines, bottom) extend over all phase candidates at given station. Several onsets can only be identified in the sonograms.

Representing the phase arrival at each station, we can form one LOC-interval stretching over all ambiguous phases where it shows distinct maxima. According to all possible phase-combinations, several regions in space show various levels of congruity. Regions of high congruity are found dominantly south of the stations, indicating the region in which the true location is expected to be found (The true set of correlated phases will yield (at least) one congruent region. Exchanging one of these phases for another phase candidate will divert the solution from that region and likely reduce the congruity: Therefore clusters of (often less congruent) solutions are expected around the true location).

Choosing the region of highest congruity, we can identify the set of corresponding phases by a forward travel time calculation to each station. Isolating this congruent patch by now only choosing the identified phases, yields a laterally and in depth well constrained region, holding the true event location.

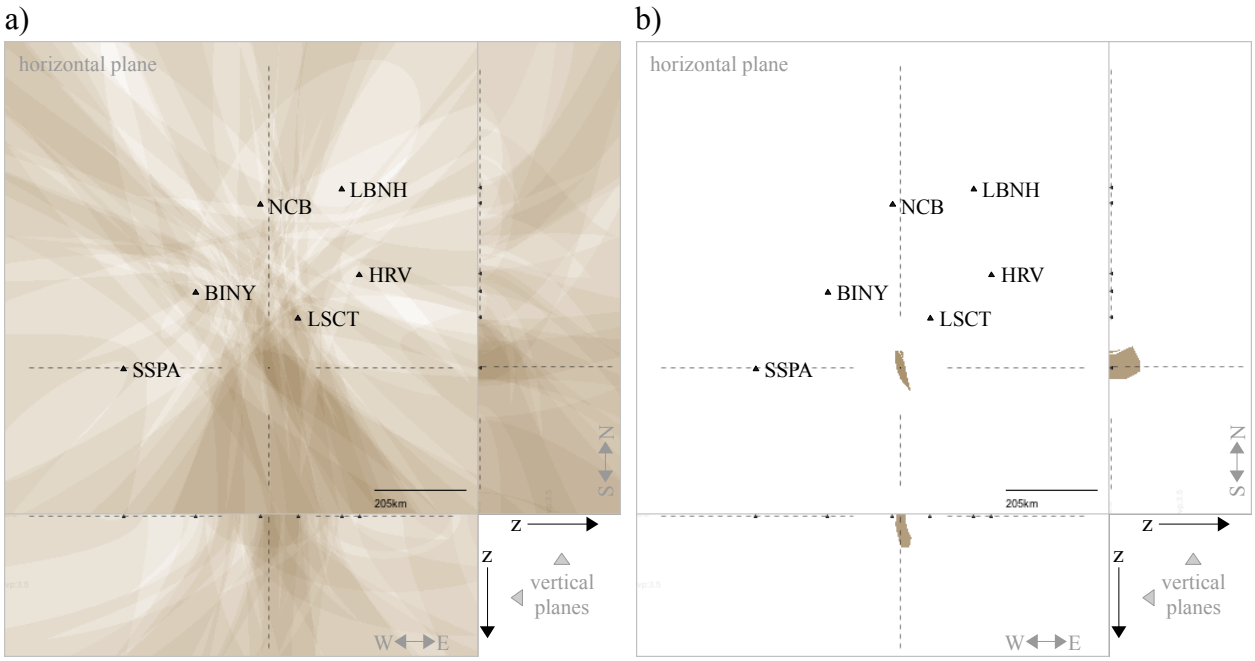


Fig. II.3.4:6: a) Display of all T-modes obtained from the given LOC-curves. The result shows a region of higher congruency south of the network and b) only showing the region of highest congruency. The ground-truth location is enclosed in the identified region.

This example is in detail outlined in chapter V.1. A multi-pick analysis may be interpreted as explicit and pick-based form of the delay-an-sum method “source-scanning” (Kao & Shan, 2004). Source-scanning uses synthetic arrival times pertaining to trial hypocenters and trial origin times to sum up the corresponding portions of signals recorded at several stations. The normalized sum,

mapped over space and time, illuminates the space time distribution of seismic energy and allows to identify seismic sources by searching for the "brightest" region – the location and origin time for which the synthetic arrival times fall on the higher event amplitudes in the signal. The advantage lies in the fact that -by only summing up the observed amplitudes- it does not rely on phase-picks and is therefore beneficial for emergent signals like tremors. The multi-pick analysis on the contrary requires- and is solely based on phase-pick data (and ignores amplitudes). The information contained in the picks (determining phase type & arrival time) allows an explicit and direct calculation of this "brightest" spot, eliminating the unknown quantity *time*. It should generally yield lower spatial uncertainties as side maxima that stem from inconsistent combinations of different phase types are avoided.

A multi-pick analysis can be used to display the macro uncertainty in two ways. In the first (type I) previously laid out, multiple phase onsets were declared as possible onsets for *one phase type* (e.g. P-phase). The second way (type II), is to declare *one phase onset* as multiple phase types (e.g. Pn *and* Pg). This would allow the proper treatment of onsets for which the association is unclear: Any method based on phase-picking faces the danger of wrong phase type association, i.e. mistakingly assigning the wrong phase type (e.g. P or S) for a detected onset. For instance, if the amplitude of the Pn-phase of a distant source falls below noise level, the following Pg phase might be wrongly misinterpreted as Pn phase. Consequently, wrong ray-paths and travel-times would be associated and the location wrongly constrained. For a S-P constraint, being based on the onsets of the questionable P-phase and the following S-phase, this error is expected to have a far stronger impact on the location than a simple mis-pick in the vicinity of the true phase. Obviously, in the given example, the suspected but invisible Pn-phase cannot be picked. However, the identified P-phase can be picked both, as Pn-phase as well as Pg-phase (type II). In the same manner, for a clear P-onset and a questionable S-onset with two possible candidates, the Level-of-Confidence curve can be raised over both S-candidates to identify both equally, as equivalent S-phases (type I). In both cases, this S-P constraint would appear as two concentric spheres of different radius (rather than one). The one sphere will correspond to the classic S-P constraint based on the true set of P- and S-onsets. The other sphere would mark the region based on the misidentified phase. This extended region consisting of two spheres, reflects the macro uncertainty due to uncertainties in the phase type association, and ensures that the constraint will enclose the hypocenter. In conjunction with several constraints, one solution will usually be disqualified while the other will be confirmed by

the solution trend of the whole data set. In this way, scattering solutions with a lower congruity are avoided and the location problem is formulated robustly according to the true state of knowledge. Again, as the solution (or the spread of solutions) depends much on the chosen velocity model, it is of fundamental necessity to properly address model uncertainties. This aspect, which is of major importance for many applications, shall be outlined in the following chapter (II.3.5).

II.3.5 **Effect of Velocity Model Uncertainties**

Previous chapters addressed the uncertainties of the time-domain based observables, by which accurate location estimates can be obtained, if accurate velocity models are used. This, however, is seldom the case and the location-error due to wrong model assumptions often outweighs the error that is caused by timing. The location estimate, which strongly depends on the chosen model, is often significantly displaced due to inaccurate models. While for events within a network, model-inaccuracies mainly affect the hypocentral depth, they displace the hypocenter estimate significantly stronger, laterally and vertically, if the event is located outside of the network (fig. II.3.5:1).

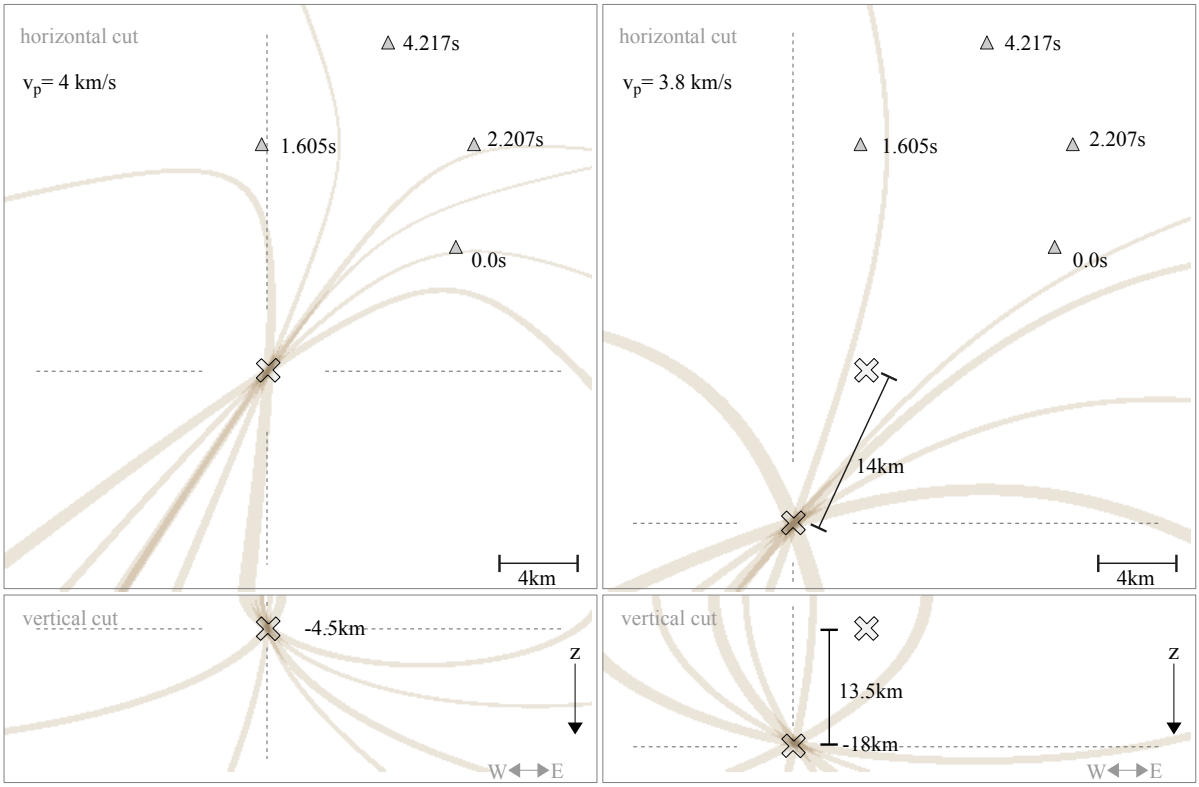


Fig. II.3.5:1: The two panels (left) and (right) show the sensitivity of the hypocenter location estimate towards model assumptions: The same arrival times (simulated for an event in a homogeneous model of vp= 4km/s) are inverted using (left) the proper model and (right) a model estimate which underestimates the true velocity by 5% (vp= 3.8km/s). The left panel therefore displays the ground-truth using the accurate (but in practice unknown) model. The right panel on the contrary represents the scenario that we may face in practice, with a location estimate based on our (always inaccurate) model choice. Depending on network geometry and source location, the discrepancy between the true source location (left) and the estimated solution based on the model (right) can be significant; In this example the discrepancy in hypocenter location is approximately 20km. The top panels display horizontal cuts showing the discrepancy in epicenter location. The bottom panels show the corresponding vertical cuts, revealing the discrepancy in focal depth. The arrival times are noted to the right of the respective station and the picking uncertainty is uniformly 0.005s.

The example shows that small inaccuracies in the model may significantly displace the hypocenter location. Unfortunately, the knowledge concerning the constitution of the sub surface is generally (very) limited, which brings us to the realization that the locations which we obtain from time based observables may well be biased by model errors. Especially for regions for which little data exists, analysts have to default to simple velocity models, particularly 1-D layered models. If enough data (events and stations) are given, Minimum-1D models in conjunction with station corrections (Kissling, 1988, 1995) can be used to find an optimal 1D-model, reducing the overall RMS misfit. As parameters are averaged over extensive areas, and may only poorly reflect the true local velocity configuration, station corrections are used to reduce the corresponding error. Yet, any derivation from the "true model" displaces the hypocenter location wherefore accurate results can only be obtained, if we reflect the model uncertainties in the solution, i.e. compute constraints that account for the model uncertainty. Geophysical prospecting additionally often only provides velocity models for one type of wave propagation, while others are inferred using theoretical relationships, e.g. regionally averaged S/P ratios. The error of such "inferred models" is therefore even higher. For a sensitivity analysis, the arrival time information could be inverted for a variety of probable models and the solution scatter be used as model correlated uncertainty. This means practically, to repeat the inversion with every model. While we will later indeed use such a post-processing approach to refine our solutions and obtain more precise results, it is not suitable for an interactive location in which ambiguous arrival times need to be evaluated within the inversion process. For this, model uncertainties have to be treated within and during the inversion.

The velocity model in conjunction with the process of wave propagation can be viewed as the operator that projects the time based input data into spatial information. As the "true" velocity model is hardly ever known, solutions consequently rely on the goodness of used approximations. If we suspect inaccuracies in our models, it is crucial for a proper location to derive a form of uncertainty measure that takes the uncertainties of the given velocity model into account, and will allow us to reflect this state of uncertainty in the constraint location and shape. Early attempts to measure and display the uncertainty in earthquake location are found in Pavlis (1986), who computed a bounding box around the location estimate to account for the model error e_{model}

$$\left|(e_{model})\right| \le s \cdot \Delta u \qquad \text{with} \qquad \Delta u = \left|(u_{true} - u_{model})\right| \tag{II.3.5-1}$$

which dimensions he estimated by multiplying the length of the corresponding ray path (s) with the maximum expected discrepancy between modeled and true slowness along the path (Δu). This, however, is only true for small perturbations for which the ray path does not differ significantly. How to determine Δu remained an open question. The general probabilistic approach of Tarantola and Valette (1982) provided for a separate treatment of velocity model and picking uncertainties. However, all implementations as Moser et al. (1992) or Lomax et al. [NonLinLoc] (2000) used a Gaussian description in a-priori density functions, combining timing and assumed model errors (Husen and Hardebeck, 2010). Pavlis showed for a simple example where the model-error yielded a bi-modal distribution, that the Gaussian assumption and combined treatment of picking and model errors can cause misleading results, concluding that picking and model errors should be treated in a different way. Husen and Hardebeck conclude therefore that "all location methods suffer from an inappropriate handling of velocity model errors". Mentioned be the work of Lin & Sanford (2001) who introduced the idea of an empirical uncertain velocity model for the case of homogenous models, which would produce correct uncertainty measures for events in large distance traveling through a homogeneous medium. Uncertainties in the arrival time were translated into velocity uncertainties, which together with lower and upper bounds of each phase's propagation velocity were used to map S-P and hyperbolic constraints for a fixed focal depth.

Based on the same concept of velocity uncertainties, that already formed the foundation for Pavlis' bounding box approach and Lin&Sanford's velocity bounds, the approach laid out in the following will allow us to obtain location uncertainties for more complex models, independent of the hypocentral distance and depth. From the interval approach for picking uncertainties (chapter II.3.1) we will derive a general concept for uncertain models in both a discrete and probabilistic context, to account for model uncertainties, independent of the model complexity. Applied to a homogeneous model, this approach could be depicted as replacing the travel-time-curve by a travel-time-band, where its shape (width) is being determined by the uncertainties of the model parameters. Similar to the other approaches (e.g. eq.II.3.5-1), this will provide minimal and maximal travel times, which will then be implemented in the formulation of the velocity-model-dependent distinct constraints.

II.3.5.1 **Uncertain Models**

The velocity models we use are our best guesses for representing the existing velocity structure. A model attributes precise velocity values to all spatial coordinates and hereby allows to map (invert) the time based data precisely to spatial locations. Such precise velocity values are, however, never exactly known. As the velocity structure is not precisely known, neither should our model be defined this way. Our data rather gives us an idea of velocity ranges, e.g. " $v_p \in [3.9, 4.1]\, km/s$ can be attributed to a region, which extends til a depth of $d \in [-2, -2.3]\, km$ ". Instead of using spuriously precise values, we should therefore attribute velocity ranges to fuzzy spatial regions, reflecting the state of knowledge much better. This defines the basic concept of "uncertain models" as used in the following.

To understand how model uncertainties affect the solution, we can evaluate the influence of these uncertainties on phase arrival times. For the simplest case of an homogenous half space the true velocity v is believed to lie in the vicinity of the estimated velocity $\bar{v}$, with a maximum deviation of δv

$$v = \bar{v} \pm \delta v \,. \tag{II.3.5.1-1}$$

For further evaluation it is useful to express this relationship using relative errors

$$v = \bar{v} \cdot (1 \pm \partial v) \qquad \text{with} \qquad \partial v = \frac{\delta v}{v}\,, \tag{II.3.5.1-2}$$

For an event being located in distance d from a sensor on surface, the travel time for the estimates is given by

$$\bar{t}t = \frac{d}{\bar{v}} \,. \tag{II.3.5.1-3}$$

Using eq. (II.3.5.1-2), the uncertainty bound travel time can be expressed as

$$tt = \frac{d}{v} = \frac{d}{\bar{v}\cdot(1\pm\partial v)} = \frac{1}{(1\pm\partial v)}\cdot\bar{tt} \,. \qquad \text{(II.3.5.1-4)}$$

As travel times tt correspond with phase arrival times t when zero adjusting the origin time $t_0 = 0$, eq. (II.3.5.1-4) could be written including arrival time uncertainties as

$$t = \frac{(1\pm\partial t)}{(1\pm\partial v)}\cdot\bar{t} \qquad \text{(II.3.5.1-5)}$$

since the relative arrival error ∂t could be expressed as

$$t = \bar{t}\pm\delta t = (1\pm\partial t)\cdot\bar{t} \qquad \text{(II.3.5.1-6)}$$

(see eq. II.3.1-1). Although a relative arrival time error doesn't make physical sense (except when used to describe time drift errors of unsynchronized data loggers) it allows us here to compare the effect of timing vs. velocity uncertainties: Timing uncertainties stand in a linear relationship with location uncertainties - velocity model uncertainties influence it inverse proportional. This means that timing uncertainties enlarge the travel-time uncertainty symmetrically, while velocity model uncertainties increase it asymmetrically, causing also a shift of the interval's mid point. Due to acting as denominator, model uncertainties naturally have a larger effect on the constraint's uncertainty (it's “width”) than arrival time uncertainties.

Already for simple and often applied 1D-layer models, the effect of model uncertainties on the travel time is complex and generally not solvable in an analytical manner. The strategy to treat uncertainties for these models will therefore lie in finding an upper (tt_{max}) and lower bound (tt_{min}) for the predicted travel times, e.g. for the homogenous half space

$$tt_{min} = \frac{1}{(1+\partial v)}\cdot\bar{tt} \quad \text{and} \quad tt_{max} = \frac{1}{(1-\partial v)}\cdot\bar{tt} \,. \qquad \text{(II.3.5.1-7)}$$

Each spatial (grid) point will hold a minimum and maximum travel time to each station, which will be evaluated in conjunction with the arrival-time-difference interval (LOC-interval): In the simplest case this would mean a check wether there is an intersection between the intervals of min&max

arrival time difference and min&max travel time difference, which would identify the evaluated spatial coordinate as a physically possible hypocenter location. For this it is hence sufficient to know the minimum and maximum travel times, which in the following we will derive from the model parameter uncertainties. It should be noted though, that if the relative error ∂v is small and similar over the whole model, eq. II.3.5.1-7 proves to be already a good first approximation also for more complex models. The solution is not exact since it neglects the different ray paths (e.g. direct, refracted head) and yields very different spatial constraint shapes, but yields comparable results for the most likely location for well behaved problems, since the mid points of the uncertainty interval are significantly stronger weighed than the borders. This simplification provides a first approximation of the model uncertainties and may be sufficient for a broad area of applications. It fails, however, to properly assess the full region of possible solutions.

Already for simple layered models the relationship between minimum and maximum travel time is complex and needs to be found computationally. The used travel time algorithm which is ray tracing 1D-velocity models (Eisermann, 2008) allows to compute several ten thousand travel time curves per second on current single processors (3GHz). The interval $[tt_{min}, tt_{max}]$ can therefore be found using a Monte Carlo analysis. The model input parameter uncertainties are hereby mapped to travel time uncertainties by calculating a sufficiently large set of random models within the uncertainty specifications. The travel time interval is given in the envelope of the set of travel time curves (see fig. II.3.5.1:2).

The 1D-velocity model layers are defined by layer velocity v and layer depth d. For the i^{th} layer the uncertainty parameters (fig. II.3.5.1:1) are now defined as

$$v_i = \bar{v}_i \pm \delta v_i \qquad \text{(II.3.5.1-8)}$$

$$d_i = \bar{d}_i \pm \delta d_i , \qquad \text{(II.3.5.1-9)}$$

based on the given model parameters $\bar{v}_i$ and $\bar{d}_i$ and their absolute errors δv_i and δd_i or generalized as

$$v_i \, \varepsilon \, [v_{i,min}, \, v_{i,max}] \qquad \text{(II.3.5.1-10)}$$

and

$$d_i \, \varepsilon \, [d_{i,min}, \, d_{i,max}] . \qquad \text{(II.3.5.1-11)}$$

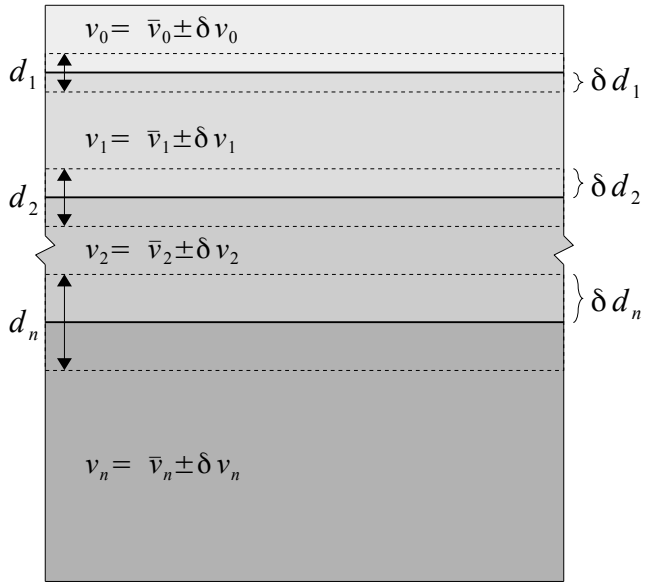


Fig. II.3.5.1:1: Uncertain 1D-velocity model definition

For each Monte Carlo analysis, a random model is created for which each layer i obtains values according to

$$v_i = \bar{v}_i + 2 \cdot (\chi_{v,i} - 0.5) \cdot \delta v_i \qquad \text{(II.3.5.1-12)}$$

and

$$d_i = \bar{d}_i + 2 \cdot (\chi_{d,i} - 0.5) \cdot \delta d_i \qquad \text{(II.3.5.1-13)}$$

with $\chi_{v,i}$ and $\chi_{d,i}$ carrying random values between 0 and 1. With $\chi_{v,i}$ and $\chi_{d,i}$, all parameters are treated as uncorrelated. For correlated parameters, all correlated parameters would be a function of a shared random variable $\chi_{v,i} = f(\chi_v)$. For instance, with not precisely known absolute layer velocities but assumed constant velocity ratios between the layers, all random-model velocities would use the same (synchronized) random value $\chi_{v,i} = \chi_v$. The same simple approach can be applied for a wide range of models, e.g. dipping layers (Eisermann, 2008), using additional fluctuating parameters as e.g. the layer's dip angle. Fig. (II.3.5.1:2) shows the travel time curve ensemble for a Monte Carlo evaluation with 20 respectively 1000 runs. A good resolution of the uncertainty interval and min and max curves is usually achieved with approximately 10^3 to 10^4

runs, which are performed within seconds on current standard hardware. Transcribing these curves onto the grid yields minimum and maximum travel times concerning the uncertain model at each grid cell.

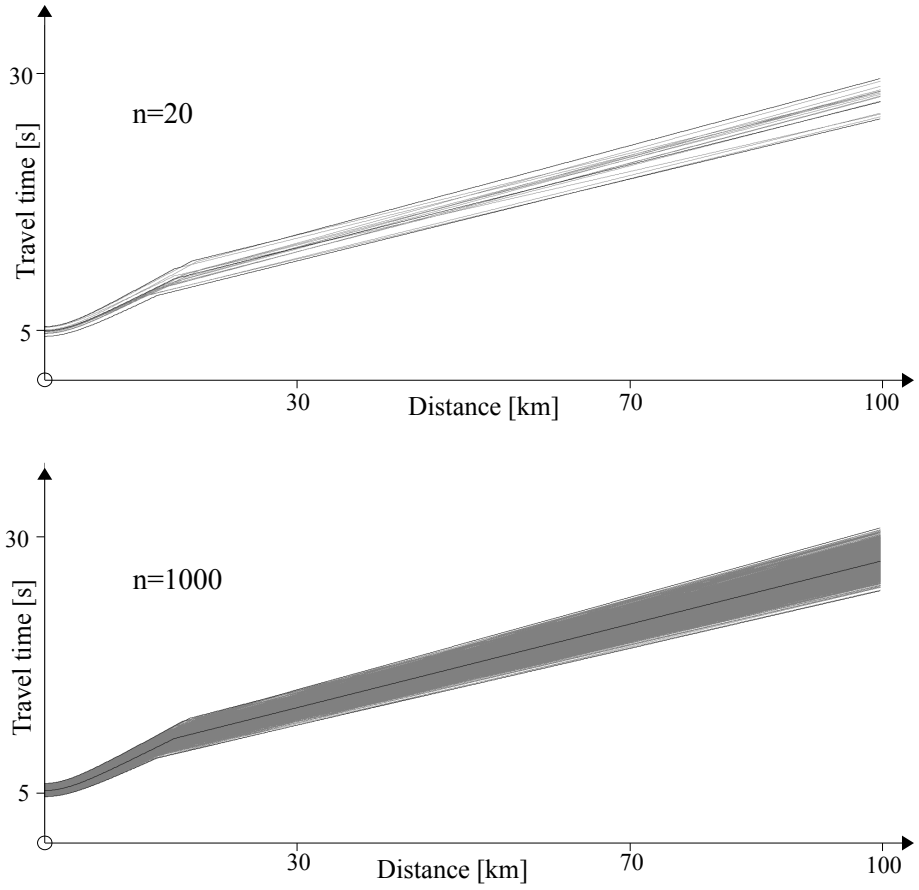


Fig. II.3.5.1:2: Travel time curve Ensemble, derived by n runs of random 1D-layered velocity models. For n=1000 the interval is sufficiently resolved delivering minimal and maximal travel times for every distance.

A cell hit count in a space time grid, plotting the travel time curves for a simulation with 10^6 runs (fig. II.3.5.1:3) shows the probability distribution of travel time curves. The maximum lies in the vicinity of the default model's travel time curve, slightly asymmetric in distance to minimum- and maximum travel time curve, as expected from eq. II.3.5.1-7. While inhomogeneities in the velocity model smooth out over longer distances, increasing the accuracy of predictions, the impact of mis-chosen models on predicted travel times increases with distance. This effect is reflected in the minimal and maximal travel time curves (fig. II.3.5.1:3) which increase in discrepancy with distance.

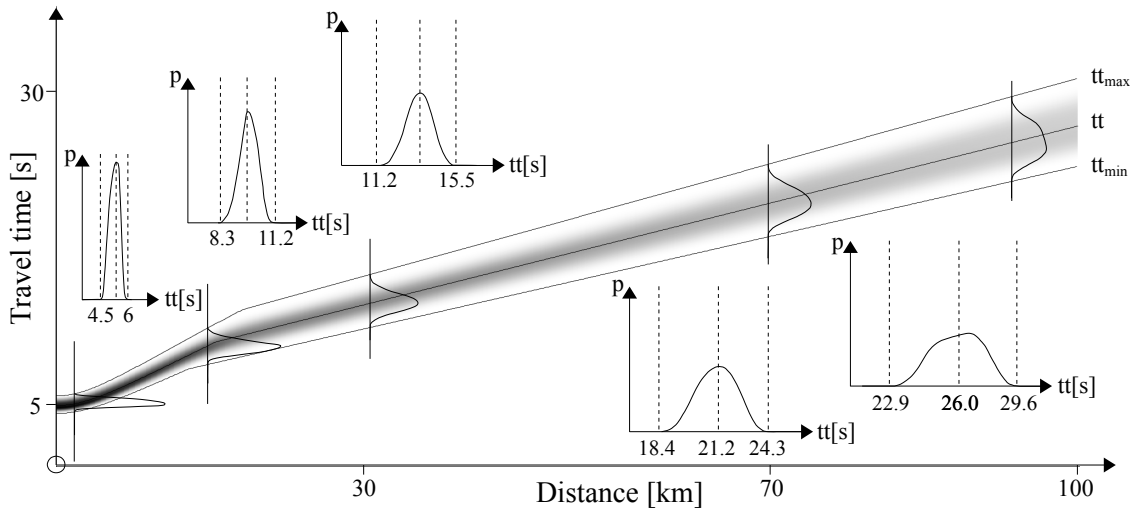


Fig. II.3.5.1:3: Probabilistic representation ($n=10^6$). Asymmetric spreading between extreme curves and curve of standard model as in the case of simple uncertain 1D-layered-model. Effect of model uncertainties grow with distance.

The concept of minimum and maximum travel times can be applied independent of the model complexity, although the calculation time and the amount of necessary runs increases with the complexity of the model. For a Monte Carlo analysis of 3D heterogeneous models, where the velocities are varied randomly according to a relative uncertainty, the calculation on standard hardware would take significantly longer and therefore needs to be rendered and stored prior to the analysis.

The obtained travel-time uncertainties provide the information required for mapping the time-based observables to space. We recall that the likelihood of a spacial coordinate to correspond with the hypocenter is (in principle) evaluated by comparing the difference of two modeled travel times $T^M = tt_B^M - tt_A^M$ with the difference of two arrival times $T^O = t_B^O - t_A^O$, observed at the station(s). The two travel times relate to the paths on which the two phases travel between the evaluated coordinate and the corresponding station(s). The previous chapters established that the observed arrival-time difference is rather described as an interval $T^O \epsilon [T_{min}^O, T_{max}^O]$ due to the uncertainties in t_A^O and t_B^O. Since we discretize space and evaluate grid cells that have a spatial extent, it is also not a single travel time value (as for a coordinate), but a (small) band of travel times $T^M \epsilon [T_{min}^M, T_{max}^M]$ proportional to the grid's spatial extent, that describes T^M for the cell. T^M serves as comparison interval for T^O at any given location. A grid cell hence qualifies as physically possible solution if

there is a overlap of both intervals, i.e.

$$T^O \cap T^M \neq \emptyset \,. \tag{II.3.5.1-14}$$

In a scenario in which only one constraint is given and this constraint is running in all it's width (which is determined by the time based observables' uncertainties) through a grid cell, then this grid cell will obtain the maximum possible likelihood value of 1 (due to the normalization of the LOC curves). In this case the T^O interval is fully contained in the T^M interval, $|T^O \cap T^M| = |T^M|$. If the constraint's width is larger than the grid cell's extent, then the T^M interval only holds a subset of the T^O interval, $|T^O \cap T^M| < |T^M|$. The cell's likelihood will be less than 1 and is found by integrating $\pi(T^O)$ (which is obtained by the cross-correlation of both involved LOC-curves) over the interval $[T^M_{min}, T^M_{max}]$.

Now, if we incorporate the uncertainties of the model, the $[T^M_{min}, T^M_{max}]$ interval will grow. Where the interval describing the cell was interpreted as equi-probable, the uncertainty interval will now also hold it's own uncertainty distribution, based on the uncertainty distributions of both involved travel times. This means that the $T^M \in [T^M_{min}, T^M_{max}]$ are varying in likelihood to occur. Every possible travel time difference that "meets" an arrival time difference obtained from the LOC-curves adds to the total likelihood of the cell to possibly contain the hypocenter, in proportion to the likelihoods of travel- and arrival-time differences. The probability for each T^M and T^O to "meet" is found in the product of their individual probabilities $p^M(T)$ and $p^O(T)$, according to eq. (II.3.3-3). The sum over all elementary pairs (fig. II.3.5.1:4) over the possible T defines the cells total likelihood.

$$p^{MO} = \sum_{\tau = T_{min}}^{T_{max}} p^M(\tau) \cdot p^O(\tau) \tag{II.3.5.1-15}$$

Using the probability densities $\pi(T^M)$ and $\pi(T^O)$, the total likelihood over the intersection of both intervals $[T^M_{min}, T^M_{max}]$ and $[T^O_{min}, T^O_{max}]$ is then given in

$$p^{MO} = \int_{T_{min}}^{T_{max}} \pi^{M}(\tau) \cdot \pi^{O}(\tau)\, d\tau \qquad \text{(II.3.5.1-16)}$$

with

$$T_{min} = max\left(T^{M}_{min}, T^{O}_{min}\right) \qquad \text{(II.3.5.1-17)}$$

and

$$T_{max} = min\left(T^{M}_{max}, T^{O}_{max}\right). \qquad \text{(II.3.5.1-18)}$$

Figure II.3.5.1:4 shows the integration over the overlap area of both distributions schematically.

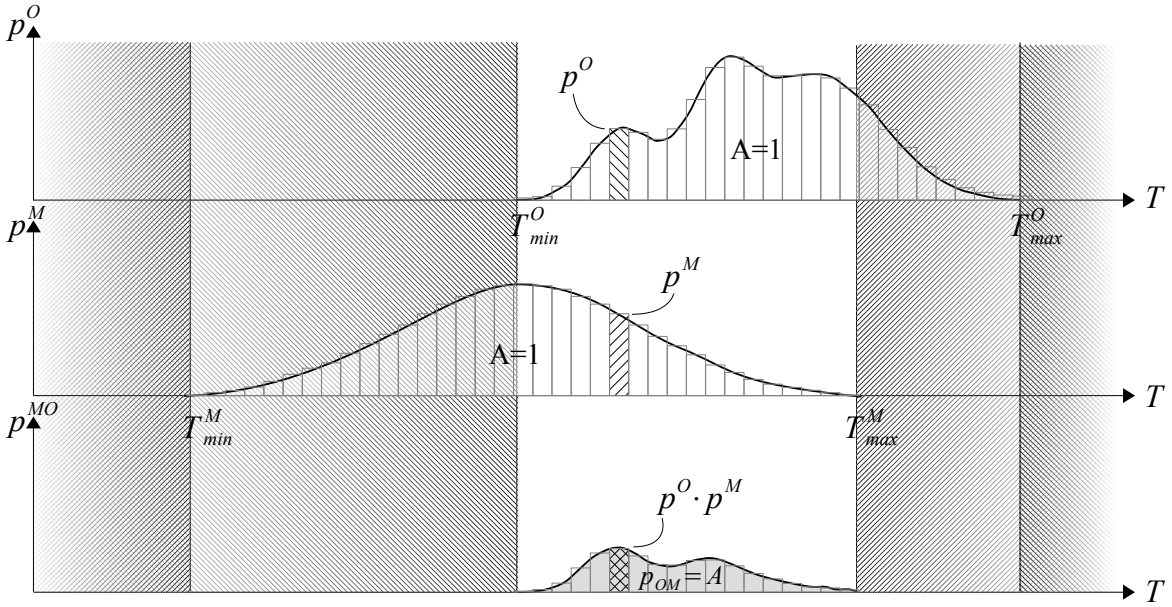


Fig. II.3.5.1:4: The cell's likelihood as constructed from model and constraint probabilities.

For a precise location, the region of highest likelihood will spread over few cells, while it will spread over several when the location problem is only poorly constrained. For a most conservative approach assessing the model-related location uncertainty, the two travel-times would be treated as uncorrelated, wherefore $p^M(T^M)$ would be obtained in the same manner as $p^O(T^O)$, which was computed by cross-correlating the likelihood distributions (LOC-curves) of both arrivals defining the constraint. For this case, $p^M(T^M)$ is constructed accordingly, by cross-correlating the distributions of both travel time ensembles. The true travel time distribution function can, however, only be obtained using a high number of runs far above $n = 10^3$ (fig. II.3.5.1:3). Also the caching of

this data for every station, phase type and grid point would require a significant amount of memory. These reasons render the evaluation of distributions in real time (on current standard hardware) difficult. What can be estimated sufficiently reliable for lower amounts of runs ($\approx 10^3$) is the lower and upper limit of the travel time ensemble, which we will focus on in the following. To represent this information we will use a flat distribution over the given interval. The effect of this simplification will be an over-estimation of the likelihoods at the interval boundaries or the mapped constraint edges, respectively. It therefore lowers the precision of the location, and results in a conservative location estimate.

Translating the model uncertainties into travel-time uncertainties between every station and grid-cell constitutes the computationally expensive part in the description of the model dependent location uncertainties. There is, however, a second aspect that defines the uncertain model as much as the uncertainties in geometry and velocity parameters do. This aspect relates to the treatment of model variations when we compute the travel-time difference between both involved phases that define the constraint. The most conservative case was already mentioned: When we define an uncertain model with e.g an average velocity of $4\mathrm{km}/s$ and an uncertainty of $\pm 0.1\mathrm{km}/s$ it is reasonable that some station may observe a local seismic velocity of $3.9\mathrm{km}/s$ while others may see $4.1\mathrm{km}/s$, or that the wave's travel path from a grid-cell of evaluation to station A may run through a region of $3.9\mathrm{km}/s$, while it runs through one of $4.1\mathrm{km}/s$ to station B (fig. II.3.5.1:5 b). In such a case we treat the travel times -and in turn the model variations- that both stations see as uncorrelated, following the same formalism that we already used within the context of arrival-time uncertainties. The travel-time difference (which is the modeled arrival-time difference) can therefore take all values between

$$T^M \in \left[tt^B_{min} - tt^A_{max}, tt^B_{max} - tt^A_{min} \right]. \qquad \text{(II.3.5.1-19)}$$

Such a travel-time difference measure is typical for class II constraints (i.e. involving two stations), inside and in close proximity of the network. This measure produces the most conservative location estimates as it holds the highest uncertainty in allowing maximum fluctuations between the models. Contrasting this, we find the case where both phases travel on similar paths. Here, both phases “see” the same model. The model variations for both phases are therefore treated as correlated. The smallest and largest travel-time differences are then given in

$$T^M \in \left[tt^B_{min} - tt^A_{min}, tt^B_{max} - tt^A_{max} \right] : \qquad \text{(II.3.5.1-20)}$$

When phase 1 experiences the smallest travel time, so does the other. The same holds for the maximum travel time: The model variation that causes the largest travel time for the one phase, will do so also for the other. This behavior is typical for some class I constraints like S-P constraints, traveling on nearly the same path (if $v_p/v_s \approx const.$). Also with class II constraints both phases may travel on a similar path, requiring the use of synchronized models. This is generally the case when the evaluated region is distant from the corresponding stations.

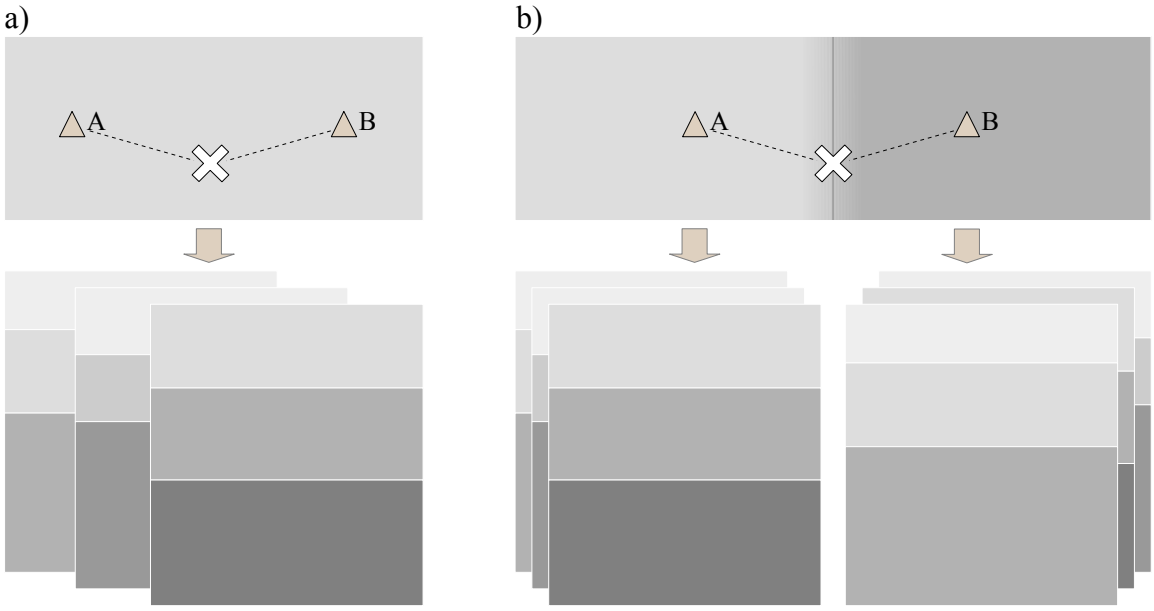


Fig. II.3.5.1:5: Correlated models (a) vs. uncorrelated models in the difference of travel-times of two phases. In a) both phases experience all model variations synchronized, while in b) each phase may simultaneously experience a different variation of the same model.

This measure produces significantly smaller location uncertainties than the uncorrelated case. This is linked to the different axes of action (fig. II.3.5.1:6): For each model variation -when treated correlated- the travel-time uncertainties shift the location dominantly parallel to (or along) the "exact" spatial constraint ["exact" referring to the "thin" constraint that is constructed when all uncertainties are ignored]. Contrary for unsynchronized (uncorrelated) model variations: Here the location is shifted perpendicular to the exact constraint. Over all model variations, synchronized models therefore hardly change the width of the constraint, while it's shape is significantly widened for unsynchronized models.

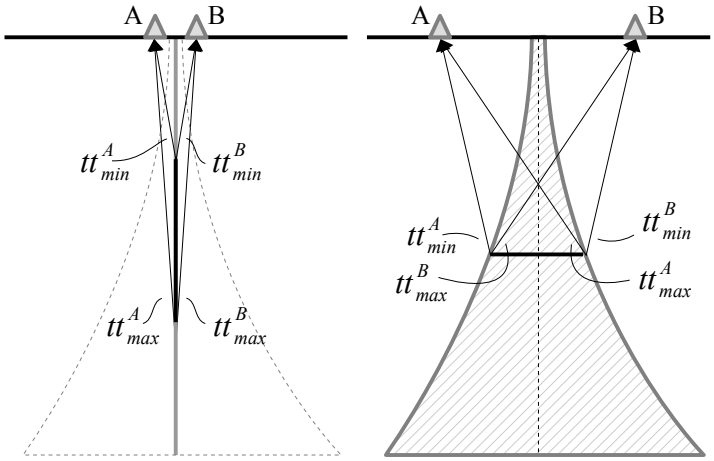


Fig. II.3.5.1:6: Correlated (a) and uncorrelated (b) travel paths for class II (i.e. two station) constraints. In far distance, both paths are correlated, running through the same model. In near distance, both paths are uncorrelated and may see different variations of the uncertain model.

For the example of an arrival time difference of 0 seconds, in case (a) the location uncertainty corresponds to a line, while for (b) it spans out a region. The minimal and maximal travel-times indicated relate to a single model variation in (a), or two variations, respectively, in (b). Over all model variations a line is stretched out in (a), while it defines a broad region in (b).

This effect makes uncorrelated models especially useful for situations where events are to be located within a network but the model is only poorly known and can hence only be approximated by 1D models. The reason for this is that uncorrelated 1D models approximate more complex ray paths than physically possible in a pure 1D structure: Generally, the applied model type must reflect all the complexity that we expect to see represented in our location estimate. If, for instance, we assume the existence of tilted layers, it is not sufficient to use variations of a 1D model, as these can not describe effects like lateral refraction. Rather, the model has to be defined as tilted-layer model with varying angles as additional uncertainty parameter. Although this is generally true, the use of unsynchronized (uncorrelated) models relaxes this requirement as it often covers the expected effects even when solely using 1D models: With synchronized (correlated) models, two stations will simultaneously only see the same model-variation. As laterally homogeneous models can never refract a ray path out of it's original plane as would occur with tilted layers, laterally homogeneous models when used synchronized will not account for laterally refracted paths. Unsynchronized

(uncorrelated) models, on the other hand, allow for fluctuations in the model velocities: Two stations may simultaneously see different variations of the model. This means, that solely using laterally homogeneous (i.e. 1D) models, both phases may travel with different velocities from any point in space to the receiver(s). Although not physically modeled, this reflects effects as refractions that are caused by the true (non-1D) structure, in the location uncertainty, causing the spatial constraint to include e.g out-of-plane locations. Treating model variations as uncorrelated therefore often yields robust location regions even when simplified models are used. Yet, with distance from a station pair, uncorrelated models largely overestimate the uncertainty. One way to overcome this downside lies in using mixed models, in which we will apply uncorrelated models in the near- and correlated models in the far-field (Chapter II.3.5.2).

In the following we will address the identification of the T^M distribution for both modes (correlates/uncorrelated) and look at how they constrain the solution space, beginning with the mode for the most conservative location estimate (i.e. most accurate, least precise) obtained by uncorrelated models. For such, the T^M interval is obtained by cross-correlating the likelihood distributions of the two travel-time intervals. Although the true distributions could be obtained by a sufficiently high amount of runs, we will simplify the evaluation for computational purposes and reduce the information to the travel-time interval limits, which can be obtained with lower amount of runs. The most conservative distribution, representing this information are flat distributions. The cross-correlation of two flat distributions results in a three segmental linear curve (fig. II.3.5.1:7) which can be solved analytically, providing for a fast computation and evaluation. Since the cross-correlation of two normalized distributions is also normalized, the maximum value of the resulting distribution is given by

$$p_{max} = \frac{f^{-1}}{\Delta^{AB} - \Delta_{min}} \tag{II.3.5.1-21}$$

with

$$\Delta^{AB} = \Delta^{A} + \Delta^{B} \tag{II.3.5.1-22}$$

and

$$\Delta_{min} = min\left\{\Delta^{A}, \Delta^{B}\right\} \tag{II.3.5.1-23}$$

where

$$\Delta^A = tt^A_{max} - tt^A_{min} \text{ and } \Delta^B = tt^B_{max} - tt^B_{min} \tag{II.3.5.1-24}$$

describe the "width" of the travel-time curve ensemble. f is the discretization factor defining the number of probability bins per time unit, usually chosen as the sample rate in consistency with the discrete arrival-time likelihood distributions. The distribution representing the model uncertainty is then given in the three segmental curve

$$\begin{aligned} p^M(T) &= p^M\left(T, tt^A_{min}, tt^A_{max}, tt^B_{min}, tt^B_{max}\right) \\ &= \begin{cases} T^M_{\text{red}} \cdot \delta_t p, & 0 < T^M_{\text{red}} < \Delta_{min} \\ p_{max}, & \Delta_{min} < T^M_{\text{red}} < \Delta^{AB} - \Delta_{min} \\ p_{max} - \left(T^M_{\text{red}} + \Delta^{AB} - \Delta_{min}\right) \cdot \delta_t p, & \Delta^{AB} - \Delta_{min} < T^M_{\text{red}} < \Delta^A + \Delta^B \\ 0, & else \end{cases} \end{aligned} \tag{II.3.5.1-25}$$

with the reduced constraint parameter

$$T^M_{\text{red}} = T - T^M_{min} \tag{II.3.5.1-26}$$

representing the adjusted difference in travel times and

$$T^M_{min} = tt^B_{min} - tt^A_{max} \tag{II.3.5.1-27}$$

being the lower limit of the interval. The derivative

$$\delta_t p = p_{max} / \Delta_{min} \tag{II.3.5.1-28}$$

describes the slope of ascending and descending segments as seen in fig. II.3.5.1:7.

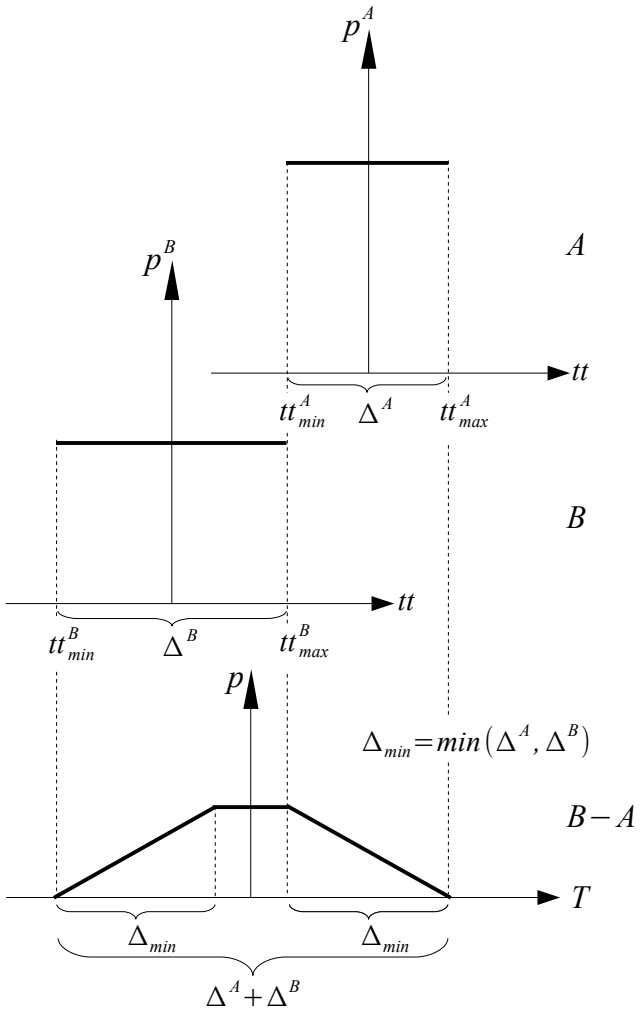


Fig. II.3.5.1:7: Three segmental analytical solution for the cross-correlation of two probability distributions.

After having obtained the T^M interval and distribution, the constraint can be mapped. To map it's (unweighed) shape it is sufficient to solely use the T^M interval according to eq. II.3.5.1-14, displaying the non-probabilistic constraint (fig. II.3.5.1:8). The example shows, that a solution that appears precise and localized, may actually become unconstrained in distance when (uncorrelated) model uncertainties of 5% are introduced. The congruent region (which now extends into distance) corresponds to possible solutions, but most of them are highly unlikely (being close to several constraints' edges). This can be quantified, applying the T^M distribution according to eq. II.3.5.1-15 in a probabilistic plot (fig. II.3.5.1:9).

a) picking uncertainty b) picking + model uncertainties

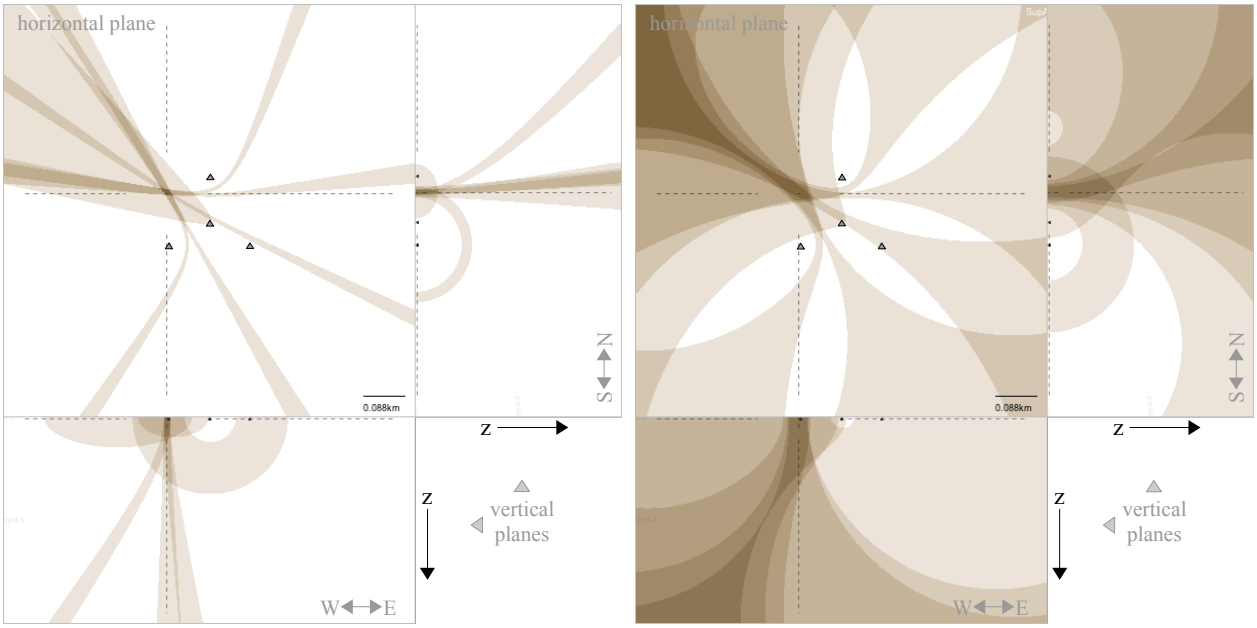


Fig. II.3.5.1:8: Comparison of uncertain hyperbolic P-P constraints where a) regards (uniform) picking uncertainties and b) regards (uniform) picking and model uncertainties. Where picking uncertainties lead to a linear growth of constraint width, model uncertainties yield a exponential growth. This reflects the fact that uncertainties in velocity affect the solutions the stronger, the longer the signal travels. The localized solution volume in (a) becomes a open volume in (b). [For this example an array aperture of 100m, P pick uncertainties of 0.0005s and a homogeneous velocity model of (4.5 ± 0.225) km/s (5% variation) was used.

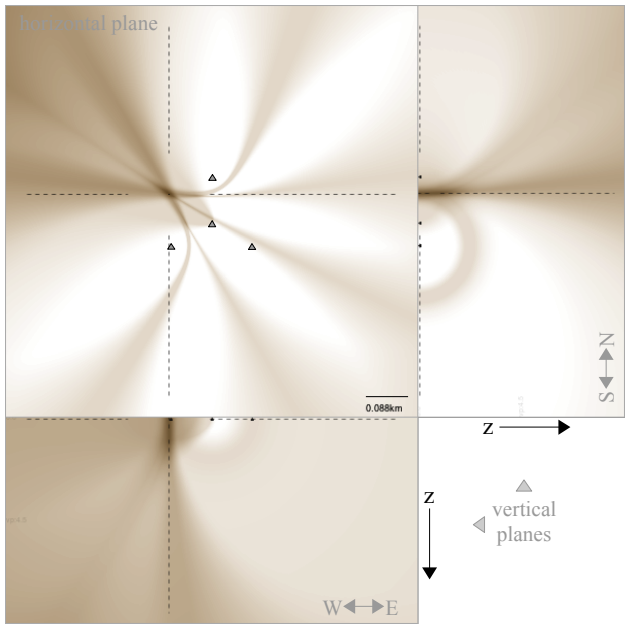


Fig. II.3.5.1:9: Probabilistic constraints, corresponding to the example of fig. (II.3.5.1:8). The most likely region is compact, despite large model uncertainties .

The probabilistic analysis shows that (although the congruent region is large) the most likely location is confined close to the true (simulated) source location. It further shows, that the constraint edges, which correspond to the limits of the T^M interval, hardly affect the most likely solution. The same must hold for the travel-time interval limits (from which the interval limits of T^M are determined). Inaccuracies in such, which may be caused by the lower amount of used model variations (causing an insufficient sampling of the set of all possible variations) therefore do not significantly affect the most likely location. If we, however, are interested in the most conservative location estimate, using uncorrelated models in a non-probabilistic analysis, it may require the additional use of class I constraints to obtain precise locations:

After introducing arrival time uncertainties, the width of azimuthal (class II) constraints will always grow with distance to the corresponding station pair (since the difference between two paths' lengths or travel times is less affected by lateral perturbations the smaller the average angle between them). In distance from both stations, a hyperbolic constraint, for instance, will therefore grow linear in width. This is the case when we use an exact model (i.e. a distinct “thin” travel time curve). As velocity model uncertainties now additionally enlarge the width of the travel-time curve ensemble with distance and we additionally allow maximum fluctuations between the two involved travel-time ensembles, the constraints width now grows exponentially with distance from both stations. For a pure P-phase based location, the impact of velocity model uncertainties may therefore be strong (see fig. II.3.5:1). Although class I constraints (e.g. S-P) are equally affected, they allow to confine the location in distance (even in a non-probabilistic analyses where congruent regions may be large (see fig. II.3.5.1:10 and II.3.5.1:11)). The constraints characteristics as laid out in chapter III.1 are amplified in presence of model uncertainties. P-P information may become unconstrained in distance. S-P constraints on the contrary tend to larger azimuthal uncertainties. The effect is amplified for both, the further the distance from the network. But as the uncertainties are bound to (perpendicular) axes, the combination of both stabilizes and confines the solution in space (fig. II.3.5.1:10). The combination of both constraint classes using a non-probabilistic analysis therefore provides for robust hypocenter regions, to obtain accurate results in uncertain environments. Figures (II.3.5.1:10) and (II.3.5.1:11) show the same example of fig. (II.3.5.1:9) for added S-P constraints.

a) picking uncertainty b) picking + model uncertainties

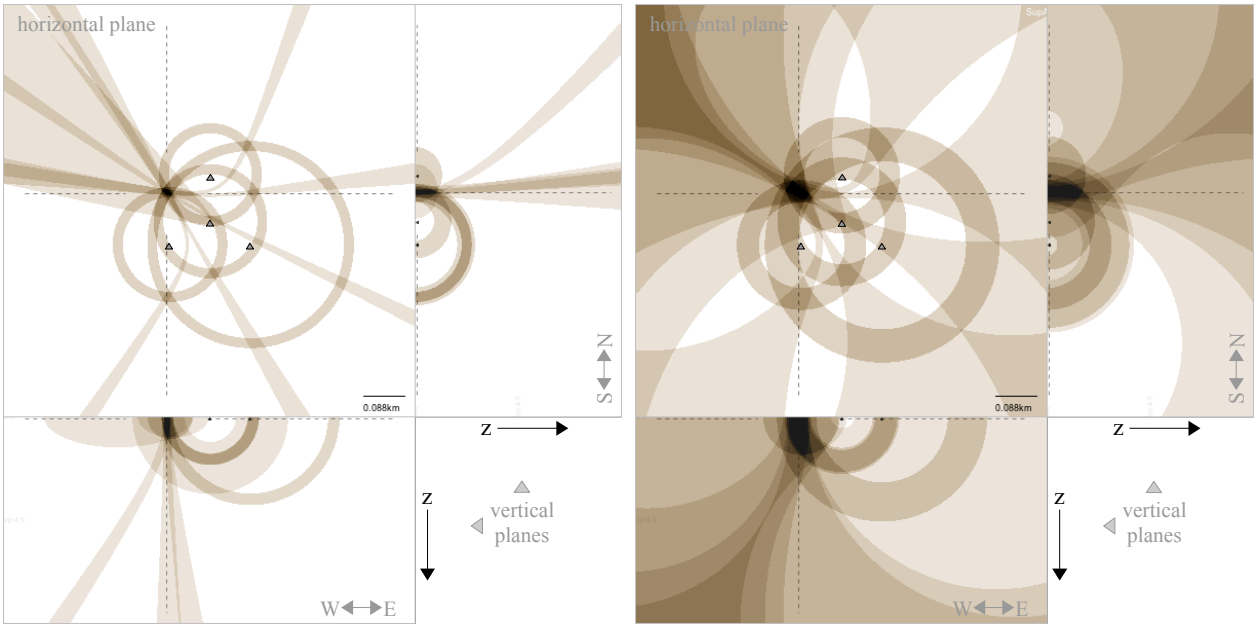


Fig. II.3.5.1:10: Location using uncertain S-P and P-P constraints for a) picking uncertainties only, b) picking and model uncertainties. The width of the S-P sphere shells and the coherent region are significantly enlarged over case (a) but by using combined information the location is well constrained compared to a purely hyperbolic location (fig. II.3.5.1:9). [The array aperture is 100m, P pick uncertainties equal 0.0005s, S pick uncertainties equal 0.001s. A homogeneous velocity model of (4.5 ± 0.225) km/s (5% variation) is used]

a) picking uncertainty, probabilistic b) picking + model uncertainties, probabilistic

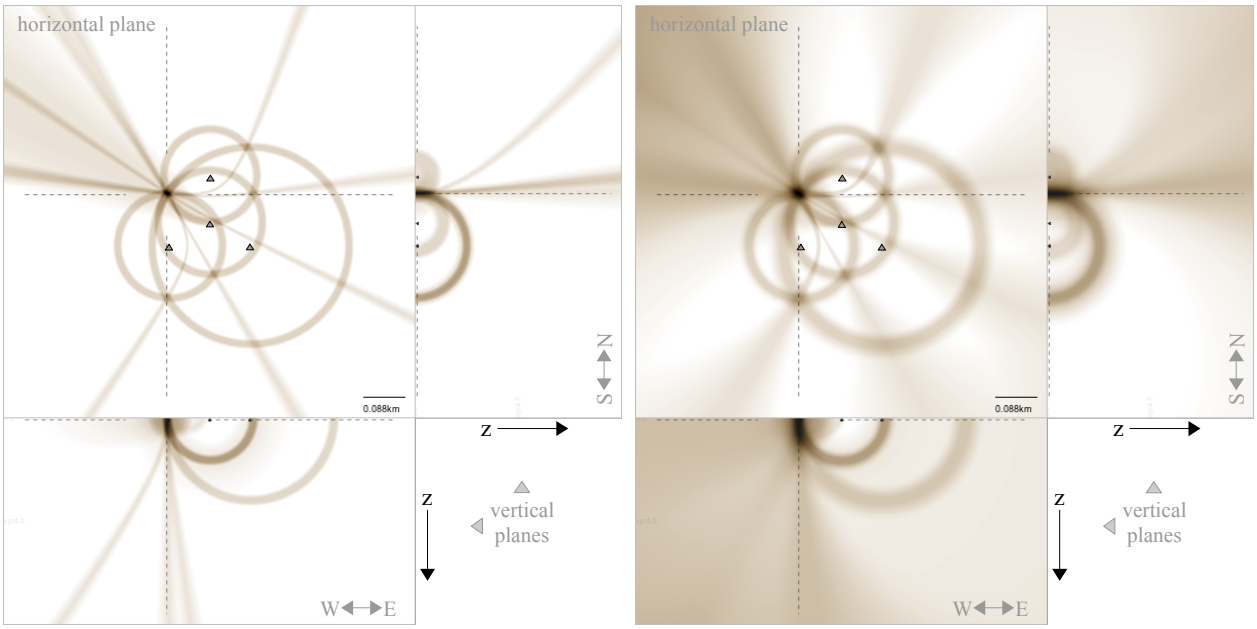


Fig. II.3.5.1:11: Location using probabilistic S-P and P-P constraints for a) picking uncertainties only, b) picking and model uncertainties for the example of fig. II.3.5.1:10.

The examples show that by the combined use of class I and class II constraints even the most conservative uncertainty measure can yield precise hypocenter locations. So far we were using

purely uncorrelated models for class II constraints to obtain robust location estimates. However, if the two stations, which define the constraint, not happen to lie on two sites of a fault, uncorrelated models will always overestimate the uncertainty in distance from the corresponding station pair. In the next chapter we will introduce mixed models, using a combination of uncorrelated and correlated models, cross-fading between both modes depending on the given scenario. This will allow us to increase the precision of the result without decreasing it's accuracy.

II.3.5.2 **Physical Correlation Correction and Model Ensembles**

When we treat model variations as uncorrelated, we do not necessarily describe a physical system (which has one true structure). Uncorrelated models allow different stations to simultaneously see different variations of the model. Depending on the scenario, this may or may not lead to physically sound travel time differences (since the travel times may have been obtained from different models), which in turn result in unrealistic solutions. Uncorrelated models allow us to approximate a heterogeneity that surpasses the (rendered paths according to the) model specifications and hereby allows a fast computation using simpler models. If, however, we treat models as uncorrelated when they are actually correlated, we allow a higher fluctuation in the travel times than physically possible, which would lead to an overestimation of the travel-time difference. This shall be laid out by looking at both modes (correlated vs. uncorrelated) in detail:

When we treat models as uncorrelated we give room to fluctuations in the fixed model-assigned velocities by allowing different phases to simultaneously see a different variation of the model. If, on the opposite, we wanted to express that exactly one of the model variations must globally represent the true velocity structure, we would treat them as correlated with the phase(s) seeing the same variation. An example of the first case would be defining an uncertain (homogeneous) model with v=(5 ± 0.2)km/s, to describe a scenario where certain regions may hold a velocity of v=4.8km/s while others may hold one of v=5.2km/s (opposed to expressing that all regions hold the same velocity, being unknown but limited to lie between v=4.8km/s and v=5.2km/s for the second case). This is reasonable when the two analyzed travel paths differ from one another, e.g. if the trial location lies in between two stations of a class II constraint. In the unknown velocity structure, one phase might truly propagate through a region of higher velocity (v=5.2km/s) while the other travels through a region of only v=4.8km/s. On the other hand, when both paths indeed are physically

coupled, allowing such fluctuations would overestimate the travel time difference and in turn the location uncertainty, as both phases in reality experience simultaneously the same (or a dependent) model variation: The model variations are “synchronized”, i.e correlated. This is the case for certain class I constraints (e.g. S-P) where both phases travel on a similar and physically coupled path. But also for class II constraints both paths can be correlated. This is the case when the trial location is in far distance: Now the phase would propagate on a very similar path to both stations. When both paths are similar, the same phase can not experience 5.2km/s and 4.8km/s simultaneously, but must rather see the same model variation, which means that the phase experiences on both paths either 4.8km/s or 5.2km/s or everything in between, synchronized. Both stations see synchronized either the shortest possible travel time or the longest, or any proportional and consistent value in between.
To understand the geometrical effect that the choice of mode (uncorrelated/correlated) makes, we will re-use the example of figure II.3.5.1:6, visualizing the effect of the path correlation on the estimated location uncertainty in the near- and in the far-field (fig. II.3.5.2:1): For the example we assume that a phase arrives at two stations at the same time with a negligible picking uncertainty. Coming from far distance, the two travel paths to station A and B see the same physical velocity model (i.e. the travel paths are correlated). Here, this causes all possible locations (due to the model uncertainty) to lie on a line (a plane in 3D).

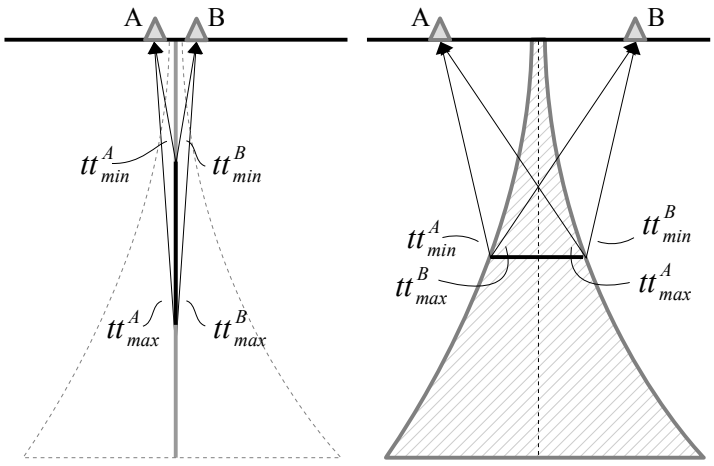


Fig. II.3.5.2:1: Correlated (a) and uncorrelated (b) travel paths for two station constraints. In far distance, both paths are correlated, running through the same model. In near distance, both paths are uncorrelated and may see different variations of the uncertain model.

For the example an arrival time difference of 0 seconds is used. In case (a) the location uncertainty lies on a line, while it stretches out a region in (b).

But for a source region in near distance, the phase can see different variations of the velocity model (i.e. the travel paths are uncorrelated). Here, the model variation causes a location uncertainty which acts perpendicular to the constraint's run, increasing the constraints width with distance. Not correcting for physical correlation would in larger distance massively overestimate the effect of model uncertainties and let the constraint grow unrealistically wide (in large distance even causing it to surround the whole network). In proximity to the network, additional constraints would be needed to constrain the solution to a smaller region (e.g. fig. II.3.5.1:10), but in larger distance the location uncertainty would remain enlarged. This overestimation will affect the solution less in a probabilistic analysis since the most likely travel time usually pertains to a model close to the default model, and the broad constraint edges would hardly be weighed. Yet, pertaining to congruent zones, the enlarged constraints would display a spurious agreement among constraints, that would bring even mispicks into (spurious) agreement with other constraints. This agreement, however, would only be due to the unphysical description. To reduce this effect, we will approximate the information of physical correlation, which will limit the lateral uncertainties in large distance: When both phases travel on the same path and are physically closely related, the modeled constraint parameter T^M may range between

$$T^M_{min} = tt^B_{min} - tt^A_{min} \qquad \text{(II.3.5.2-1)}$$

and

$$T^M_{max} = tt^B_{max} - tt^A_{max}\,, \qquad \text{(II.3.5.2-2)}$$

$$T^M \in [tt^B_{min} - tt^A_{min}, tt^B_{max} - tt^A_{max}]\,, \qquad \text{(II.3.5.2-3)}$$

both seeing the same variation of the velocity field. The corresponding likelihood distribution over the interval will be approximated as flat. Since the T interval is given for uncorrelated travel paths in

$$T^M \in [tt^B_{min} - tt^A_{max}, tt^B_{max} - tt^A_{min}]\,, \qquad \text{(II.3.5.2-4)}$$

a mixed form

$$T^M \in [tt^B_{min} - (tt^A_{min} \cdot cf + tt^A_{max}(1-cf)),\ \ tt^B_{max} - (tt^A_{max} \cdot cf + tt^A_{min}(1-cf))] \qquad \text{(II.3.5.2-5)}$$

can be formulated that yields eq. (II.3.5.2-4) and (II.3.5.2-3) as end-members, with cf being the correlation factor, describing whether travel paths are probabilistically (to be) treated as physically correlated ($cf = 1$) or uncorrelated ($cf = 0$). Eq. II.3.5.1-5 now allows to linearly fade between both end-members to describe a mixed state when none of the end-members seems to be appropriate to use ("A" and "B" stand for both involved phases). A correlation factor of one ($cf = 1$) may be used to describe scenarios with fixed velocity ratios without perturbations (i.e. the velocity model is structurally identical with the subsurface configuration, except for an unknown globally constant velocity scaling factor). For very estranged models of P and S velocities (very inhomogeneous vp/vs-ratios) both phases may significantly differ and intermediate cf -values might yield an appropriate estimate to use. The same holds for one station P_{dir}- P_{ref} constraints, where the phases travel on very different paths, yet are physically correlated as they both strongly depend on the layering (structure) of the model. For class II constraints (e.g. the P-P constraint) between two stations this is more complex and, as previously laid out, depends on the distance between trial location and both stations: For short distances, both stations may see different models ($cf = 0$), while in larger distance the model variations have to be treated as correlated ($cf = 1$). The relative distance

$$r = d_{xyz} / D^{A,B} \tag{II.3.5.2-6}$$

can hence be used to control the correlation factor, with $D^{A,B}$ being the distance between both stations A and B and d_{xyz} being the distance from the station's midpoint to the evaluated location. For short distances ($r \leq r_{near}$), both rays are expected to be capable of seeing different variations of the same model, as paths are independent. For large distances ($r > r_{far}$), wave-paths are very similar and both stations should see the same model variation. In between, both intervals are cross-faded

$$cf = \begin{cases} 0, & r \leq r_{near} \\ (r - r_{near}) / (r_{far} - r_{near}), & r_{near} < r < r_{far} \\ 1, & r \geq r_{far} \end{cases}. \tag{II.3.5.2-7}$$

The values for r_{near} and r_{far} stand in relation to the dominant frequency of the evaluated signal, which determines the width of the fresnel-zone ("banana-doughnut", Born-Fréchet kernel theory). Lower dominant frequencies (larger wavelengths) correspond to a larger correlation-radius, lowering r_{near} and r_{far}, while higher dominant frequencies increase their values. In the following,

we will use static values for r_{near} and r_{far}, with r_{near} chosen to correspond to an average angle of 10° between both ray paths, and r_{far} corresponding to a narrower 5° angle. This causes a quick initial growth of the constraint's width in the near field (accounting for velocity perturbations), which stabilizes in the intermediate field and gains precision (decreases in width) in larger distances (fig. II.3.5.2:2).

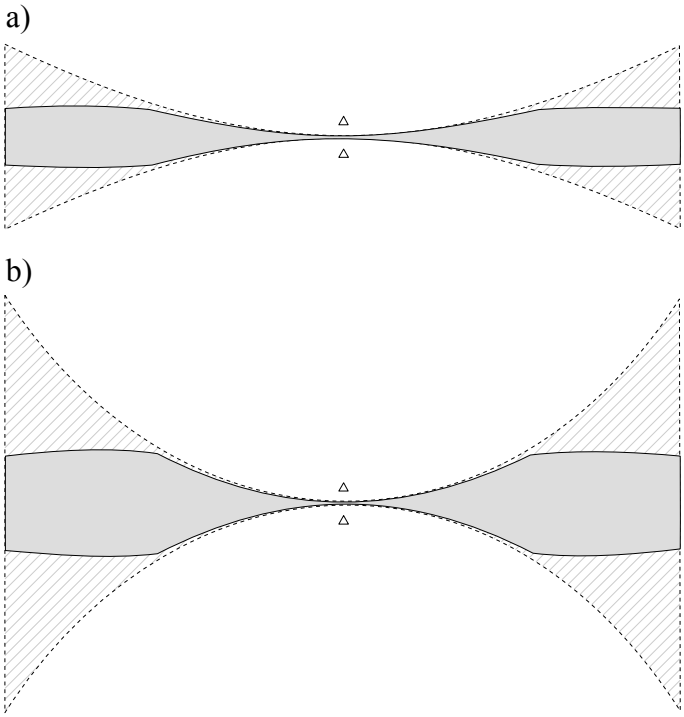


Fig. II.3.5.2:2: Model correlation correction for a) 1% model uncertainty and b) 2% model uncertainty. The hatched area corresponds to uncorrelated measures, which overestimate the uncertainty in distance. The correction limits the region in distance (solid area). The correlation factor increases with distance from stations according to eq. II.3.5.2-7. The curves correspond to two stations in 2km distance in a homogeneous velocity field of 4.5km/s. Picking uncertainties are 0.005s.

Additional azimuth-dependent correction terms could be derived that would allow for further corrections, narrowing down estimated uncertainties for strongly bent hyperbolas for trial locations in intermediate distances to the stations. Here the signal may travel partly in regions of correlated- and partly in regions of uncorrelated paths. This behavior, however, is complex and will not be regarded in the work presented. For the real-time application, rather than looking for physically likelier cases increasing the correlation factor and the precision, we continue to use lower correlation factors, supporting accuracy (To support higher degrees of precision, we will introduce model-ensembles the next sub section).

After having obtained measures to include the model uncertainty in the constraint mapping, we will apply such in the following example, showing the analysis of a gas field event (the full analysis can be found in chapter V.1). With picking uncertainties being applied but model uncertainties disregarded, no region of constraint agreement is found (fig. II.3.5.2:3).

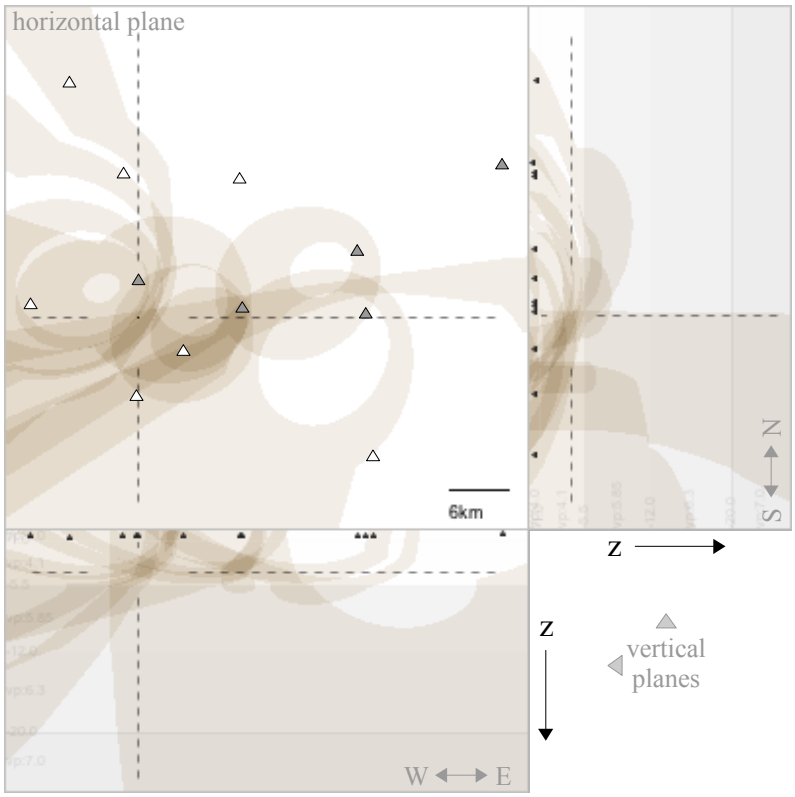


Fig. II.3.5.2:3: The example shows 12 surface stations that registered an event in a gas field. For clarity only the constraints of a 5-station subset (dark grey filled triangles) are shown. Although picking uncertainties are applied and widen the constraints, no region of common agreement is found. The color intensity scales with the amount of agreeing constraints.

This changes when we include model uncertainties. A 3% uncertainty in the P-velocity model is sufficient to bring all information into agreement. Since the whole region is in proximity to the stations, model variations are treated as uncorrelated. Looking at an isolated subset of 7 stations, one can see the corrected constraints' run, as well as the region of full congruity in which all constraints overlap (fig. II.3.5.2:4 a&b).

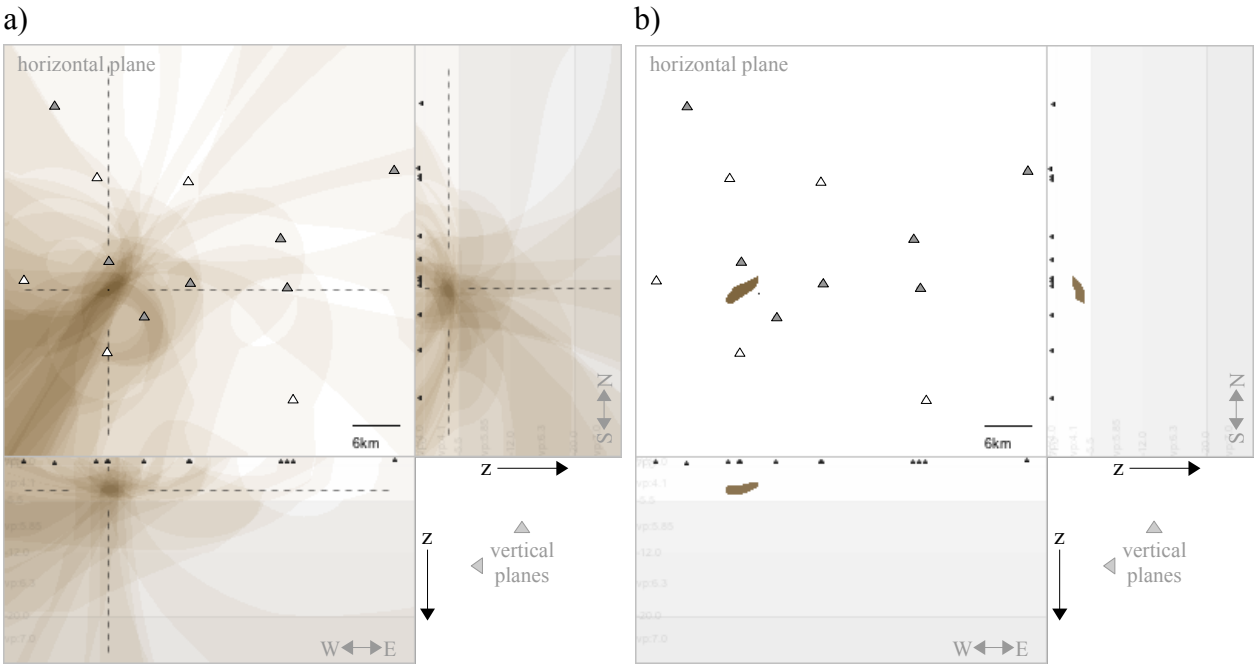


Fig. II.3.5.2:4: Applying a 3% model uncertainty brings all information into agreement. Panel a) shows the constraints for a 7-station subset (dark grey filled triangles) and panel b) the corresponding congruent region with 100% agreement.

When plotting the 66 constraints of all 12 stations, the map takes the character of a continuous likelihood map, showing a well constrained hypocenter region.

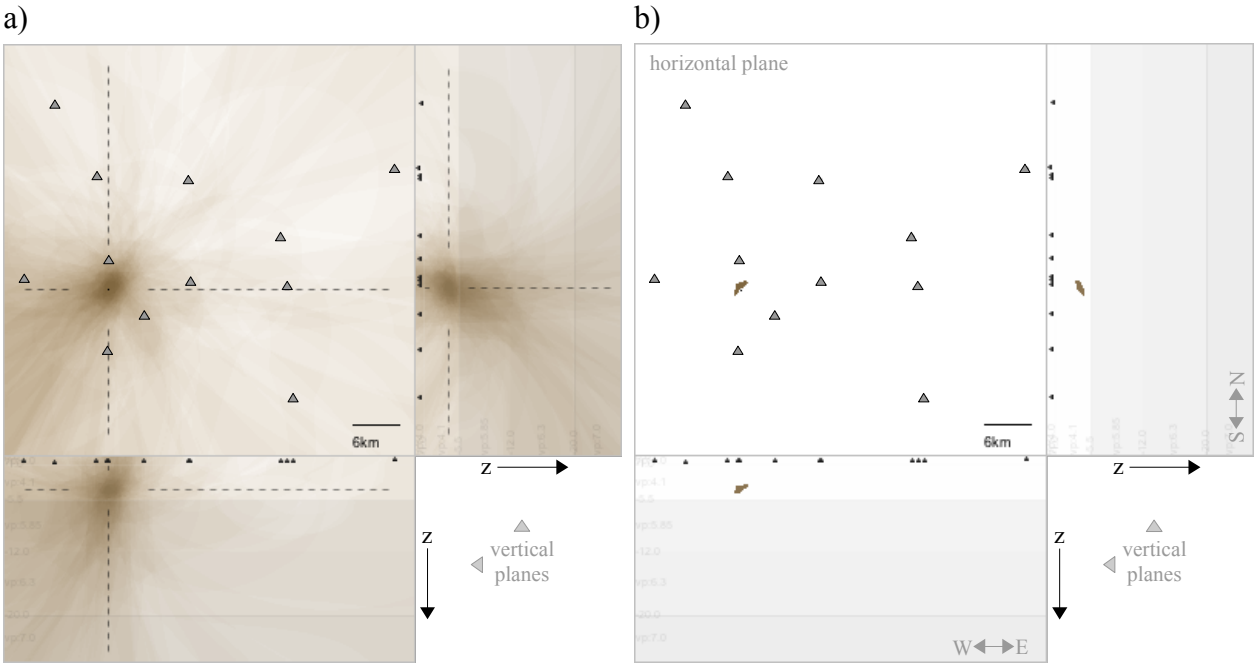


Fig. II.3.5.2:5: Adding the remaining constraints reveals a clear and precise maximum. Panel b) shows the corresponding congruent region of 100% agreement.

As powerful as these uncertainty measures are, they are only a rough proxy for the true uncertainty: Even when we use correlated models, they are only treated correlated per individual constraint, but not between the constraints. This means that one location (point) that is marked by two constraints, may have seen one model variation in the one constraint and a second variation in the other, which should not be the case for correlated models. This means that the effect of model uncertainties on the solution region will always be overestimated. To display the exact constraint- (and location-) uncertainty based on the defined model uncertainties, we need to utilize model-ensembles.

MODEL-ENSEMBLES

Uncertain models yield robust uncertainty measures which can be applied in real time, as the computationally expensive part of Monte-Carlo simulations (i.e. the travel time rendering) are performed prior to the analysis. They do, however, overestimate the effect of the described model uncertainty on the location uncertainty and only provide upper bound approximations of the location uncertainty, both when used uncorrelated or correlated. Not only because the correlation of travel-time differences is only valid per constraint and not globally for all, but also because we take for granted that the travel-times that two stations "see" are linked to the same model variation: In the correlated case, for instance, we calculate the maximum possible travel-time difference between two stations by subtracting both their largest travel-times, assuming they relate to the same model. This, however, is only strictly true for laterally homogeneous models under surface networks and is not a given in heterogeneous environments. We will conclude this chapter on model uncertainties with a short look on the (computationally more expensive) approach of model-ensembles, which will allow us to a) describe uncertainties with highest precision, b) treat model variations completely correlated over all constraints, and c) have travel-time differences linked to individual models, hence accurately mapping the exact effect of model uncertainties unto the location. The underlying idea shall be laid out in the following:

In previous chapters we used *one* uncertain model (it's parameters' mean values believed to represent the real velocity structure to a good degree) in which we accounted for imaginable inaccuracies by allowing parameter perturbations (velocity uncertainties). Although this strategy is reasonable -since the true velocity field is *unknown*- there is at least one (unknown) true and exact model (free of uncertainties) that links the data to the true hypocenter and origin time. We could

therefore (theoretically) test for all possible models that confirm the given structural subsurface information, calculating all constraints for each model to identify the small fraction of models that produce congruent regions. Merging the found regions would then display the true model-dependent location uncertainty. Stacking all constraints computed over all models would provide a map indicating the most likely solution in respect to the model uncertainty. Technically, such a method repeats the fundamental location procedure (i.e. mapping the arrival-time-difference based constraints for a given velocity/travel-time model) for all predefined models in the ensemble. The difference to the method of uncertain models, which uses the same model description and likewise renders all Monte-Carlo generated models, is that constraints need to be computed for every model of the ensemble (while they are only computed once using minimum and maximum found travel-times in the method of uncertain models). Since this is computationally too expensive to be run in real-time, this method mainly serves to refine the solution in a post-processing step using more accurate uncertainty estimates. Obviously, the location scheme can not be rerun for each of the (infinite) possible models. When we define the set of models in the ensemble, the dimensionality of the problem is too high to be sampled sufficiently, even in the simplest case because of sampling/discretization problems: For instance, to express a velocity range in a certain layer, one might use ten models increasing the velocity with each model at one tenth of the range starting from the minimal velocity. For minute picking uncertainties, none of these models may actually bring all constraints into agreement (if the true velocity lies in between the sampled instances). Instead of defining the models as exact models (which would result in a location that corresponds to totally correlated models) we may define them instead as uncertain models in which small uncertainties are assigned to account for the transition of one sampled model instance to another, allowing for small fluctuations in each of the ensembles' models.

Fig. II.3.5.2:6 compares the methodologies of uncertain models and model ensembles for the simple case of a homogeneous model, the velocity known within an error margin of five percent. In the framework of uncertain models the travel-times obtain a ± 5% uncertainty, leading to broadened spatial constraints, which in turn broaden the congruent region. Model-ensembles determine the solution region more precisely, by evaluating an ensemble of models with different velocities, treating picking uncertainties as usual. For each model, the congruity regions and probabilistic constraints are computed. These regions and the most likely solutions are stacked on summation grids. The stacked constraints again represent probabilistic constraints respecting picking and model uncertainties, showing the clusters of locations in space for which the respective models explained

the observed phase arrivals best. The merged congruent regions reveal the true model-uncertainty-dependent location uncertainty.

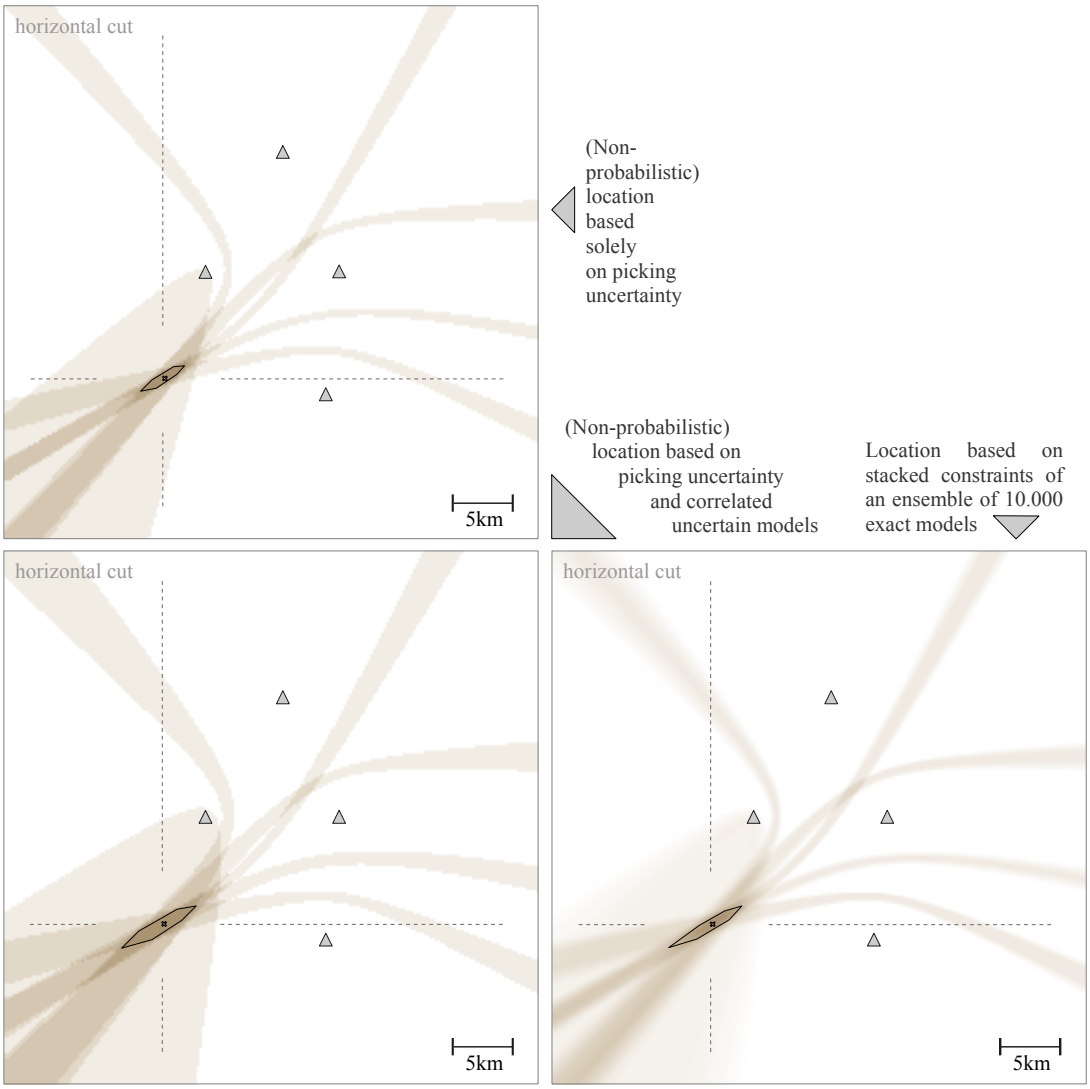


Fig. II.3.5.2:6: Uncertainty measures. The top panel shows constraints and congruent region (black outlined polyhedron) based on a picking uncertainty of approximately 0.05s. In the bottom left panel a 5% (synchronized/correlated) model uncertainty is added. The right panel shows the stacked constraints (as in the top panel) over 10.000 model variations, hence guarantying complete correlation among constraints. The congruent region is therefore thinner as the corresponding proxy using (synchronized) uncertain models.

Since the models are now fully correlated between all constraints, the congruent region obtained from the model-ensemble is narrower than the one obtained from uncertain models. Mapping the model uncertainties in this manner can hence increase the precision of the location significantly, especially when uncertainties are very large. This is demonstrated in fig. II.3.5.2:7, where we assume the mean velocity only to be known with a 50% uncertainty. The uncertain-model's

congruent region is here extremely large. Yet, using the given information of homogeneity and complete model correlation in model-ensembles reduces the location uncertainty drastically (fig. II.3.5.2:7, dashed polyhedron).

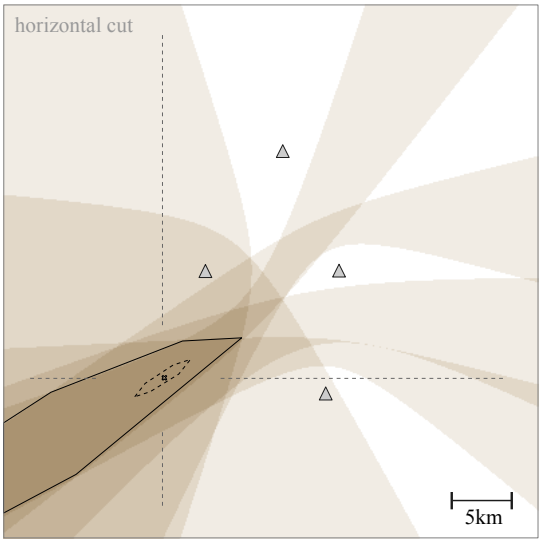


Fig. II.3.5.2:7: Comparing uncertain models and model-ensembles: One set of arrival-times leads to two significantly different location uncertainties when the the uncertainty in velocity is large (here 50%). For uncertain models the congruent region is significantly larger (solid polyhedron), since models are only treated correlated per constraints, overestimating the location uncertainty. Model-ensembles, which keep the correlation valid amongst all constraints, describe the location uncertainty precisely.

The interesting point is that the dashed region not only corresponds to a 50% but 100% uncertainty in velocity - in fact, marks the region of all possible hypocenter locations for a completely unknown velocity. Solely the information that the velocity structure is homogeneous is sufficient to constrain the solution to the confined dashed region (fig. II.3.5.2:8): When we look at the models similar to the true model ($1.0 \cdot v_p$) we find that for this source location only models with values between 97% and 104% of v_p are able to bring the constraints into agreement. Merging the solutions of all models yields the dashed region of congruency of fig. II.3.5.2:7. Since it is only this small band of models that confirms the arrival-time data, the region of possible locations would not increase even when extending the range of v_p to $v_p \in [0, \infty]$. This small elongated region would increase (in width) if we allowed for inhomogeneities in the model, which we could express simply by describing each of the ensembles' models as uncertain model of small uncorrelated uncertainty.

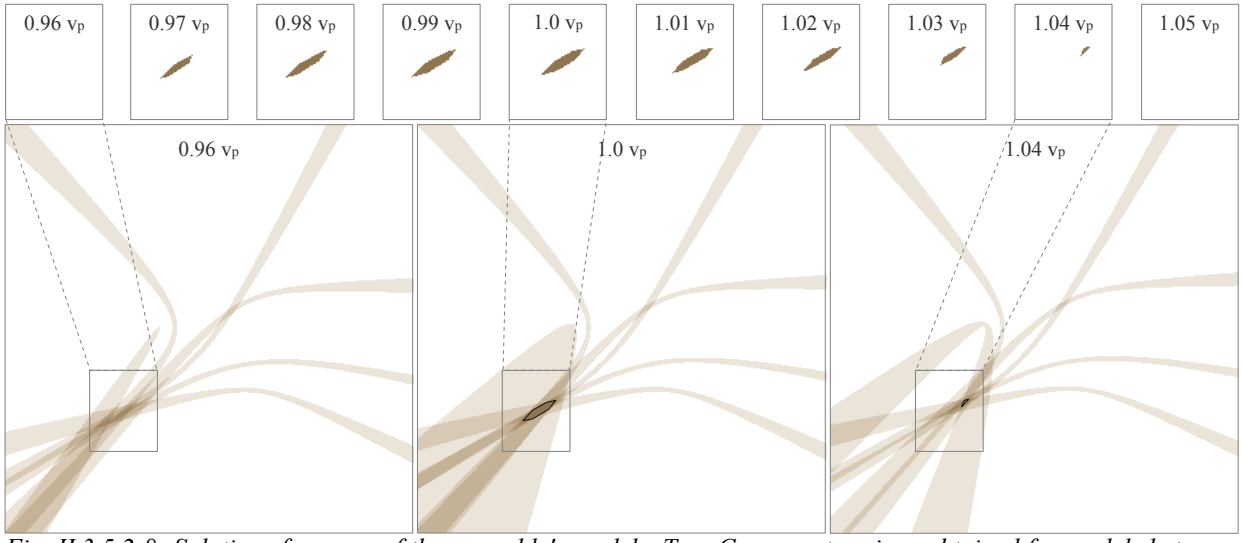


Fig. II.3.5.2:8: Solutions for some of the ensemble's models. Top: Congruent regions obtained for models between [0.96, 1.05]·v_p*. Only a small band of model parameters lead to constraint agreement. Merging these solutions displays the total location uncertainty. Bottom: Constraints for certain models: Left: Velocities too slow do not bring information into agreement. Middle: constraints for the true model. Right: Largest velocity that still yields a small region of congruency.*

The example shows that we can expect more precise location uncertainty estimates by using model ensembles, whenever the model uncertainties become large. This, for instance, is the case when the model uncertainty includes varying geometries and hence large fluctuations in velocity. A common example for 1D models are unknown layer depths, in which proximity the velocity fluctuates between two layer velocities.

For real time analysis, uncertain models provide a robust measure of location inaccuracies, yet tend to overestimate location uncertainties, when using model variations synchronized concerning the wave paths, between pairs of phase arrivals. A post processing using model-ensembles allows to obtain more accurate uncertainty estimates and hence more precise location uncertainties. It is, however, computationally more expensive, evaluating only tens of 1D-models per second: Evaluating the common amount of 1000 model variations requires therefore up to a minute computation time, rendering this method as inapplicable for real-time/interactive schemes. Complex models require larger ensembles of Monte-Carlo-generated models.

Having estimating the effect of model uncertainties unto the model-dependent location constraints (and in turn the location uncertainty), we will conclude the general topic of uncertainty, looking at the construction of the model-independent constraints of back-azimuth beams, before we will go over to the technical aspects and problems of the constraint computation in the following chapters.

II.3.6 **Array Beam Uncertainty**

The scheme of distinct constraints, which is applied in this thesis to constrain hypocenter regions, can as well be used to constrain any other parameter. In this chapter, it will be used to constrain the array back-azimuth interval, providing robust opening angles for the back-azimuth beam as a function of the data quality.

The total uncertainty of the array back-azimuth estimate relies on, however, two other significant aspects. The **first**, although neglected in the following, relates to the given velocity structure. The concept of a "Back-Azimuth" (BA), which is the direction to the source -measured at a seismic station- only makes sense for laterally homogeneous media. Structures as e.g. dipping layers can refract the ray out of the onsite measured direction – an effect well known in optics, commonly leading to an underestimation of an object's depth under water. The related error can only be corrected if sufficient knowledge about the velocity structure is given. In that case, the ray may be traced back through the medium, starting with a vector in the direction of the back-azimuth, following the gradient of the travel-time field obtained from the given velocity model. This generates a complex back-azimuth constraint for laterally inhomogeneous models. However, this detailed model information is seldom given. In practice, simple BA correction functions are applied, which are obtained by comparing observed and catalog BA. These 360° periodic functions can e.g. appear in the form of a sinusoidal curve for dipping layer structures below the array. In this work, sufficiently lateral homogeneous models will be assumed and a fixed angular uncertainty of 5° being assigned to cover the impact, small structural inhomogeneities may exert.

The **second** aspect is linked to the equation deriving the slowness estimate, which requires planar station layouts (which will be assumed to be given or otherwise to have been corrected for) and plane wave fronts, which are only given for events in larger distance. The methodological error that arises for events in near distance will be evaluated in chapter III.1.1.

The last aspect, being the subject of this chapter, is linked to the signal-correlation in (small-aperture) arrays: An aperture too small to capture the signal's dominant wavelength prohibits a proper cross-correlation (or would lead to wide maxima with a *f-k*-analysis, respectively). An aperture too large, on the other hand, will lead to a large variability in the waveform between the array stations, causing the quality of the cross-correlation to drop (or the semblance to be low,

respectively) affecting the estimate's accuracy. Further, if the signal is correlated in the time domain, a sample rate too small will – especially with small-aperture arrays or steep incidence angles – cause large azimuthal inaccuracies due to the poor discretization. Eq. II.1.2.3-6 provides an unique back-azimuth estimate based on the two time shifts between three traces. However, all these aspects add errors to the estimate, which need to be covered by a proper parameter uncertainty to yield congruent hypocenter regions and accurate results. This measure will now be established using the concept of distinct constraints as in previous chapters, only that here the slowness-space is being constrained, to identify the uncertainty interval of the back-azimuth parameter. This will determine the beam opening angle based on the given waveform- and geometrical data, covering the errors pertaining to low data quality, poor sampling/discretization and short station-station distances in small-aperture arrays. Where constraints, based on the difference of two arrival times, were used to constrain a congruent solution region in space, here will the two time shifts between a set of three stations be used to constrain a congruent region in the back-azimuth/slowness phase space, constraining the feasible back-azimuth interval, which defines the direction and opening angle of the back-azimuth beam. We will therefore

a) jackknife the dataset (of array traces)

to form

b) uncertain (slowness) constraints

that

c) narrow down the solution in (slowness) space

to

d) identify the feasible region for the (back-azimuth) estimate,

which provides us with the *data-based* uncertainty for the back-azimuth location constraint. Since the slowness estimator (eq. II.1.2.3-6) requires the coordinates of three stations and the time shifts between the three corresponding waveforms, the array stations will hence be permuted into triple-groups, exactly solving eq. II.1.2.3-6 for the slowness (a). The constraint's (triple-group) uncertainties are determined by the uncertainty of the time-shifts (taking the station coordinates as error-free). The time-shifts are usually obtained via a cross-correlation of the waveforms, since this provides a very accurate and precise measurement of the time difference - given that the waveforms are similar! Cross-correlation factors (i.e. the height of the maximum peak in the cross-correlation

of the normalized waveforms) significantly smaller than one, wide maximum peaks, close side peaks of similar height, are all indicators that the estimate may not be accurately determined. In such a case, a "feature"-based correlation provides better estimates, only focussing on the correlation of a single peak or trough rather than the whole wave-train. This is often the case with very weak signals (typical for Nanoseismic Monitoring) where the coherency drops over neighboring wave crests. Of special interest is the correlation of the first phase arrival, since this traveled on the fastest path and is hence the favorable segment for the directivity analysis, i.e. the derivation of the back-azimuth. This avoids the mixing of different phase-arrivals in the wave-train, in which later phases as propagating from the source and earlier phases traveling on slightly different (longer) paths superpose, hence yielding slightly different back-azimuth estimates. We will therefor focus on the most "pure" i.e. the first arriving phase. Even this first oscillation, observed on several traces (stations), will slightly vary in length, which means we can correlate the rising ramp, the peak, or the decreasing segment and will in all three cases obtain slightly different time-shifts (see the first phase in the P-phase of fig. II.3.6:1).

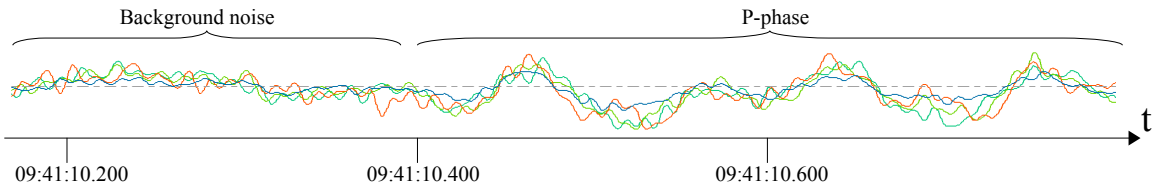


Fig. II.3.6:1: Overlay of four array traces, shifted according to the cross-correlation lag. The approx. first 0.5s of the low-SNR P-phase are used as correlation time window [09:41:10.4, 09:41:10.9].

As we are interested in the direction of the first arrival of energy, we would therefore attempt to obtain the time-shifts of the first ramp. Both standard approaches, signal cross-correlation and *f-k*-analysis, however, weigh the maximum (crest) significantly stronger than the phase onset. In noise conditions (as in fig. II.3.6:1), where a cross-correlation is not suitable, it may even be favorable to use direct onset picking rather than cross-correlation, weighing the onset of energy stronger than the later maximum (crest) of it. Transition to a picking based approach is straight forward, since the offsets obtained from cross-correlation can also be translated into a set of correlated onset times t^*, fixed to the beginning of the emerging signals. The t^* can therefore directly be obtained through picking and the uncertainties manually be assigned, following chapter. II.3.1. Where the cross-correlation would have yielded correlated uncertainty distributions, individual picking keeps them

uncorrelated, which describes the data in these high-noise conditions better. For the shown example of dissimilar waveforms (fig. II.3.6:1), the full wave-train cross-correlations of the array signals would yield discrepant sub-solutions in the Jackknifing scheme (fig. II.3.6:2, white vectors), with none pointing towards the true direction. Manual correlation (manual alignment of the array traces) is highly ambiguous, yielding strongly varying back-azimuth estimates. The f-k-analysis (fig. II.3.6:2, map) confirms the Jackknifing trend and shows the maximum energy for the stacked waveforms in direction of approx. 295°.

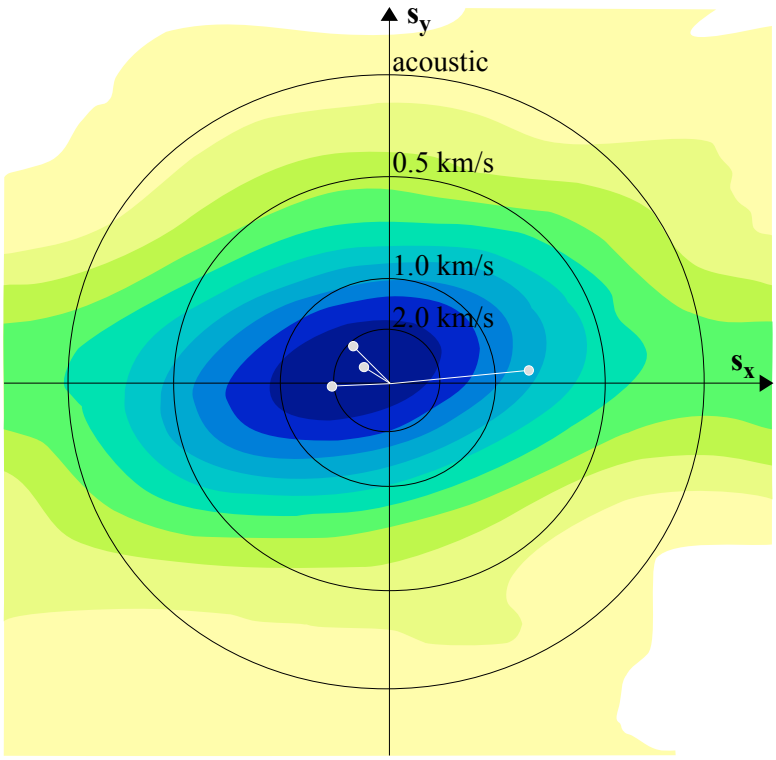


Fig. II.3.6:2: For the given lags, most of the four sub-solutions (white lines) indicate seismic velocities, yet diverge strongly in back-azimuth. The fk-analysis shows a maximum at approx. 295° with seismic velocities.

To obtain better results, we will now focus on "correlating" solely the first trough instead of the full signal. The uncertainties for the correlation are determined manually by the uncertainties of the trough's onset (fig. II.3.6:3). Due to low signal similarity and noise contamination the uncertainties will be treated as uncorrelated, following the standard formalism used in previous chapters.

Chapter II.1.2.3 laid out how a constraint (triple-group) for azimuth and slowness can be constructed for a given set of three array stations $\{A, B, C\}$, based on equation II.1.2.3-6.

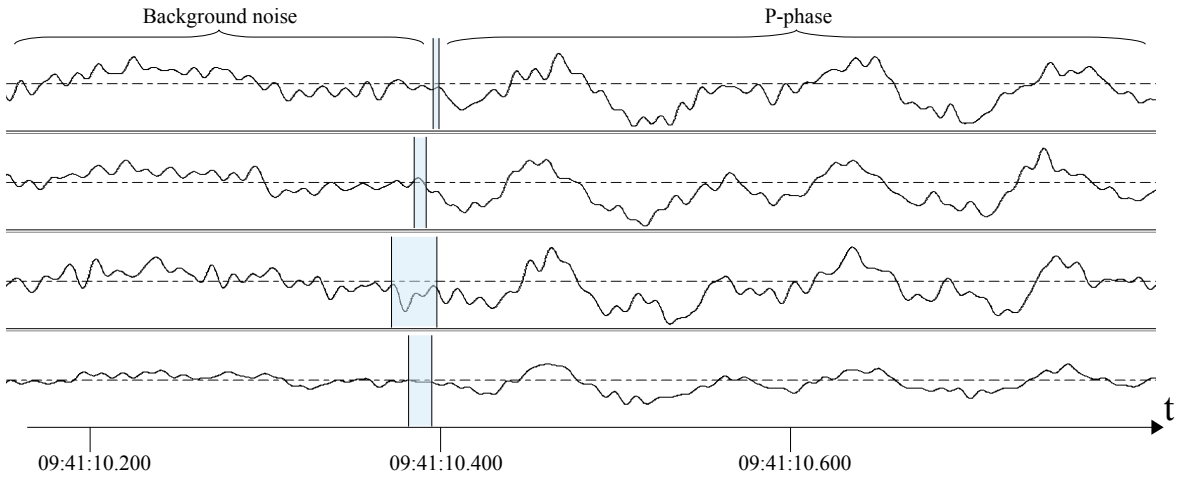


Fig. II.3.6:3: Weak P-phase on four small-aperture array stations. The P-phase is being identified by its higher dominant frequency. A global cross-correlation yields erroneous results, as the signal loses phase from one crest to the next. Instead the onset of the first trough is being used to correlate the four array traces. Due to the low SNR the onset is uncertain (blue windows mark the possible onset intervals).

Each of such constraints results in a point in the azimuth-slowness phase space, if uncertainties are neglected. For several constraints these points will usually scatter, the scatter indicating the agreement and robustness of the given estimates. Following the approach of chapter II.3.1, we will now project the uncertainties in arrival-times obtained from fig. II.3.6-3 into uncertainties in the phase-space to compute congruent regions identifying slowness and azimuth values consistent with all measured time differences: The uncertainties in the time domain are projected into phase-space by computing the set of points that is given by the slowness vectors obtained for every possible permutation of the set of $t_{|i|} \in [\underline{t}_{|i|}, \bar{t}_{|i|}]$. This yields a slowness map,

expressing the likelihood $p^{A,B,C}_{\begin{pmatrix} sx \\ sy \end{pmatrix}}$ that a certain slownesses vector $\begin{pmatrix} sx \\ sy \end{pmatrix}$ is constrained:

$$p^{A,B,C}_{\begin{pmatrix} sx \\ sy \end{pmatrix}} = \sum_{\underline{t}_A}^{\bar{t}_A} \sum_{\underline{t}_B}^{\bar{t}_B} \sum_{\underline{t}_C}^{\bar{t}_C} \begin{cases} 1, & \boldsymbol{S}(t_A, t_B, t_C) = \begin{pmatrix} sx \\ sy \end{pmatrix} \\ 0, & else \end{cases} \qquad \text{(II.3.6-1)}$$

with

$$\underline{t}_{|i|} \text{ and } \bar{t}_{|i|} \qquad \text{(II.3.6-2)}$$

representing the lower and upper limit of the onset times ($t^* \in \left[\underline{t}_{|i|}, \bar{t}_{|i|} \right]$). $\boldsymbol{S}$ is the function computing the slowness vector based on the three onsets and station locations (eq. II.1.2.3-6). Fig.

II.3.6:4 shows, that contrary to the error free scenario where the solution is defined by one point in phase space obtained from $\boldsymbol{S}(t_A^*, t_B^*, t_C^*)$, here a solution region is constrained in slowness-space by $p^{A,B,C}$.

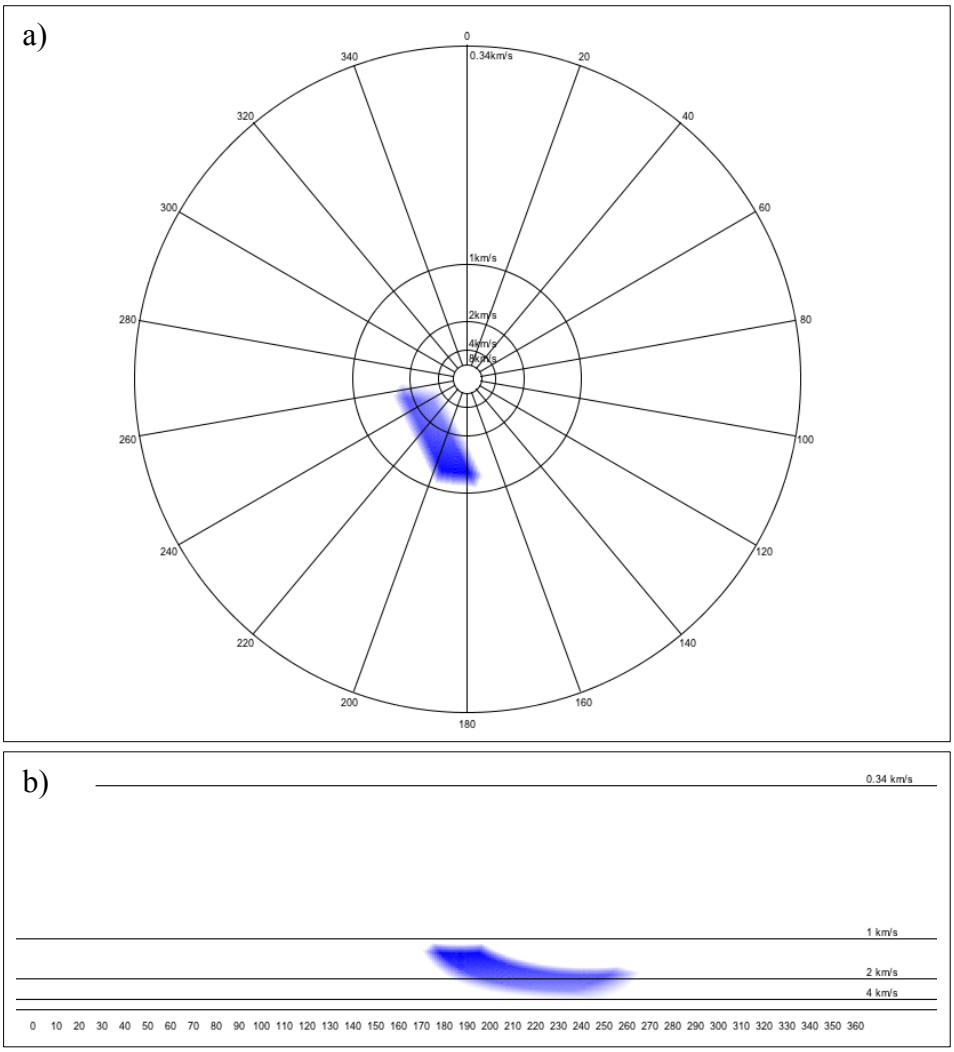


Fig. II.3.6:4: Slowness $p_{\binom{sx}{sy}}$ *-map displaying all solutions in azimuth/slowness phase space for given cross-correlation/picking uncertainties in a) Polar coordinates b) Cartesian coordinates. The back-azimuth interval for this subgroup stretches from 175° to 260°.*

Each triple-group permutated from the set of array stations yielded one constraint in the azimuth-slowness phase space. The sum of all subgroup likelihoods

$$P_{\binom{sx}{sy}} = \sum_{A=1}^{n} \sum_{B=A+1}^{n} \sum_{C=B+1}^{n} p_{\binom{sx}{sy}}^{A,B,C} \qquad \text{(II.3.6-3)}$$

identifies the azimuth-/slowness values most likely for the full array, with n being the number of array stations.

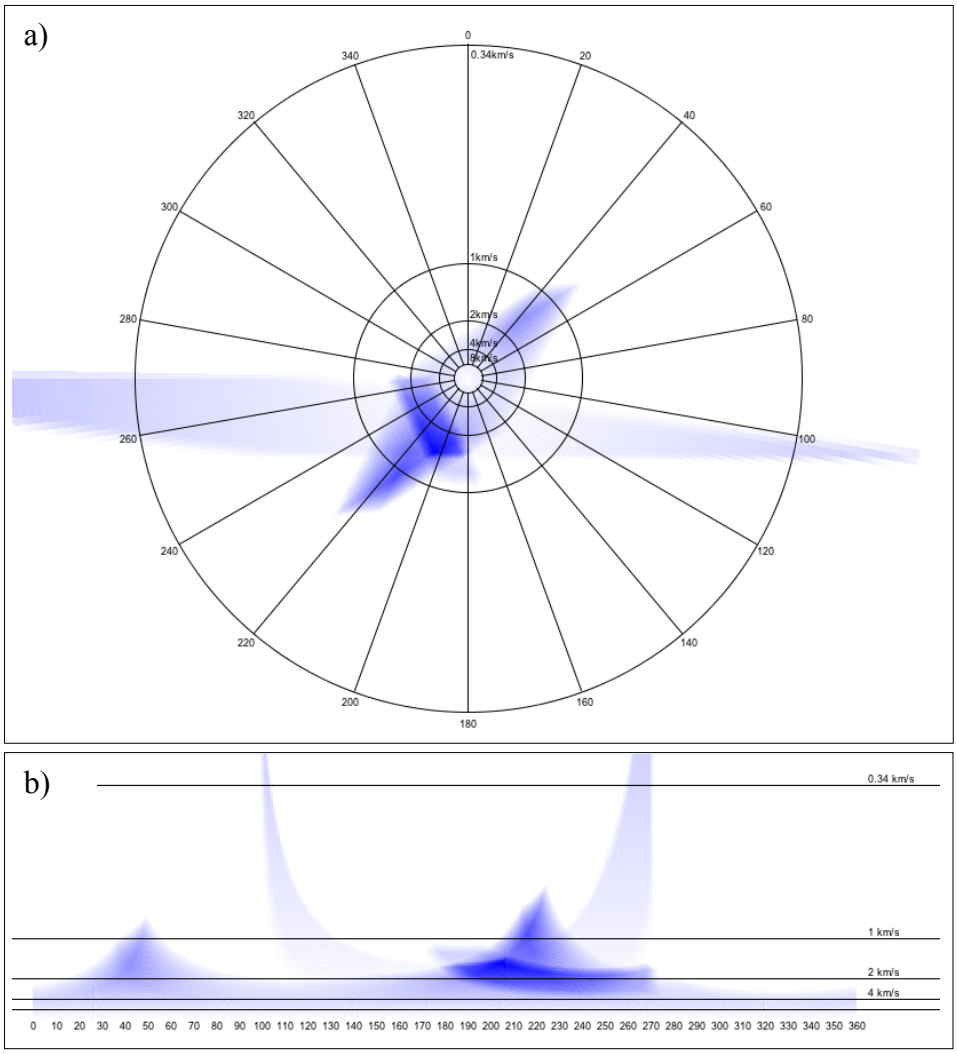


Fig. II.3.6:5: Using all four triple groups generated from the four array stations {{1,2,3}, {1,2,4},{1,3,4},{2,3,4}} yields four constraints for the azimuth/slowness phase space.

Here, however, we are interested in the congruent region, spanning out the back-azimuth interval. We can therefore mask eq. II.3.6-3 to identify the intersection of all constraints as region with $P^m_{\binom{sx}{sy}} = 1$:

$$P^m_{\binom{sx}{sy}} = \frac{1}{g} \underbrace{\sum_{A=1}^{n} \sum_{B=A+1}^{n} \sum_{C=B+1}^{n}}_{g \text{ triple groups}} , \begin{cases} 1, & p^{A,B,C}_{\binom{sx}{sy}} > 0 \\ 0, & else \end{cases} \tag{II.3.6-4}$$

with g defining the number of triple groups

$$g = \frac{1}{6}\left(n^3 - 3n^2 + 2n\right). \qquad \text{(II.3.6-5)}$$

This identifies the azimuth and slowness values, that are consistent with all time differences and uncertainties in the array:

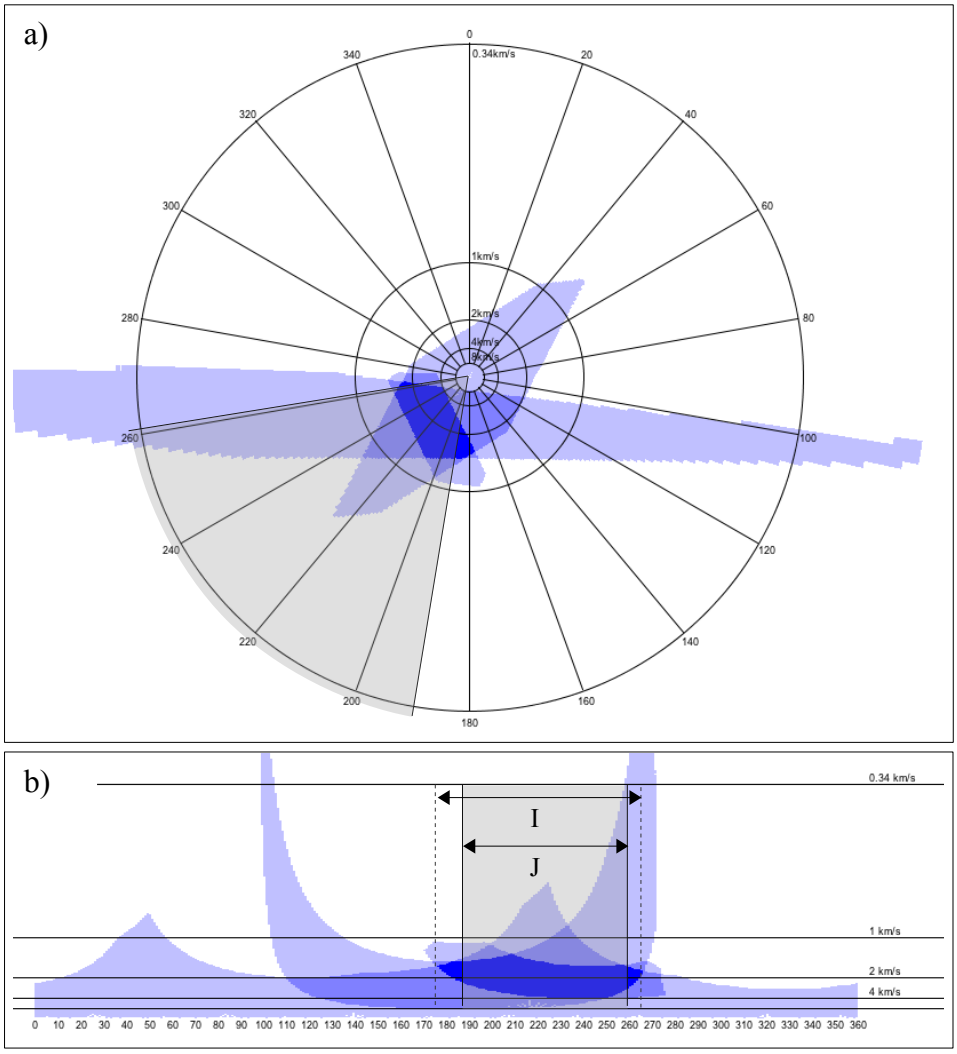


Fig. II.3.6:6: Non probabilistic display of fig. II.3.6:5, identifying the congruent region (dark blue) in which all slowness constraints agree (back-azimuth Interval I). Any slowness vector from within this region will cause arrival time differences, conform with the assigned uncertainties on all traces. As the local p-wave velocity at the array is known to be approximately 2.2km/s, only slowness values lower than 1/(2.2km/s) are physically possible. The congruent region under this slowness threshold is limited to the interval J (grey zone) and defines the signal-based beam uncertainty.

Fig. II.3.6:6 shows that the congruent region for the given uncertainty intervals on the four noisy

signals traces spans out a large back-azimuth interval (I) of $BA \approx [176, 265]^\circ$. This interval displays the signal-based uncertainty that would define the minimum width of the beam constraint. Since the low SNR signal is followed by a S-phase signature, it can be safely assumed that the signal is a seismic signal. The average local P-velocity is approximately 2.2km/s. This allows to narrow down the back-azimuth interval to $BA \approx [187, 262]^\circ$ (resembling "the beam", fig. II.3.6:6: Grey cone) since the observed slowness defines the lower bound of the signal's propagation velocity (Based on the incidence angle, the apparent velocity can only be higher than the absolute wave velocity).

The signal stems from an air-plane crash approx. 20km South-West of the small aperture array, with a back-azimuth of approx. 253°. While automatic (or manual) cross-correlation of the signal in fig. II.3.6:3 produced highly fluctuating back-azimuth results, using uncertainty windows around the phase arrival provides a quantitive measure for the back-azimuth uncertainty, identifies the right solution. Hereby, the back-azimuth beam's opening angle is brought into dependency to the data quality and array geometry. By defining a minimum uncertainty of e.g. one sample, sampling/discretization effects are naturally covered. Coherent signals, a well distributed array geometry and a locally homogeneous velocity structure yield narrow beam angles while the angle widens when these qualities are diminished. The example of fig. II.3.6:6 lacks both, coherent signals as well as structural homogeneity. In combination with the modification for near distances where the plain wave assumption is violated (chapter III.1.1), a robust back azimuth constraint is formed. With good data, cross-correlation is the method of choice to provide precise time-shifts [Note: in that case, the sensitivity analysis (eq. II.3.6-1) needs to be performed in a correlated manner], but when the signal quality drops in low SNR conditions, a "feature"-based correlation, even using the "picking" of the onset of energy becomes useful to constrain the (less-precise) solution robustly in slowness-space.

This chapter concludes the translation of time based uncertainties into constraint uncertainties. The following chapter will address the technical aspects of mapping these time-based constraints to space, in the context of a direct search formalism.

II.4 **Constraint Evaluation in a Sampled Region**

One of the complications that arise when only few stations are given are ambiguous solutions (chapter III.2). These multiple, equally possible hypocenter locations (which are not due to errors in the parameters but due to equivalent mathematical solutions to the underlying system of equations), may cause significant misinterpretations of the source, if one of the solution candidates is overseen and/or the false one is trusted to mark the hypocenter location. One fundamental, but technically complex demand for the analysis are therefore global and complete solutions, revealing all existing solutions both a) extensive solution regions in large distances as well as b) compact regions near by. This, however, is not trivial: Complex velocity models and multiple cross-sections are only some of the aspects that make up the complexity of the constraint-based real-time location analysis and necessitate the use of global direct search methods a.k.a. grid searches. In such, the solution space is sampled and evaluated at a set of coordinates (e.g. points of a regular grid). But whenever space is sampled, solutions can be overlooked: The direct search method is limited in two ways, in a) the extent of the search region (the grid size) and b) the evaluation point density (the grid resolution). For complete solutions, it requires a) extensive search regions and b) high evaluation point densities, to identify compact hypocenter locations and evaluate the solution neighborhood. Yet, both aspects increase the computation time and hinder real-time analysis. The task of the following chapters is therefore to derive methods that satisfy both requirements in a real-time capable manner. The importance of these aspects can be easily demonstrated with the following example taken from the field of real-time location: When one of today's running Earthquake-Early-Warning codes was tested on the 2015 Ml 4.2 Dead-Sea event, using the four first triggering stations, it mis-located the earthquake by roughly 200km (fig. II.4:1) although the event was situated in between the first triggering stations - in only (approx.) 10km epicentral distance. This overestimation of epicentral distance would have lead to a significant overestimation of the magnitude and in turn to a false alert. How could it have produced such a wrong location estimate? The cause is found in the algorithm's attempt to identify the hypocenter by the grid point of lowest residuum sum value. Two complications caused this attempt to fail, one fundamental one, being an inherent problem of the arrival-time based location, and a second technical one. The first complication was that the set of arrival times yielded two possible locations. This circumstance of ambiguous locations (multiple exact solutions) is covered in detail in chapter III.2, which shows: Based on P-arrival times only, the

true location cannot be identified. The algorithm, only expecting one solution, also identified only one of the locations.

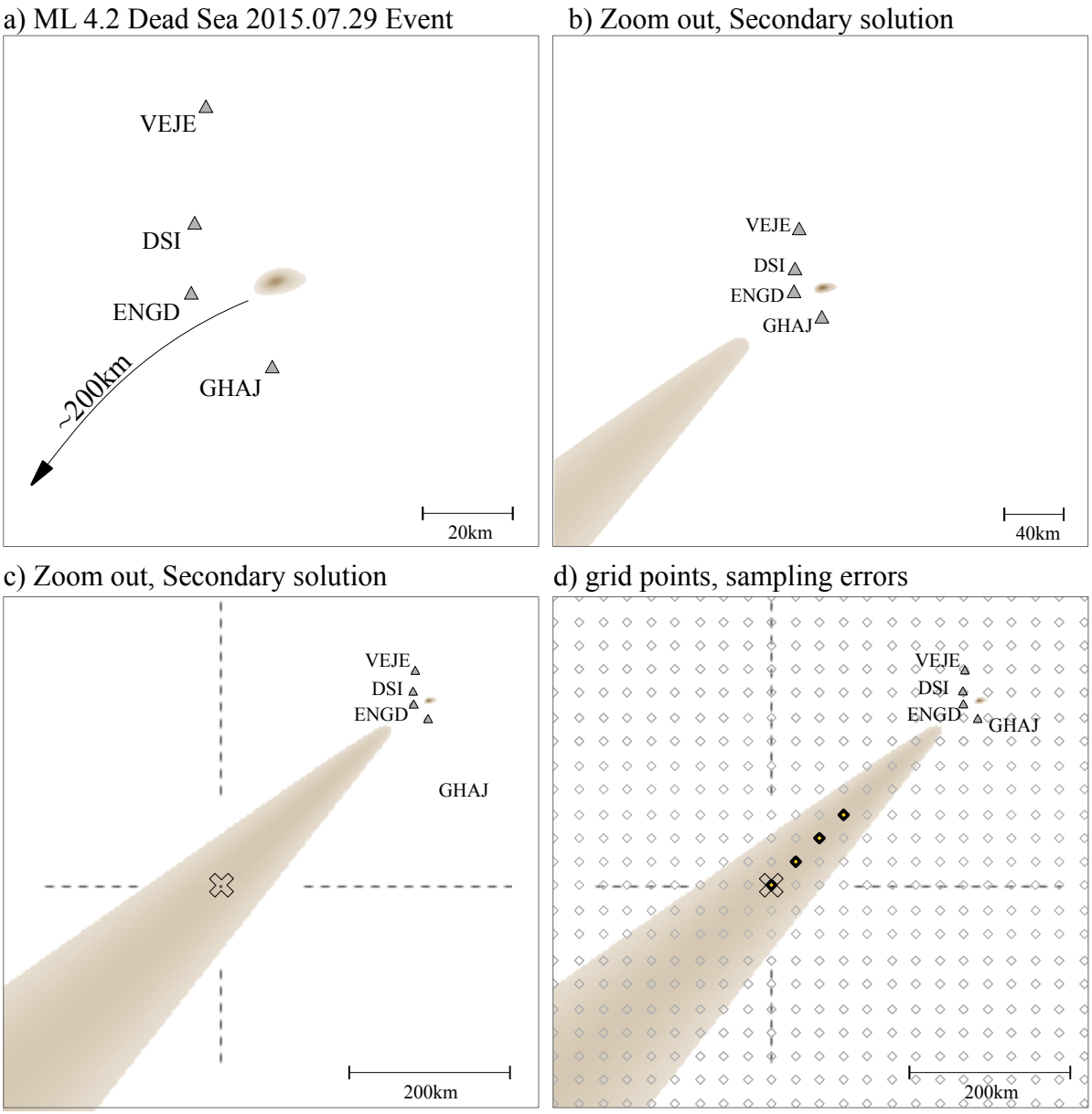


Fig. II.4:1: Falsely identified hypocenter location: a) The true hypocenter location is marked by a minimum in the residuum plot. The algorithm, however, located the event approx. 200km south-west. b) Zooming out reveals a secondary, less distinct, minima. c) Zooming out even further shows the location chosen by the the algorithm. d) Viewing the evaluation grid explains the mis-chosen hypocenter: The compact primary solution falls between grid points in such a way that grid points in the center of the secondary solutions reach lower residuum values than the points in proximity to the true location. The four grid points of lowest residuum values are marked yellow

In this case, one of the two solutions was even less likely than the other as it required larger uncertainties in the input parameters to render it an exact solution. However, the algorithm chose

this solution, being the wrong solution. Now, the answer is found on the technical side: Dynamically refined grids (e.g. oct-trees) are technically able to identify all minima, but even the commonly applied Oct-tree algorithm can overlook sharp and compact minima that are situated in between grid points. Exactly this was the second complication, that surfaced in the example of fig. II.4:1: A secondary minimum in larger distance from the network existed and the primary solution with lowest residua sum was situated between two grid points, so that grid points at the secondary minimum yielded lower residua values than those in proximity to the primary and true solution. The example emphasizes the importance of grid schemes that provide *complete* solutions and do not neglect part of the true solution set. In the following we will tackle the problem of completeness on both ends of the scale: the limited extent of the grid and it's resolution.

In the simplest case, the three dimensional space is discretized into an equally spaced multi dimensional grid. To increase the grid's extent, we can introduce spatial transformations to increase the grid cell size with distance, effectively reducing the grid resolution in distance of the network, improving performance and allowing to cover larger regions. This approach does not increase computational cost during the analysis, since the spatial (forward) transformation is only performed during the initial grid-rendering procedure (assigning travel-times to all grid points); and the back-transformation is only required whenever a region is declared as hypocenter and the geographical coordinates become of interest. This procedure is, however, only helpful if we can form a method that solves the second aspect: to provide complete solutions even when the grid resolution is low. The following chapter will focus on this aspect, while the spatial transformation will be covered in detail in chapter II.4.2. To derive a method that can analyze *between* grid points and hence provide complete solutions independent of the resolution, we will start looking at the fundamental concept of a grid point based direct search:

To map constraints, which are based on observed arrival-time differences, into space, it requires a comparative quantity at each grid point, that connects time and space. The same way that velocity describes and connects changes in space and time, the velocity model connects the time defined constraints with locations in space. For constraints which are based on arrival time differences, the comparative quantity can be established by calculating travel times from each grid cell to all stations. This is due to the equivalence of travel time $tt(\bar{\boldsymbol{x}}, \bar{\boldsymbol{s}})$ and arrival time $t^O(\bar{\boldsymbol{s}})$ for a zero adjustment concerning the origin time $t_0 = 0$,

$$t^O(\bar{\boldsymbol{s}}) = t_0 + tt(\bar{\boldsymbol{x}}, \bar{\boldsymbol{s}}) \qquad \text{(II.4-1)}$$

with $\bar{x}$ being grid cell- and $\bar{s}$ station coordinates. Differences in onset times Δ^O in general are equal to differences in travel times Δ^{tt}, since

$$\Delta^O = t_B^O - t_A^O = (t_0 + tt_B) - (t_0 + tt_A) = tt_B - tt_A = \Delta^{tt} . \quad \text{(II.4-2)}$$

With cached travel times at each grid cell, the comparative value based on given models can be evaluated in real time at each grid cell for every constraint due to the simplicity of required operations. Cached travel time tables can nowadays be found in many applications inverting time based data e.g. seismic imaging or location schemes (Lomax, 2009). In the following chapter we will now derive a direct search approach that evaluates grid cells rather then grid points.

II.4.1 **Grid Discretization**

When we compute location constraints on the points of regular or irregular grids, resolution is a critical aspect. A resolution too high causes the analysis to be computationally expensive, a resolution too low may significantly falsify the result in regions where the constraint's width is comparable to the grid spacing (aliasing). In regions where the constraint's width drops below the grid spacing the constraint evaluation may be fully suppressed (fig. II.4.1:1).

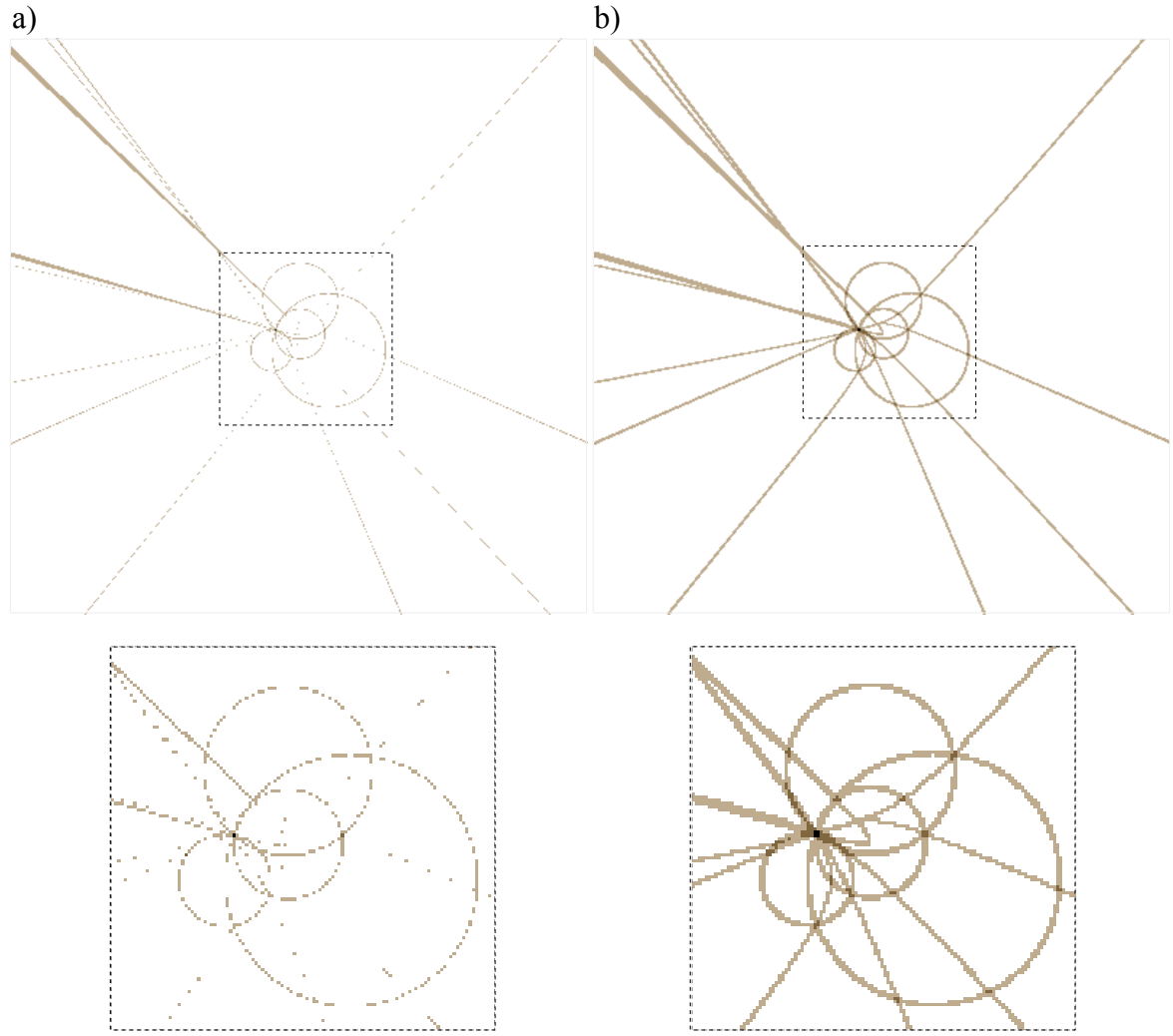


Fig. II.4.1:1: Comparison of evaluation a) on grid points and b) over grid cells. A point based approach may underestimate the true value, not identifying the global solution (a). A simple interval based approach may overestimate the true value, but will identify the global solution, independent of the grid resolution. The central region is shown magnified (below).

With several constraints, these regions are distributed differently for every constraint depending on the station layout, which makes the effect unto the solution mapping unpredictable. As solutions can

completely be missed, a point vise computation of constraints is unfavorable. The problem can, however, be solved by evaluating the constraints (or more accurately: integrating their likelihood distributions) over the grid cell rather than solely using the values on the grid points. This way, constraints and the global and secondary solutions are identified even when the grid resolution is too low for grid points to "see" them. For readability, the following derivation will be shown disregarding model uncertainties, whose interval based formulation follows the same pattern. The difference between the constraint parameter T^O (i.e. the arrival-time difference) and the modeled value T^M (i.e. the travel-time difference) at grid point (x, y, z) defines the constraint deviation V (i.e. the discrepancy between modeled and observed quantities)

$$V = T^{OM} = T^M - T^O \,. \tag{II.4.1-1}$$

A region is marked as part of the constraint if the deviation is less than the constraint uncertainty. Now, if we attempt to integrate over grid cells rather than to evaluate on grid points only, we need to estimate the range of constraint deviation over the grid cell's extent. This range can be approximated by the values of T^M at the surrounding grid points (fig. II.4.1:2).

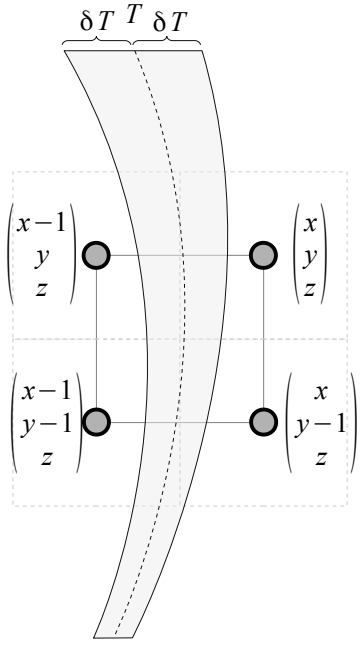


Fig. II.4.1:2: Constraint evaluation over a grid cell. Where a grid point holds one travel-time value, the grid cell holds a range of travel-times. This interval can be approximated as stretching from the minimum to the maximum travel-time found at the cells corners.

As the grid is evaluated along horizontal and vertical cross sections (in interactive location) as well in full 3D (for automated analysis), different sets of surrounding grid points are needed. Let $\Theta^{XY}_{(x,y,z)}$ define the set of values at surrounding grid points of index (x,y,z) for horizontal cross sections

$$\Theta^{XY}_{(x,y,z)} = \left\{ T^{M}_{(x,y,z)}, T^{M}_{(x-1,y,z)}, T^{M}_{(x-1,y-1,z)}, T^{M}_{(x,y-1,z)} \right\} \tag{II.4.1-2}$$

$\Theta^{XZ}_{(x,y,z)}$ and $\Theta^{YZ}_{(x,y,z)}$ the sets for vertical cross sections

$$\Theta^{XZ}_{(x,y,z)} = \left\{ T^{M}_{(x,y,z)}, T^{M}_{(x-1,y,z)}, T^{M}_{(x-1,y,z-1)}, T^{M}_{(x,y,z-1)} \right\} \tag{II.4.1-3}$$

$$\Theta^{YZ}_{(x,y,z)} = \left\{ T^{M}_{(x,y,z)}, T^{M}_{(x,y-1,z)}, T^{M}_{(x,y-1,z-1)}, T^{M}_{(x,y,z-1)} \right\} \tag{II.4.1-4}$$

and $\Theta^{3D}_{(x,y,z)}$ the set for the full grid cell

$$\Theta^{(3D)}_{(x,y,z)} = \left\{ \Theta^{XY}_{(x,y,z)}, \Theta^{XY}_{(x,y,z-1)} \right\} . \tag{II.4.1-5}$$

The lower bound $\underline{\Theta}$ of signed deviations at grid index (x,y,z) can now be estimated by the minimum value of the relevant set $\Theta_{(x,y,z)}$. The upper bound $\overline{\Theta}$ is approximated by the maximum value of the set.

$$\underline{\Theta} \approx \ min\left(\Theta_{(x,y,z)}\right) \tag{II.4.1-6}$$

$$\overline{\Theta} \approx \ max\left(\Theta_{(x,y,z)}\right) \tag{II.4.1-7}$$

The likelihood that this cell is constrained by the arrival-time difference is then given by

$$p_{(x,y,z)}\left(\underline{\Theta} \le T \le \overline{\Theta}\right) = \int_{\underline{\Theta}}^{\overline{\Theta}} p_{df}(T)\, dT = \sum_{\tau=0}^{\lfloor (\overline{\Theta}-\underline{\Theta})/\delta t \rfloor} p_{mf}\left(\underline{\Theta} + \tau \cdot \delta t\right), \tag{II.4.1-8}$$

with δt being the sample length. As this evaluation has to be performed at every grid cell for each constraint, the summation of likelihoods over the given interval would become computationally

expensive. This can be avoided deriving the local constraint likelihood from a cumulative distribution function P instead, which is defined as

$$P(X) = \int_{-\infty}^{X} p(x)\,dx = p(x \le X) \tag{II.4.1-9}$$

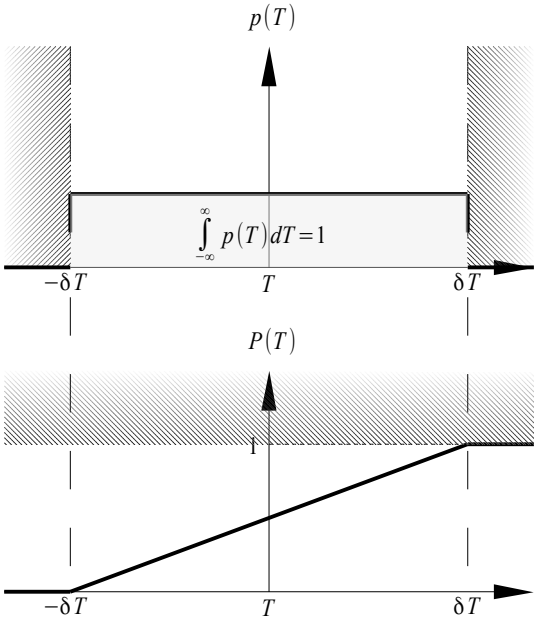


Fig. II.4.1:3: Cumulative distribution function

The likelihood over a given interval is then obtained via the difference of P at upper and lower bound

$$p(a \le x \le b) = \int_{a}^{b} p(x)\,dx = P(b) - P(a)\,, \tag{II.4.1-10}$$

replacing the summation in eq. II.4.1-8 by

$$p_{(x,y,z)}(\underline{\Theta} \le T \le \overline{\Theta}) = P(\overline{\Theta}) - P\underline{\Theta}\,. \tag{II.4.1-11}$$

The cumulative likelihood function has to be calculated and stored only once per constraint, making this formulation favorable for realtime calculation as it eases computation cost, but as the likelihood of constraint fulfillment is evaluated over the spatial extent of a grid cell, it causes the undesirable behavior that areas of true equal likelihood show different apparent likelihoods (fig. II.4.1:4). This is due to the discretization of space: Where the constraint in its uncertainty passes through one single grid cell, the likelihood is at its maximum. Where in larger distances the constraint surface widens and runs over several cells, the likelihood per cell drops below that value. This area dependent likelihood is misleading as points with constant constraint deviation should obtain the same weight e.g. the maximum for points with zero deviation at the constraint's center. This would be the case if infinitesimal small grid cells could be used. To compensate this effect that is caused by numerical discretization, the maximum value that could occur in one imaginary infinitesimal cells is ascribed to the grid cell in whole. Instead of eq. II.4.1-11 the following formulation will be used to evaluate the likelihood of the constraint at each grid cell:

$$p_{(x,y,z)}\left(\underline{\Theta} \le T \le \overline{\Theta}\right) \equiv \; p_{mf}\left(\Theta_m\right) \tag{II.4.1-12}$$

with

$$\Theta_m = \underset{\underline{\Theta} \le \Theta \le \overline{\Theta}}{arg\,max} \; p_{mf}\left(\Theta\right). \tag{II.4.1-13}$$

Figure (II.4.1:4) shows the effect of integrated likelihoods vs. the use of the maximum value occurring within a grid cell. While integrated likelihoods lead to pseudo solutions, the maximum value approach correctly represents the likelihood distribution.

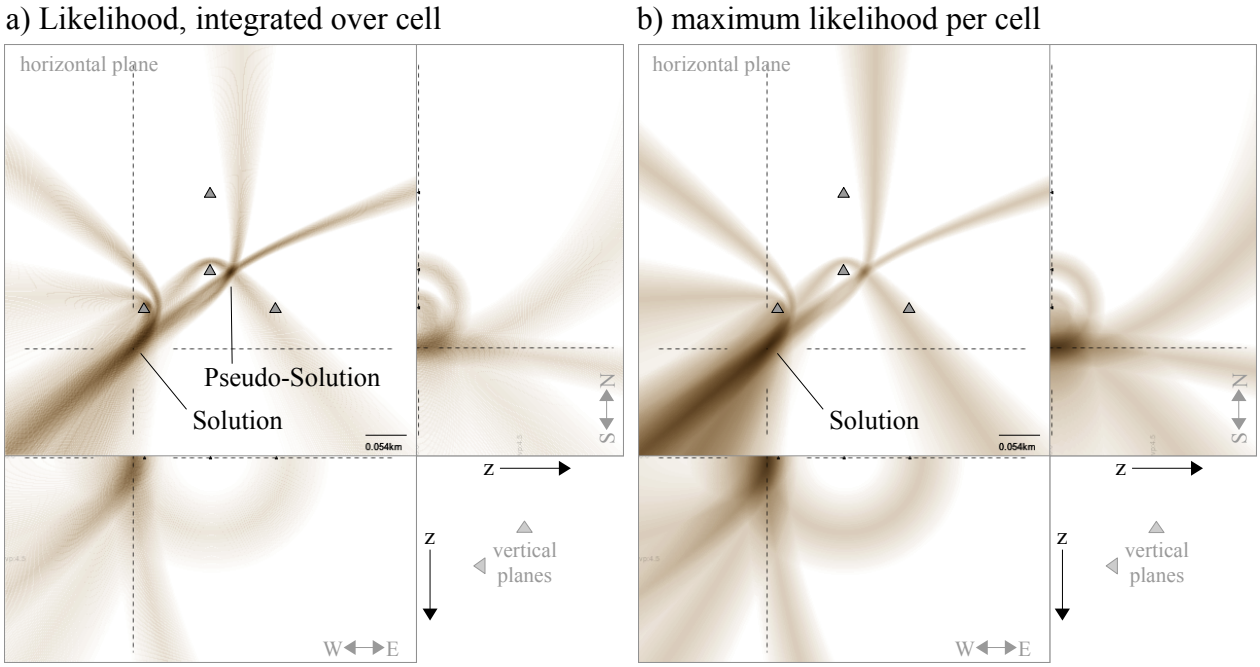


Fig. II.4.1:4: Probabilistic representations. a) Likelihood that a constraint will lie in a certain grid cell, obtained by integration of likelihoods over the grid cell interval. Where a constraint's width is comparable to the grid cell size, the likelihood may integrate up to the maximum value of one. This value is not reached in places where the growing constraint is much larger than the grid cells. This effect of grid discretization causes pseudo-solutions as constraints "fade out" with distance. b) Using the maximum value occurring in a grid cell as representative likelihood corrects for this discretization, displaying the likelihood resolution independent.

This grid-*cell*-based concept avoids the problems that surfaced in the example of the mis-located hypocenter of fig. II.4:1 at the beginning of the chapter: Fig. II.4.1:5 shows the congruent region in correspondence to the minimum residuum of fig. II.4:1 for different zoom levels and varying grid resolutions. The congruent region, spanned out by P-phase-based hyperbolic constraints, confirms the (neglected) solution of the residuum based approach. The extensive secondary solution that was identified as hypocenter by the grid-point approach, turns out to exceed the assigned uncertainties (the constraints do not overlap) and is therefore naturally neglected. The identification of the compact hypocenter solution is independent of the grid resolution: Even with extremely low grid resolutions the hypocenter is (with lower precision) accurately located.

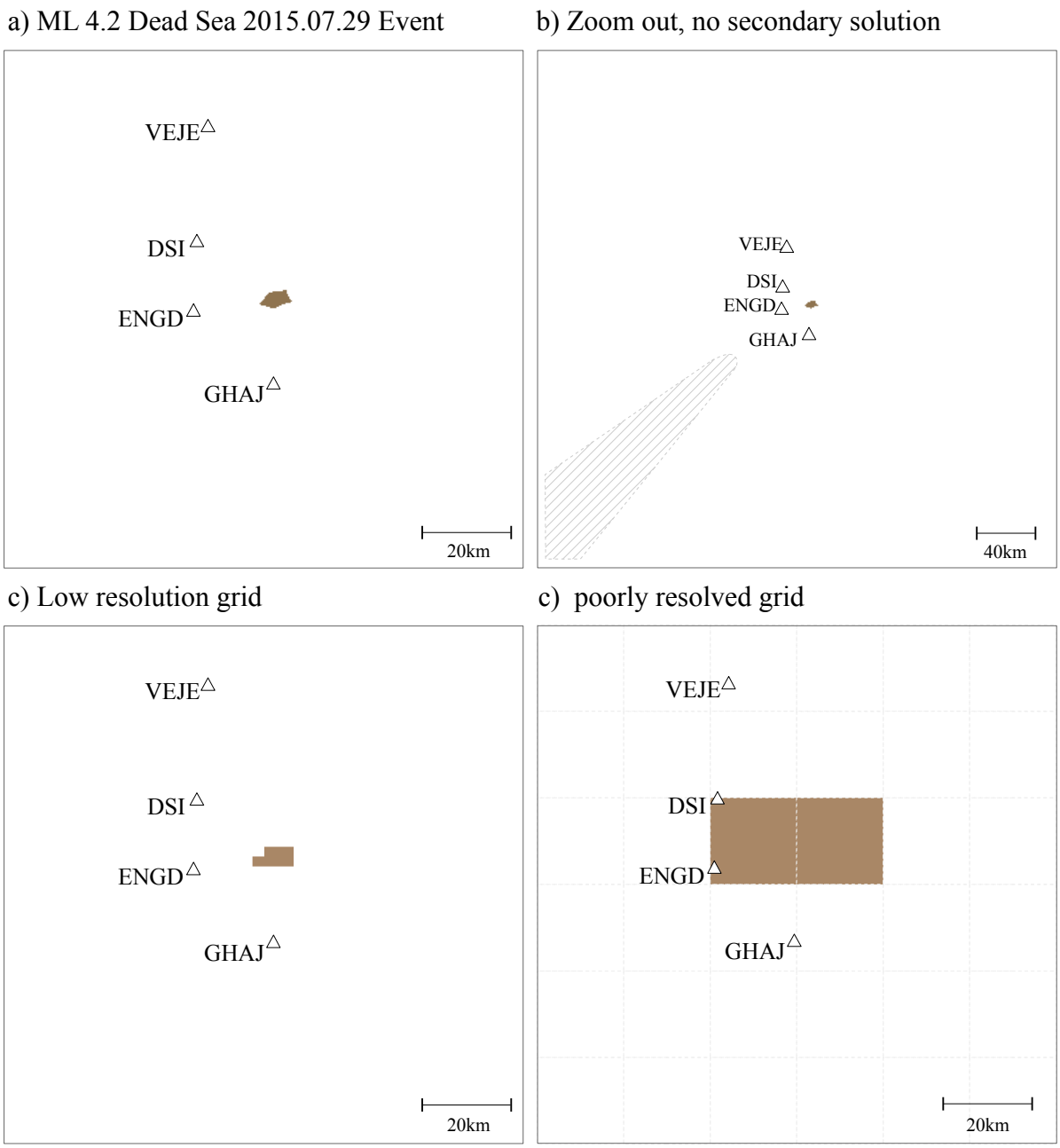


Fig. II.4.1:5: Constraint based solution corresponding to the example of fig. II.4:1. Panel a) shows the congruent region formed by the hyperbolic P-P constraints, corresponding with the minimal residuum of fig. II.4:1a). b) Zooming out, it becomes obvious that the secondary solution of low residuum (dashed region) does not lie within uncertainty bounds (the constraints do not overlap in this region) and is therefore naturally neglected. Using the grid cell based approach, even lowering the grid resolution correctly reproduces the solution (c) and even for extremely poor resolved grids the solution is successfully identified.

Where the grid-*point* based evaluation of residua sums missed the compact primary minimum, the grid-*cell* based constraint evaluation identifies even most compact location regions with low grid resolutions. This provides internal solution completeness and allows in the following to apply a variable grid resolution to comply with the second aspect of completeness: A large grid extent to identify possible ambiguities in larger distance.

II.4.2 **Travel-Time Grids and Grid Organization**

Direct search grids are often initialized with a constant grid cell size and subsequently locally refined around found minima (e.g. see the Oct-tree scheme). This serves to identify all local minima that could not be found using the initial grid point spacing, eventually identifying the global solution. Such an approach is, however, not favorable for a real-time constraint-based analysis (e.g. location algorithms in earthquake early warning systems or interactive visualization), especially when complex velocity models are used. Looking at the organization and the computation of the direct search grid tells us why:

If we can disregard a) uncertainties and constraint congruity, b) a probabilistic analysis, c) complex velocity models, d) three dimensional station distributions and e) the use of vertical cross sections of the solution space, constraints can be computed extremely fast using semi-analytical approaches (Eisermann, 2008). For each of item a-e), the iterative semi-analytical approaches either lack the capability or lack robustness, which reflects in a significantly increased computation time. If any of item a-e) are required, a real-time computation and analysis of constraints can only be achieved by evaluation on multidimensional grids, either in full or at given cross sections (eg. xy, xz, yz), using a simple, fast and robust direct search. The direct search utilizes two groups of three-space-dimensional grids, each holding additional dimensions for the various information: One group contains grids holding the results of the constraint evaluation (i.e. the “mapped” constraints). Every calculation during the analysis is performed on such. They consist of three spatial dimensions and one additional dimension for each constraint type to allow a separate representation and evaluation of the independent constrain groups and supplementary residua sum attractors. The computation of velocity-model-dependent constraints is based on the travel times of all phases between corresponding stations and all grid points. These travel times can be calculated prior analysis and stored in cache matrices, guarantying realtime applicability independent of the complexity of used velocity models. These travel-time grids make up the second group of grids. The grid's dimension is determined by the three spatial dimensions and twice the amount of evaluated phases, as minimal and maximal possible arrival times are required. Since the analysis requires the travel-times between any given coordinate and all n stations, n of such grids exist.

The travel-time of the analyzed phases is one of the fundamental quantities in the location inversion problem. Its calculation is a forward problem and can be performed using analytical, raytracing or

waveform solving schemes, depending on the model's complexity. While analytical solutions can only be obtained for simple models as the homogenous space, they allow most precise estimates, in travel time and amplitude. Ray tracers allow a fast computation of travel times but fail for estimates of signal energy or amplitude as only a thin ray bundle ("banana model" of high frequency approximation) within the wave field is evaluated (Dahlen et al., 2000). The amplitude relations of diffracted waves are highly complex and for refracted waves only solved for special cases (Berry et al., 2013). For heterogeneous (3D) models wavefront solvers are used (Podvin and Lecompte, 1991), which leave hardly restrictions for the velocity field but suffer in precision (and often require soft velocity transitions). In this work, all are utilized, depending on the model requirements. The travel time calculation in parallel and dipping layer models is performed using raytracing algorithms described in detail in Eisermann (2008). Per default, picks are interpreted as first arrivals of a given wave type, having travelled on the fastest path. In heterogeneous models this requires the computation of direct and refracted wave paths. Reflected paths are neglected, as they cannot yield first arrivals, being always slower than either the direct or refracted path. While the refracted path's travel-time can be found analytically for parallel layer models, the travel time pertaining to the direct path between two points is an iterative process (since the ray's starting angle is unknown). Hence, the calculation of the fastest travel time between two points may require many iterations and would create a high computational effort for the calculation at every single grid point. This calculation, however, can be sped up since the continuous travel time curve can be computed very efficiently for parallel (1D) layer models: For the first point corresponding to a vanishing epicentral distance, the ray's starting angle is zero (pointing vertically downwards from the station). The travel time for this point can therefore be obtained in a single iteration. Also, all following travel times for growing epicentral distances can usually be computed in a single iteration: The starting angle for the next ray is hereby slightly increased based on linear extrapolations from travel times of previous epicentral distances. This allows a sampling of the travel time curve with a nearly constant step size in epicentral distance. The computational cost concerning the complete travel time curve until the maximal epicentral distance occurring in the grid, is minimal: It is comparable to the effort of a single travel time calculation between two random points. After the travel time curve for a certain depth is computed, it is interpolated onto the grid points according to the epicentral distance of each grid point to the station. This is done for all depths over the grid. Travel time grids for parallel models are computed within fractions of seconds. For dipping layer models, additional ray leveling algorithms have to be applied, which increase the computational cost (Eisermann, 2008) and the

time needed for travel time caching. For fully heterogeneous models the finite-difference code of Podvin and Lecompte (1991) (also used in NonLinLoc (Lomax et al., 2000)) was ported, which allows an extremely fast travel time computation in high contrast models with sharp first-order velocity contrasts (up to a ratio of 1:10). The scheme yields first arrival's travel times, taking into account all propagation modes as diffracted and refracted (head) waves. The scheme is based on solving the Eikonal equation

$$(\nabla t)^2 = s^2 \tag{II.4.2-1}$$

(with $\boldsymbol{s}(\boldsymbol{x})$ being the slowness and $t(\boldsymbol{x})$ the arrival time at location $\boldsymbol{x}$), by systematically applying Huygen's principle in treating neighboring grid points as secondary sources, approximated by stencils of multiple propagating local plane wavefronts. This scheme provides fast and robust results for high contrast slowness fields, given that the grid resolution is sufficiently high (Nyquist) and the analyzed wavelength sufficiently small (ray approximation). The computation of a travel time grid is computed within seconds. However, to derive a Monte-Carlo estimation of the model dependent travel-time uncertainties, it requires thousands of rendering processes. With every change of the spacial analysis window or of velocity model attributes, cache matrices have to be recomputed. For this reason, dynamically refined grids are problematic in real-time applications, especially when complex models are used. Exempt of the need of re-rendering are attribute changes concerning phases with dependent travel-times, e.g. varying S-velocities when they are derived from the P-velocities over S/P-ratios (if the ratio is globally valid over the whole model). In these cases no recalculation is required as the travel time of one phase is simply a function of the unaltered phase travel time $tt^P_{x,y,z}$

$$tt^S_{x,y,z} = tt^S_{x,y,z}\left(tt^P_{x,y,z}\right); \tag{II.4.2-2}$$

In this case applies

$$tt^S_{x,y,z} = \gamma \cdot tt^P_{x,y,z} \quad \text{with} \quad \gamma = \frac{v_s}{v_p}. \tag{II.4.2-3}$$

In such cases the model modification can be treated within the real time calculation of constraints,

allowing a modification of the phase's velocity model in real time. This is often used to evaluate the correlation between S-P and P-P information when no detailed S-model is given.

As changes in the grid geometry require the re-rendering of all travel-times (multiple times for model uncertainties), a static (i.e. not changing) grid resolution is favorable for real-time schemes, given that it allows to yield complete solutions. The previous chapter showed that internal completeness can be achieved by integrating likelihoods over the grid cell's travel-time-difference interval. Having guaranteed completeness for the lower end of the scale ("overcoming the problem of lower limits for the grid cell size"), it remains to fulfill the demand for complete solutions (i.e. identifying all possible hypocenter locations) also on the upper end of the scale (i.e. overcoming the problem of upper limits for the grid extent). Both increase the total amount of grid cells which is limited by memory and computation power. A common phenomenon that appears, when only a small number of stations is given, is an ambiguous solution (Chapter III.2); One of the solutions is usually found within the network, while the other is often located further outside. For this reason, automatic schemes often misinterpret teleseismic events as local events (or vice versa), as arrival time differences and amplitude ratios are comparable for both. This aspect is of special importance for fully-automated Earthquake Early Warning systems, which base their location estimates on few stations only. Disregarding either of the ambiguous solutions would be problematic: An incorrectly chosen near-by solution underestimates magnitudes and may lead to missed alarms, while an incorrectly chosen distant solution overestimates magnitudes and may in turn cause a false alarm, beside the fact that predicted arrival-times for strong shaking are misestimated for both. For automated schemes it is therefore important to have grids covering extensive areas. As the spatial constraints grow with distance ("in width"), the grid resolution may be reduced with distance without causing the precision to be significantly decreased, allowing to expand the grid's extent without additional computational cost. As the grid is often analyzed on orthogonal cross-sections (e.g. solutions may be searched in certain bands of hypocentral depth), the grid cell position (and consequently size) is not set in relation to the cartesian- but a component wise distance from the network center ("distance per axis") (fig. II.4.2:1).

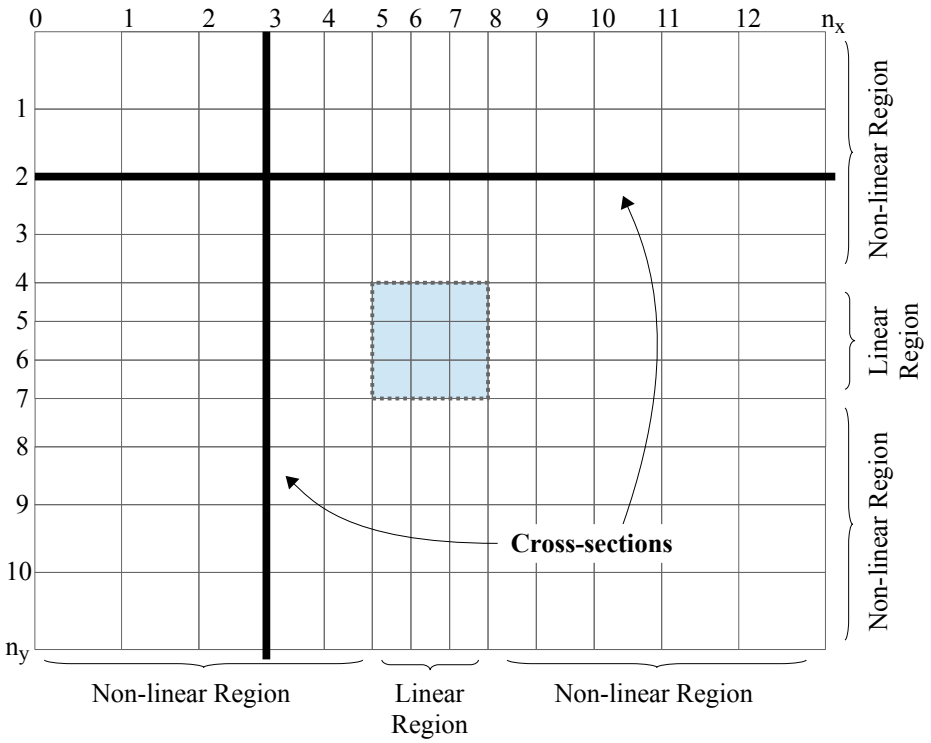


Fig. II.4.2:1: Grid of variable resolution. Constant resolution in vicinity of the network (blue cells). Reduced resolution outside with varying grid cell size. Vertical cross sections (bold black lines) hold the same third coordinate like regular grids.

A general form for the mapping of coordinates $\boldsymbol{P}$ to grid points with index $\boldsymbol{I}$ can be written as

$$\boldsymbol{P}_i(\boldsymbol{I}_i) = \boldsymbol{C}_i + F\left(2\cdot\boldsymbol{I}_i/\boldsymbol{n}_i - 1\right)\cdot\Delta_i\,, \tag{II.4.2-4}$$

with $\boldsymbol{C}$ being the coordinate at grid center, $\boldsymbol{n}_i$ being the number of grid cells per coordinate axis i and Δ_i the quarter grid size length on axis i. The mapping function F determines the spacing between grid points. The scaling, individual for each coordinate axis, assures planar cross sections corresponding to the analyzed axes. Regular Grids are described by F being the identity function. Here, F will be formulated to yield a linear behavior close to the interval center and to follow a power function towards the boundaries with

$$F(\gamma) = \begin{cases} 2\gamma\,, & \gamma < 0.5 \\ 2\gamma + [2\cdot(2\gamma-1)]^3\,, & else \end{cases}. \tag{II.4.2-5}$$

This choice of F provides an inner window corresponding to a regular grid, but yields a five times

larger grid extent than regular grids through the peripheral interval, to allow the identification of distant secondary solutions (fig.II.4.2:2).

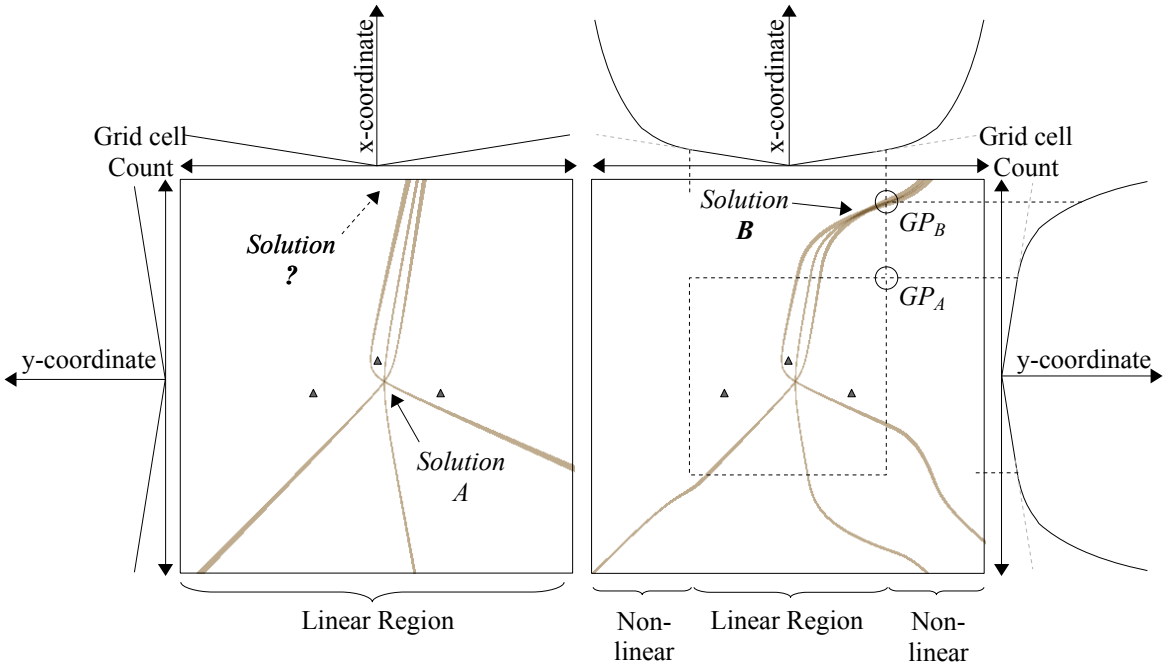


Fig. II.4.2:2: Stretched Grid. Left) Regular grid of constant resolution. Secondary solutions lie often outside of the analysis window. Right) Central region of constant resolution, outer region with grid point coordinates being assigned by a power function. Grid point A (GP_A) still lies in the linear regime, while GP_B lies in the non-linear extension zone. This functions as a wide angle mirror increasing the field of view: Secondary solutions are identified.

Fig. II.4.2:3: De-skewed visualization of fig. II.4.2:2 showing the increase in grid cell size with distance from the detecting stations.

The different modes for central and peripheral intervals cause e.g the first 15km of depth to be analyzed in high resolution, decreasing from there until e.g. 100km depth at the grid's boundary. In this manner, computational resources are assigned to where they are needed: Local solutions are precisely analyzed using a high resolution, while distant- or solutions in depth that are less compact are evaluated less detailed, keeping the relative error of estimates comparable over the whole grid extent. The mapping of coordinates to grid points occurs at the time of travel time computation during the grid initialization and does therefore not affect the real time analysis through additional computational cost. The use of a variable grid spacing allows to cover a larger region with comparable computational effort and supports the aspect of solution completeness on the large scale. After having derived a direct search scheme that allows to identify all solutions of interest, the following chapter will lay out the formal description of all constraint classes in this variable sized grid-*cell* based direct search evaluation.

II.4.3 **Fundamental Constraints in a Grid Formulation**

Previous chapters served do derive the general formulation of the different constraints and their uncertainties, both in picking and in model. This chapter serves as library, listing the detailed equations for the evaluation of constraints, adapted to the evaluation on grids in a direct search scheme. For the interactive analysis, constraint types are composited and plotted in certain color channels, which allows to comprehend the individual impact of the different constraint types e.g P-P, S-P, P-P-P separately on the global solution. Each color intensifies near the solution, showing the internal agreement of each constraint type. Where all constraint groups are in agreement, the basic colors add up to white, marking the solution by a bright shine. v_{xyz} is the value for the constraint channel at grid cell (x, y, z). cc stands for the total number of constraints per channel while $nc^{(A-B)}$ identifies the total number of constraints of type (A-B). $\underline{\Theta}_{(x,y,z)}$ and $\overline{\Theta}_{(x,y,z)}$ are lower and upper bound of the modeled constraint parameter (i.e. travel-time difference) of all grid points of grid cell (x, y, z), including the effects of model uncertainty. The modeled constraint parameter at grid point (x, y, z) is given by

$$\Theta_{xyz} = tt^{B}_{xyz} - tt^{A}_{xyz} \qquad \text{(II.4.3-1)}$$

while the observed constraint parameter based on onset times is defined as

$$T^{(B-A)} = t^B - t^A \,. \tag{II.4.3-2}$$

For S-P constraints, $T^{(B-A)}$ becomes $T^{(S,i-P,i)}$, with i being the station index. The number of constraints $nc^{(S-P)}$ equals the number of stations with detected P- and S phase. The same pattern applies for all evaluated phase differences at a single station, e.g. the difference between direct phase and refracted phase $T^{(P,i-P_{ref},i)}$. For hyperbolic constraints, $T^{(B-A)}$ becomes $T^{(P,i-P,j)}$ for P-phase based - or $T^{(S,i-S,j)}$ for S-phase based hyperbolic constraints, respectively, with $i \neq j$ being obtained by permutation of stations into groups of two. The total number of hyperbolic constraints is given by

$$nc^{(P,1-P,2)} = \frac{1}{2}\left(n^2 - n\right) \tag{II.4.3-3}$$

(accordingly for $nc^{(S,1-S,2)}$) with n being the number of stations that detected the corresponding phase. Constraints may be analyzed in either a probabilistic (p) or non-probabilistic (n) manner. In both cases they might be evaluated either solely using picking uncertainty (P) or additionally model uncertainty (PM). The general formulation for constraints based on the time difference between different phases/stations is listed in the following.

Constraints using picking uncertainty (P)

-Non-probabilistic-

The standard comparative form which is valid only for infinitesimal small grid cells

(P-n) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} \mathrm{I}\left(\left[tt_{xyz}^{B} - tt_{xyz}^{A} \right] - \left[t^{B} - t^{A} \right],\ \delta t^{A} + \delta t^{B} \right) \qquad \text{(II.4.3-4)}$$

with

(P-n) $$\mathrm{I}(v, \delta v) = \begin{cases} 1, & v < |\delta v| \\ 0, & else \end{cases} \qquad \text{(II.4.3-5)}$$

becomes

(P-n) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} \mathrm{I}\left(\underline{\Theta}_{xyz},\ \overline{\Theta}_{xyz},\ T^{(B-A)},\ \delta t^{A} + \delta t^{B} \right) \qquad \text{(II.4.3-6)}$$

with

(P-n) $$\mathrm{I}(\underline{m}, \overline{m}, v, \delta v) = \begin{cases} 1, & [\underline{m}, \overline{m}] \cap [v - \delta v, v + \delta v] \neq \{\} \\ 0, & else \end{cases} \qquad \text{(II.4.3-7)}$$

using a grid cell size independent interval formulation.

-Probabilistic-

Discrete grid-cell size dependent form:

(P-p) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} P^{O(A-B)}(\underline{\Theta}_{xyz}) - P^{O(A-B)}(\overline{\Theta}_{xyz}), \qquad \text{(II.4.3-8)}$$

where $P^{O(A-B)}$ is the cumulative (eq. II.4.1-11) of the constraint likelihood distribution p obtained from the cross-correlation of both phases' LOC functions.

Grid-cell size corrected form:

(P-p) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} p\left(\underset{\underline{\Theta}_{xyz} \le \Theta_{xyz} \le \overline{\Theta}_{xyz}}{arg\,max}\ p(\Theta_{xyz}) \right) \qquad \text{(II.4.3-9)}$$

Constraints using picking and model uncertainty (PM)

-Non-probabilistic-

(PM-n) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} \mathrm{I}\left(\underline{\Theta}_{xyz}, \overline{\Theta}_{xyz}, \underline{T}^{(B-A)}, \overline{T}^{(B-A)} \right) \quad \text{(II.4.3-10)}$$

with

(PM-n) $$\mathrm{I}(\underline{m}, \overline{m}, \underline{v}, \overline{v}) = \begin{cases} 1, & [\underline{m}, \overline{m}] \cap [\underline{v}, \overline{v}] \neq \{\} \\ 0, & else \end{cases} \quad \text{(II.4.3-11)}$$

The congruent region obtained from truncated residua (according to eq. II.3.2-5) becomes

(PM-n) $$\overline{R}_i^P = \mathbf{\Lambda}\left(\left| t_i^P - t_{i,xyz}^{P,M} \right| - \delta t^P - max\left(tt_{xyz}^P - \underline{tt}_{xyz}^P, \overline{tt}_{xyz}^P - tt_{xyz}^P \right) \right) \quad , \quad \text{(II.4.3-12)}$$

with Λ being the ramp function (eq. II.3.2-6).

-Probabilistic- (PM-p)

(PM-p) $$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} \bar{\boldsymbol{P}}^{M(A-B)} \cdot \bar{\boldsymbol{P}}^{O(A-B)} \quad \text{(II.4.3-13)}$$

with

(PM-p) $$\bar{\boldsymbol{P}}^{O(A-B)} = p^A * p^B \quad \text{(II.4.3-14)}$$

and $\bar{\boldsymbol{P}}^{M(A-B)}$ being calculated according to eq. (II.3.5.1-25)

(PM-p) $$\bar{\boldsymbol{P}}^{M(A-B)} \text{-|} \quad p^M\left(T, tt_{min}^A, tt_{max}^A, tt_{min}^B, tt_{max}^B \right) \quad \text{(II.4.3-15)}$$

$\bar{\boldsymbol{P}}^{M(A-B)}$ and $\bar{\boldsymbol{P}}^{O(A-B)}$ are vectors containing the values derived from the discrete likelihood distributions p^A, p^B and p^M (fig. II.3.5.1:4). Both vectors are defined over the parameter interval $[T_{min}, T_{max}]$ according to eq. (II.3.5.1-17) and (II.3.5.1-18). Using the same resolution, they consequently contain the same number of elements n_e. The grid cell volume corrected form uses

(PM-p)
$$v_{xyz}^{(A-B)} = \frac{1}{cc^{(A-B)}} \sum_{i=0}^{nc^{(A-B)}} \bar{\boldsymbol{P}}^{M(A-B),C} \cdot \bar{\boldsymbol{P}}^{O(A-B),C} \tag{II.4.3-16}$$

with

$$C = \underset{c \in [1, n_e]}{arg\,max}\, \bar{\boldsymbol{P}}^{M(A-B),c} \cdot \bar{\boldsymbol{P}}^{O(A-B),c} \tag{II.4.3-17}$$

Instead. Here, C and c denote column number C resp. c of the probability vectors (i.e. the vector's c[th] component).

Back azimuth Beam

The back azimuth obtained by small aperture arrays of particle motion is velocity model independent. Contrary to other constraints, in it's basic form it therefore does not rely on computed travel times but is determined purely geometrically

-Non-probabilistic- (n)

(n)
$$v_{xyz}^{Beam} = \frac{1}{cc} \sum_{i=0}^{n(s^P)} \begin{cases} 1, & B \le 1 \\ 0, & else \end{cases} \tag{II.4.3-18}$$

with

(n)
$$B = \frac{1}{\Delta\Phi} \cos^{-1}\left(\begin{matrix} x \\ y \end{matrix}\right) \cdot \hat{n}^B \tag{II.4.3-19}$$

Here, $\hat{n}^B$ describes the back azimuth unit vector and $\Delta\Phi$ the angular uncertainty or maximum expected discrepancy being half the beam opening angle. Further, $\Delta\Phi$ may be a function of epicentral distance and array geometry as described in chapter III.1.1.

-Probabilistic- (p)

(p)
$$v_{xyz}^{Beam} = \frac{1}{cc} \sum_{i=0}^{n(s^P)} \begin{cases} 1-B, & B \le 1 \\ 0, & else \end{cases} \tag{II.4.3-20}$$

This chapter concludes the formal derivation and description of constraints. Over the last chapters we have mapped picking and model uncertainty intervals to spatial constraints, advancing the accuracy of hypocenter estimates in uncertain environments: Congruent regions provide maximum bounds for the hypocenter region, and a probabilistic analysis provides the corresponding most likely location. A multi-pick analysis was introduced to provide robust estimates when ambiguous phase onsets exist, to increase accuracy where mis-picks are expected. Evaluating constraints over grid-*cells* of varying size allows to identify sharpest locations independent of the grid resolution, near by and in far distance. In the following chapter we will build on these qualities to understand the constraint characteristics, identifying their advantages and disadvantages, and how they can be combined to obtain optimal hypocenter estimates. This will lead us to the important topic of ambiguous locations, which we may face if we attempt to locate with few stations based on P-phase arrivals. We will tackle the question, why they happen and if we can predict where the second solution may be found.

III Constraint Precision and Location Ambiguity

Given the different constraint types and a way to map their uncertainties, one can analyze their ability to constrain hypocenter regions. Mapping the precision is one way to visualize the advantages and disadvantages of each constraint type or their combinations. It further serves as means to design networks yielding precise locations in the area of focus, respecting the type of analyzed data expected to become available. The precision depends strongly on the network geometry and the amount of given stations. When the number of stations decreases to a critical limit, solution ambiguities arise. It is of fundamental importance to understand why, when and where they arise to yield a correct location and event interpretation. In the following two chapters these points will be outlined.

III.1 Constraint Precision

“Accuracy” and “precision” are closely linked terms that describe the quality of our results. While “accuracy” describes the degree of the result's truthfulness (“the true value is contained in the estimated interval”), “precision” pertains to the related uncertainties, the result's vagueness (“the size of the interval”). An accurate result with low precision may be meaningless; An inaccurate result of highest precision renders the precision spurious and the result misleading. This will generally be the case if we ascribe a higher precision to the input parameters (e.g. picking and model) than they truly hold, yielding inaccurate results. Just as accurate results (e.g. locations) depend on accurate assumptions (e.g. picks, models), so does the result's precision depend on the precision of the input parameters - but further on the methodology's ability to constrain a given location. As different constraint classes usually yield high precisions in different regimes, their combination is useful to stabilize the location quality over extended regions. The precision mapping hence also provides us with more than just an estimate of the preciseness of location that may be expected in certain regions and depths (making it an useful tool for network layout planning) but also allows us to visualize and study the advantages and disadvantages of each constraint type (or combinations of several types) concerning their ability to locate. The different constraint types vary strongly in character: Class I constraints, like S-P (or P-P_{ref}) constraints, hold the advantage that their spatial resemblance is finite, limiting the possible hypocenter region. They vary strongly radially but little in distance. Class II constraints (as well as back-azimuth constraints) on the other

hand resemble infinite bodies, wherefore it (usually) requires at least two constraints to obtain a limited region. General exceptions are class II constraints based on parameter *ratios* between two stations (e.g. amplitude ratios), which yield non-concentric spheres around those stations. Back-azimuth constraints are additionally fully unconstrained in focal depth, as the information of incidence angle is often neglected. To understand how precise the various constraints actually constrain the location inside or outside the network or in depth, we will map their precision for a simple network geometry.

This is achieved by quantifying the feasible region (which describes the set of possible locations, chapter II.3.2) given by the intersection of all used constraints, for all possible hypocenter locations of interest. Since the input parameters are treated to behave ideally (i.e. we do not need to deal with mis-picks), this region will be defined by a congruency of one ($c=1$). To evaluate the precision at any location, we will use the travel-times to all stations to generate a set of (ideal, non-perturbed) synthetic arrival times, based on which we will construct the constraints. The estimate is therefore void of errors due to inaccurate picking or models. Using expected values for the picking uncertainty, the location uncertainty can be assessed by quantifying the feasible region. The mapping over the whole grid is performed by repeating this scheme at many grid points. Using a spatial interpolation, the final map is generated for several depths, stretching over the region of interest.

The feasible region could be described by it's volume, but it is more meaningful to quantify it separately, for epicentral spread and vertical extent. Straightforward measures to describe the epicentral spread would be given in its area (a) or the distance between the furthest feasible location and the evaluated epicenter location (b). However, the area (a) does not give information about the shape (i.e. the distribution of solutions around the true location) and the maximum distance (b) gives no information about the surface. In the following, we will therefore use the moment of area (m_A) with reference point at the evaluated location, as it quantifies the distribution of the epicentral surface. The moment m_A weighs each surface element by its distance to a given axis. In this case the axis will be chosen to lie vertically, perpendicular to the plane of analysis, located at the point of the epicenter. It will therefore weigh each surface element dA according to it's distance d to the true epicenter

$$m_A = \int_A d \cdot dA \qquad \text{(III.1-1)}$$

and describe the distribution of the epicentral surface. Epicentral surfaces with large maximum distances but small area (e.g. elongated ellipses) are weighed similar to surfaces of smaller maximum distance but of larger area (e.g. circle). For comparability the correlation between m_A, radius and area of a circular surface is given in the following (table III.1.1, fig. III.1:1).

Radius (km)	Area (km^2)	MoA (km^3)
0.001	$\pi \cdot 10^{-6}$	$2.094 \cdot 10^{-9}$
0.01	$\pi \cdot 10^{-4}$	$2.094 \cdot 10^{-6}$
0.1	$\pi \cdot 10^{-2}$	$2.094 \cdot 10^{-3}$
1	$\pi \cdot 10^{0}$	$2.094 \cdot 10^{0}$
10	$\pi \cdot 10^{2}$	$2.094 \cdot 10^{3}$
100	$\pi \cdot 10^{4}$	$2.094 \cdot 10^{6}$

Table III.1.1: Correlation between m_A and the surface of a circular surface of a given radius.

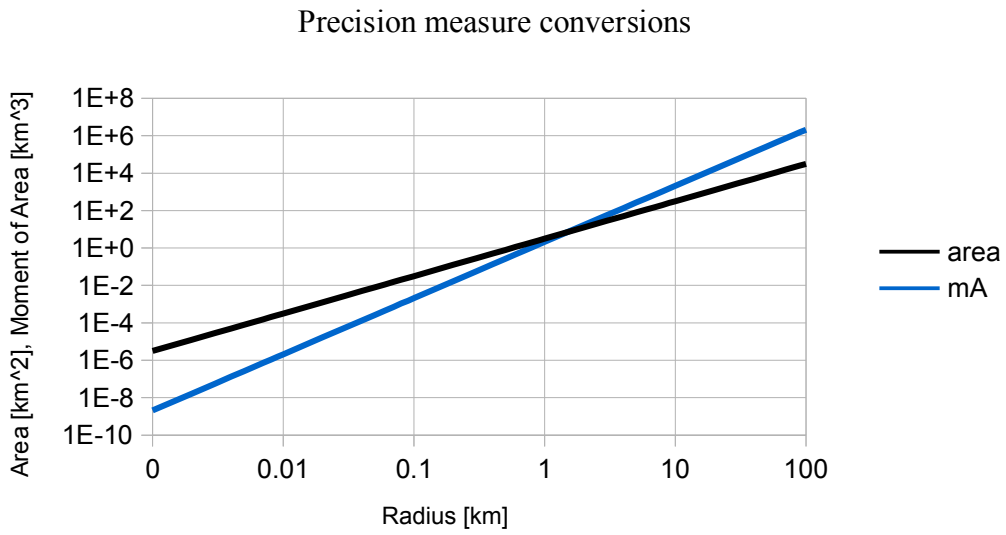


Fig. III.1:1: Correlation between m_A and the surface of a circular surface of a given radius.

To evaluate a network's precision in locating events, several combinations of individual constraint types have to be considered, as different phase types and information might become available during the analysis. In single station networks, classic multilateration (S^A-P^A) is applied as well as hyperbolic location (P^B-P^A), as well as both in combination. At a small aperture array, (S^A-P^A) distance information is coupled with the array back azimuth beam estimate. In small aperture array networks also the beam individually finds application when the S-phase is not identifiable or does not exist. When small-aperture arrays are not available, at times the less precise three-component

single-station back-azimuth estimate is used (chapter IV.1). This one shows the same characteristics as P-P-P estimates, however, generally less precise (holding a typical beam-opening angle of 20°-30°). In the following, we will look at precision maps for the following six constraint groups (fig. III.1:2):

A) P-P-P beams,

B) S-P,

C) P-P,

D) S-P and P-P combined,

E) S-P and P-P-P beams combined,

and all constraints combined:

F) S-P, P-P and P-P-P beams.

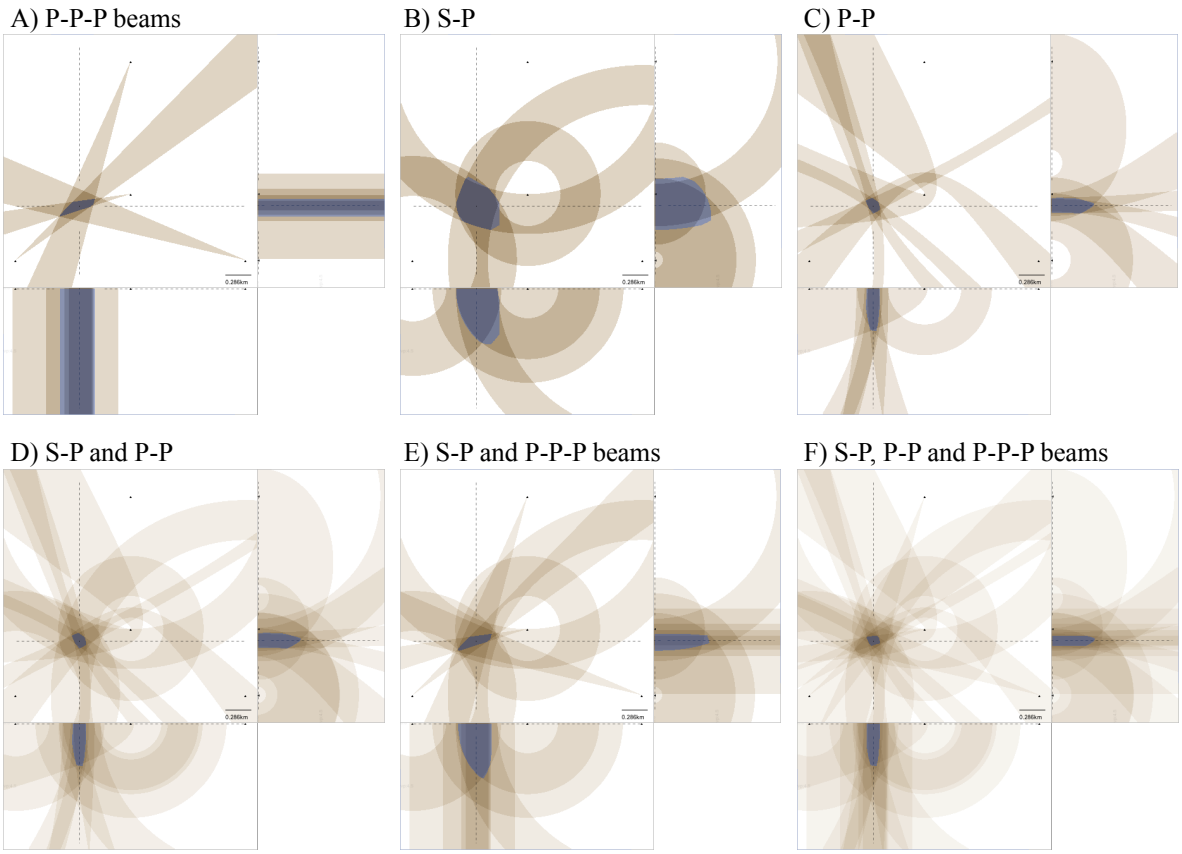


Fig. III.1:2: Cross section display of the constraints of groups A-F. The surface projection of the three-dimensional feasible region, which is used to quantify the precision, is superposed in blue. The feasible regions appear larger than the constraint intersection in the cross sections, as the intersection extends over a range of depths.

The precision map is generated by simulating epicenters at every location of the map. For each epicenter the phase arrival times are computed for all stations as input data for the subsequent computation of the feasible region as constrained by each group of fig. III.1:2. The region's moment of area is used to quantify the precision in the map. The color scale used in fig. III.1:3 - III.1:8 spans over the full precision range observed among the constraint groups for the given network geometry. The individual group's range is indicated by white triangles in each figure. The used network example (see fig. III.1:3) is taken from a Nanoseismic Monitoring campaign at Mont Terri, Switzerland, using a small network consisting of four SNS small aperture arrays. The network extends over an area of roughly 6km times 1.5km, being strongly elongated. The region of surveillance reaches down to a depth of -4km. Precision maps are generated for sources in depths of 0km, -1km, -2km, -3km and -4km. The local velocity in the top four kilometers averages around 4.5km/s. The onset detection uncertainty is chosen as 0.0125s for P onsets and 0.025s for S onsets. The time differences in P-P-P beams are obtained using cross-correlation of the waveforms at the three stations. The beam uncertainty (opening angle) is dependent on the array geometry and strongly affected by the plain waveform assumption in the vicinity of the array. For simplicity, here the opening angle is set to 10° which is valid for hypocentral distances larger then three aperture lengths. The detailed analysis of array beam precision follows in chapter III.1.1. For the simulated event location it is assumed that all stations detect the event. Although weak signals may in practice not be detected at distant stations, the result should be comparable nonetheless since the constraints' “width” grows with distance: Given a certain constraint type, more distant stations may stabilize the location but will hardly increase the precision.

The constraint groups' precision is displayed in the following six figures.

(A) Precision of P-P-P beams

Fig. III.1:3 shows, that back-azimuth beams yield a high precision inside the network. Outside, especially along the major axis of the network, the epicentral area grows rapidly, leaving the region quickly unconstrained.

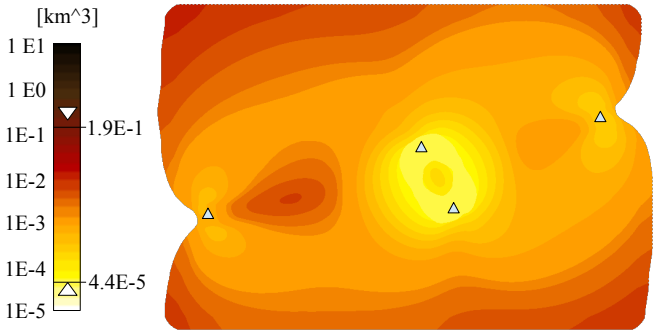


Fig. III.1:3: A) ***P-P-P Beam precision****. Triangles mark the four SNS locations, each consisting of central station and a surrounding small aperture tripartite array. The closer the epicenter location to a pair of arrays, the higher the precision. The color scale holds the full precision range observed in fig. III.1:3 - III.1:8, the white triangles mark the range observed in the individual figure.*

We find that the closer two stations are, the higher is the location precision in the area between them. The precision pattern depends on the network geometry and on the goodness of the back-azimuth estimates ("beam opening angle"). Contrary to array estimates, particle-motion estimates do not rely on a plain-wave assumption and are therefore applicable even in the immediate near-field. Particle-motion estimates are, however, of less precise. Back-azimuth information is stable against variations in the velocity model, given it is laterally homogeneous. If the velocity structure is known, the incident angle may be used to constrain the depth. Otherwise the depth remains unconstrained.

(B-D) Precision of S-P and P-P constraints

Location based on S-P spheres yields the highest precision close to the stations for near surface sources (fig. III.1:4). In that case the sources are approximately in one plane with the stations, and the sphere of the closest station is compact, yielding a small epicentral surface around that station. This local peak in precision, however, appears only in large sparse networks or regions of low seismic velocity. This is due to the fact that the S-phase can only be distinguished from the P-coda and be picked after a certain travel time.

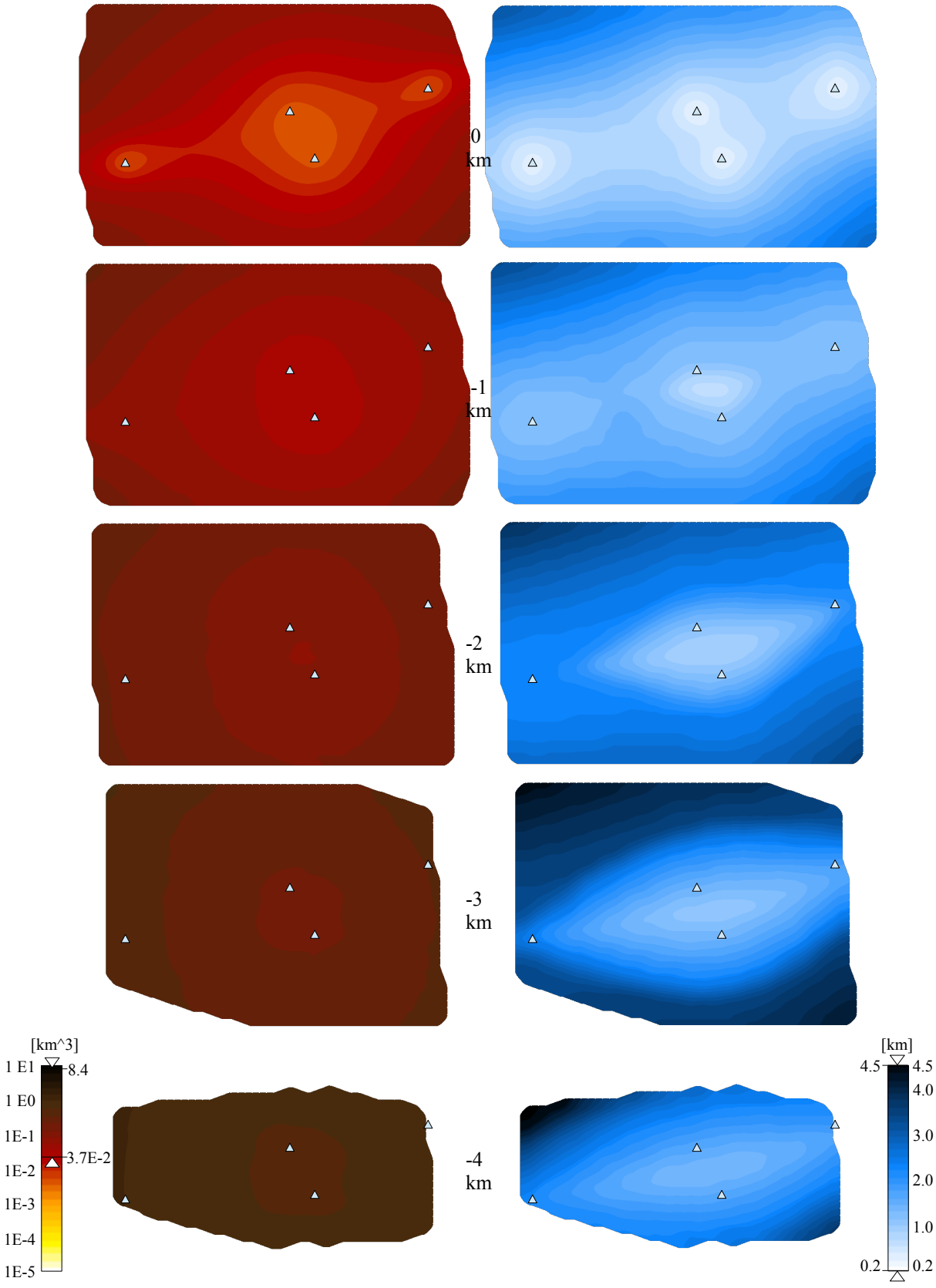


Fig. III.1:4: B) ***S-P precision.*** *Left – horizontal precision, Right: vertical precision Simulation depth ranging from first to last row between 0km and 4km.*

The horizontal precision decreases quickly with depth, showing the highest precision towards the network center: The deeper the source, the more it is located at the bottom of the constraining spheres where their horizontal overlap is the largest - but their vertical extent the smallest possible. Contrary to the behavior of other constraints, this causes the vertical precision to stabilize with depth, at times even to increase (see fig III.1:6). This counter intuitive behavior is due to the fact that contrary to other constraints, which grow wider with distance, the sphere shell's thickness is independent of the sphere radius (when disregarding velocity model uncertainties). The highest precision is again found at the network center. While beam and hyperbolic constraints show an extreme decrease in horizontal precision close to the stations leaving the network (in elongated networks at the stations lying on the major axis), the precision of S-P constraints only decreases slowly. Yet as the picking uncertainty for S-onsets is higher than for P-onsets, hyperbolic location generally yields a higher horizontal precision inside the network (fig. III.1:5).

Hyperbolic location shows the highest horizontal precision at the network center. The vertical precision quickly decreases with depth while the horizontal precision does so only slowly, contrary to the behavior of S-P constraints. Combination of S-P and P-P constraints combines the strengths of both methods and increases vertical and horizontal precision and stabilizes the location at the network edges (fig III.1:6). Outside the network the congruent zones of S-P spheres form a sliver, well constraining the distance, yet only poorly the azimuth. The P-P hyperboloids on the contrary form a beam-like congruent zone that constrains the azimuth but not the distance. Combination of both constraints merges complementary information and is therefore of special importance for locating sources outside the network, leading over to the field of seismic arrays.

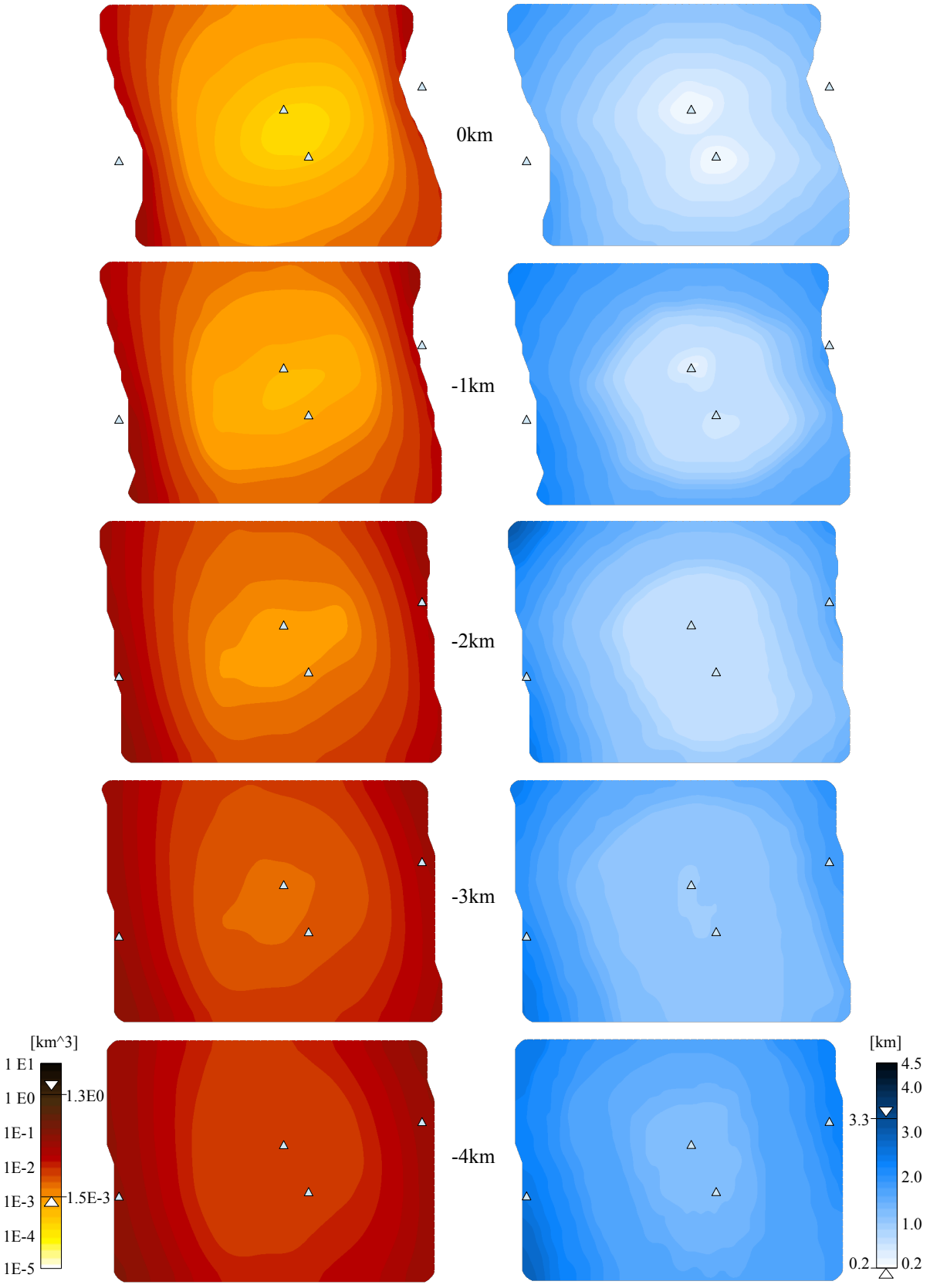


Fig. III.1:5: C) ***P-P precision****. Left – horizontal precision, Right: vertical precision Simulation depth ranging from first to last row between 0km and 4km.*

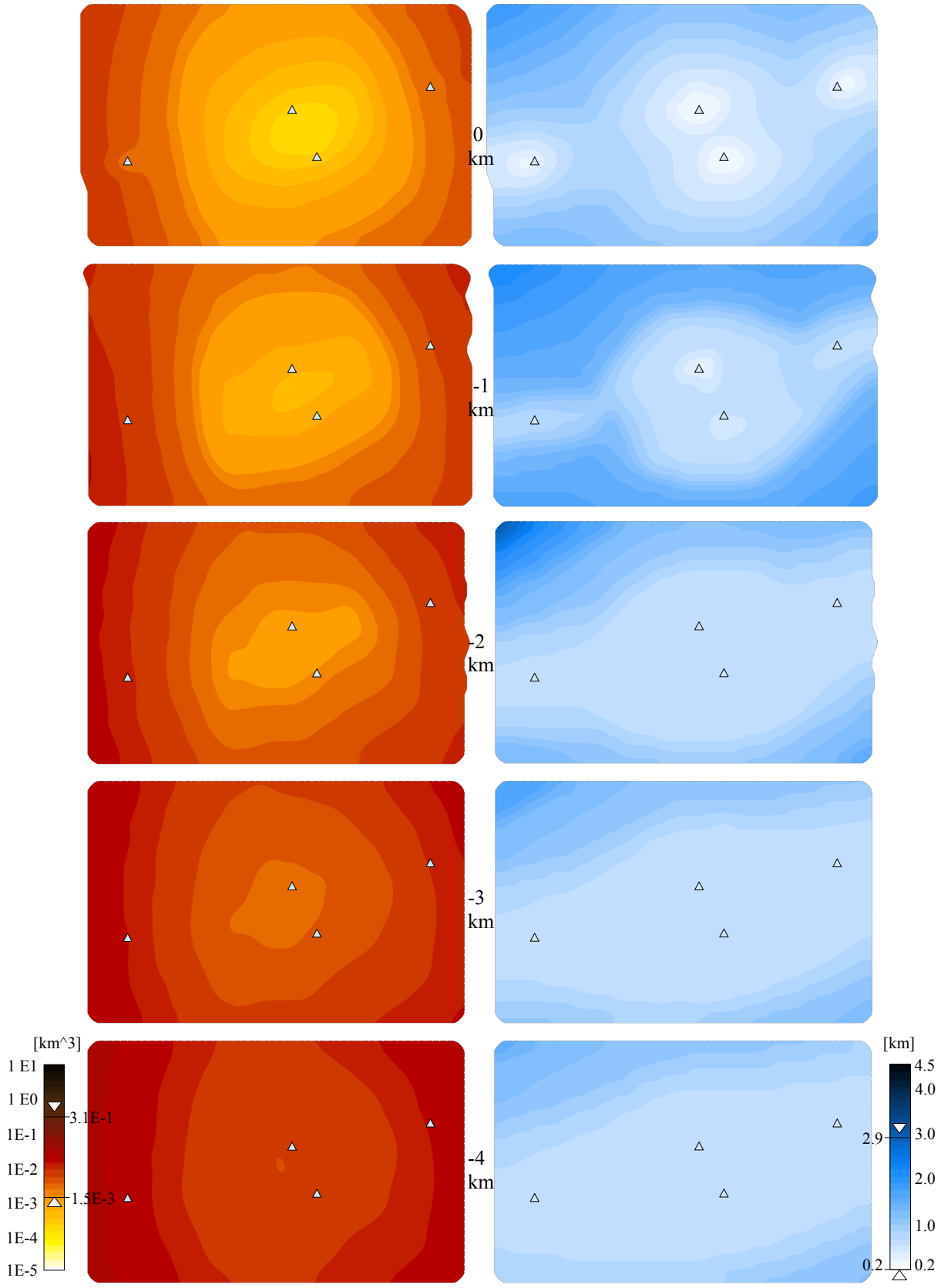


Fig. III.1:6: D) ***S-P and P-P precision****. Left – horizontal precision, Right: vertical precision Simulation depth ranging from first to last row between 0km and 4km.*

E) Precision of S-P and P-P-P beam constraints, combined

Similarly advantageous to combining S-P and P-P constraints outside the network, is the combination of S-P and back-azimuth constraints, as obtained from seismic arrays (fig. III.1:7). Around the network-edge, P-P constraints may offer a slightly higher precision than back-azimuth beams, but the latter is less affected by velocity model errors, which adds robustness to the estimate. Additionally, this combination benefits of the high horizontal precision of the back-azimuth constraints and reduces the vertical uncertainty of the pure S-P constraint by 50%, despite the fact that P-P-P constraints by themselves not even constrain the depth.

In larger distance, P-P constraints again gain importance, since errors in the assumption of lateral homogeneity introduce azimuthal errors, which the travel-time dependent P-P constraint are less prone to.

F) Precision of S-P, P-P and P-P-P beam constraints, combined

The highest precision is obviously obtained when all constraints (P-P, S-P and back-azimuth) are available and the congruent region can be based on all constraint types combined (fig. III.1:8). In this case, the horizontal precision (m_A) ranges between $2.8 \cdot 10^{-5}$ km^3 and $1.3 \cdot 10^{-1}$ km^3. This can be compared to (circular) epicentral surfaces with areas ranging between $1.7 \cdot 10^{-3}$ km^2 and 0.5 km^2 which corresponds to maximal epicentral uncertainties between $2.3 \cdot 10^{-2}$ km and 0.4 km, i.e. roughly 20 to 400 meters. The vertical extent of the hypocentral volume ranges up to 2.86 km. The minimum value falls below the grid-spacing and can therefore not properly be identified. The applied grid resolution with a grid cell height of 200 m technically limits the possible resolution for the lower estimate to 200m. All quoted precision values pertain to the situation of a perfectly known velocity structure.

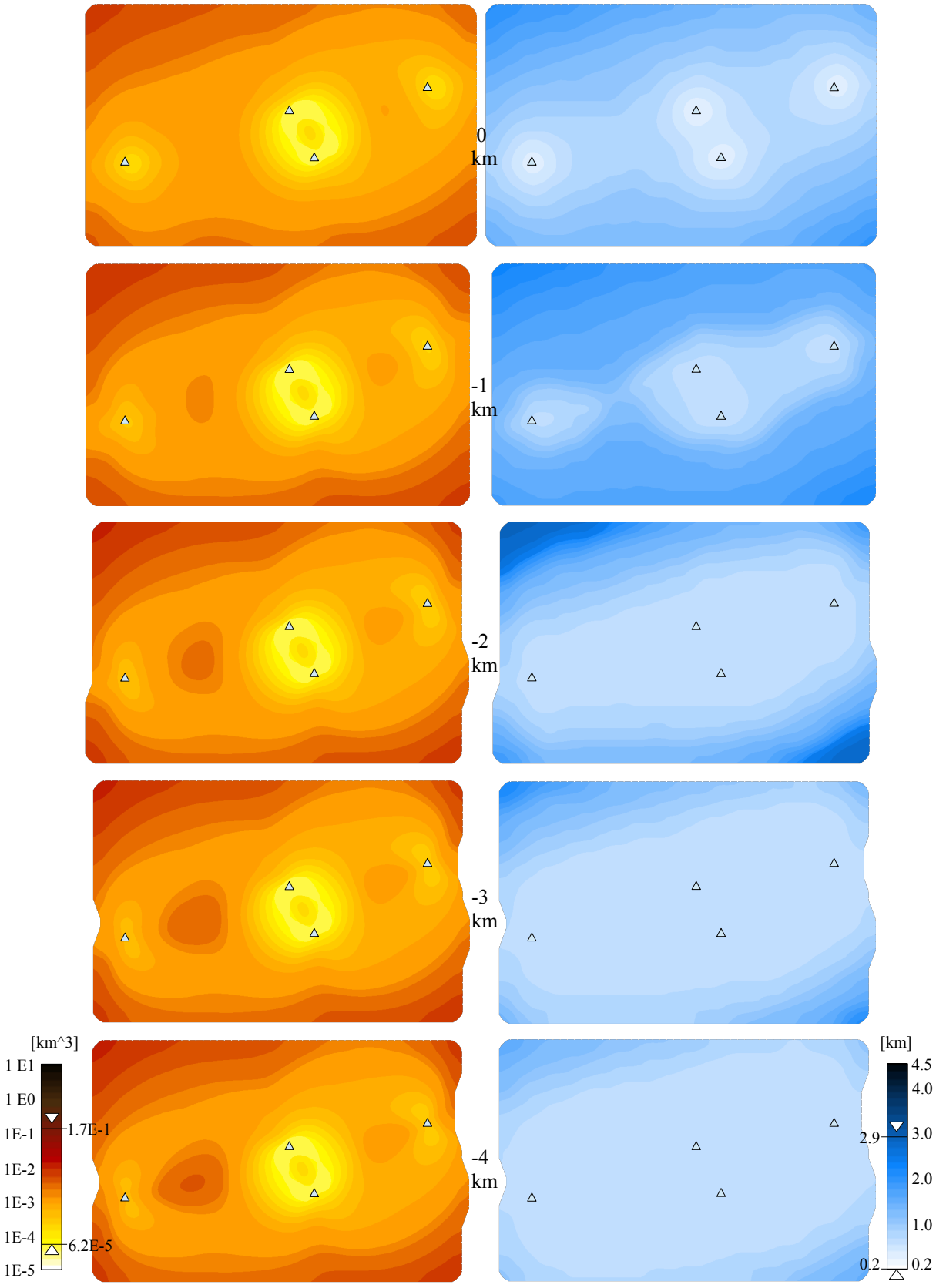


Fig. III.1:7: E) ***S-P and P-P-P Beam precision****. Left – horizontal precision, Right: vertical precision Simulation depth ranging from first to last row between 0km and 4km.*

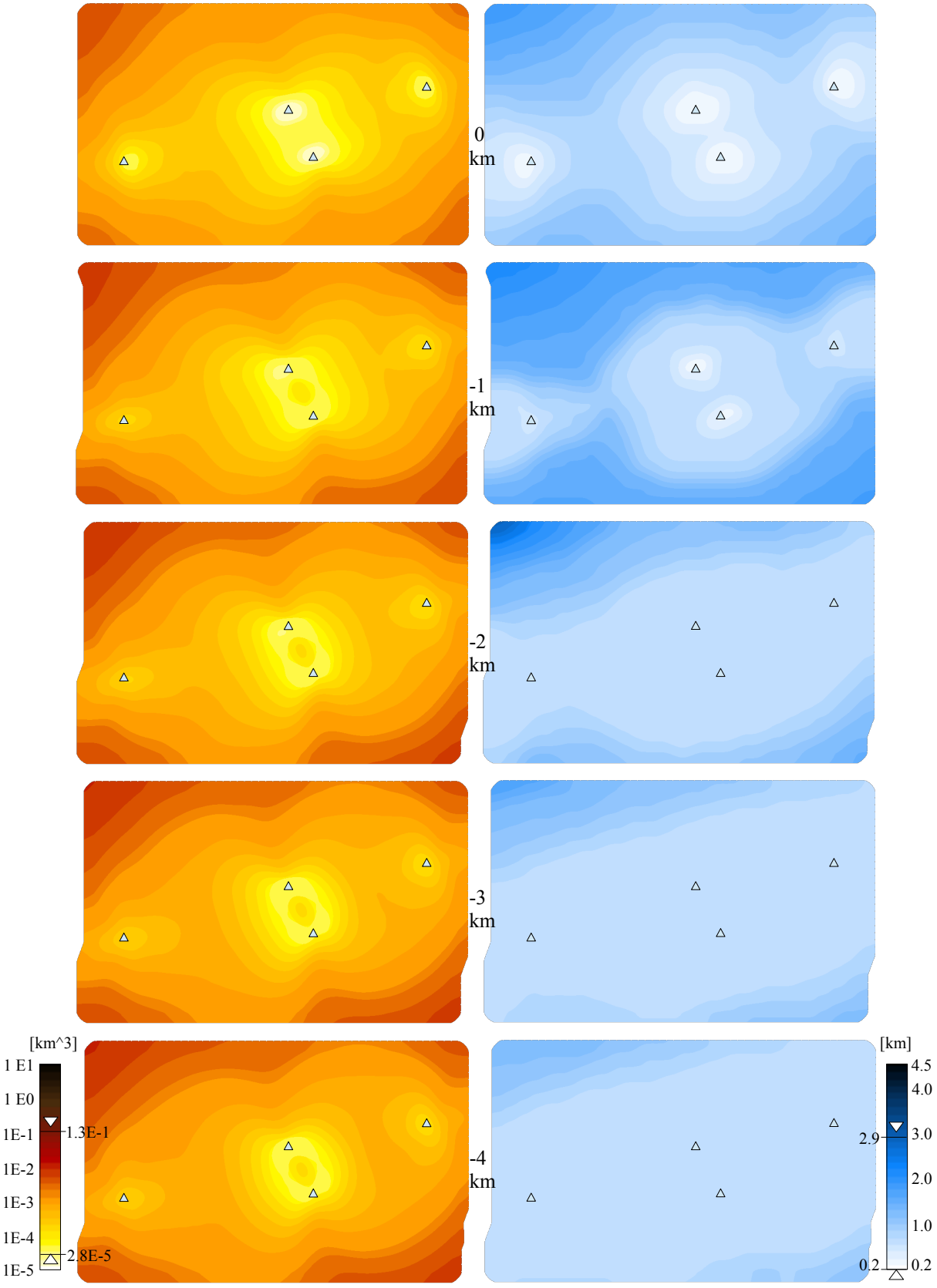


Fig. III.1:8: F) ***S-P and P-P and P-P-P Beam precision****. Left – horizontal precision, Right: vertical precision. Simulation depth ranging from first to last row between 0km and 4km.*

These precision maps were composed to display the methodological limitation to the precision that can be achieved by application of the various constraints. To obtain the pure methodological limitation, the velocity structure was treated as perfectly known and model uncertainties were switched off. To obtain realistic precision estimates, model uncertainties need to be considered. As this network example is compact, such uncertainties would have predominantly affected the estimates close to the network edge and outside the network. The back-azimuth precision estimate was technically treated as obtained from single three-component-stations, since the back-azimuth source was linked to one single coordinate. The precision map of small-aperture array networks will, however, show a similar character (yet more precise), despite two major differences in the approach: Small-aperture arrays consist of distributed stations analyzing plane wave fronts. This means on the one hand, that back-azimuth estimates will suffer at close distances where the plain wave assumption is violated. For sources in it's vicinity, on the other hand, the array takes over the character of a network, being able to be analyzed using S-P and P-P constraints. The distinction between array and network is solely given by the correlation radius of the signals (Bormann, 2012). If a group of sensors lies within the signal's correlation radius it is called 'array', otherwise it resembles a 'network'. Now, for intermediate source distances, the array may either be still used as network or already as array providing back-azimuth information which is less dependent on the velocity model assumptions than the P-P and S-P constraints used in a network analysis. For these intermediate cases the precision and errors of back-azimuth estimates become of interest and shall be analyzed in the following chapter.

III.1.1 **Precision of Tripartite Array Back-Azimuth Beams**

Seismic arrays deliver important information as slowness and coherency measures and provide source-azimuth information. The back azimuth beam constrains the location robustly as it is derived independent of the velocity model for laterally homogeneous media, as the local velocity cancels out of the equations. The equation for the back azimuth beam, which computes the direction of the event's incoming signal, is derived assuming incoming plane wave fronts.

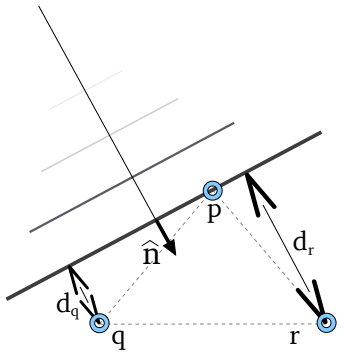


Fig. III.1.1:1: Tripartite array and incoming wave front. Plane wave assumption.

This assumption is correct for distant sources (fig. III.1.1:1), but is violated for shorter epicentral distances. The error due to the violation is reported to be less than 7.1 degrees for equilateral tripartite arrays and epicentral distances larger than the arrays inter station distance (aperture) (Bregman et al., NDC). This determines a practical back azimuth beam opening angle to roughly 15 degrees, given the absence of lateral inhomogeneities. For an equilateral triangle configuration the estimate accuracy further increases with source depth. This is, however, not the case for all array configurations given in the field as e.g. in the Jackknifing analysis of 4-station small-aperture-arrays used in Nanoseismic Monitoring. These consist of a central station, surrounded by three (equally spaced) satellite stations in an aperture of approx. 100m. Three stations are required to obtain a back-azimuth (or slowness) estimate (eq. III.1.1-2). With 4 given stations, 4 three-station sub-arrays can be formed, each providing one back-azimuth estimate that can be analyzed in Jackknifing manner. The sub-array that consists of the three satellite stations has the shape of a equilateral triangle. The other three of the four evaluated sub-arrays are, however, increasingly collinear, describing isosceles triangles with an obtuse angle of 120°. For these, the depth behavior is very

different and the 15° opening angle rule is not valid. The solutions of the four sub-systems are therefore not equally trustworthy, a fact that needs to be considered since Jackknifing reconstructs the global solution based on the solutions of the sub-systems. Chapter II.3.6 derived sub-solution beam-opening angles for each sub-system in dependance of its data quality, using a sensitivity analysis based on eq. III.1.1-2. This equation, however, relies on the plane-wavefront assumption, which if violated, renders the result inaccurate. In the following, we will therefore analyze the error arising from this violation, in intermediate distances, for various array geometries and focal depths, to obtain estimates for the minimal distance from which the error becomes negligible and the estimate reliable. It further allows us to provide appropriate uncertainty terms to cover the associated error, if the back-azimuth estimate is used for shorter distances. The exact solution for the array's back-azimuth vector (chapter II.1.2.3) was derived as

$$\hat{n} = \left(\gamma^2 + 1\right)^{-1/2} \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \tag{III.1.1-1}$$

with

$$\gamma = -\frac{r_y \cdot (t_p - t_q) + q_y \cdot (t_r - t_p) + p_y \cdot (t_q - t_r)}{r_x \cdot (t_p - t_q) + q_x \cdot (t_r - t_p) + p_x \cdot (t_q - t_r)} \tag{III.1.1-2}$$

and $\vec{p}, \vec{q}, \vec{r}$ being the station locations, and t_p, t_q, t_r the corresponding onset times, which appear only in differences and are obtained by cross-correlation of the given waveforms. As reference point serves the geometrical center of the array

$$\vec{C} = \frac{1}{3} \begin{pmatrix} p_x + q_x + r_x \\ p_y + q_y + r_y \end{pmatrix}. \tag{III.1.1-3}$$

Due to the plain wave front assumption, the beam vector shows a angular variation in accuracy, depending on the geometry of the distributed array stations. As the assumption only holds for far distant sources, the accuracy suffers for sources near by - with well distributed array configurations dominantly for shallow source depths. The associated error shall be mapped in the following.

Given a source location, we can compute all involved travel-times to provide the onset parameters t_p, t_q, t_r for eq. (III.1.1-2) (setting the origin time to $t_0 = 0$). This allows to calculate the

corresponding back-azimuth vector $\hat{n}$ for a given station geometry using eq. (III.1.1-1). The angular discrepancy between true back-azimuth $\tilde{n}$ and the tripartite array based back-azimuth $\hat{n}$ is given by

$$\epsilon = \cos^{-1} \hat{n} \cdot \vec{n} \qquad \text{(III.1.1-4)}$$

and is mapped in fig. III.1.1:2 for different array geometries and simulated hypocenter depths.

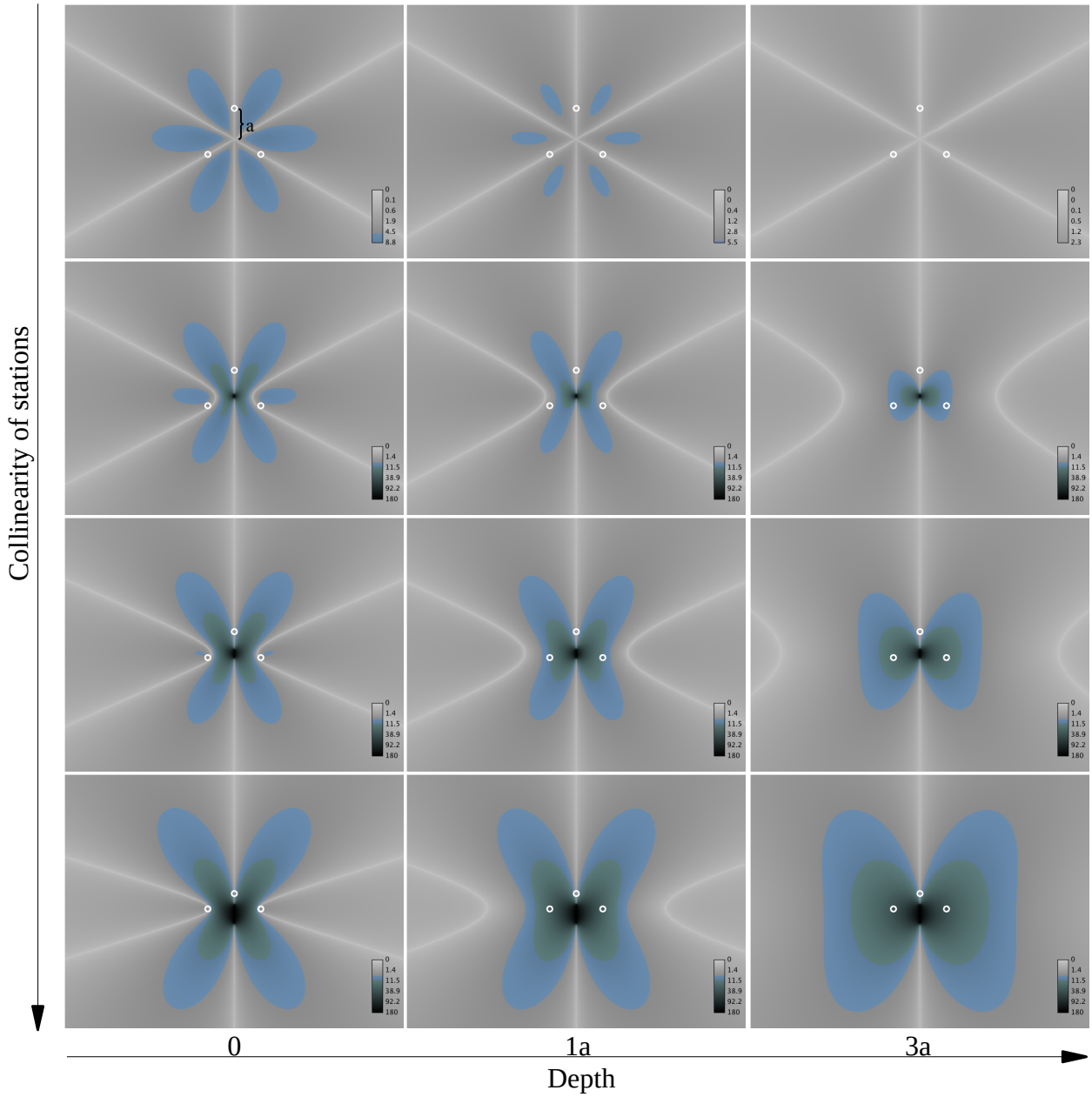


Fig. III.1.1:2: Angular precision of tripartite array geometries over depth [surface level, 1x aperture, 3x aperture]. Gray areas signify errors under 5 degrees, blue mark areas above 5 degrees and dark green areas signify errors above 10 degrees. The first row corresponds to the largest sub-group configuration of a four-partite array. The last row to the smallest sub-group, formed by the central station and two outer array stations.

An equilateral configuration (fig. III.1.1:2, top row) generally holds the highest accuracy with maximum errors not exceeding 10°, the error decreasing with source-depth. The error is below 5° for distances over three aperture lengths (fig. III.1.1:3) - and if the source lies below a depth of two aperture lengths (which is usually the case), the error does not exceed 2.5°, independent of epicentral distance. For this case, narrow beam opening angles of 5° suffice to constrain the possible epicenter location. For the general case (independent of depth) an opening angle of 10° suffices for hypo-central distances above three aperture lengths.

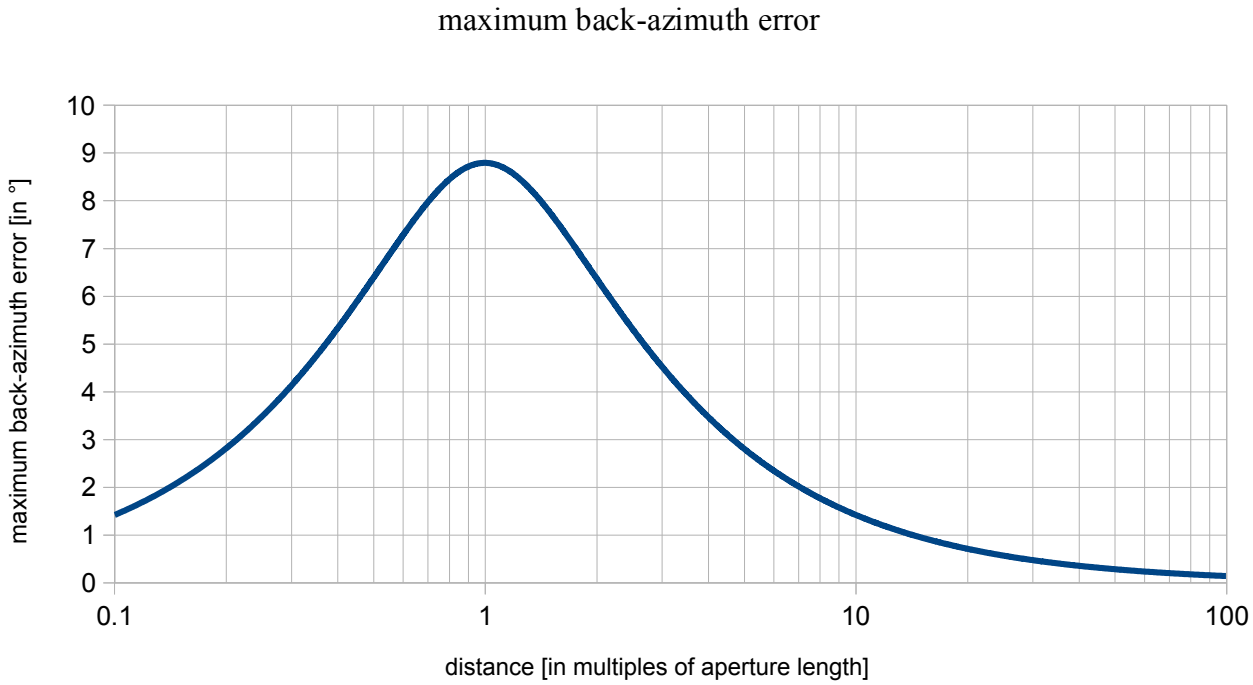


Fig. III.1.1:3: Maximum error of equilateral tripartite array back-azimuth estimates. This graph pertains to the array configuration of fig. III.1.1:2, top row and plots the highest observed discrepancy between true and calculated back-azimuth values for events in different distances (i.e. for a given distance that discrepancy is chosen that is highest amongst all azimuth directions). Highest discrepancies are given for sources lying on a circle around the triangle centroid with a radius equal to one aperture (with the back-azimuth beam originating from the triangle's geometrical center). To avoid large errors, the back-azimuth constraint should only be used for epicentral distances above several apertures, where the plane-wave assumption is sufficiently valid. Instead, the beam could also be convoluted with this distance-dependent discrepancy curve, yielding a beam width that varies with distance, computed for the given array geometry (i.e. the graph above would determine the minimum beam width (in degrees) at a given distance).

For an equilateral configuration, the three medians of the triangle (lines that run from each vertex through the centroid of the triangle) form three high accuracy axes, for which the back-azimuth estimate is always exact. For non-equilateral configurations the medians are asymptotes to the exact regions, which distance themselves with depth (fig. III.1.1:2). The estimate error increases with the

collinearity of the layout, yielding errors up to 180° in the vicinity of the array center. With increasing depth the error (usually) decreases, except for strongly collinear layouts (fig. III.1.1:2, bottom row). For such, the error even increases with depth, but the extent of the inaccurate region converges with depth (fig. III.1.1:4).

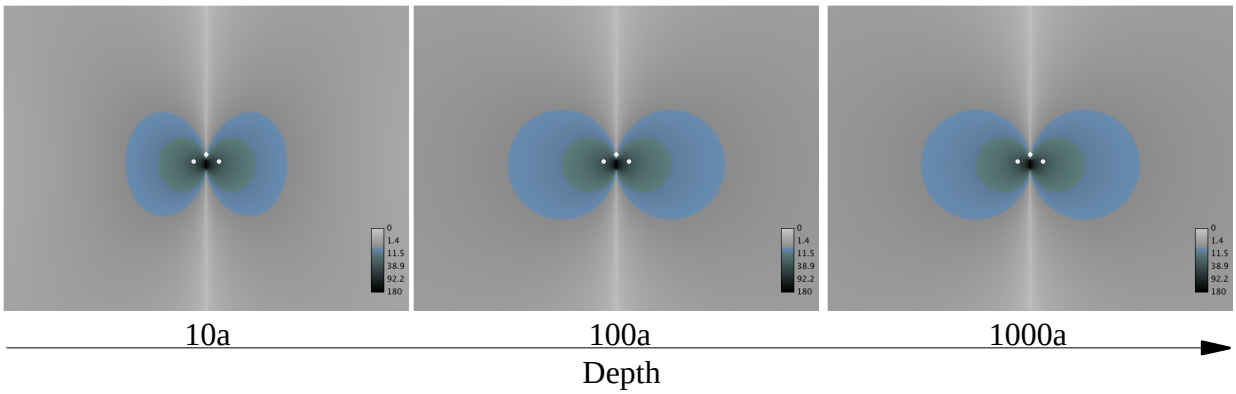


Fig. III.1.1:4: Angular error of a tripartite array with center-array station configuration (fig. III.1.1:2 bottom row) for larger depths. The extent of the low accuracy zone (blue and green areas) is confined, not exceeding a maximum distance of approx. 3.7 aperture lengths for the error>10° zone (green) and approx. 7.5 aperture lengths for the error>5° zone (blue).

The error remains below 5° for sources with epicentral distances larger than approx. 7.5 aperture lengths. This value increases with collinearity. Aperture length is here defined as the longest distance between the centroid and any station. At 3.5 lengths the maximum error reaches 10° and 40° at 1 aperture length. This has practical implications for the analysis of events in intermediate distances (e.g. in hill-slope monitoring, locating (near-)surface events in distances of 1-10 aperture lengths): With four-partite arrays being analyzed in a Jackknifing manner, significant weight should be put on the back azimuth estimate given by the sub-array consisting of the three outer (satellite) stations rather than on the three other estimates that include the central station, holding obtuse 120° angles.

The described errors reflect the inherent methodological accuracy. Chapter II.3.6 already provided uncertainty measures to cover errors based on the signal quality relying on eq. (III.1.1-1). The accuracy analysis showed that the estimate of four-partite arrays can be trusted for distances larger then approx. 20 aperture lengths. Here, the maximum error of eq. (III.1.1-1) is less than 1°. For shorter distances down to 7.5 aperture lengths, the estimate can be biased with an error of up to 5° In this intermediate field, with distances lying between 7.5 and 20 aperture lengths, it is sufficient to add 5° to the signal-based uncertainty for the beam opening angle to obtain accurate results. If the

estimate, however, is used to locate closer events, the more complex distant-dependent uncertainty curve, following fig. III.1.1:3, would need to be applied, shaping the back-azimuth with a non-linear varying width. Inhomogeneities in the velocity model, distort the wave front and display another source of error for the back-azimuth estimate. Tilted layers for instance redirect the back azimuth beam from it's initial direction, depending on azimuth and incidence angle. This aspect is hard to incorporate as the detailed model-information is usually not given, but as the absolute error grows with distance, this aspect hardly affects estimates in near distance.

Using appropriate uncertainty measures, the intersection of two (or more) linear beams will always yield a unique solution. This is different for the standard (non-linear) P-onset based location. Over the previous chapters several examples where shown in which an unique set of arrival times counter-intuitively led to multiple, equally possible hypocenter locations, predominantly when only few data was given. P-onset based location is the one mostly applied and plays the dominant role in real-time location as applied in earthquake early warning systems, today. Yet, it is right there where all conditions are met for these ambiguities to arise, which have the potential to strongly falsify the location estimate. A well known example are signals of teleseismic sources, which are often mis-interpreted and mis-located as local sources appearing to be situated inside the network. The following chapter will address why, when and where these ambiguities arise.

III.2 **Location Ambiguities in Hyperbolic Location**

An event surrounded by three stations can easily be located. So, we may think - and this is true, if the right data is available: Three S-P constraints could be constructed, which intersect (in an error-free world) in one common point: the hypocenter location. The location could likewise be identified by three (or even two) back-azimuth beams, which would intersect in the same point. The location could also be solely based on P-arrivals: This would lead to three hyperbolic constraints, which likewise would intersect in a common point in the middle of our small network. A location based on the arrival-times of P-phases is the dominant location technique used today, since P-phases can be detected most precisely and also arrive first, which makes them the phase of choice for real-time location schemes. It is, however, also the one technique which may render locating this event a bit more complex than we may have thought: A shift from the network center towards one of the stations is sufficient to cause a second solution to appear (fig. III.2:1).

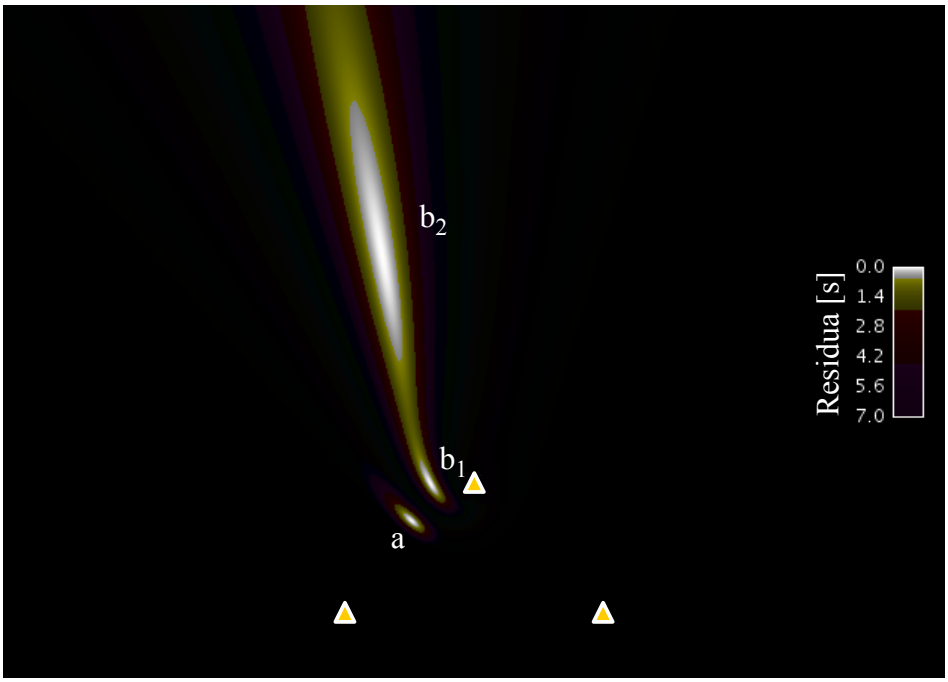


Fig. III.2:1: Ambiguities in a residual map. For a set of arrival-times corresponding to location a, one clear residual minimum is found. For onsets that yield a solution lying in b_1 or in b_2, two minima b_1 and b_2 exist. Non-global-search methods risk to find and present only one of the two possible locations.

This solution is exact, but less precise, being situated outside of the network - in a way that not only

the travel-time differences are the same for both solutions – also the amplitude ratios are similar, so that it may be difficult to identify the true solution solely based on the P-arrivals. A network location based on P-arrival times is a non-linear problem - even in the simplest case of a homogeneous (half-) space due to the non-linear relationship between travel-time and hypocenter coordinates. It therefore shouldn't surprise us that more than one solution may be possible. It comes, however, counter-intuitive and it is not straightforward why and where these ambiguities arise. While the effect is the same whether we use a residua based approach or hyperbolic P-P constraints, it can be analyzed more illustrative for the latter:

The time difference between one phase arriving at two stations constrains the source to lie on a hyperbola (in 2D), which led to the name “hyperbolic location”. A hyperbola is formed with each pair of stations and displays a spatially one dimensional solution like the classical S-P circle. Contrary to the circle on which the origin time t_0 is constant, t_0 varies on the hyperbola and can thus be interpreted as the parameter of the curve. The permutation of all given stations into pairs yields a set of hyperbolae. With three stations, three hyperbolae are generated - usually intersecting in one common point. For shallow events this would give sufficient, but non-redundant data, to allow a rough determination of the hypocenter. Yet, situations are observed where the three hyperbolae intersect in two points (fig. III.2:2) or even cease to intersect at all.

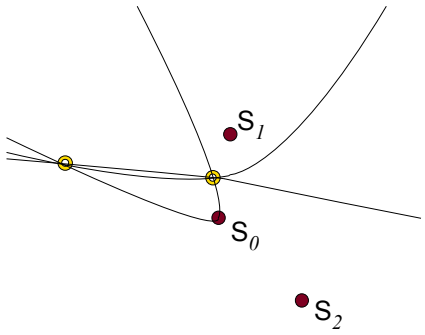


Fig. III.2:2: Three station network (red circles) in a 2D location scenario. Three detected P-onsets constrain three hyperbolae which intersect in two locations.

Assuming error-free data, a location can be constrained uniquely if the corresponding system of equations is overdetermined concerning the underlying model assumptions, as is the case when many stations with P-detections are given. When the amount of available stations drops to a certain

limit, the set of equations becomes exactly determined. This limit is reached with 4 stations due to the 4 dimensional model vector, consisting of three space- and one time dimension. One is tempted to assume that the 4 P-arrivals uniquely determine the four unknowns (origin time and hypocenter), but this would be true only for a set of linear equations, which is not given for this approach, which is interdisciplinary referred to as TDOA (i.e. Time Difference Of Arrival). The low amount of stations doesn't only lead to an increase in location uncertainty, but may also cause secondary solutions to arise. Contrary to an uncertainty that stretches the solution in space and time, these ambiguities are exact and distinct solutions, similar to those that are expected with a linear station setup, where the degenerated solution is only constrained in distance around the axis of stations, but no longer fixed to one spatial coordinate. Indeed we will find, that this case is only a special case of the more complex problem.

Most single event location approaches rely on the arrival time differences between the first arriving phases at different stations (TDOA). Hence, this effect was not only observed in seismology. The first observation of this effect reaches back to as far as 1928 when Iglada observed multiple solutions while locating via TDOA. Cete reaffirmed this observation in 1977 and Rindorf (1981, 1984) proved that maximal two solutions can arise for the TDOA approach. Chan (1994) showed the same in signal processing and Spiesberger (2001) observed the effect in bio-acoustic application of multi-lateration. Although this effect was recurrently observed, it stayed hardly noticed (e.g. in Swanson et al., 1992) or analyzed by the seismological community, since usually far more than four stations are given and used. Consequently, literature often quotes the minimum amount of stations with P-detections too low to yield only unique solutions, e.g. three in two-dimensional problems (Ruff,2001, Pujol,2004). For the standard situation where more than the minimum amount of stations is given, the secondary solution is usually disqualified in being inconsistent with some of the P-arrivals. This is, however, only the case if the stations are not located on a common hyperboloid (Xiaochun, 2008), which would render the geometrical information redundant and allow ambiguities to arise with any amount of stations. This was in type observable in the opening example of chapter II.4: P-detections on 4 stations and a fixed focal depth rendered the three-dimensional problem overdetermined, yet, two distinct solutions were found in the residual plot. Looking at the 4 stations confirms that they can all be fixed to one hyperbolic curve. This demonstrates that ambiguities can arise even when not expected and how important it is to mould algorithms to properly treat ambiguities. This is crucial for real-time location schemes based on P-arrivals as used in earthquake early warning systems, today (Allen et al., 2009) that locate based on

few arrivals only. But also where due to high effort and cost only few stations can be deployed (e.g. European Rosetta comet mission) or due to weak seismic signals only few network stations register the events (Walter&Joswig, 2009), this ambiguity may arise and must be treated properly to be either resolved (if possible) or taken into account as additional uncertainty in location.

In the following, we will identify the boundary conditions that lead to this degeneration and then analyze the relationship between the two locations. This will allow us to identify the regions that yield unique- and those that lead to ambiguous solutions, a knowledge which is also useful for network planning. It will further enable us to predict the approximate location of the secondary solution. We will find that although this degeneration is a boundary effect in the time domain (i.e. that for well distributed stations it only occurs for large arrival-time differences), it actually affects large areas in the space domain, with at least 1/3 of all possible hypocenter locations having a secondary solution. To understand when and where the ambiguity occurs, the solution will first be derived analytically in two space dimensions in an homogeneous environment. Following this, we will map the ambiguities to identify these zones for different network layouts in two and three dimensions.

III.2.1 **Deciphering Hyperbolic Location**

The problem of determining the hypocenter location and origin time can be expressed as the search for the intersection of hyperboloids. This search in 3-space dimensions (in a four-dimensional problem) can be reduced to a search in two space-dimensions (x,y,t_0), the search for the intersections of hyperbolae. In this three dimensional problem the system is exactly determined with three stations, forming three hyperbolae. To resolve the true nature of the ambiguity, the system is treated as ideal (void of uncertainty), meaningly infinitely thin hyperbolic curves. Each one individually displays the possible epicenters over the corresponding origin time which may be interpreted as the parameter of the curve. These objects are hence three dimensional (x,y,t_0) but are projected into two dimensional curves. Expanding them back into their three dimensions in space-time-diagrams will allow us to understand the source of the ambiguity and will provide for the mapping of the solution space.

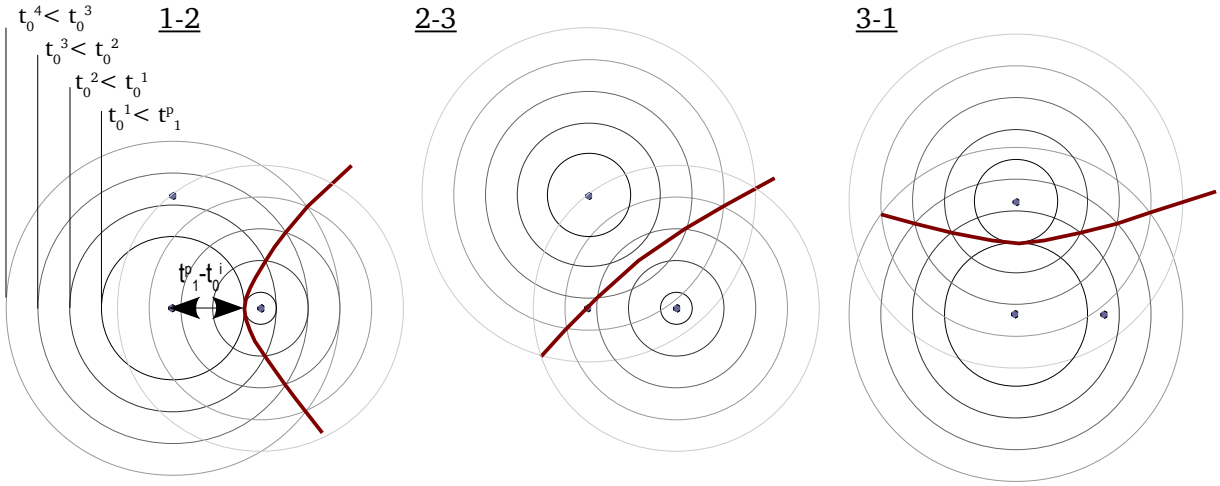


Fig. III.2.1:1: Three hyperbolae as created from the intersections of growing hypothetical travel-time circles: A trial origin time determines a travel-time between origin-time and the P-arrival t^p, which translates into circular spatial constraints around the both stations of the P-P constraint. Decreasing the (unknown) origin-time causes the circles to grow. The intersection of both circles over all origin-times forms the hyperbola. The hyperbola does therefore display all epicenter solutions that are possible for all trial origin-times. The permutation of 3 stations into pairs yields 3 hyperbolae. Their intersection marks the global solution, identifying epicenter and the true origin-time.

In the following the epicenter and origin time are represented by the vector

$$\underline{e}=(x,y,t_0) \tag{III.2.1-1}$$

while $\underline{s}_i$ represents the position (x_i, y_i) and recorded arrival t_i^p of station i

$$\underline{s}_i=(x_i, y_i, t_i^p)\,. \tag{III.2.1-2}$$

The constitutive equation

$$l_i=l(P,\underline{C},tt_i)=0\,. \tag{III.2.1-3}$$

connects the wave's travel time tt_i

$$tt_i=t_i^p-t_0 \tag{III.2.1-4}$$

between hypocenter and i^{th} station over the model $\underline{\underline{C}}$ and the fastest travel path $P(r_{xy})$ with the epicentral distance r_{xy}. P is a function of the different ray paths

$$P=\{D, L, R\}\ , \tag{III.2.1-5}$$

with D denoting the direct, L the reflected and R the refracted paths. For now we will assume a homogeneous half space, therefore $\underline{\underline{C}}$ will be a constant velocity $\underline{\underline{C}}\rightarrow c$ and the ray path P=D a direct one. Equation III.2.1-3 becomes then

$$l_i^D=\sqrt{\underbrace{(x_i-x_0)^2+(y_i-y_0)^2}_{r_{xy}^2}+\underbrace{(z_i-z_0)^2}_{Z^2}}-c\cdot tt_i=0\ . \tag{III.2.1-6}$$

The epicentral distance r_{xy} is therefore accordingly

$$r_{xy}=\sqrt{(c\cdot tt_i)^2-Z^2} \tag{III.2.1-7}$$

The focal depth affects the solution via the vertical distance Z from station i,

$$Z=z_i-z\ , \tag{III.2.1-8}$$

while Z is used as external parameter that will be left constant during the search for the planar solution.

The hyperbola displays all possible spatial solutions concerning the arrival-time difference between two stations over all possible origin times, which decreases along the extending branches. For each pair of stations, a sequence of decreasing possible trial origin times t_0^j is being assumed. The corresponding travel time tt_{ij} to station i is found in the difference between arrival time and trial origin time, $tt_{ij}=t_i^p-t_0^j$. Using eq. (III.2.1-7), the epicentral distance r_{ij} is derived for each of the

two stations, which results in a circular, degenerated solution around each station (fig. III.2.1:1). The two circles' intersections mark the points in the right distance r_{1j} and r_{2j} to both stations and build -in sequence of decreasing trial origin-times t_0^j - the hyperbola. The permutation of three stations into pairs yields three hyperbolae, which (usually) intersect in one common point, but may also have two or no point of intersection. These points of threefold intersection ("triplepoints") display the solution which is consistent with the information of all three stations (fig. III.2.1:2). Fig. III.2.1:1 and III.2.1:2 show that the hyperbola corresponds to the sequence of intersections of two growing $(t_i^p - t_0^j)$ -circles. The triplepoint, which is the intersection of three hyperbolae, can likewise be expressed as the intersection of three circles with the same t_0^j (fig. III.2.1:3).

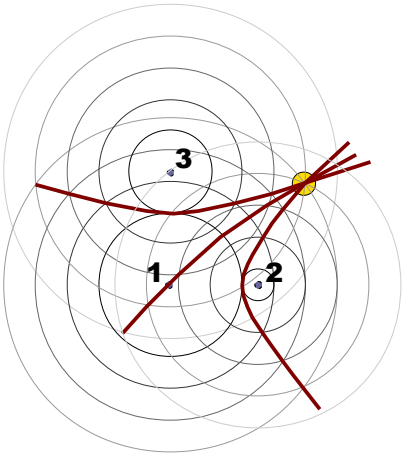


Fig. III.2.1:2: Fusion of Fig. III.2.1:1: (1-2), (2-3) and (3-1). The triplepoint is the intersection of three hyperbolae, and at the same time of three circles constructed from the same trial origin time (see fig. III.2.1:3)

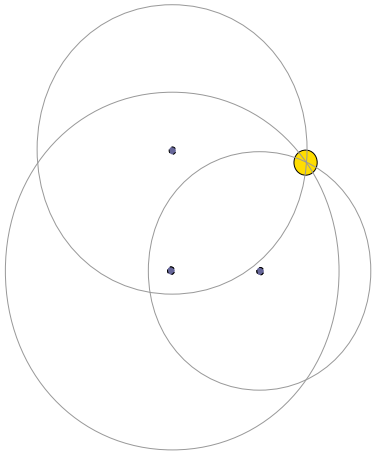


Fig. III.2.1:3: A triplepoint -the intersection of 3 hyperbolae- is also the intersection of 3 circles with the same t_0^j .

Using this circular representation, we will display the system of fig. III.2.1:2 in it's three dimensions in a space-time-diagram with the axes x, y and t_0. Beginning with two of the three stations, $\underline{s}_1$ and $\underline{s}_2$ (which provide the basis for one hyperbola), both data sets and constitutive equation are visualized (fig. III.2.1:4): With $Z = 0$, the constitutive equation (eq. III.2.1-6) takes the form of a cone, touching the station vector with it's tip (which points out that no solution can exist later than the earliest P-onset, which is evident since the event has to occur before it is recorded). With

decreasing origin time (note that the vertical axis is inverted) the radius of the cone grows as the phase is given more time to travel. Cutting out a x-y-plane for any given t_0^j reveals the correlation with the previously used circular description.

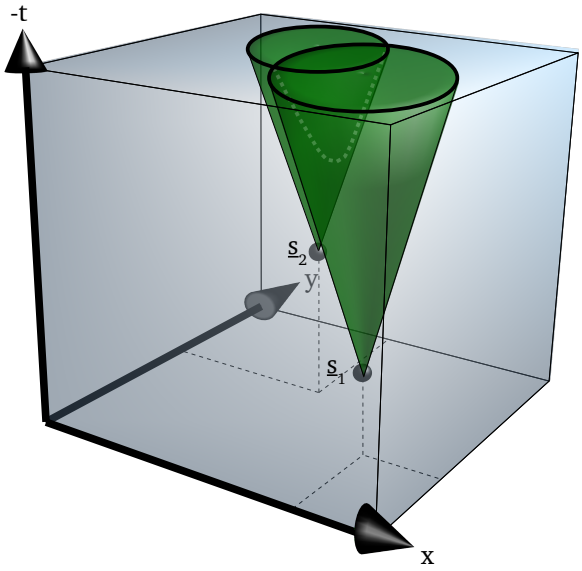


Fig. III.2.1:4: Space-time diagram. Points represent the data vectors $\underline{s}_1$ & $\underline{s}_2$, the cones display the constitutive equation for Z=0. The two cones intersection curve represents the hyperbola (dashed line).

The hyperbola appears in the diagram as the cone-cone intersection (dashed curve, fig. III.2.1:4). A projection into the x-y-plane would display its well known form.

The cone's slope m for Z=0 is given as

$$m = \frac{\partial r_{xy}}{\partial tt_i} = \frac{c^2 \cdot tt_i}{\sqrt{(c \cdot tt_i)^2 - Z^2}} \overset{Z=0}{=} c$$

and is equal to the seismic velocity c. With an event laying in another depth than the analyzed x-y-plane ($Z \neq 0$), m is no longer constant, deforming the cone's tip (fig. III.2.1:5).

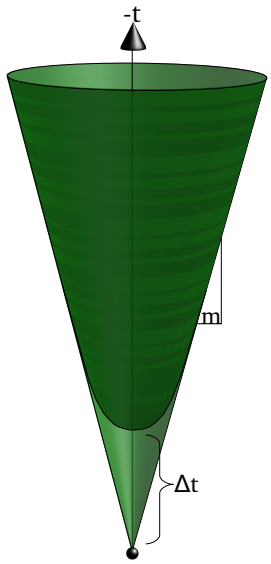


Fig. III.2.1:5: Cone deformation for Z≠0.

Instead of the tip touching the data vector, the "cone" starts rounded and vertically shifted. This shift (i.e. a time offset) can be derived from eq. III.2.1-7 as

$$\Delta t_i = c^{-1} \cdot Z \,, \tag{III.2.1-9}$$

which is the time that the wave needs to travel from the hypocenter to the evaluated x-y-plane. With decreasing time (tt_i increasing) the deviation to the undisturbed cone becomes negligible

$$r_{xy}^{Z\neq 0}(tt \gg 0) = r_{xy}^{Z=0} \tag{III.2.1-10}$$

For simplicity, we will continue to use cones with $Z=0$. Having expanded the hyperbolae to it's three dimensions as the cone-cone intersection curve, the triplepoints may now be constructed and analyzed by adding the third data vector with its cone into the diagram. This will allow us to directly see that ambiguous solutions are no strange phenomena but a natural part of hyperbolic constraining. As the intersection of three hyperbolae marks the triplepoint(s), so does the intersection of the three cones. Since the intersection of the third cone with either, cone 1 or cone 2,

has to lie on the intersection curve of the first two cones, we will only display the intersection curve (i.e. the three-dim. hyperbola) and the third cone for better visibility (fig. III.2.1:6): By placing $\underline{s}_3$ and it's cone strategically, we can indeed yield between zero and two intersections. The simplest case for ambiguous solutions is found by placing the third cone anywhere on the imaginary line connecting $\underline{s}_1$ and $\underline{s}_2$. The third cone will intersect both branches of the intersection curve symmetrically, which corresponds to a linear station layout: For any time-difference (i.e. the vertical shift between the $\underline{s}_i$) two intersections exist. Moving $\underline{s}_3$ off the axis still yields intersections on both branches, but no longer symmetrically: One intersection will move closer to the apex of the intersection curve, the other further away. This is the typically observed behavior of hyperbolic ambiguity, with one solution being near-by (and usually inside) the network, and a secondary distant solution outside.

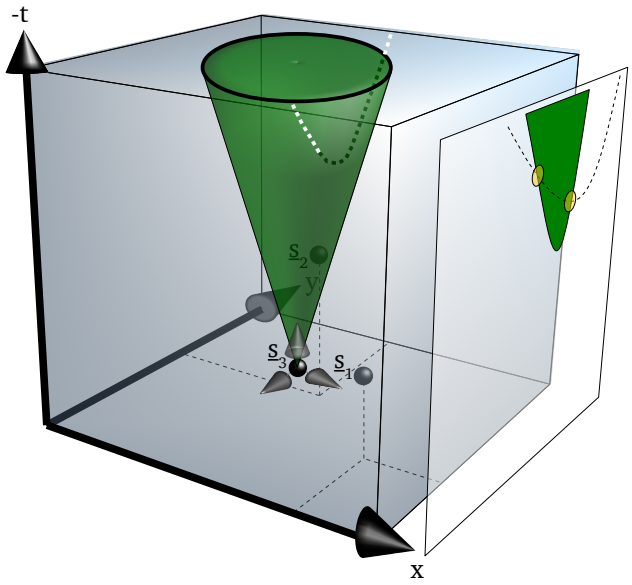


Fig. III.2.1:6: Possible intersections of three cones. By shifting $\underline{s}_3$ along the three axes, 0..2 intersections can be achieved between the third cone and the (dashed) intersection curve between cone 1&2, (besides the case of infinite solutions when two cones (and stations) are collocated). The example shows the case of 2 intersections. A projection into the y-t-plane shows the two intersections (yellow circles).

Moving $\underline{s}_3$ further off the axis causes the intersection close to the apex to actually cross over to the other branch, while the distant intersection on this branch simultaneously ceases to exist (due to the branches becoming increasingly steep with decreasing origin-time). This correlates with the

intuitive, general case with only one solution. For a stronger tilt of the intersection curve (which is given with larger vertical distances between $\underline{s}_1$ and $\underline{s}_2$, i.e. larger arrival-time differences Δt^p, fig. III.2.1:7) the third cone might not intersect at all with the intersection curve.

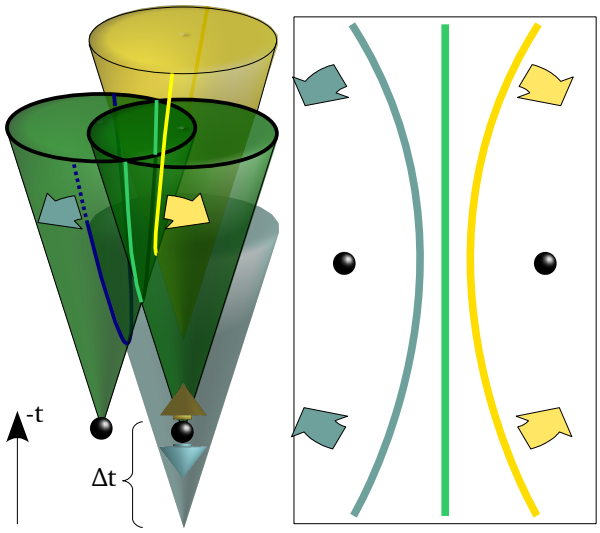


Fig. III.2.1:7: Cone-cone intersection curve increasing curvature with increasing arrival-time difference Δt^p (i.e. increasing vertical shift): Lifting the right cone to earlier arrival times (yellow) increases the curvature around the right station; Lowering it to later arrival times than the left cone, increases the curvature around the left station.

For plausibility, the $\underline{s}_i$ mustn't lay within another cone. This is the graphical equivalent to the plausibility condition that the arrival-time difference of two station can't be greater than the (longest) time a signal travels on the slowest ray path that connects both stations: Since all cones hold the same slope m, no intersection would exist. Using space-time-diagrams, we witnessed that hyperbolic location yields either no, one or two solution(s) (with a homogeneous model). We will now use the underlying circular description of triplepoints to derive an analytical solution, which will allow us to map the regions of ambiguity.

III.2.2 **Analytical Solution for the Homogeneous Space**

The number of triplepoints can be quantified analytically by using the circular representation (in which the problem is analyzed per fixed origin time), with the attempt of calculating all possible intersections of the three circles. To do so, we will again start to calculate the intersection between two circles and then continue to evaluate if the third circle also intersects this point.

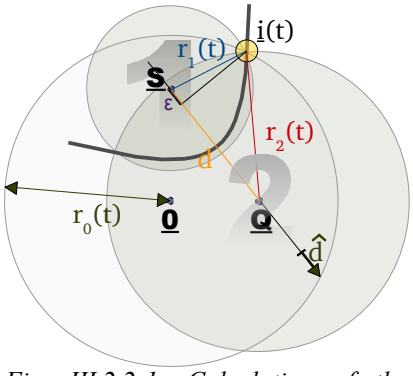


Fig. III.2.2:1: Calculation of the triplepoint as intersection of three circles. $\underline{S}=\underline{p_1}-\underline{p_0}$, $\underline{Q}=\underline{p_2}-\underline{p_0}$, with p_i being the position of station i. The intersection (marked yellow) is projected on the axis connecting station 1&2.

The intersection of the two circles can be projected onto the axis connecting both station. The distance from the projected point to the first station is given in

$$\varepsilon=\left(r_1^2+d^2-r_2^2\right)/\,2\mathrm{d}\,;\quad r_i=\sqrt{c^2\cdot\left(t_i^p-t_0\right)^2-Z^2}\,, \tag{III.2.2-1}$$

with r_i being the radius of station i and d being the distance between both stations. The intersection point between both circles (that constructs the hyperbola) can then be identified as

$$\vec{i}_{l/r}(t)=\vec{S}+\varepsilon\cdot\hat{d}+\sqrt{r_1^2-\varepsilon^2}\cdot\begin{pmatrix}-\hat{d}_y\\ \hat{d}_x\end{pmatrix}, \tag{III.2.2-2}$$

with $\vec{S}$ being the coordinate of station 1 and $\hat{d}$ being the normalized direction vector, pointing from station 1 to station 2. This intersection point is a triplepoint if the third circle also intersects this point, when

$$\left|\vec{i}_{l/r}(t_0)\right| \stackrel{!}{=} r_0(t_0) \,. \tag{III.2.2-3}$$

This condition yields the following equation

$$f(t_0) = \frac{\begin{array}{c} S_x^2 + S_y^2 + 2t_0\left(a(\vec{S}\cdot\hat{d}) + c^2(t_0^p - t_1^p)\right) \\ + c^2\left((t_1^p)^2 - (t_0^p)^2\right) + 2\mathrm{b}(\vec{S}\cdot\hat{d}) + 2\gamma\cdot(\hat{d}\times\vec{S})\cdot \end{array}}{\sqrt{t_0^2(c^2-a^2) - 2t_0(c^2 t_1^p + ab) + c^2 t_1^p - b^2 - Z^2}} \quad , \tag{III.2.2-4}$$

$$\stackrel{!}{=} 0$$

with

$$a = \frac{c^2}{d}(t_2^p - t_1^p) \tag{III.2.2-5}$$

and

$$b = \frac{c^2}{2\mathrm{d}}\left((t_1^p)^2 - (t_2^p)^2\right) \quad . \tag{III.2.2-6}$$

The gamma factor

$$\gamma = \langle 1, -1 \rangle \tag{III.2.2-7}$$

determines the side (left/right) of the hyperbola branch. With $f(t_0)$ being of second order, $f(t) \stackrel{!}{=} 0$ provides up to two solutions per branch as also for the whole system. The united curve's run (both curves for both $\gamma = \pm 1$), which describes the complete hyperbola, shows that there can be 0, 1 or 2 solutions (fig. III.2.2:2), meaningly 0, 1 or 2 triplepoints in a system of three stations (see eq. III.2.2-8), depending on the position and orientation of $f(t_0)$. For a linear layout, $f(t_0)$ becomes linear and the solutions for $\gamma = \pm 1$ degenerate to be identical equal. There exist therefore only two cases: two symmetrical triplepoints (one on each branch) or none.

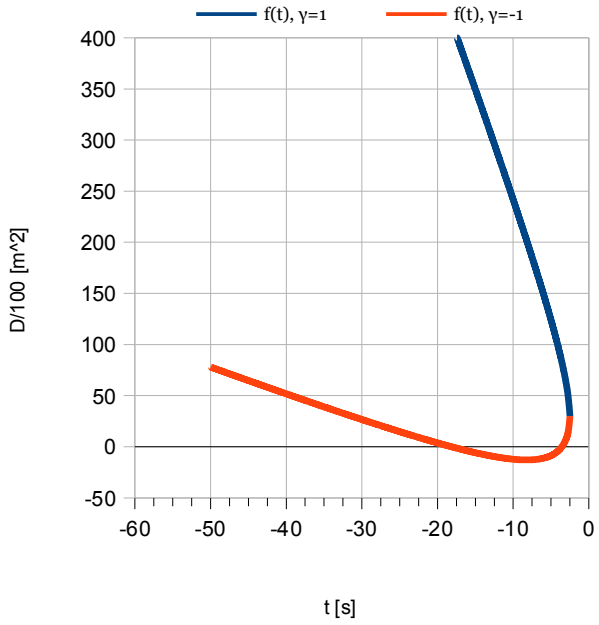


Fig. III.2.2:2 f(t) with two roots, being constructed of two segments corresponding to the two hyperbola branches.

Sorting eq. III.2.2-4 concerning the origin time yields the quadratic equation

$$t_0^2+\frac{2\left(c^2 t_1^p+ab+\alpha\beta\right)}{\alpha^2+a^2-c^2}\cdot t_0+\frac{c^2\left(t_1^p\right)^2-b^2-\beta^2-Z^2}{c^2-a^2-\alpha^2}=0 \qquad \text{(III.2.2-8)}$$

with

$$\alpha=\alpha(\gamma)=\frac{a\left(\vec{S}\cdot\hat{d}\right)+c^2\left(t_1^p-t_0^p\right)}{-\gamma\left(\hat{d}\times\vec{S}\right)} \qquad \text{(III.2.2-9)}$$

and

$$\beta(\gamma)=\frac{S_x^2+S_y^2+c^2\left(\left(t_1^p\right)^2-\left(t_0^p\right)^2\right)+2b\left(\vec{S}\cdot\hat{d}\right)}{-2\gamma\left(\hat{d}\times\vec{S}\right)} \quad . \qquad \text{(III.2.2-10)}$$

Solving this equation for the origin time yields the explicit form

$$t_0^{1/2}=\frac{c^2 t_1^p+ab+\alpha\beta}{c^2-a^2-\alpha^2}\pm\sqrt{\underbrace{\left(\frac{c^2 t_1^p+ab+\alpha\beta}{\alpha^2+a^2-c^2}\right)^2}_{\aleph}-\underbrace{\frac{-c^2\left(t_1^p\right)^2+b^2+\beta^2+Z^2}{\alpha^2+a^2-c^2}}_{\eth}} \quad . \qquad \text{(III.2.2-11)}$$

To avoid singularities in a linear setup, the term $\hat{d}\times\vec{S}$ should in practice be factored out of the terms α and β, but is here left unchanged for better readability. Note that γ vanishes in all occurrences [$\beta^2(\gamma)=\beta$, $\alpha^2(\gamma)=\alpha$, $\alpha(\gamma)\beta(\gamma)=\alpha\beta$], making all distinction void between left and right branch. This shows again that max. two triplepoints may exist in a system of three stations. With the discriminant

$$D=א-ת \tag{III.2.2-12}$$

a first measure for the maximum number of solutions N_{max} is given:

$$\begin{aligned} &D<0 \text{ : no solution } (N_{max}=0)\\ &D=0 \text{ : max. one solution } (N_{max}=1)\\ &D>0 \text{ : max. two solutions } (N_{max}=2). \end{aligned} \tag{III.2.2-13}$$

This, however, only serves as upper limit for the amount of solutions, since eq. III.2.2-11 allows for a-causal events. The true number of solutions may therefore be less when those are filtered out. The term Z^2 in eq. III.2.2-11 can only act negatively within the root (compare with eq.III.2.2-15). Therefore, if no solution/triplepoint (representing the hypocenter for a given focal depth) exists within the plane which is being expanded by the station locations ($Z=0$), no solution will exist at all (for any $Z\neq 0$): No hypocenter can be determined based on the given data and model.
If eq. III.2.2-11 has a solution, back substitution of $t_0^{1/2}$ into eq. III.2.2-1 and III.2.2-2 provides the location $\vec{T}_n$ of the corresponding n^{th} triplepoint

$$\vec{T}_n=\vec{S}+\frac{(r_1^2+d^2-r_2^2)}{2\mathrm{d}}\cdot\hat{d}+\sqrt{r_1^2-\frac{(r_1^2+d^2-r_2^2)^2}{4\mathrm{d}^2}}\cdot\begin{pmatrix}-\hat{d}_y\\ \hat{d}_x\end{pmatrix} \tag{III.2.2-14}$$

with

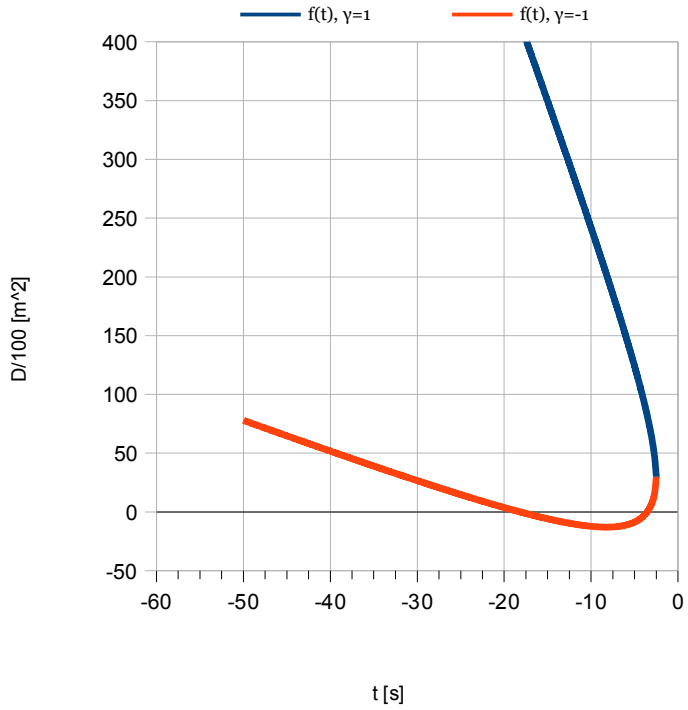
$$r_i=\sqrt{c^2\cdot(t_i^p-t_0^n)^2-Z^2}\,, \tag{III.2.2-15}$$

t_0^n being either $t_0^n=t_0^1$ or $t_0^n=t_0^2$, collectively for all r_i .

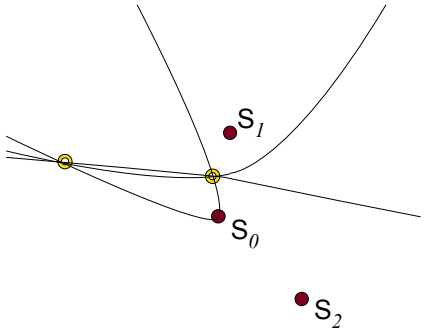
Examples for the three cases that are possible with $f(t_0)$ (eq. III.2.2-11): No, one or two solutions, are presented in the following (fig. III.2.2:3 - III.2.2:5). Following this, we will use eq. III.2.2-11 and III.2.2-14 to map the critical regions that lead to ambiguities.

Two ambiguous solution

a)



b)



S_0=(0,0)km t_0^p= 1.3 s

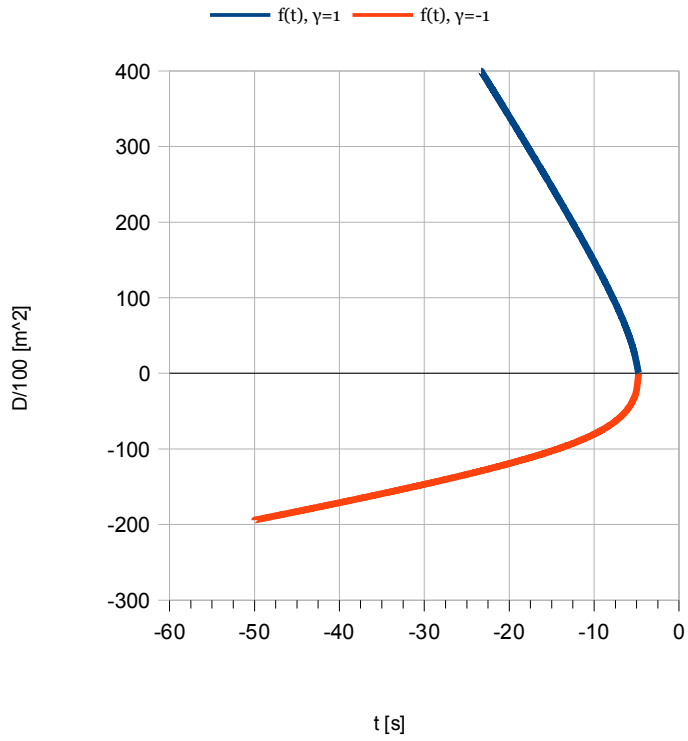
S_1=(15,100)km t_1^p= 2.0 s v_p= 10km/s

S_2=(100,-100)km t_2^p= 14.63 s

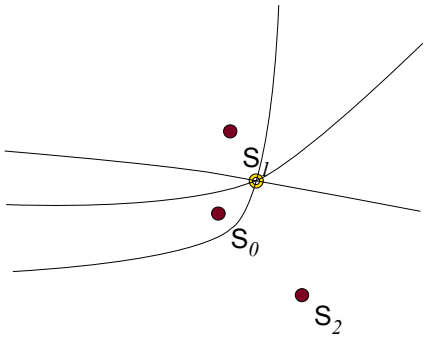
Fig. III.2.2:3 Two solutions. Panel a) shows the curve run of f(t) with two roots. Panel b) shows the three hyperbolae constraining the corresponding two ambiguous locations (yellow circles). The station positions (S_i, red circles) and arrival times are listed above.

One unique solution

a)



b)



S_0=(0,0)km t_0^p= 1.3 s

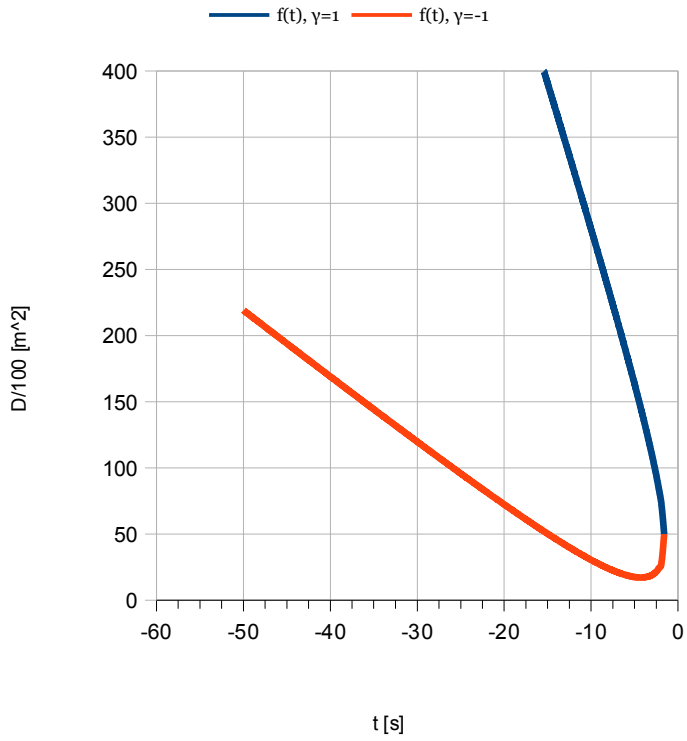
S_1=(15,100)km t_1^p= 1.7 s v_p= 10km/s

S_2=(100,-100)km t_2^p= 10.33 s

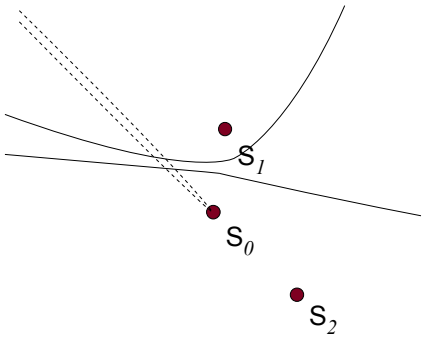
Fig. III.2.2:4 One solution: Panel a) shows the curve run of f(t) with one zero crossing. Panel b) shows the three hyperbolae constraining the corresponding unique location (yellow circle). The station positions (S_i, red circles) and arrival times are listed above.

No solution

a)



b)



S_0=(0,0)km t_0^p= 1.3 s

S_1=(15,100)km t_1^p= 2.0 s v_p= 10km/s

S_2=(100,-100)km t_2^p= 16.63 s

Fig. III.2.2:5 No solution. Panel a) shows the curve run of f(t) with no zero crossing. Panel b) shows two hyperbolae with no intersection. The dashed curve corresponds to the third hyperbola as would exist with slightly altered t^p. The station positions (S_i, red circles) and arrival times are listed above.

III.2.3 **Mapping Ambiguities**

Having obtained an analytical form to identify the maximum amount of possible solutions and their locations, we may now identify the boundary conditions that lead to ambiguous solutions and can map the regions of ambiguity. To do so, we will perform a 3dim. grid search varying the three P-arrival times for a given network geometry. The three dimensional parameter space can be reduced to a two dimensional one by expressing two of the arrival times relative to the third. Solving eq. (III.2.2-11) and determining D (eq. III.2.2-12) for all possible arrival time differences yields the fig. III.2.3:1 with the horizontal axes $\left(t_1^p - t_0^p\right)$, $\left(t_2^p - t_0^p\right)$ and the discriminant $D = א - ת$ assigned to the ordinate.

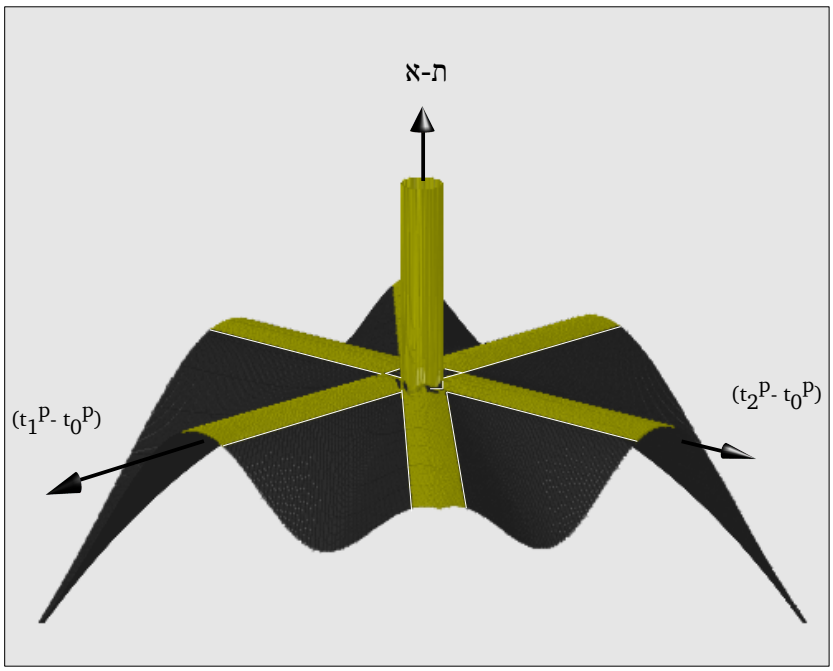


Fig. III.2.3:1: Mapping of D for all possible arrival time differences. Dark gray indicates regions with no solution (D<0); Yellow indicates regions of maximal two solutions (D>0). Both regions are separated by a fine line (white) of maximal one solution (D=0).

Areas of max. two solutions are highlighted yellow, areas of no solution in dark gray. These areas are separated by lines of maximal one solution (white).

The condition of plausibility demands that the arrival time difference between two stations cannot be greater than the longest physically possible travel-time (slowest path) from one station to the other. Requiring this condition for every station pair,

$$\left|t_1^p - t_0^p\right| \leq \left|\vec{s}_1 - \vec{s}_0\right| / v_p \quad \text{(III.2.3-1)}$$

$$\left|t_2^p - t_0^p\right| \leq \left|\vec{s}_2 - \vec{s}_0\right| / v_p \quad \text{(III.2.3-2)}$$

$$\left|t_2^p - t_1^p\right| \leq \left|\vec{s}_2 - \vec{s}_1\right| / v_p \quad \text{(III.2.3-3)}$$

with $\vec{s}_i$ standing for the three station coordinates and v_p for the corresponding seismic velocity

shrinks the valid area unto the central singularity of fig. III.2.3:1. Eq. (III.2.2-11) also disregards causality: It solves equally for events that have not yet happened (i.e. for an origin time after one of the involved P-arrivals) as for those that occurred in the past (i.e. an origin time before the first P-arrival). Filtering out a-causal solutions now yields the true number of solution (Fig III.2.3:2 d).

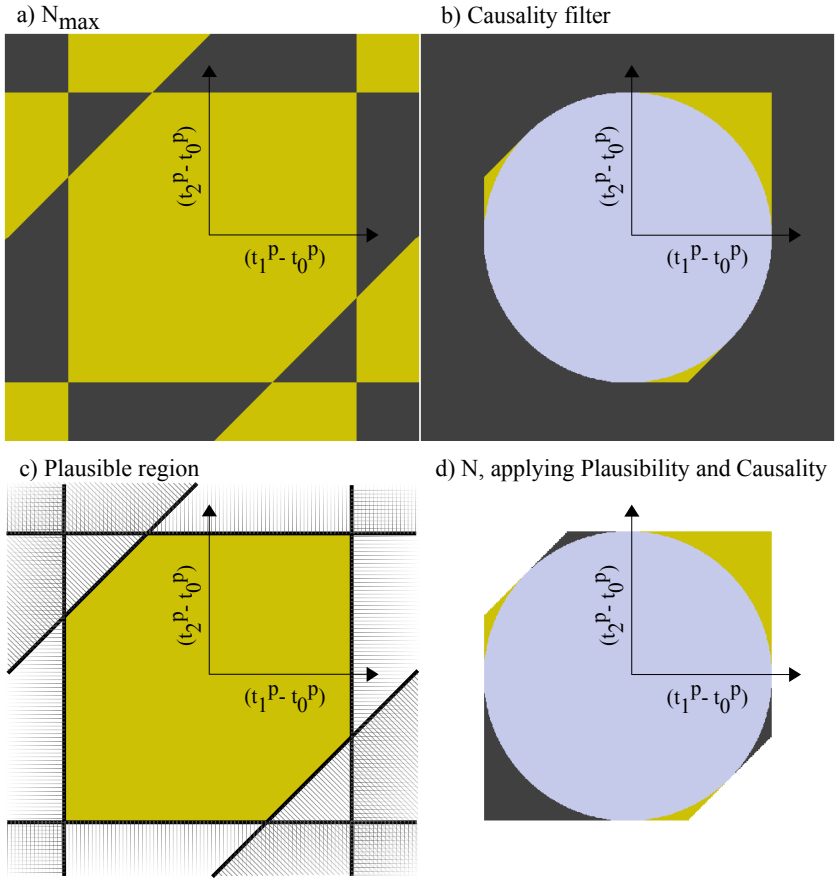


Fig. III.2.3:2: Top view of the central region of fig. III.2.3:1, mapping D. The Color coding denotes the amount of existing solutions for the given onset parameters: Yellow: Two solutions, light blue: One solution, Dark Gray: No Solution.

Panel a) shows the maximal amount of solutions based on the determinant D and corresponds to a top view of fig. III.2.3:1's central region.

b) Solely requiring causality reduces the amount of solutions in most regions.

c) Solely requiring plausibility masks the valid region close to the origin of the coordinate system (small arrival-time differences). Areas outside yield no solution: Areas with horizontal hatching are disqualified by violating eq. III.2.3-1. Areas with vertical hatching violate eq. III.2.3-2. Areas with diagonal hatching violate eq. III.2.3-3 and are therefore disqualified.

d) Requiring plausibility and causality yields the true number of solutions for the given arrival-time parameters. Even within the plausible parameter interval, there exist combinations of arrival-time differences that do not yield a solution (raising the question which physical principle is violated there).

Fig. III.2.3:2 (d) shows that the case of ambiguous or non-existing solutions is a boundary effect in the time domain, occurring only with extreme P-arrival time differences. Solving eq. (III.2.2-14), the corresponding locations can be plotted. Examining the solution in the space domain reveals that,

although ambiguity is a boundary effect in the time domain, significant parts in space are actually ambiguous (fig. III.2.3:3).

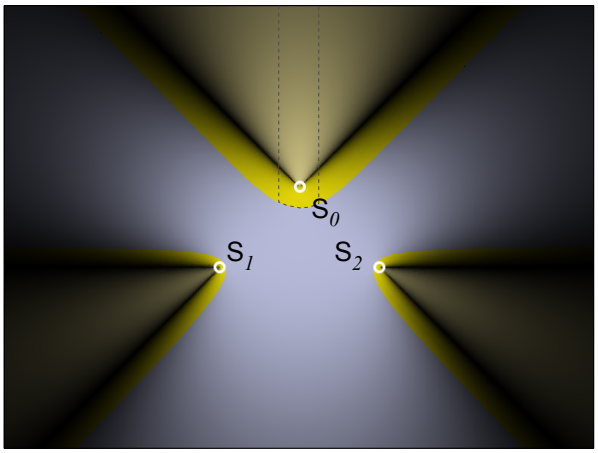


Fig. III.2.3:3: Space diagram marking regions with one solution (blue gray) and two solutions (yellow). For each point in a dark yellow zone a corresponding solution exists in the enclosed light yellow zone. The shading indicates the location precision: The brighter the color, the more precise the location. This distribution is valid for events with depth << station distance. For deeper events see fig. III.2.3:7.

Between 1/3 and the whole space are bound to ambiguities (depending on the network geometry): For every solution being found in a dark yellow area, a second one exists in the enclosed light yellow area, and vice versa. This displays an additional source of uncertainty, requiring additional information to identify the true solution. Let us now see how the two ambiguous solutions relate to each other. We will therefore look in detail into the dashed region of fig. III.2.2:3.

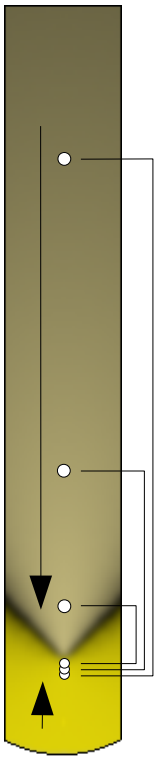


Fig. III.2.3:4: Zoom into the dashed slice of fig. III.2.3:3, showing three examples of sets of ambiguous locations (white circles). See text below.

Starting at $t_0^p = t_1^p = t_2^p = 0$, the solution is located at the center of fig. III.2.3:3. Decreasing t_0^p moves the solution vertically towards station S_0 and in the time diagram (fig. III.2.3:2d) towards the upper right corner. As soon as we enter the yellow area (fig. III.2.3:3 & III.2.3:4) the one location estimate splits into two. One solution will slowly continue to move toward the station while the other approaches fast from infinity (fig. III.2.3:4). When the upper right corner of fig. III.2.3:2d is reached, both solutions will join into one. With further growing onset differences, parameters leave the plausible space and solutions cease to exist.

Well separated ambiguous solutions can easily be distinguished and resolved by additional measures like S-P information (even if the S-onset can only be picked poorly) or particle motion, possibly pointing at the true solution. For solutions close to the border of the plausible region, however, the two solutions draw near to each other. This prevents a discrimination based on S-P-information or particle motion due to their lower precision, which practically means that we may observe two solutions and have to accept both (similar to the meta uncertainty in a multi-pick

analysis). Low accuracy zones are observed close to each station, extending outwards along the connection axes of two stations (fig. III.2.3:3 and III.2.3:5). This can be understood observing the ambiguous solutions in fig. III.2.3:5: For a small area around any given axis (red ellipse), all corresponding solutions (white-blue circles) are spread along the axis. An event having been found to lie close to one of the axes will therefore always appear to be blurred elongated to it.

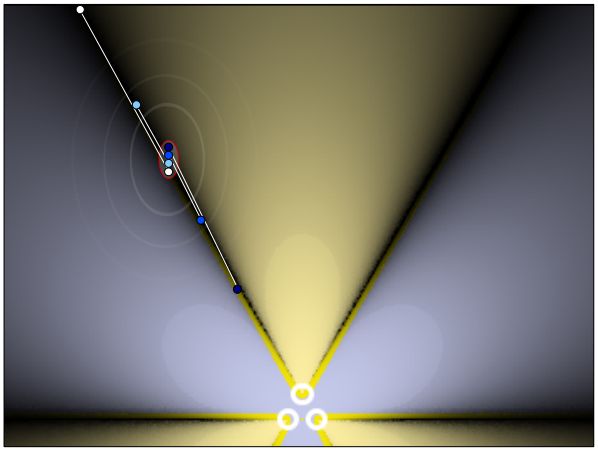


Fig. III.2.3:5: An event location with the uncertainty of the red ellipse has corresponding solutions along the whole axis. This causes locations close to the axes to be of higher uncertainty, appearing blurred along the axis.

The area of degeneration depends strongly on the geometry of the network layout, but is independent of the seismic velocity. The same is true for the relative position between two ambiguous solutions – A solution at location A will always have a sister-solution at point B, independent of the seismic velocity. The area enclosed by the connection axes (low accuracy zones) can hereby always be determined as degenerated (fig. III.2.3:6).

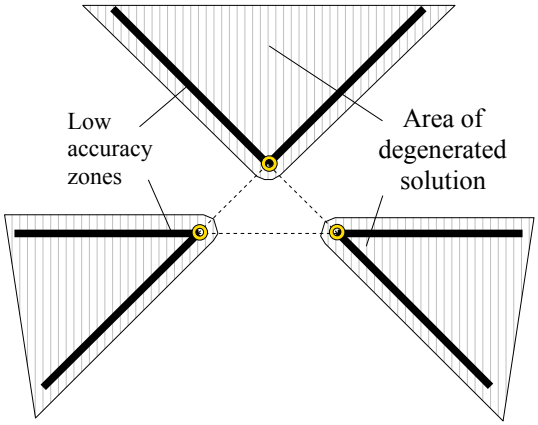


Fig. III.2.3:6: Ambiguous zones. Areas enclosed by the low accuracy axes are always degenerated.

It is therefore beneficial, if an event is located in between stations (e.g. the side of the triangle, the network edge), rather than being located close to a station (i.e. the corner of the triangle). This representation is valid for events whose depth is significantly smaller than the station distance. For deeper events, the cone-shaped regions distance themselves from the network while their boundary belts (bright yellow) widen.

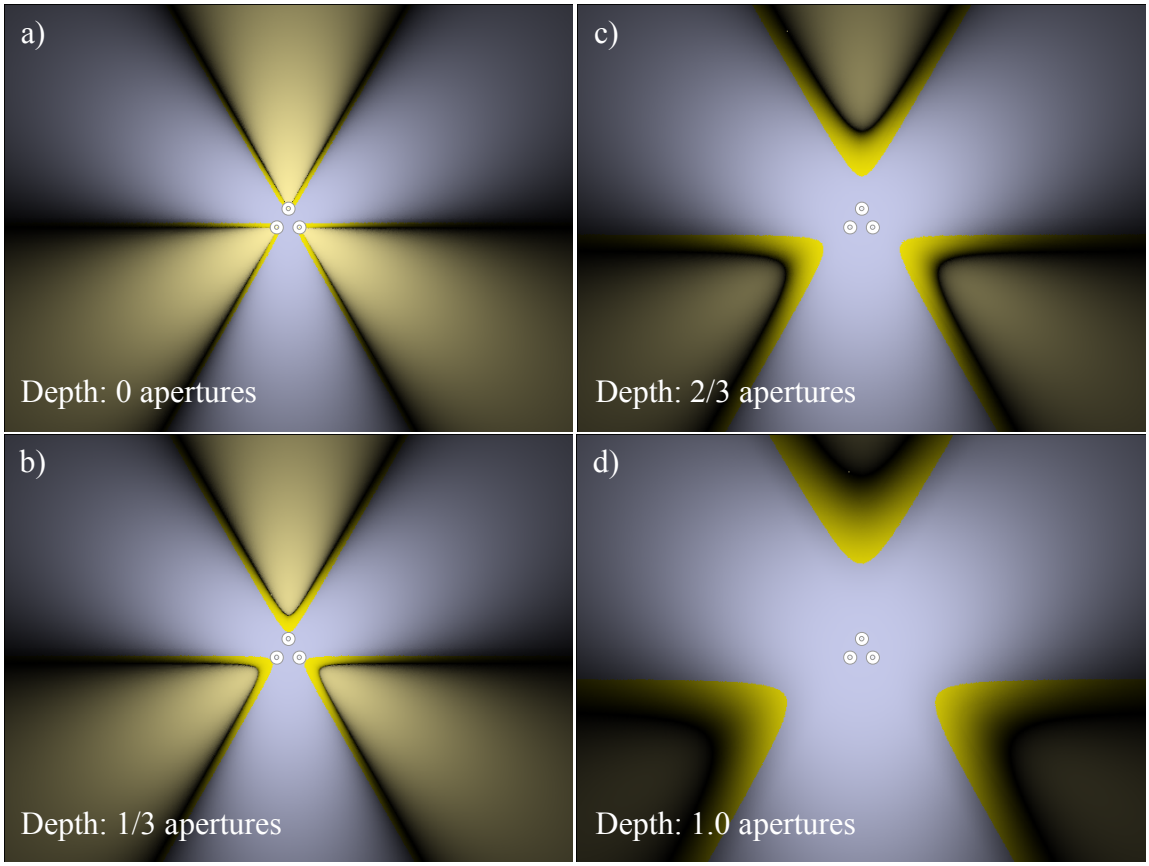


Fig. III.2.3:7: Ambiguous regions for different hypocenter depths relative to the network aperture. The deeper the event, the further is the ambiguous region located from the network and the thicker the (dark yellow) rim. The latter means that regionally, both locations are found in intermediate distances, while for the standard case (panel (a): station distance>seismogenic depth) the majority of locations within the light yellow region maps to secondary solutions very close to the station.

For asymmetric network layouts the three ambiguous regions hold different widths (fig. III.2.3:8) - the wider the region, the slower it distances itself with depth. In the following, we will look at different station distributions, transitioning from an ideal to a linear layout. Mapping both, spatial and temporal representation, we will see that it is not only the area between the low accuracy axes (which are always ambiguous), for which the arrival-time based location approach may yield ambiguous location estimates (fig. III.2.3:8).

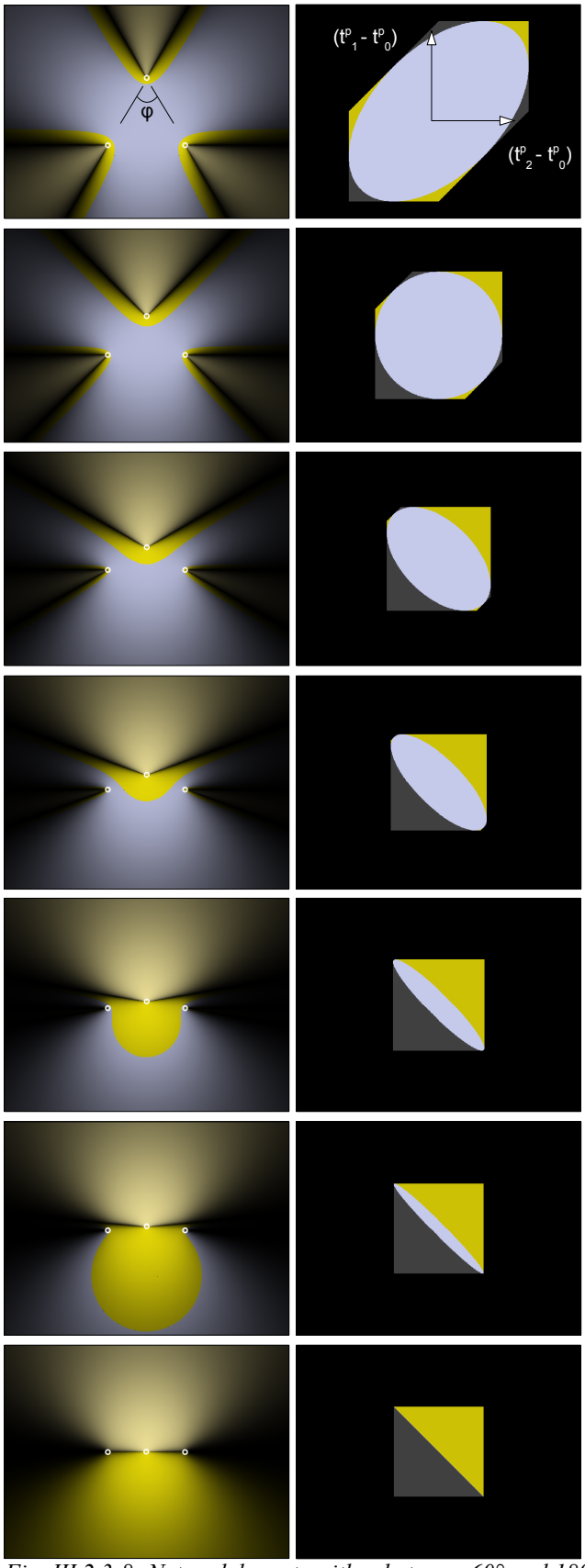
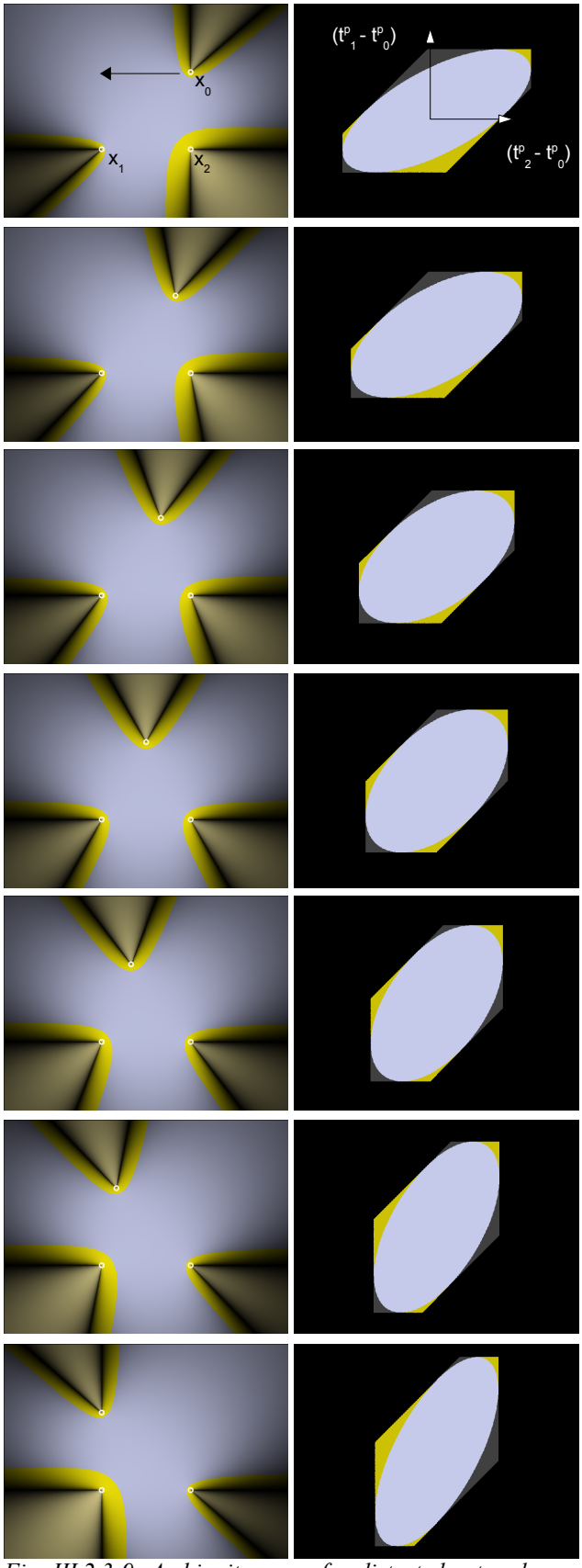


Fig. III.2.3:8: Network layouts with φ between 60° and 180°. Left: Spatial representation, Right: Temporal representation (black indicates the non-plausible parameter space).

In an uniformly distributed layout, ambiguous solutions only appear for extreme arrival-time differences. With φ increasing, ambiguities start to occur also for smaller differences. Likewise does the ambiguous region in the space domain (only) occupy about one third of space for an uniformly distributed layout, but quickly increases to full space when the station distribution reaches the resemblance of a linear layout. During this transition, the ambiguous region bulges into the network and at φ=140° already fills most of it. The temporal representation shows that for flatter network geometries (larger φ's) ambiguous solutions start to occur already with smaller onset time differences. For a linear layout the zone of unique solutions has disappeared and every arrival-time difference yields either an ambiguous solution or none. In the following transition (fig. III.2.3:9) the network geometry is distorted asymmetrically, shifting the north station on a East-West axis. For well distributed stations, the low accuracy axes (constructed following the scheme of fig. III.2.3:6) enclose the majority of ambiguous solutions. Flatter geometries (φ=120°, fig. III.2.3:8) widen the band around this zone.



$$x_0 = \frac{x_2 + x_1}{2} + D \cdot (x_2 - x_1)$$

D= +0.50

D= +0.33

D= +0.17

D= 0.0

D= -0.17

D= -0.33

D= -0.50

Fig. III.2.3:9: Ambiguity zones for distorted network geometries. The horizontal position x_0 of the upper station is moved from x_2 to x_1.

The opening angle of the cone-shaped ambiguous regions varies over the transition. It's tip is always pointing toward areas that are of unique solutions and the larger the region of ambiguity is, the larger is the region of unique solutions on the opposite side.
The effect of location ambiguities stresses the importance of global search methods. Evaluating the solution space solely at sample points (in iterative methods) or grid points (in direct searches) risks to identify only one of both solutions, possibly basing the event interpretation on the wrong location. Due to the larger spatial extent of the distant solution, algorithms will also tend to iterate toward that solution, if not prevented by an initial trial location close to the near-by solution (see fig. III.2:1). The analyzed case in 2-space-dimensions is a typical case for Earthquake Early Warning systems, which attempt to locate based on a minimum amount of stations, therefore initially even fixing the hypocentral depth (to the dominant depth of the catalog). In sparse networks the early location estimate may remain the official estimate over a significant time. This renders the topic of ambiguities so important, especially since no (known) algorithm today attempts to resolve the ambiguity but only chooses the most attractive (but possibly false) one. We will conclude the topic of hyperbolic ambiguities with mapping the ambiguous regions for the three-(space-)dimensional case with 4 given stations, tackling a standard scenario: locating low magnitude events.

III.2.4 **Ambiguities in 3D Environments**

To avoid ambiguities and ensure unique solutions in two-(space-)dimensional problems (e.g. a known focal depth or shallow events in hill slope monitoring with planar station distributions), it requires four given stations. The general situation is, however, that one needs to deal with unknown hypocenter depths and at times non-planar network geometries (e.g. surface stations on strong topography, borehole stations or subsurface mine monitoring). In good natured three (space-)dimensional scenarios the hypocenter is fixed with four stations. However, this is not the general case: For a one dimensional space (i.e. a line) it requires three stations to rule-out ambiguous solutions. For a two dimensional space (i.e. a plane), four stations are required. Following this pattern confirms what Xiaochun (2008) showed formally: that 5 stations are required to avoid ambiguities in three space dimensions (given that they don't lie on a common hyperboloid). For weak events, however, this amount of stations might not always be given, e.g. for a four-partite

small-array detecting a local event. In this case, it is useful to know which regions are prone to ambiguities and if the location estimate is trustworthy. Starting with a three-dimensional station distribution in full space, a grid search identifying ambiguous solutions (fig. III.2.4:1) reveals a pattern similar to the two dimensional case (fig. III.2.3:3).

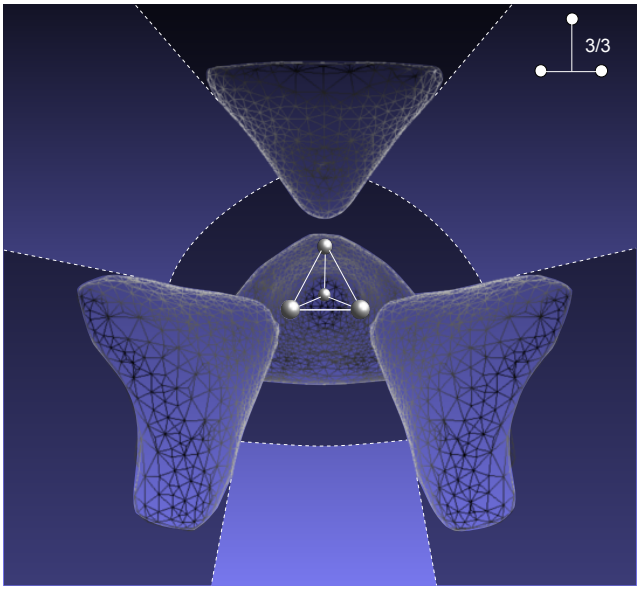


Fig. III.2.4:1: Equally distributed 4-station setup in full space. Within the four cones, solutions are ambiguous. The cones are pointing to the geometrical center of the four stations. Dark blue areas mark extrapolated ambiguous areas (white dashed lines).

Fig. III.2.4:1 shows a symmetrically distributed set of four stations with ambiguous cone-shaped regions, similar to fig. III.2.3:3. The four stations form a tetrahedron whose corners likewise point to regions of ambiguity while its faces are oriented to regions of unique solutions. The four cones are pointing to the geometrical center of the station distribution and are symmetrical, similar to the 2dim. case. This changes, when we distort the network geometry: When lowering the top station, the cone of the approaching station widens while the others narrow, collectively creating a larger region of ambiguity. Fig. III.2.4:2 displays this scenario, with the top station being lowered to two thirds of it's original height.

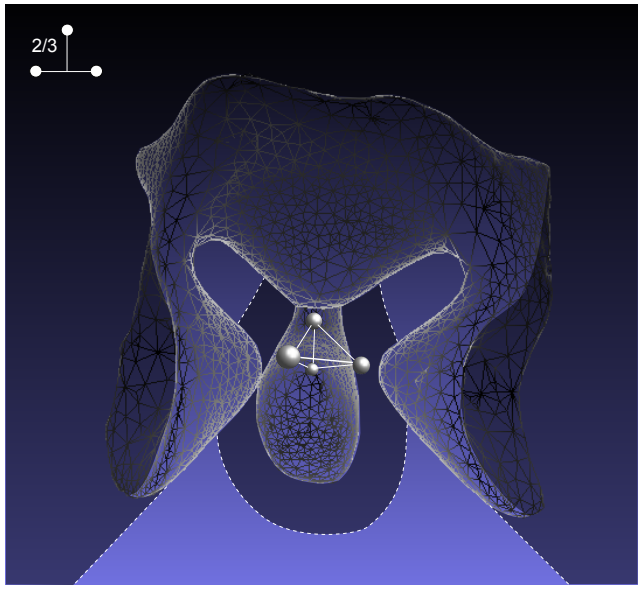


Fig. III.2.4:2: The center station is lowered to two thirds of the initial height of Fig. III.2.4:4. Similar to the 2dim. case, the widening of the upper ambiguity cone can be observed.

If we continue to lower the station, the three dimensional geometry will eventually collapse into a two dimensional one (once the central station reaches the plane of the other three stations): describing a planar network. Alike the end-member of the 2dim. case, this will render all solutions to be ambiguous - the ambiguous region filling the full space. However, like for the 2dim. case, also here the corresponding solutions are symmetrical - mirrored through the plane containing the stations.

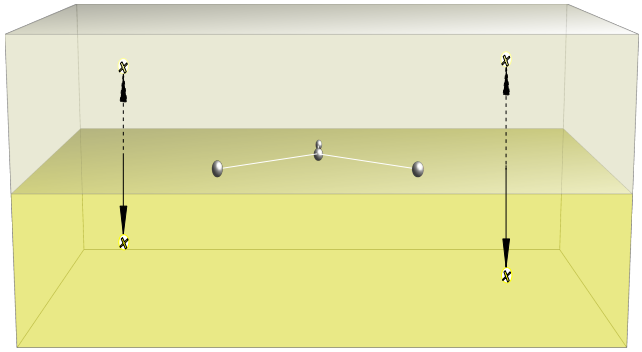


Fig. III.2.4:3: All solutions become ambiguous when the 3dim. station setup collapses into a 2dim. plane, the secondary solution being mirrored to the other side of the plane. For the wide class of applications where stations are located on surface and no strong topography is involved, the secondary solution will always be located above surface (but not mirrored due to different media and velocities) and can be neglected. However, this also means that a sonic bang can appear as shallow earthquake.

This means for the wide class of applications that use surface networks (with no significant topography), that the secondary solution will always be located above surface (although not "mirrored" due to media with different velocities) and can hence be neglected. In this case, four stations are sufficient to always yield unique location estimates. [Note: On the other side it also means that acoustic sources (as jets or thunderstorms) can be "projected" into the subsurface, appearing as shallow seismic events (shallower, due to the higher seismic wave speed). Practically, this is only problematic for high sources directly above the network, or for networks on e.g. sedimentary basins, in which the seismic velocity is only a few times the acoustic velocity, as otherwise the large arrival-time differences would be inconsistent with the seismic model, and could not be matched.]

Given a stronger topography or a subsurface three-dimensional station geometry, four stations are likely do yield ambiguities. The extent of the problem can easily be visualized by turning fig. III.2.4:2 upside down (fig. III.2.4:4). This resembles the geometry of a four-partite array, where the central station is installed in a borehole to yield a better SNR and detection quality: But the majority of solutions situated in the space right underneath the network would be ambiguous (while the open space above would yield unique solutions).

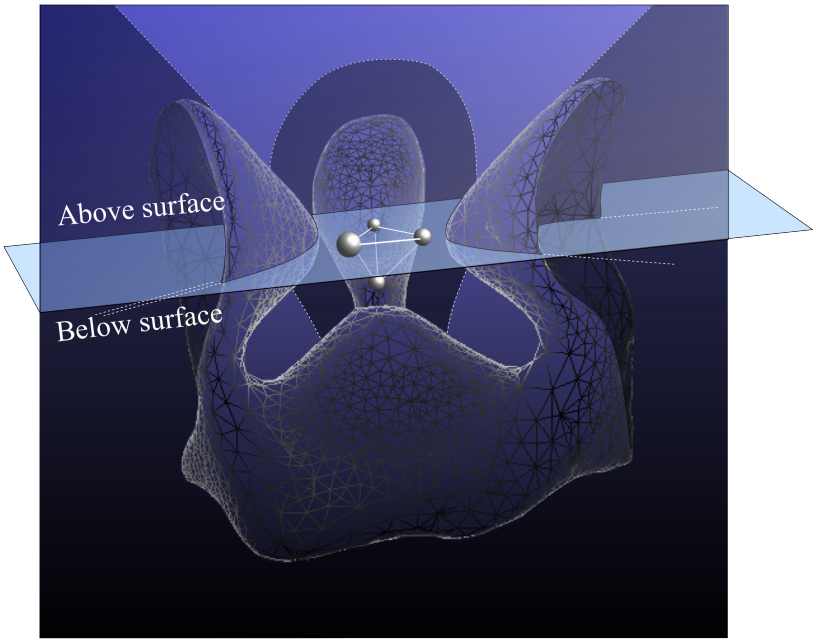


Fig. III.2.4:4: A central borehole station added to three surface stations in a homogeneous half-space. The majority of locations derived would be ambiguous.

This forms a dilemma, as a sunk central station would indeed allow to lower the detection threshold,

but unfortunately also decreases the location accuracy. For three dimensional distributions (as in mine monitoring), the distribution should be oriented to face (i.e. not point at) the focus area to provide there for unique solutions.

We can conclude the following concerning a location that is based on the arrival-times of a single phase type:

For the amount of stations that is generally described as the "required minimum" (three in 2D, four in 3D) we find a significant part being ambiguous. In 2D, this is the case for at least one third of all solutions. These ambiguous zones do not only lie outside of the network, but also bulge into it and grow in size for flatter station distributions. To solely obtain unique solutions in two dimensional problems four stations are required, which must not lie on a common hyperbola. In the general case of three dimensional problems, at least five stations are required (not lying on a hyperboloid). However, for the wide class of applications where the station setup is planar on surface and no strong topography is involved, four stations suffice. When two distinct solutions are found, the ambiguity can usually be resolved using S-P distance- or particle motion back-azimuth information. Additionally, if the stations are part of a larger network and the event is solely detected on a small subset because of it's low magnitude, then the distant solution can often be neglected, as it should have been detected at other network stations first (→ Voronoi cells).

When only three or four stations can be deployed due to high effort, they should be oriented facing (i.e. not pointing at) the region of interest to cover the focus area with zones of unique solutions. Signals from regions outside the network being "pointed at" by the network (fig. III.2.3:6) will always lead to location ambiguities. It should be noted that this two-fold solution was obtained analyzing a homogeneous half space. In case of refractions caused by e.g. a layered subsurface, the amount of ambiguous locations might even scale up with the number of refracting layers. This can be inferred from observations as in fig. III.2.4:5, in which a solution (Point 1) appears with two other solutions (Point 2 & 3) although all lie in a region that can only yield unique solutions (compare fig. III.2.3:6). The other two solutions stem from faster travel-paths running through deeper regions of higher velocity in a two-layer (over half-space) model. With every existing refracting layer-boundary another solution is generated in larger distance. For naturally ambiguous solutions this limits the total amount of solutions per set of arrival-times to $2 \cdot n$, with n being the amount of refractors.

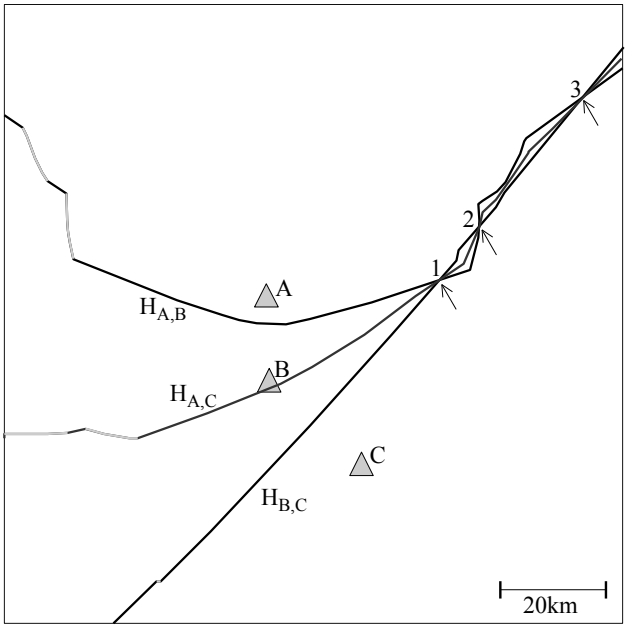


Fig. III.2.4:5: Multiple ambiguities for a single triple-group caused by two refractors at 10.6km and 15.5km depth increasing the near surface velocity of 3.5km/s to 5.7km/s and 6.4km/s. The changes in the curvature (highlighted in grey on the left side) of the three hyperbolas ($H_{A,B}$, $H_{A,C}$, $H_{A,B}$) are caused by the velocity changes around the refracting layer boundaries and mark the regions in which one travel-path-type changes to another (direct to once-refracted and once-refracted to twice-refracted) - corresponding to the points in the travel-time curve whose first derivation is discontinuous. These changes in curvature provide for additional crossing-points i.e. for the existence of additional ambiguous solutions. Through the existence of the refractors the amount of solutions exceeds the limit of 2 (which was valid for homogeneous models). Here, all solutions are located in a region that is classified as zone of unique solutions. The two solutions (2 & 3) are merely "echoes" of solution 1, due to additional travel-paths that move in higher velocities below the refractors, allowing constraint agreement in larger distances.

When only few stations are given one needs to be aware of ambiguities to continue the location process until the inapplicable solution may be disqualified and excluded. Even more so for (automatic) algorithms, which need to be laid out to expect, (possibly) resolve - but in any case correctly evaluate these equally possible location candidates. Especially in earthquake early warning this is crucial, as only identifying one of both locations may cause major estimate errors resulting in false or neglected warnings.

IV **Constraints for Real-Time Location**

The term “real time location” refers to the evolving and subsequent constraining of hypocentral location from the time of first detection onward, long before the complete set of information (i.e. the complete waveform data of all stations) is received. As at this time information is sparse, the early location estimate is prone to ambiguities; not only in location, but also in detection as only fractions of seconds of waveform data are available as base for the detector, raising the risk for onsets to be wrongly picked or wrongly associated by the event associator. All these aspects call for a robust locator, making the derived constraint based approach well suited. Especially concepts like the multi-pick evaluation become useful here, to increase robustness when multiple pick candidates exist. Starting with rough estimates at the time of first detection, the solution is subsequently refined with every arrival of new phases - or by the information that can be extracted from the delaying of such. A real-time analysis is generally applied to improve reaction time: Time sensitive issues as e.g. pressure control in fracking or gas storage processes require a real-time analysis due to the risk of induced seismicity. Real-time location, however, in it's extreme form is applied in Earthquake Early Warning systems (EEW), where location and magnitude estimates are required in matter of seconds to improve warning time and decrease the blind zone (i.e. the region surrounding the hypocenter for which no alert can be issued in time), before the later S and (most destructive) surface waves arrive at population centers. As the fault plane of large magnitude events grows over several- to tens of seconds, early magnitude estimates usually saturate at M_L 6.0 and can only serve as minimum magnitude estimate. This, however, is sufficient for warning as the magnitude threshold for warning is usually fixed to M_L 5.0. The process between detection, event association, and finally parameter (i.e. location and magnitude) estimation is complex and this chapter will focus solely on the aspect of real-time hypocenter location. Inevitably there exists a trade off between accuracy and stability on the one side and time available for location on the other. The more time is given, the more information can be gathered, the range of applicable methodologies grows, gradually increasing the precision of early estimates of e.g. event location and magnitude, eventually later focal mechanism or extended source and propagation history. Early in time, most information is not yet available and predictions might turn out inaccurate or wrong. Todays running EEW codes usually start after 4-6 detections and under- (or over-) estimate epicentral distance and magnitude in the initial stadium. Accounting for uncertainties, especially of the one of the model, is

of major importance for real-time location, especially for events close to- or outside the network edges. Using the uncertainty based constraint approach enables us to report valid uncertainty intervals for the reported parameters, yielding e.g. minimal and maximal hypocentral distances or min. Magnitude, rather than over- or underestimates.

Good predictions require a set of reliable, stable and early available information on which the automated processing can be based on. Most automated systems rely on the analysis of several arrived P-phases, due to the accuracy and confidence with which these can be automatically identified. However, all fundamental constraints presented in earlier chapters are useful for real time location, constraining the hypocentral region further once the corresponding phases are detected. Equal Differential Time (EDT) P^1-P^2 constraints can precisely constrain the location, but might only become available with time, as they require the detection of at least a second station for generating the constraint, and at least three to four detections to yield unique solutions (also using focal depth estimates). Depending on the event-station geometry and network density this might require time. Some constraints are of special importance as they require few detections, yet provide robust location estimates (fig. IV:1): Back-azimuth information can be retrieved from the first detection on and allows to resolve ambiguities that usually occur in the early stage. It hence stabilizes the early location with fewer stations and allows to shorten the warning time. S-P constraints provide the important estimate of hypocentral distance. Yet, S-P constraints require the A) more complex and error-prone detection of the S phase which B) only arrives later in time, depending on the epicentral distance. We will therefore introduce using the absence of the S-phase to derive lower bounds for the hypocentral distance (chapter IV.2). The delaying of P-phases at other stations, on the other side, provides the estimate of the maximum distance. These three methods are applicable instantaneously with the first detection and reduce the possible solution interval with time. Fig. IV:1 displays the four constraint classes, that will be used for the real-time scheme.

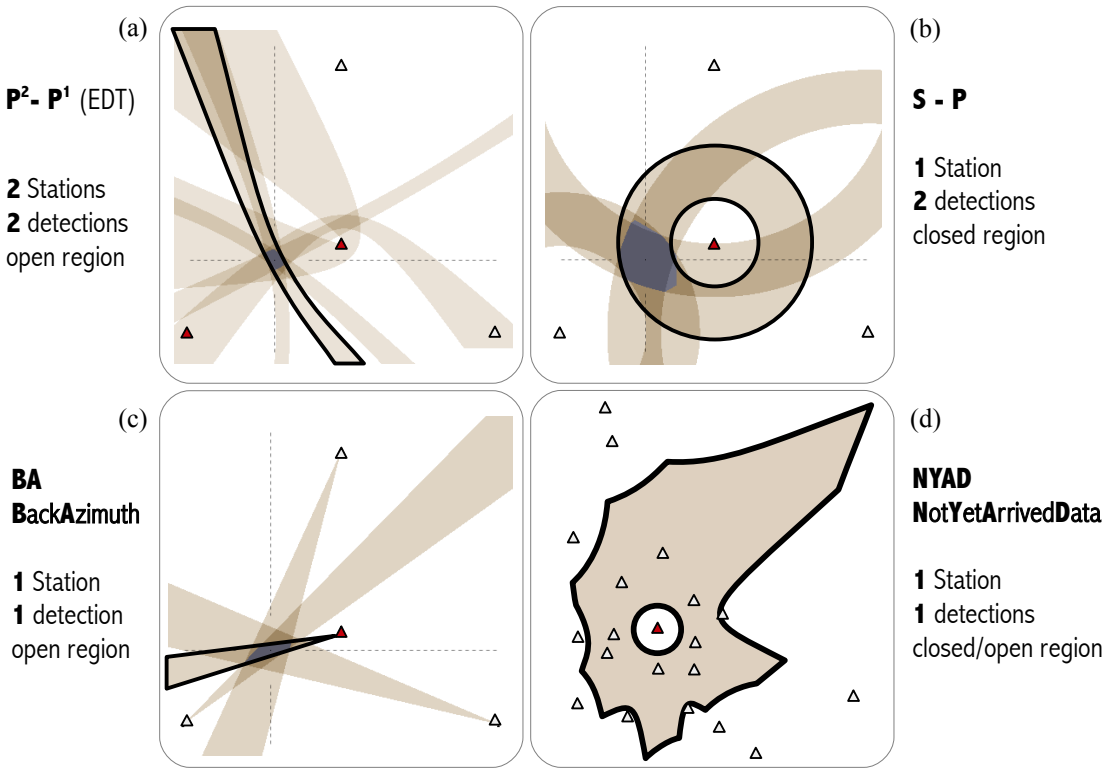


Fig. IV:1: Realtime location constraints. Comparison of amount of detections, detected stations and constraint characteristics. Back-azimuth and NYAD constraints can be used from the first detection on.

It is not only due to location issues that EEW systems require detections at 4-6 stations before raising an alarm: Usually, at least two or three detections at different sites are required to gain the confidence that the signal belongs to a seismic event and is not of anthropogenic (or e.g. acoustic) origin. Here, small-aperture-arrays present a powerful solution to increase confidence with fewer detections, due to their ability to characterize and evaluate the signal. Small aperture arrays do not solely measure at a single point but -consisting of distributed stations- are able to sample the local wavefront, which allows for quality checks based on slowness, inter-station amplitude ratios and signal coherency. For location, they yield the valuable back-azimuth (BA) information, which in combination with Not-Yet-Arrived-Data concepts (i.e. shrinking “Voronoi cells” and growing No-S exclusion zones, chapter) allow initial location estimates in fractions of seconds that improve those of today's real-time location schemes (like e.g. PRESTO, ELARM-S) (Satriano et al., 2011, Kuyuk et al., 2014). Small-aperture-arrays are therefore ideal measurement units for hybrid onsite-regional systems, which function on the local level as individual and independent warning systems (e.g. at nuclear reactors, factories), but further report their data instantaneously to real-time-analysis

centers, where the data from all arrays is combined for regional wide EEW. However, the vast majority of networks today consists only of distributed single (three-component) sensors. There have been attempts early on to derive back-azimuth information from the three-dimensional particle motion recorded on single stations (Ruud et al., 1988), based on the fact that the seismic phases propagate in a polarized manner. These approaches, however, were not reliable enough for EEW as they often yielded inaccurate results. However, since back-azimuth information is important to resolve the early ambiguities, a new single station approach, yielding trustworthy estimates, will be derived in the following chapter.

IV.1 **Particle Motion Back-Azimuth**

While most standard real-time location schemes today rely solely on P-phase based location, providing first proper location estimates after 4-6 detections, the addition of back-azimuth information stabilizes the early location and allows to obtain robust results already after two detections. Below four detections the hypocentral depth is usually fixed to e.g. 10km depth. With three available P-phases the solution is often ambiguous and the three EDT's pose a dependent system. This means that for all physically possible inter-station time differences a solution is provided. Without redundancy, there exists no quality control: Even three uncorrelated random picks in a narrow time window, may produce a reasonable looking location. Using back-azimuth information, already after two detections three independent constraints are available, two back-azimuth beams and one EDT. This allows a quality control based on constraint congruity, showing whether constraints agree or not. Fig. IV.1:1 shows an early location example, comparing P-based hyperbolic location and back-azimuth constraints. The hyperbolic location shows two possible locations after three detections, one closer and one more distant to a main population center. What the figure doesn't show, is that a small change in the modeled velocity will shift the locations, with the three hyperbolic constraints still intersecting with the same maximum congruency: In this early stage (where no redundancy exists), congruency is meaningless. The picture is fundamentally different when we use back-azimuth information: Not only is the ambiguity resolved (clearly identifying the distant region as solution), also three independent constraints are given, which allow to use the constraint congruity as parameter indicating the trustworthiness of the result – both after only two detections.

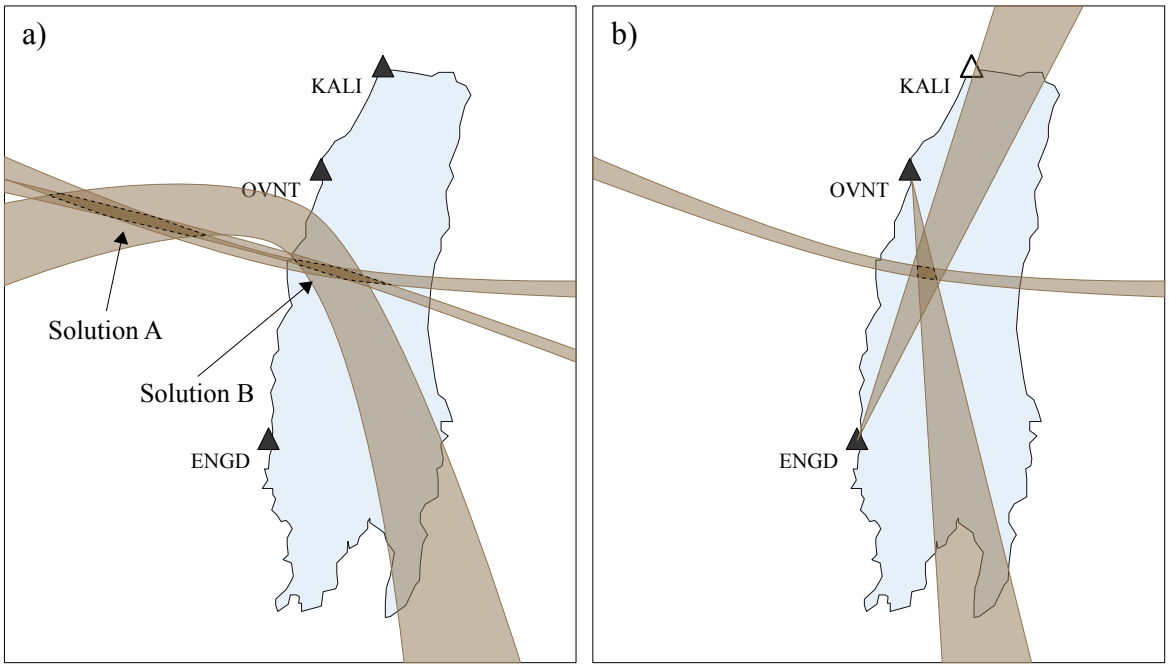


Fig. IV.1:1: Analyzing three P-arrivals a) based on P-onsets only: two solutions exist, given by three dependent constraints (EDTs) b) Implementing back-azimuth information, a unique solution is given by three independent constraints already with two detections (BA).

Ruud et al. showed 1988 that this information may also be obtained analyzing the particle motion of the incoming P-wave recorded at three component single stations. Contemporaneously, Nakamura (1988) developed an automatic algorithm computing single station back azimuths in real time. Yet, Lockman and Allen (2005) pointed out the significant amount of estimates with errors too high to still render this method reliable for predictions in earthquake early warning applications. Alessandrini et al. (1994) demonstrated that the data yielding best results is found at the onset of the first P-wave arrival, while later parts of the signal are not linearly polarized due to effects of local scattering at heterogeneities. Later signals might further not travel on direct paths, may be of different wave type and therefore yield incorrect azimuth values. Accordingly, Noda et al. (2012) found that the accuracy can be increased by limiting the waveform data to the signal between detection and first maximum on the 1-2Hz bandpass filtered vertical trace. This fact would allow an almost instantaneous constraining of location as the required data is usually obtained in 0.2-0.3 seconds. Yet, Lockman and Allen's observation remained confirmed, as 40% of estimate errors still ranged between 30 and 180 degrees. Eisermann et al. (2015) shows, that bad estimates can be filtered out by requiring congruity between different back-azimuth estimators, leaving reliable back azimuth estimates for automatic real time location. The methods being used range between approaches being based on a single sample, an evolving moving average, and a principal components analysis.

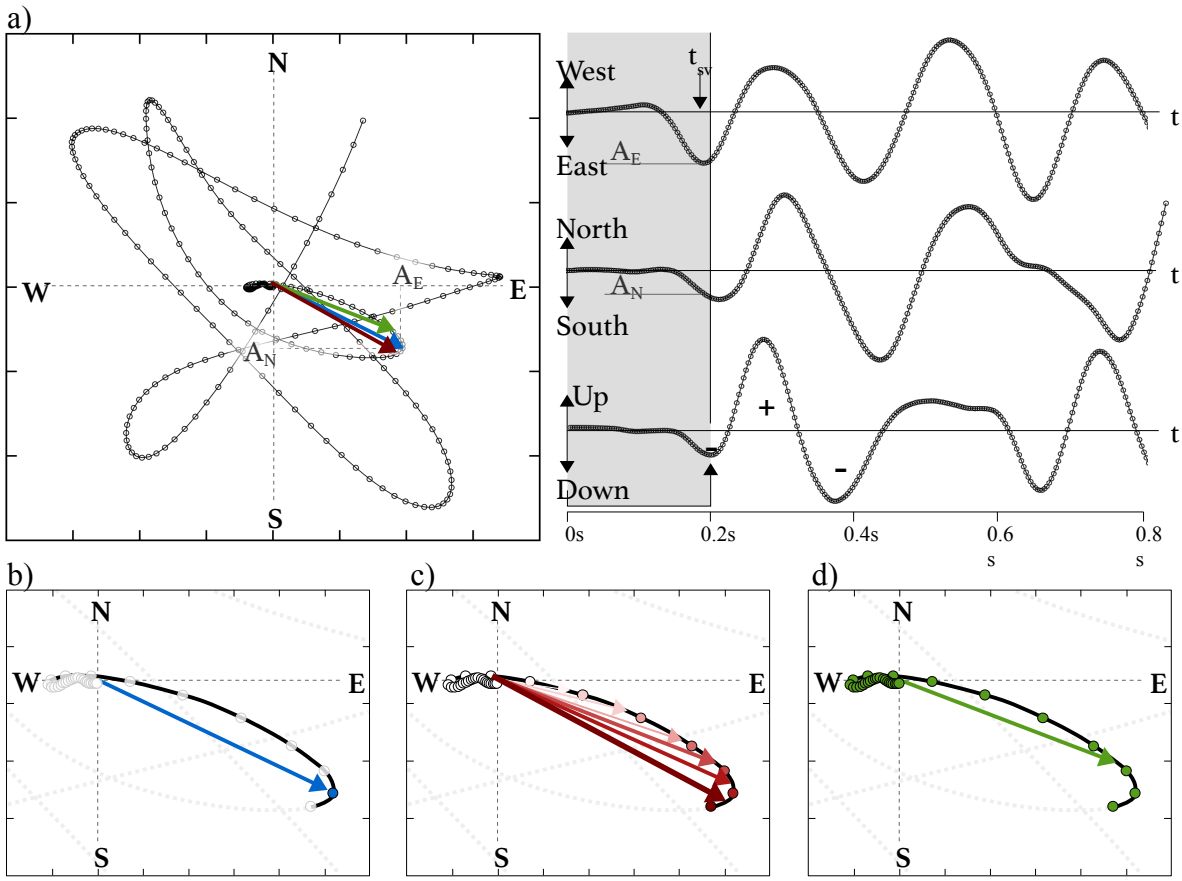


Fig. IV.1:2: a) Horizontal projection of particle Motion and projection into EW-, NS, and Vertical plane (right). The gray zone marks the variable analysis window defined by detection and first significant maximum on the vertical trace. b) Single Value method: The particle motion direction is obtained from the values at the one sample of highest horizontal amplitude. The sign of the vertical velocity amplitude solves the directivity ambiguity. c) Moving average method: The direction is evaluated at each sample and averaged with previous estimates in a way that the latest samples receive the highest weight. d) A two-dimensional Principal Component Analysis is used to analyze the trend of direction.

While the overall performance of each method is good, each method produces a significant number of wrong estimates. However, robust and accurate estimates tend to show similar values across the methods, while outlier values tend to fluctuate. This allows to use consistency between the methods as a filter to eliminate bad estimates. The three methods shall be qualitatively laid out in the following (quantitative details are described in appendix 1).

Method 1: Single Value (SV). The direct approach for the extraction of the back azimuth from particle motion relies on the peak amplitude values of the horizontal traces (Bormann, 2012), and hence will be identified in the following as peak single value (SV) method. The back-azimuth Φ is

obtained from the NS-EW amplitude ratio at one of the maxima of the horizontal traces. The time window for evaluation will be defined according to Noda et al. (2012) from detection of the P-phase to the first maximum of the vertical trace. The amplitude values A_E (NS-trace) and A_N (EW-trace) are chosen at the point where one of the two traces reaches the most significant maximum (fig. IV.1:2, see t_{sv}). For physical consistency, both amplitude values need to be estimated at the same sample $\bar{i}$ on both NS and EW traces. The back-azimuth, computed from the horizontal amplitudes, is geometrically bound to a 180 degree (forward or backward) ambiguity. This ambiguity can be resolved looking at the vertical amplitude using the fact that the corresponding particle motion vector cannot point outside the lower half space, as the event has to lie below surface. If the vector does, it needs to be flipped to point toward the event. This flipping condition pertains to p-waves and is applicable for surface stations in absence of strong topography. It is valid, independent of the source mechanism (strike slip, explosions and implosions) and its radiation pattern, which determines the signal polarizations. Although it is solely based on the amplitude values of one single sample, the SV approach delivers robust back-azimuth estimates with few outliers for high SNR (fig. IV.1:5), but outlier population increases with decreasing SNR.

Method 2: Moving average (MA). The approach proposed by Nakamura (1988) can be described as a continuously applied form of the SV approach. Back-azimuth values are computed at every sample, derived from a cumulative smoothed average over the horizontal traces. The flipping condition is applied at every sample. Noda et al. (2012) showed that for PCA-based back-azimuth estimations, accuracy can be improved by using a variable time window ending at the first maximum of the vertical trace (Noda et al., 2012). Applying this dynamic window, also the estimates of this MA approach improves: Figure IV.1:3 displays how the realtime back-azimuth error decreases with shorter data window lengths [3s and 0.6s, dynamic window]. This is advantageous for real-time EEW, as shortest time intervals of on average 0.25s yield the most accurate results. While the original approach averages over all smoothed sample-wise back-azimuth estimates, applying Noda's modification, the last estimate of the moving average at the maximum of the vertical trace was found to yield the highest accuracy.

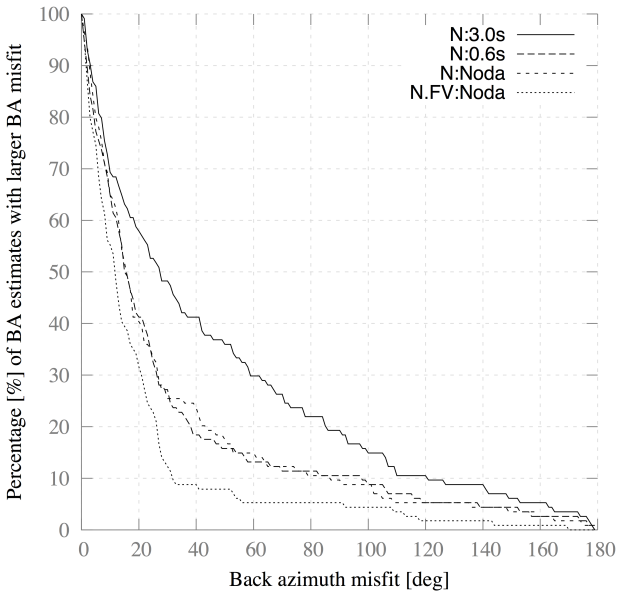


Fig. IV.1:3 Cumulative discrepancy for MA (N) estimates for three different time windows (3s, 0.6s, dynamic window according to Noda). The accuracy of estimates increases with shorter time windows. Using the last estimate (N.FV) at the maxima of the vertical trace yields a higher accuracy than the standard averaging approach.

For low magnitude (M<4.0) events, a significant improvement of back-azimuth estimates is obtained by applying a muting condition for samples with amplitudes under noise level,

$$S_{hor} < N_{hor} \ \vee \ S_{vert} < N_{vert} \,, \qquad \text{(IV.1-1)}$$

avoiding the chaotic influence of noise. The independent condition for the vertical SNR is important to avoid large errors at zero crossings: The flipping condition is contingent on the sign of the vertical amplitude. It is hence prone to errors in proximity of zero crossings, being able to falsify the back-azimuth value by 180 degrees, if erroneously applied. Zones of small Z-amplitudes, which are especially sensitive to sign flipping (e.g. even solely through the windowed offset correction) are therefore excluded from the analysis.

Method 3: Analysis using Principal Component Analysis (PCA). A principal component analysis (PCA) is an eigenvalue decomposition that transforms the data into a new coordinate system, such that the data on the first axis holds the largest variance. Subsequent orthogonal components are computed in sequence of decreasing variance. Since the P-polarized particle motion should hold the largest variance on the axis pointing towards the hypocenter, the axis of the back-azimuth can be

obtained from the the orientation of the first Eigenvector. A two-dimensional PCA was performed on the horizontal data, analyzing all non-muted samples (eq. IV.1-1) of the N-S and E-W traces within the time window selected according to Noda et al. (2012). As the PCA only determines the axis, but not the direction, there is a 180° ambiguity in direction that can be resolved by evaluating the segment-wise direction at all samples to find the dominate direction of the particle motion on this axis (a "segment" being defined as the vector between two subsequent samples of the horizontal data). The elementary flipping condition (using the vertical amplitudes) is thereby applied at every segment.

To obtain good results, signal samples used should lie significantly above noise level in order to limit the arbitrary influence of noise on the back-azimuth estimates. Analysis of a South California data set has shown that most outliers indeed hold a SNR smaller than 10. The amount of bad back-azimuth estimates reduces significantly, when only data is analyzed that hold a SNR_{vert}>10 and SNR_{hor}>10. Signals of interest for EEW originate usually from nearby locations, have stronger amplitudes and thus should naturally fulfill these conditions. The three optimized back-azimuth calculation techniques (SV, MA and PCA) provide similar levels of accuracy (fig. IV.1:4) while using different computation techniques with individual flipping conditions. Each estimator produces outliers, the values, however, tend to vary between the methods. With many signals the three waveform components initially oscillate coherently but later loose coherence. In other cases they gain coherency after initial fluctuations, emerging from the noise. In the first case SV is expected to yield more accurate back azimuth estimates, in the latter case MA should do. This can be understood, remembering that SV delivers an estimate based on the amplitude ratio of one single sample only which usually lies before the end of the interval. MA averages over the full window, but gives more weight to the latest samples at the interval's end. PCA weights each sample of the interval equally. Similar estimates of all three methods indicate therefore lasting coherent waveform oscillations, providing for trustworthy estimates. This can be used to construct a filter, based on the similarity of the three estimates, only trusting the result when estimates are similar within a certain error margin: Requiring the three independent methods to agree within 4° filters out all outliers of the dataset. (fig. IV.1:4, bottom panel).

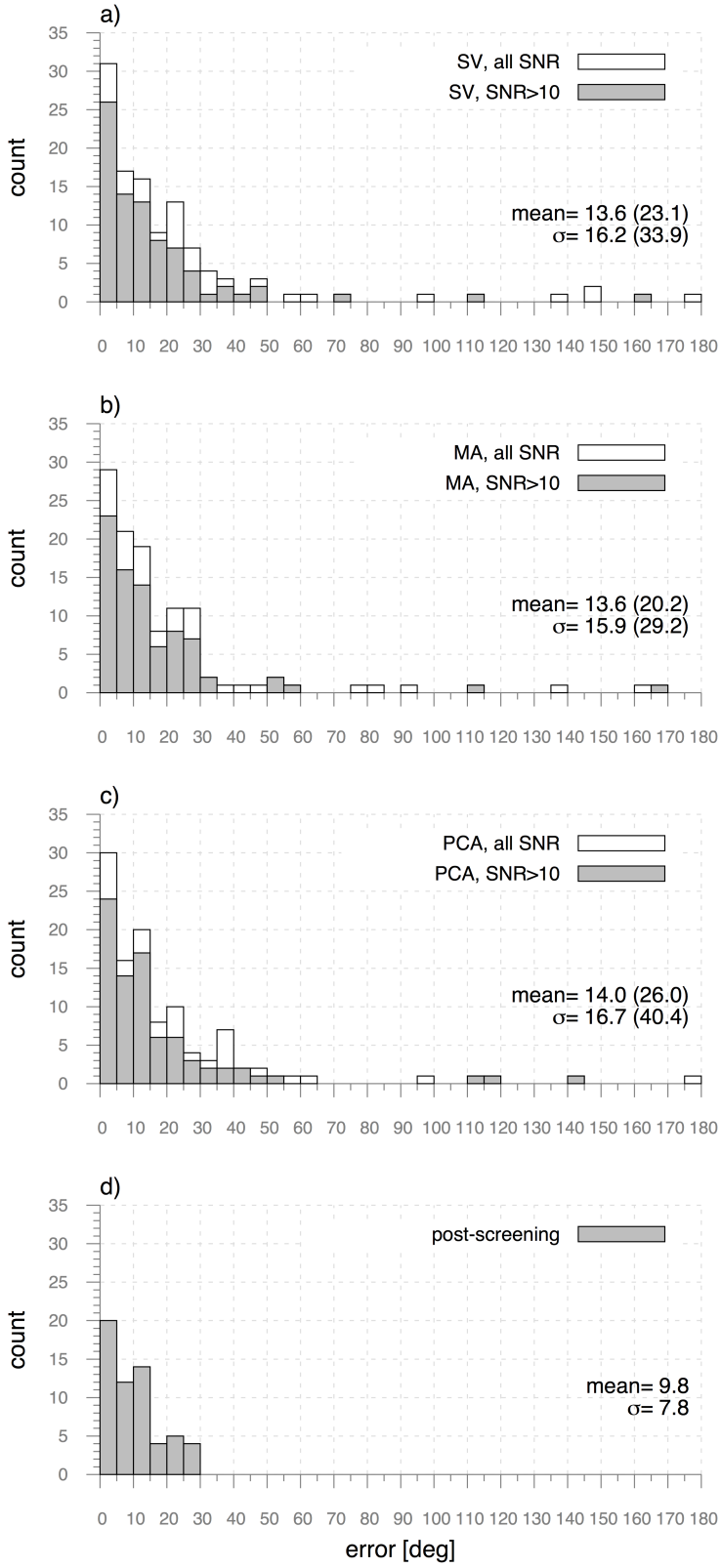


Fig. IV.1:4: Histograms for SV, MA, PCA and post-screened estimates. White histogram bars include low SNR data, while grey exclude these (values in parentheses).

The individual methods have errors with a mean between 13.6° and 14.0° and a standard deviation between 15.9° and 16.7° with outliers up to 180°. Requiring a SNR>10 and a 4° method-coherency removes all outliers. The coherency threshold is obtained by evaluation of the existing catalogue; A value too high will filter out good estimates and a value too low will allow inaccurate estimates to increase. For this choice of method-coherency the resulting mean error lies at 9.8° with a 7.8° standard deviation, errors not exceeding 30° (fig. IV.1:5).

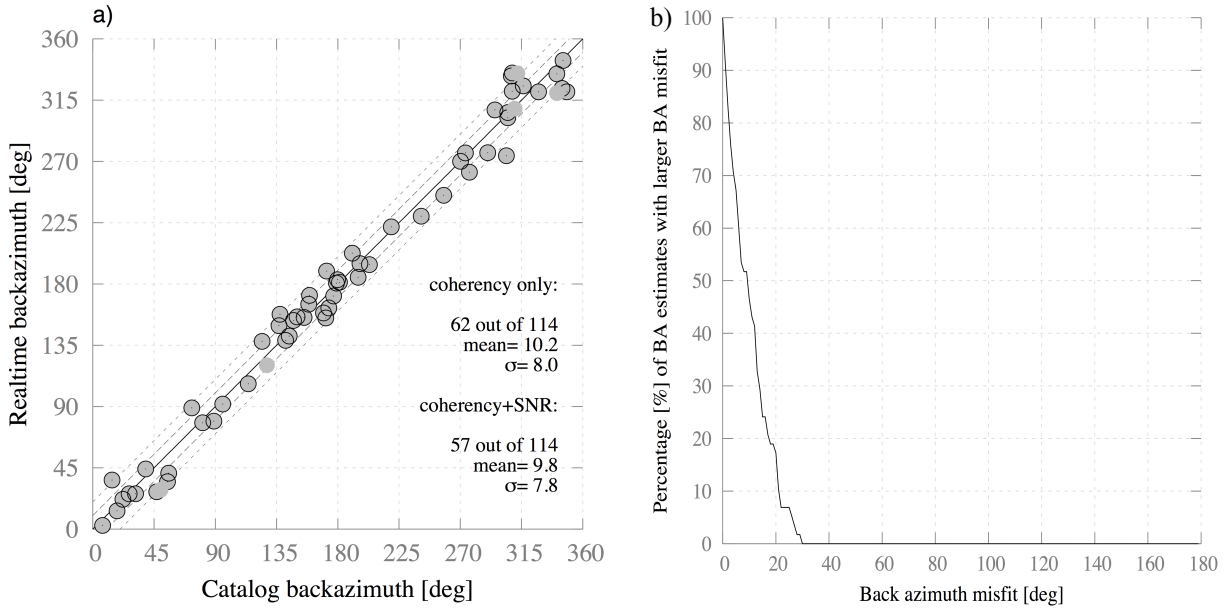


Fig. IV.1:5: (a) Post-Screening real time back-azimuth over Catalog back-azimuth. Black circles denote estimates that passed SNR and coherency filter. Only requiring method coherency allows five more estimates to pass (grey disks), yet makes the system prone to wrong estimates for less strict coherency angles. The coherency filter predominately filters low SNR estimates. 75% of data over SNR threshold passes the coherency filter. (b) Cumulative discrepancy for coherent estimates. Contrary to fig. IV.1:3, a narrow peak is given, ending at 30 degrees.

The data window required for analysis averages to roughly a quarter second. 75% of the significant data relevant for Earthquake Early Warning applications with SNR>10 pass the filter. These results show, statistically speaking, that for seven out of ten events reliable back-azimuth azimuth estimates are available to speed up and stabilize the early location. In the other cases the scheme will fall back to a location solely based on NYAD cells (next chapter), EDT and S-P regions.

Stations that produce a significant amount of bad estimates for catalog events should be excluded for back azimuth estimations. Problems may be related to strong local topography or surrounding lateral velocity variations, in which the particle motion would have to be back traced through the model. Seismometer orientation is a critical factor that must be taken into account. Niu and Li

(2011) evaluated the China Digital Seismic Network (CDSN), the largest permanent seismic network worldwide, towards rotation of seismic traces. They found that 270 of 803 evaluated broad band stations had either a misorientation of the horizontal traces larger than 8 degrees, or falsely named components or flipped polarities. The transportable array of the USArray, valued as one of the best-installed networks worldwide, was set up using an interferometric fiberoptic gyroscope for sensor orientation. Yet, Ekstrom and Busby (2008) found that 7.4% of stations were misoriented by more than 7 degrees. This stresses the fact, that sensor orientation is an important parameter, that can not be neglected when using single station back azimuth estimates in real time location.

Today's realtime location schemes rely on the P-phase detection of at least 4 stations. Introducing BA information into the schemes yields already three independent location constraints (2 BA, 1 EDT), given only two detections. Two of the three are being independent of the local velocity. Ambiguous solutions that often occur in the early stage of location are in most cases resolved by BA information. Beside the stability aspect for the early solution, it is also the time gain that renders realtime BA information valuable, especially for sparse networks. Using BA information, the largest gain in time is given when events are located distant from the network station's Voronoi cell edge or corners, meaningly lying significantly closer to one station than another. Events close to Voronoi cell corners result in at least three detections within a short time and therefore often yield fast locations solely using P-onset information. Depending on the variance of the network latency (which may slow down the detection of later phase arrivals), back-azimuth information can significantly speed up the location process also in these cases, as it becomes available with the data of the first detection.

IV.2 **Not-Yet-Arrived-Data**

Horizontal back azimuth information may be obtained from arrays or three dimensional particle motion. Alessandrini et al. (1994) showed that three dimensional back-azimuth/incidence vectors may at times even successfully be traced back through the 3D-medium, which may open the door for early hypocenter depth estimates. Nonetheless, with only one back azimuth information available, the constraint mapped into space is infinitively long, linearly widening with distance. Once second and third detections occur, enough information is given to constrain a limited region using S-P, P-P or additional back azimuth information. As this additional information may only

arrive with time, it is useful to apply hypothetical constraints to constrain the location faster. So far, location constraints were derived from seismic phases *having arrived* at one or multiple stations. In this section, the location will be constrained by the *absence* of arrivals at one or multiple stations. Rydelek and Pujol (2004) applied the concept of Not-Yet-Arrived-Data (NYAD) proposed by Horiuchi (2003), to cap inherently infinite EDT (P-P) curves which were obtained after two stations triggered. Here, in the same manner, the NYAD cell can be used to cap the infinite back azimuth beam. The NYAD concept translates the information that the event's signal has arrived at some stations, but not yet at other ones, to a geometric region constraint. Accordingly the event can only lie within a certain region, that depends on the network geometry and the time difference between first detection and the latency corrected real time. Satriano et al. (2008) applied the concept from the time of first detection on, using the NYAD region as initial location constraint, which is shrinking with time. The uncertainty of picking and velocity model, however, were generally neglected for NYAD computations, likewise was the impact of latencies underestimated and neglected in previous studies. The EDT uncertainties, if considered, were treated empirically, attempting a one parameter correction to account for both, picking and velocity model uncertainties. But as model uncertainties affect the constraint geometry differently than picking uncertainties, the computed EDT regions did usually not reflect the true region based on given information. The same holds for the computed NYAD region. In the following, we will derive a corrected form of the Not-Yet-Arrived-Data (NYAD) concept, which will then be used to cap the otherwise infinite back-azimuth beams - already in the earliest stage of real time location: with only one detection. In absence of three component stations or small aperture arrays, the NYAD region itself remains the first region constraint.

For now, the concept shall be laid out for a scenario disregarding uncertainties (fig. IV.2:1): Given are two stations at time $t = t^{P,1}$ when station 1 triggers for a P-onset. If station 2 would have triggered in the same moment,

$$t^{P,2} = t + \varepsilon, \quad \varepsilon \to 0 \tag{IV.2-1}$$

an EDT hyperboloid could be constructed between both stations (taking the shape of a plane), on which the hypocenter would be determined to lie on. Considering a sampled (gridded) space, the simple equation defining the EDT surface (neglecting uncertainties) is given in

$$\left(tt_{xyz}^{P,2} - tt_{xyz}^{P,1}\right) - \left(t^{P,2} - t^{P,1}\right) = 0\,, \tag{IV.2-2}$$

with $tt_{xyz}^{P,i}$ being the P-phase travel-time between grid-point coordinate $[x, y, z]$ and station i; and $t^{P,i}$ the P-arrival time at this station. As station 2 might trigger at this moment, but also up to a certain point later in time, the possible area for location is not only given by the surface of the hyperboloid but by the whole inner region of the hyperboloid that encloses the first triggered station. This is expressed by the inequality derived from eq. (IV.2-2)

$$\left(tt_{xyz}^{P,2} - tt_{xyz}^{P,1}\right) - \left(t^{P,2} - t^{P,1}\right) \geq 0\,. \tag{IV.2-3}$$

This inequality may also be evaluated at later points in time, with t in eq. (IV.2-1) denoting the current real time. A latency independent form is given by

$$\left(tt_{xyz}^{P,2} - tt_{xyz}^{P,1}\right) - \left(\tilde{t}^{P,2} - t^{P,1}\right) \geq 0\,, \tag{IV.2-4}$$

with $\tilde{t}^{P,2}$ being the last timestamp of station 2 that arrived at the computation center. As time progresses without the second station triggering, $\tilde{t}^{P,2}$ grows, wherefore the region, defined by the inequality, shrinks. The region's boundary bends convexly around the first station, according to the geometrical behavior of EDT constraints (fig. IV.2:1, top right panel). Using the last arrived timestamps (rather than the current true time) has the advantage that varying latencies and interrupted data flow will not lead to inaccurate results but will just delay the shrinking process.

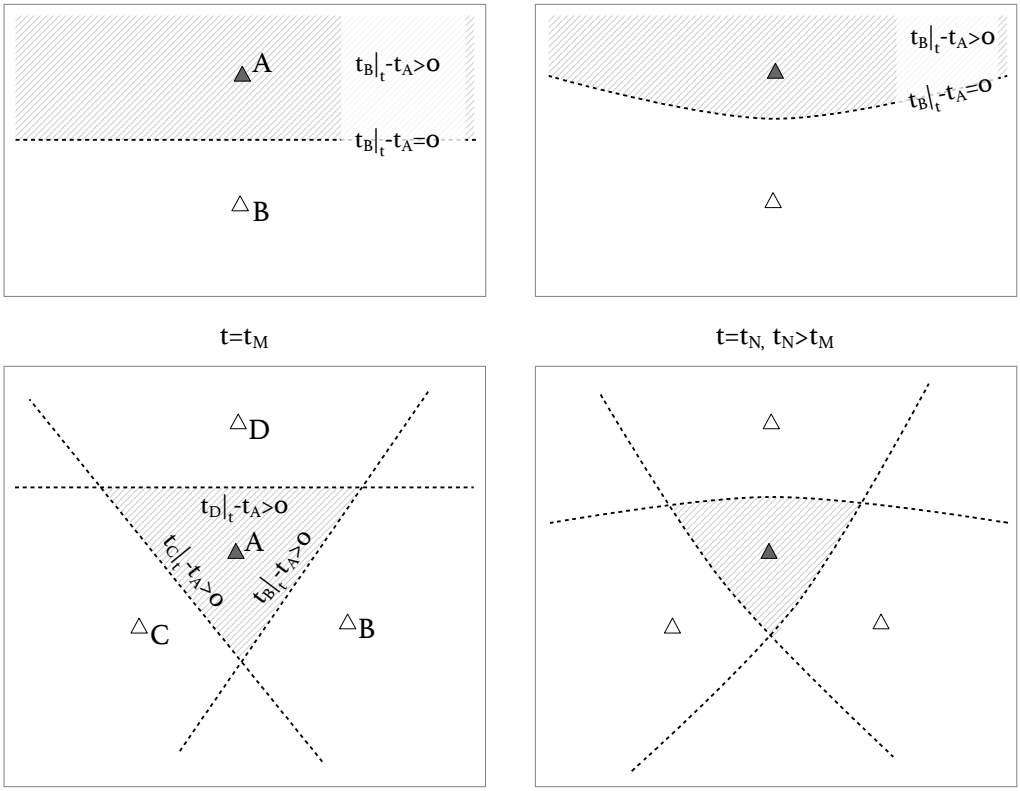


Fig. IV.2:1: NYAD concept (2D horizontal cut). Top: Two station example. Bottom: Multi station example. Left: At time of first detection. Right: advanced time before second detection.

If several stations are given, the region containing the hypocenter will fulfill all inequalities that are composed using all station pairs formed between the first and every other station. As time progresses, this region shrinks in size and provides increasingly precise location estimates. The region is bounded by a set of hypothetical EDT constraints corresponding to time differences between the first station's detection time stamp and the current time stamp of the other stations. The binary probability ("mask") for a location to fall into this region at any time is then given by

$$V_{xyz}(t) = \begin{cases} 1, & \underbrace{\bigwedge_{s=2}^{D} \left[tt_{xyz}^{P,s} - tt_{xyz}^{P,1} = t^{P,s} - t^{P,1} \right]}_{a} \wedge \underbrace{\bigwedge_{s>D}^{S} \left[tt_{xyz}^{P,s} - tt_{xyz}^{P,1} \geq \tilde{t}^{P,s} - t^{P,1} \right]}_{b} \\ 0, & else \end{cases} . \qquad \text{(IV.2-5)}$$

D is the number of stations with detections, while S is the total number of stations in the network. The logical "AND" iteration starts at index 2, comparing the first triggered station's arrival time (index 1) with those of all other triggered and not-yet-triggered stations. Term a accounts for

stations that already detected (index 2 ... D) while b evaluates the inequalities determined by stations without detections (index D+1 ... S). Without latency fluctuations, the region would start out as Voronoi cell at time of first detection (Cua, 2005; Cua&Heaton, 2007). In practice, station-, network- and system latency exists and cause the cell to start out as concave region, possibly several times larger than the Voronoi cell. The time $\tilde{t}^s$ which corresponds to the last time stamp received from station s lags behind the true time t by the (prior unknown) total system latency t^s_{lat}.

$$t = \tilde{t}^s + t^s_{lat} \tag{IV.2-6}$$

The processing time at station and computation center may usually be neglected, but significant latency variance is caused by network data transport delay and data packet length.

$$t^s_{lat} = t^s_{delay} + t^s_{packet} \,. \tag{IV.2-7}$$

The packet length causes a unpredictable variation in the latency, since a packet must be completely filled with data before it is being sent. Contrary to the transport latency, its concrete value is unpredictable, but smaller then the packet length. Depending on the station- and event location, one station might record the event's signal onset at the beginning of a packet while another station might receive and pack it as last sample of the packet. In the latter case no latency would be caused by packeting, while in the first the full packet length would add to the amount of total latency. The maximum packeting delay possible is therefore given in the packet length t^s_{packet} itself. The packet length in national networks often compasses several seconds but may be expected to average around one second in near future. The South California seismic network began to update this length from 5 seconds to 1s starting in 2011. A reasonable network transport delay averages in national networks today around 5 seconds (Brown et. Al, 2011), but varies strongly, with values of 10 seconds not being uncommon. The station latency could be predicted for the next incoming samples, extrapolating the transport latencies of the last received samples. Although simply using the last available time stamp (eq. IV.2-4) avoids this source of error, in the following we will use the predicting form to demonstrate the impact of latency on the NYAD region. Term b of eq. (IV.2-5) including latency variation then becomes

$$b = \bigwedge_{s>D}^{S} \left[tt^{P,s}_{xyz} - tt^{P,1}_{xyz} \geq \left(t - t^s_{delay} - t^s_{packet} \right) - t^{P,1} \right] . \tag{IV.2-8}$$

If every station just experienced the same latency ($t^s_{lat} = t_{lat}$), the NYAD cell would again start out as Voronoi cell, as not only timestamp $t^{P,1}$ would be observed t_{lat} seconds after true time t , but also all other stations' data – the location would hence merely be delayed by t_{lat} seconds. However, since t^s_{packet} and t^s_{delay} denote uncertainties in latency (i.e. a latency variance rather then a fixed constant delay) the maximum cell volume corresponding to this uncertainty is a multiple of the constant-latency-scenario (fig. IV.2:2), even encompassing other stations. This enlarged cell displays the worst case scenario, with all stations lagging the assumed maximum latency behind the station of first detection. In practice, the true cell will therefore be more compact: Some EDT boundary segments will be closer to the outer cell, others closer to the Voronoi cell. Neighboring stations with a smaller latency then the one of the first trigger will even generate EDT boundaries that will concavely surround the epicenter –lying inside the Voronoi cell– at the time the first trigger arrives.

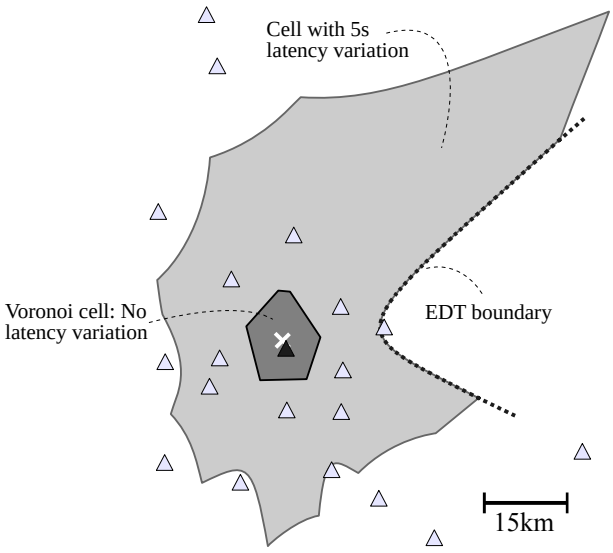


Fig. IV.2:2: Effect of latency variation on a NYAD cell (light grey). In a latency-free scenario the NYAD cell starts out as Voronoi cell (light grey), enclosing the station of first detection (black triangle). Latency variance enlarges the initial cell to a concave volume, constructed by hypothetical EDT surfaces. Now also other stations (grey triangles) may be encompassed, as the wave might already have arrived at those stations but the signal not yet have been received at the data center. The epicenter is marked by a white x.

Data of neighboring stations with much smaller latencies may arrive significantly earlier, such that

even phase pickers could trigger out of order (as the arrival of the P-phase to a later station could be observed at the data center before its arrival at the first). This circumstance caused many to be skeptical towards the NYAD approach, as the algorithm would have to interpret this later arrival as first arrival and consequently choose the “wrong” NYAD cell. Although this is true, the consecutive deduction that the event would be mis-located because of that, is wrong. It would only be true if we disregarded the latencies (e.g. by using the true time instead of the time stamp of the last arrived sample). A simple geometric proof (fig. IV.2:3) shows that although the wrong cell would indeed be initially chosen, the true location will yet be contained in that (due to respected latencies) larger concave cell.

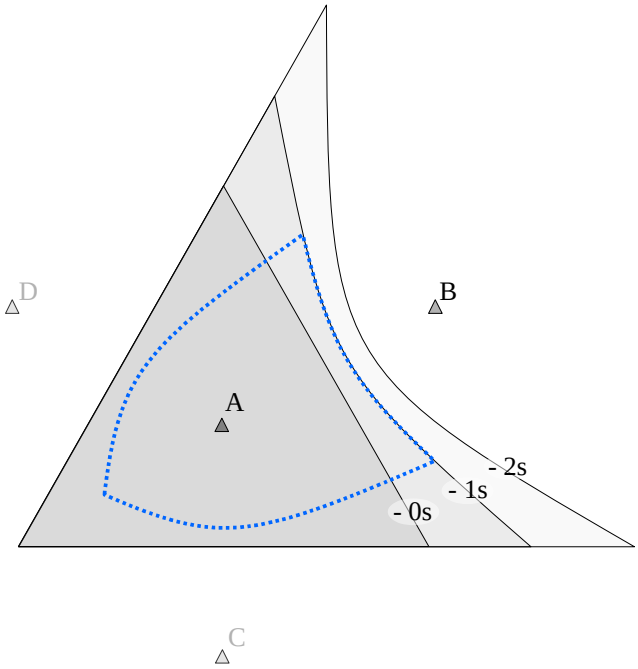


Fig. IV.2:3: Geometric proof showing that a (due to latencies) wrongly chosen "first triggered station" does not lead to mis-locations, if the NYAD cell is based on the last timestamps having arrived at the data center:

The figure shows different NYAD cell sizes for a minimal network (consisting of four stations A,B,C and D) in which the data center received a trigger (only) from station A. The dark grey triangular cell (ending at the ***-0s*** *edge towards B) corresponds to a realtime configuration in which no station experiences significant latencies and only A has triggered so far. A cell ending at the hyperbolic* ***-1s*** *edge (consisting of the dark grey and intermediate gray region) would be given if station B lagged 1 second behind. Due to the latency we cannot know if B actually did trigger in the last second – at the same time as A or even up to a second earlier. The cell ending at the* ***-2s*** *edge (all grey regions) would be given if B lagged 2 seconds behind. We will chose exactly this case for the proof:*

Let be given that A,C and D show no significant latencies – their data arrives in true real-time. Only station B lags 2 seconds behind. Let also station B be the one that is closest to the event and experiences the first trigger. Let station A experience the next trigger, 1 second after B. However, since A has no latencies, its P-arrival will be received first at the data center and consequently be interpreted as the first trigger. The NYAD cell is consequently constructed around A instead of B, whose P-trigger will only arrive 1 second after the one of A at the data center (since B's trigger occurred 1 second prior to A, but the transmission took 2 seconds). Since B lags 2 seconds behind, the initial NYAD cell will be ending at the (-2s) edge (largest cell configuration). We remember that the NYAD edges are constructed from hypothetical EDT surfaces ("the EDT would be here if the trigger was reported right now"). E.g. the (-2s) edge between A and B corresponds to an actual EDT that would be constructed if B triggered exactly 2 seconds earlier than A: With the 2-seconds latency both triggers would arrive simultaneously at the data center and the EDT would be generated based on the two trigger timestamps. Since (in this scenario) B experienced the trigger only one second before A, ***the (-1s) edge corresponds to the true EDT curve*** *that* ***will be*** *constructed once the trigger of B has eventually arrived.* ***The epicenter location must therefore be located on the (-1s) edge, <u>which lies inside the initial cell</u>.*** *The exact evolution is as follows:*
At the instance of time when A's trigger is received, the cell ends at the (-2s) edge (largest cell configuration). The trigger of B is received 1 second later: By this time the hypothetical EDT edge moved to the (-1s) position. Also the edges related to C and D moved towards A, taking now a convex shape. The NYAD region hence shrunk to the dashed blue region. Now B's trigger is received and the location is constrained to the EDT corresponding to the blue (-1s) edge towards B. ***As the NYAD edges shrink monotonically with time and the initial cell contains all later cells and all possible future EDT segments (containing the epicenter), any event whose phase arrival occurred at B not earlier than "the time corresponding to the latency" before A, must consequently lie within the initial NYAD cell.*** *We know that it did not occur earlier than that, as it would otherwise have arrived first at the data center.* ***The initial cell's prediction concerning the location was hence accurate.***

An “incorrectly” chosen cell will hence only lead to a less precise location constraint until the time the true first arrival eventually appears. Once two P-arrivals are given, the NYAD cell instantaneously collapses into an EDT segment.

Usually, NYAD calculations are based on a discrete velocity model, like empirical minimum-1D-models. While for EDT constraints often at least a one parameter uncertainty was added to account for picking and model uncertainties (Satriano et al., 2008), this was fully neglected for the computation of the NYAD cells (although they are constructed from hypothetical EDT segments). Like for EDT's, the impact of model uncertainties unto the NYAD region is large. These uncertainties are even amplified in depth, where EDT uncertainties occupy significantly more space than in the plane spanned out by the network stations (NYAD cells are typically calculated for an assumed hypocentral depth of 10km). To derive an uncertainty corrected from, we will reformulate eq. (IV.2-5) to include picking and model uncertainties. Term a, corresponding to fixed NYAD edges given by EDT's obtained from stations already having detected, then becomes

$$a = \left[T_{max}^{M[1,s]} \geq T_{min}^{O[1,s]} \right] \wedge \left[T_{max}^{O[1,s]} \geq T_{min}^{M[1,s]} \right] \tag{IV.2-9}$$

with the minimum and maximum travel time values

$$T_{xyz,min}^{M[1,s]} = tt_{xyz,min}^{P,s} - tt_{xyz,max}^{P,1}, \tag{IV.2-10}$$

$$T_{xyz,max}^{M[1,s]} = tt_{xyz,max}^{P,s} - tt_{xyz,min}^{P,1}, \tag{IV.2-11}$$

and onset values

$$T_{min}^{O[1,s]} = \left(t^{P,s} - t^{P,1}\right) - \left(\delta t^{P,1} + \delta t^{P,s}\right), \tag{IV.2-12}$$

$$T_{max}^{O[1,s]} = \left(t^{P,s} - t^{P,1}\right) + \left(\delta t^{P,1} + \delta t^{P,s}\right). \tag{IV.2-13}$$

Term b, corresponding to the stations not yet having triggered, becomes simply

$$\left[\, T_{max}^{M[1,s]} \geq T_{min}^{O[1,s]} \right] \tag{IV.2-14}$$

with the same minimum and maximum travel time values but different hypothetical onset values

$$T_{min}^{O[1,s]} = \left(\tilde{t}^{P,s} - t^{P,1}\right) - \left(2 \cdot \delta t^{P,1}\right), \tag{IV.2-15}$$

$$T_{max}^{O[1,s]} = \left(\tilde{t}^{P,s} - t^{P,1}\right) + \left(2 \cdot \delta t^{P,1}\right). \tag{IV.2-16}$$

As the later station's P-onset is only a hypothetical one, its picking uncertainty is assumed to be similar to the one measured at the first station. If the current time is chosen to be used rather then the time stamp of the last sample arrived, $\tilde{t}^{P,s}$ needs to be replaced according to eq. (IV.2-6). Eq. (IV.2-5) including travel time and picking uncertainties is finally given in

$$V_{xyz}(t) = \begin{cases} 1, & \underbrace{\bigwedge_{s=2}^{D} \left[\, T_{max}^{M[1,s]} \geq T_{min}^{O[1,s]} \right] \wedge \left[T_{max}^{O[1,s]} \geq T_{min}^{M[1,s]} \right]}_{a} \;\wedge\; \underbrace{\bigwedge_{s>D}^{S} \left[\, T_{max}^{M[1,s]} \geq T_{min}^{O[1,s]} \right]}_{b} \\ 0, & else \end{cases}. \tag{IV.2-17}$$

NYAD would be sensitive to wrongly identified first stations if latencies are not correctly addressed. Due to varying latencies, later detecting stations may arrive first in the data center and appear as station of first detection, in which case a wrong bounding NYAD region would be chosen and analyzed. However, when latencies are properly treated (e.g. by simply using the time stamp of the last arrived sample rather than current true time), no false information is generated. Initially, still the wrong cell may be chosen due to latencies, but it would always enclose the true hypocenter due to it's extended size.

NOT-YET-ARRIVED-S

NYAD information allows to constrain the maximum hypocentral distance at the early stage of the location process. This allows to cap otherwise infinite constraints like back-azimuth beams or

EDTs. This maximum distance is, however, not the most important information for early warning systems. The minimum distance, which translates via the observed amplitudes to a minimum magnitude, would be of much higher value, as its exceeding of the $M_L5.0$ threshold is sufficient to raise a warning. The real-time magnitude estimate is calculated from the peak amplitude prior to the S-phase arrival (Lior et al., 2015) – the maximum value *observed so far:* Since larger amplitudes may still be recorded afterwards, the magnitude estimate may still grow, but not decrease. The estimate based on a minimum distance would therefore provide a true lower bound for the magnitude, allowing to warn at earliest times. However, there is no proxy for a minimum distance in EEW, today.

To obtain this information, we can apply the concept of Not-Yet-Arrived-Data, which uses the delaying of a P-phase at different stations, to the delaying of the S-phase at a single station. Like the shrinking NYAD cell evolutionary constrains the location based on inter-station hyperbolic P^1-P^2 constraints, we can use the delaying of the S-phase at one station to construct a growing S-P sphere - as evolutionary version of a S-P location: When a station detects a S-phase, the time difference to the earlier arrived P-phase allows to construct a (“circular”) S-P constraint that surrounds the station. From this constraint (that due to uncertainties resembles a “band” surrounding the station) can in turn the minimal and maximal hypocentral distance be derived. This standard use of S-P constraining is in particular useful in cases where the faster P-phase doesn't quickly reach the required amount of stations to constrain the hypocenter, or in cases where the event is located outside the network (where EDTs fail to resolve epicentral distances precisely).

However, even *before* the S-Phase arrival, valuable information about the minimum hypocentral distance can be obtained, if it is certain that the S-phase has indeed not yet arrived: Counting the seconds between lightning and thunder in a warm summer thunderstorm allows the estimation of the distance (in meters) to the electric discharge by multiplying the counted seconds with approx. 350. While still counting the seconds, the minimum distance could be updated every second, in calculating the distance as if the thunder arrived at this very instant of time. In the same manner a growing hypothetical S-P-constraint can be constructed at any given time before the S-arrival, defining a minimum hypocentral travel-time and -in turn- the minimum hypocentral distance to the event. This constraint will be referred to as Not-Yet-Arrived-S-Phase (NYAS). As it constrains only the minimum travel-time, the spatial resemblance of the hypothetical S-P constraint will be a “band” of “infinite thickness”, holding an inner radius but being unconstrained in the outer radius: Every location in larger distance than the inner radius (i.e. the minimum distance) is a possible

hypocenter location. This constraint forms therefore a spherical exclusion zone concerning the possible hypocenter location around the station, which grows with time. NYAS and NYAD together -NYAS serving as inner boundary, and NYAD as outer boundary- form a location constraint that is shrinking with time (fig.IV.2:4).

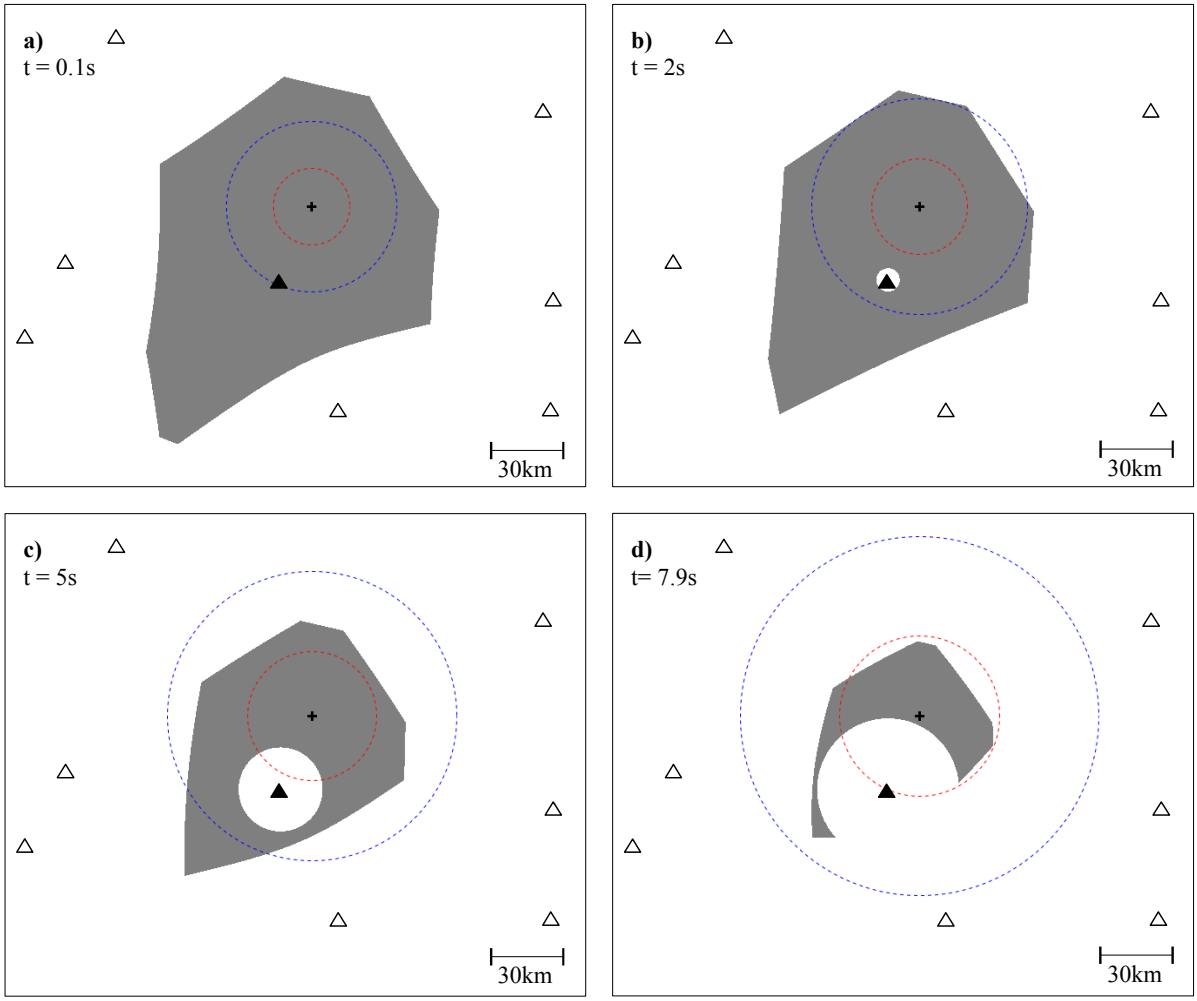


Fig. IV.2:4: NAYD – NYAS region for a near-surface event. a) The P-wave (wavefront indicated by blue circle) propagating from the epicenter (black plus) reaches the first station (black triangle). Snapshot taken 0.1s after first detection. The S-wavefront (red circle) will arrive 8s later. b) 2 seconds later the hypothetical S-P constraint spans out an spherical exclusion zone with a 5km radius. c) 5s after first detection the minimum hypocentral distance grew to 17km. d) 0.1s before the S-phase would arrive, providing a hypocentral distance of approximately 33km, the NYAS minimum distance grew to 30km.

The minimum hypocentral distance can be derived from the minimum travel time obtained from a hypothetical S-P-constraint. For a grid point (x,y,z) in space, the standard S-P constraint (neglecting uncertainties) is defined as

$$\left(tt^{S}_{xyz} - tt^{P}_{xyz}\right) - \left(t^{S} - t^{P}\right) = 0\,, \qquad \text{(IV.2-18)}$$

with t^P being the time of P-phase arrival at a given station and t^S being the time of the corresponding S-phase arrival at the same station. tt^P_{xyz} and tt^S_{xyz} are the travel times of the P- and S-phase between the station and a given point (x,y,z) in space. Before the S-phase arrival, t^S is still unknown but must be later than the current time t, or more correct respecting latencies, later than the timestamp $\tilde{t}$ of the last received data package. $\tilde{t}$ lags behind real time t by the latency t_{lat}, again consisting of packaging and transport latency

$$\tilde{t} = t - t_{lat} = t - t_{delay} - t_{packet} . \tag{IV.2-19}$$

The region defined by

$$\left(tt^S_{xyz} - tt^P_{xyz}\right) - \left(\tilde{t} - t^P\right) < 0 \tag{IV.2-20}$$

is therefore excluded from the solution space of possible hypocenter locations and can hence be used to define a minimum hypocentral travel-time. For the special case of a homogeneous velocity model, the hypocentral distance is given in

$$d_{min} = v_p \cdot tt_p = v_p \cdot \frac{\tilde{t} - t^p}{\left(v_p / v_s - 1\right)} = \frac{\tilde{t} - t^p}{\left(1/v_s - 1/v_p\right)} . \tag{IV.2-21}$$

Including picking and model uncertainties the constraint can be formulated as

$$C_{xyz}(t) = \begin{cases} 1, & [\, T^M_{max} \geq T^O_{min} \,] \\ 0, & else \end{cases} \tag{IV.2-22}$$

with the maximum value for the modeled travel time difference at grid point (x,y,z),

$$T^M_{xyz,max} = tt^S_{xyz,max} - tt^P_{xyz,max} \tag{IV.2-23}$$

and the minimum value for the hypothetically observed arrival time differences

$$T^{O}_{min} = (\tilde{t} - t^{P}) - (\delta t^{S} + \delta t^{P}) . \quad \text{(IV.2-24)}$$

For the common case that tt^{S}_{xyz} is obtained from tt^{P}_{xyz} using vp/vs ratios, then $tt^{S}_{xyz,max}$ is given by using the maximum possible value for vp/vs $tt^{S}_{xyz,max} = (v_s / v_p)_{max} \cdot tt^{P}_{xyz,max}$. Eq. (IV.2-23) then becomes

$$T^{M}_{xyz,max} = ((v_s / v_p)_{max} - 1) \cdot tt^{P}_{xyz,max} . \quad \text{(IV.2-25)}$$

The picking (or detection) uncertainty for the S-phase δt^{S} depends on the performance of the used detection-algorithm: The approach of NYAS requires the ability to know that the S-phase has not yet arrived - or ideally, to detect the S-phase arrival in real-time in a reliable manner. Since the scheme aims for the determination of the minimum travel-time, late detections of the S-phase would be critical as they lead to an overestimation of the travel-time. Early detections, on the other side, are uncritical as they only lead to an underestimation of the minimum - rendering the estimate for the lower bound of the travel-time (and distance) less precise but still accurate. Advances in automatic S-phase detection, like autoregressive prediction (Küperkoch et al. 2012) of the S-Phase, allow a confident picking with error margins of ± 1s, albeit not completely in real-time as data after the S-arrival is required to identify changes in the data's statistical properties. A small percentage of S-phases is not picked at all. Although today's near real-time S-phase detectors are not yet able to provide the immediate response NYAS requires, the following section shall show that the absence of S-wave signatures can be determined confidently in real-time (trigger less than 0.5 seconds after the S-arrival). As early detections are uncritical, but late detections must be avoided, the algorithm aims to “drop-out” as early as S-phase similar signatures are detected (even if they do not stem from a S-phase). The next section shall show that simple proxies tracking the quantities and changes in polarization, amplitude, rectilinearity and dominant frequency allow a drop-out usually less than 0.2s after the S-phase arrival, with approx. 80% dropping out no earlier than 1 second before (fig. IV.2:5). Setting the picking/detecting uncertainty δt^{S} to 0.5s therefore allows to obtain robust real-time estimates for the minimum S-P interval.

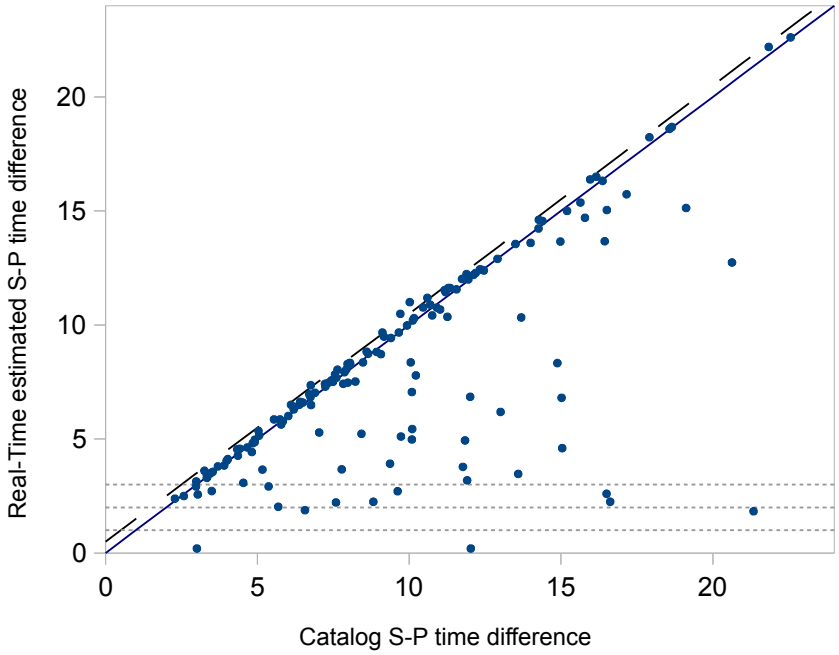


Fig. IV.2:5: Real-time S-P estimate over catalog S-P time interval. Using the described detector concerning the absence of S-wave signatures, for most cases the drop-out occurs at the time of the S-arrival. Adding a half-second picking uncertainty (dashed diagonal) is sufficient to avoid significant over-estimations. Early drop-outs occur frequently, but very little within the first two seconds (second fine dashed line) after which the first magnitude estimates are calculated (requiring distance estimates). Since the scheme doesn't yield late S-picks, it allows to provide accurate lower bounds for the S-P interval and in turn the hypocentral distance.

Real-time detections are bound to be late, as a phase has to arrive before it can be detected and a certain amount of post-arrival data has to be analyzed. Some characteristics may additionally only become pronounced enough with time, e.g. when the S-phase emerges from a strong P-coda. In such a case, e.g. the polarization will start to change from the time of the S-phase arrival but may reach the trigger threshold only seconds later, when the amplitudes grew sufficiently. One key strategy to reduce delays to a insignificant minimum and yield real-time detections is to put the reporting of hypothetical S-P time-differences on hold whenever the drop-out condition is expected to be approached. If, following, the drop-out condition is reached, the algorithm marks a potential S-phase as having arrived. As the last reported hypothetical S-P time-difference was given at the time when the algorithm switched to “on-hold”, the true S-P interval is properly determined even when the drop-out condition is reached late. If, however, the parameters fall back below their thresholds before the “drop-out” condition is reached, the reporting of now larger S-P time-

differences is resumed. This is typical for phases in the P-coda showing indications of S-phase characteristics (e.g. rise in amplitude), but not reaching the drop-out condition - the reporting is put on hold until the phase passed by and the parameters fell back significantly under their thresholds. The detailed detector scheme, monitoring the absence of S-phase energy shall be constructed in the following, beginning with the strongest parameter: the polarization.

Polarization. This parameter is the most important indicator for an S-phase arrival as the change in polarization is (usually) the most robust and fastest to track and often sufficient to yield good S-arrival time estimates. It will be computed as follows:
Given a three component velocity seismogram with traces X, Y and Z, the initial data window that reaches the highest rectilinearity following the P-arrival is used to rotate the components into Eigen-space using a Principal Component Analysis. The window is maximally allowed to extend to 0.3 seconds past the P-phase arrival, which corresponds to a ~2.5km hypocentral distance. Since the seismogenic depth itself is generally deeper, no S-phase arrival is expected in this interval. The rotated trace $V^{\parallel}$ will carry the significant part of the energy of the just recorded initial P-wave. As the S-phase will follow a similar path, its energy is expected to appear on the two perpendicular traces $V_1^{\perp}$ and $V_2^{\perp}$. Similar to the H/V ratio, here the relationship between the Energies E on $|V^{\perp}|$ and $|V^{\parallel}|$ will be analyzed, where $|V^{\perp}|$ is being defined as

$$|V^{\perp}| = \sqrt{(V_1^{\perp})^2+(V_2^{\perp})^2}\,. \tag{IV.2-26}$$

Here, we will evaluate the difference between $E^{\perp}$ (S-polarization) and $E^{\parallel}$ (P-polarization), computed over a sliding window of one second interval

$$E_{[LS]}^{\perp-\parallel} = E_{[LS]}^{\perp} - E_{[LS]}^{\parallel} = \sum_{s=LS-w}^{LS} (V_{[s]}^{\perp})^2 - (V_{[s]}^{\parallel})^2\,, \tag{IV.2-27}$$

with [ ] (brackets) indicating the sample index, LS indicating the last available sample and w the 1 second window length in samples (i.e. the sample rate). For larger amplitudes on the perpendicular polarized traces, $E^{\perp-\parallel}$ will be positive. Since the energy scales with the squares of

the amplitudes (and since those are significantly higher during the S-arrival than when emerging from the noise during the P-arrival), $E^{\perp-\parallel}$ shows a clear distinct positive peak at the S-phase arrival (fig. IV.2:6-IV.2:9). $E^{\perp-\parallel}$ is, however, not normalized and therefore unreliable for realtime picking. For automated picking and quality control a normalized difference-over-sum measure E_R will be used, comparing the energy of both components

$$E_R = \frac{E^{\perp} - E^{\parallel}}{E^{\perp} + E^{\parallel}} = \frac{E^{\perp-\parallel}}{E^{total}} \,. \qquad \text{(IV.2-28)}$$

E_R displays the normalized energy difference between both components $|V^{\parallel}|$ and $|V^{\perp}|$, ranging from -1 to +1. Starting with $E_R \sim -1$ at the strongly polarized P-arrival, it rises with time to eventually approach $+1$ at the arrival of the S-phase. If polarization was the only parameter evaluated, the drop-out condition could be defined as $E_R = 0.7$ and the on-hold condition as $E_R = 0.5$. Reliable pre-S timestamps (as the interval devoid of S-energy) would then be obtained by delaying the update of $\tilde{t}$ ("the growth of the exclusion zone") when E_R exceeds 0.5, continuing to update when it falls back below 0.5, and fully stopped (drop-out) once it exceeds 0.7. The detection uncertainty of 0.5s makes up for late detections, effectively delaying the report by a half second. Figures IV.2:6 - IV.2:9 demonstrate this scheme on the data of a M_l 4.2 Dead Sea event recorded at different sites with hypocentral distances ranging between 14km and 101km.

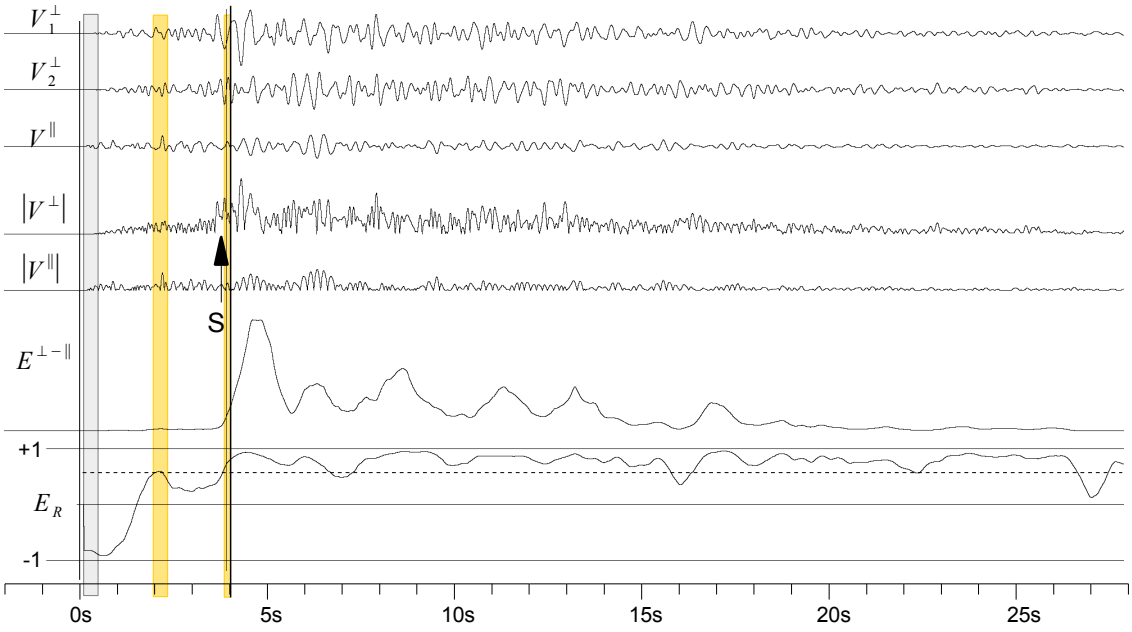


Fig. IV.2:6: Ml4.2 event in 14km hypocentral distance. The gray bar marks the early data used to rotate the traces into the Eigensystem. Orange bars mark times where the reporting would be put on-hold due to $E_R > 0.5$. The growth of the NYAS-sphere is put on-hold for 0.4s prior the arrival of the S-phase at approx. 3.7s. The S-phase arrives at the beginning of the second on-hold phase, which is followed by the final drop-out condition ($E_R > 0.7$). With the assigned S-detection uncertainty of 0.5s, the S-P interval is accurately determined.

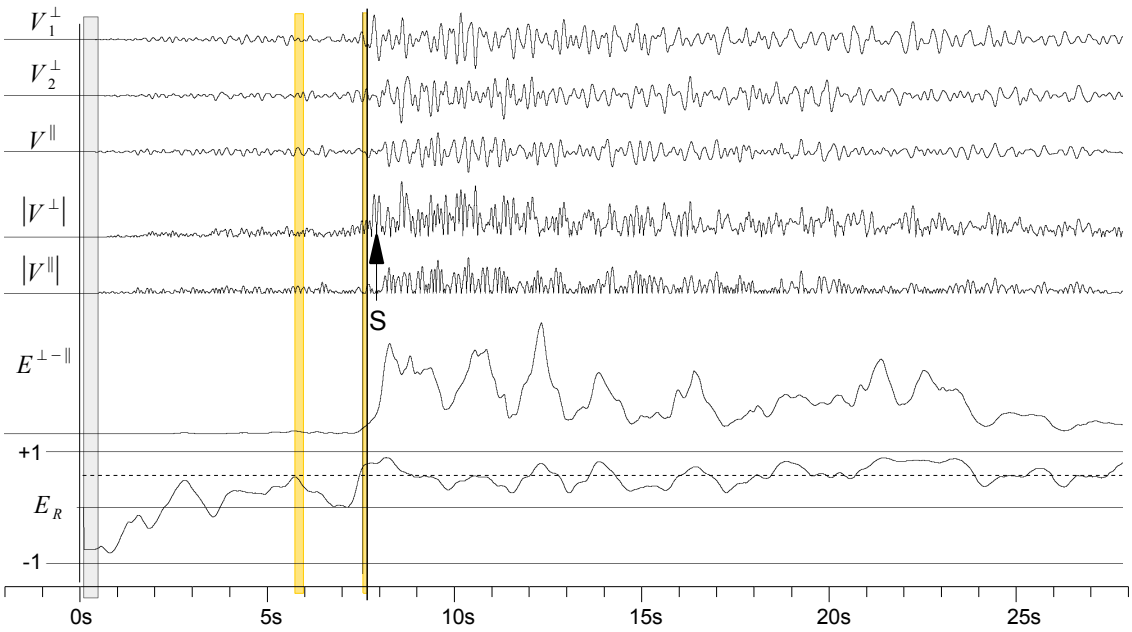


Fig. IV.2:7: Ml4.2 event in 47km hypocentral distance. The gray bar marks the early data used to rotate the traces into the Eigensystem. Orange bars mark times where the reporting would be put on-hold due to $E_R > 0.5$. The growth of the NYAS-sphere is delayed (on-hold) for 0.2s prior the arrival of the S-phase at approx. 7.7s. The algorithm marks the end of the S-P interval 0.2 seconds before the arrival of the significantly higher amplitudes of the S-phase.

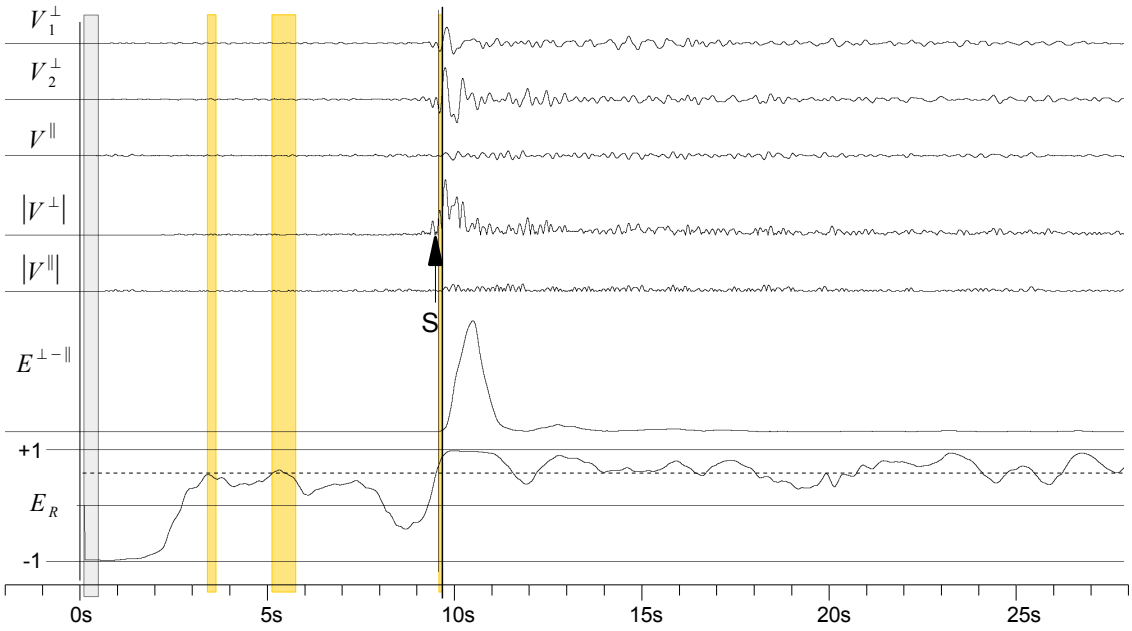


Fig. IV.2:8: Ml4.2 event in 53km hypocentral distance. The gray bar marks the early data used to rotate the traces into the Eigensystem. Orange bars mark times where the reporting would be put on-hold due to E_R>0.5 . The growth of the NYAS-sphere is delayed (on-hold) for 0.9s in total prior the arrival of the S-phase at approx. 9.4s. The time of drop-out (E_R>0.7) agrees with the S-phase arrival, but is slightly late. Subtracting the assigned S-detection uncertainty of 0.5s, the S-P interval is accurately determined.

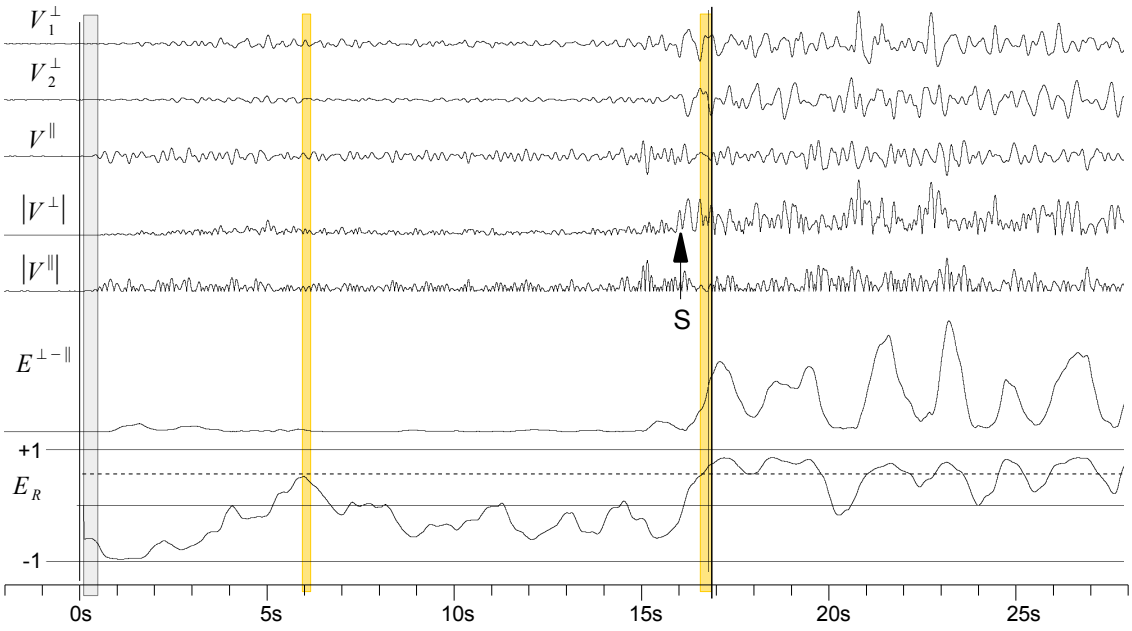


Fig. IV.2:9: Ml4.2 event in 101km hypocentral distance. The gray bar marks the early data used to rotate the traces into the Eigensystem. Orange bars mark times where the reporting would be put on-hold due to E^R>0.5. The growth of the NYAS-sphere is delayed (on-hold) for 0.4s prior the arrival of the S-phase at approx. 16s. Despite the lower SNR conditions, the E^R curve still identifies the S-energy. However, the first higher amplitudes arrive shortly before the drop-out of the algorithm, which in this case relies on the 0.5s detection uncertainty to properly determine the S-P interval.

These timelines show that a simple detector could be based solely on the polarization. However, as the last example (IV.2:9) demonstrates, may relying on a single parameter lead to late drop-outs when the S-phase characteristic is not as pronounced - alike for amplitudes, whose increase is typical for S-phase arrivals, albeit not always observed with short epicentral distances. A late (or missed) drop-out can only be avoided by lowering the parameter's threshold, e.g. drastically for the amplitude criteria, where we can only demand it not to decrease. For a single parameter, this would of course often lead to very early drop-outs. To avoid this, we will now add other parameters and require their mutual fulfillment - meaningly, all (or most) parameters need to exceed their thresholds for a drop-out. This condition of a mutual fulfillment of multiple parameter criteria "pushes" the otherwise early drop-out back to later times, as earlier phases do exhibit some but not all characteristics of S-phase arrivals. It hereby provides the stability that is needed to apply individual parameter thresholds that are low enough to always be met (avoiding late picks).

The sum of parameters exceeding their respective thresholds θ forms the S-potential ϕ ,

$$\begin{aligned}\phi = & \underbrace{[E_R > \theta_E] + [\Delta_t E_R > \theta_{dE}]}_{polarization} \\ & + \underbrace{max([R > \theta_R], [|\Delta_t R| > \theta_{dR}])}_{rectilinearity} \\ & + \underbrace{[A_C > \theta_A]}_{amplitude} \\ & + \underbrace{max([\Delta_t f_0 < \theta_{df}], [|\Delta_t \Delta_t f_0| > \theta_{ddf}])}_{frequency}\end{aligned} \tag{IV.2-29}$$

using the Iverson bracket notation

$$[C] = \begin{cases} 1, & if\ C\ is\ true \\ 0, & otherwise \end{cases} \tag{IV.2-30}$$

being based on the energy ratio E_R , the rectilinearity R , the amplitude contrast A_C , the dominant frequency f_0 , their change rates (Δ_t) and their corresponding detection thresholds θ . The *max* function serves as logical "OR" for parameters where both, a high value or a significant change over time signify the S-arrival. With the used parameter set the S-potential ranges in values between 0 and 5. Where the first example (fig. IV.2:6-IV.2:9) demonstrated the "drop-out" and "on-

hold" concept concerning a single parameter, it will now be applied to ϕ. The drop-out occurs when $\phi = 5$. The algorithm switches to "on-hold" when approx. two thirds of parameters exceed their individual thresholds, when $\phi = 3$, indicating the possibility of an approaching S-phase arrival. The additional parameters shall be laid out in the following.

Polarization change. Since the S-phase arrival is always accompanied by a positive trend in E_R we will require as first additional criterion that the smoothed change rate $\Delta_t E_R$ has to be positive. The smoothed change rate Δ_t measures the difference in arithmetic mean between two neighboring sliding windows to avoid high frequency fluctuations that would occur using a simple finite difference (representing ∂ / ∂_t) between neighboring samples. We further use a significant shorter second window (approx. 0.1s long, compared to a 1s long first window) ending at the latest sample available (LS) to detect property changes faster.

$$\left(\Delta_t E_R\right)_{[LS]} = \frac{sr}{w1+w2}\left(\left[\frac{1}{w1}\sum_{s=LS-w1-w2}^{LS-w2} E_{R,[s]}\right] - \left[\frac{1}{w2}\sum_{s=LS-w2}^{LS} E_{R,[s]}\right]\right) \qquad \text{(IV.2-31)}$$

The quotient of the sample rate (*sr*) and both window lengths in samples (*w1* & *w2*) takes account of the time interval over which the change is being observed.

Rectilinearity. Both, arriving P- and S-phases are linear polarized (in a homogeneous medium) and should exhibit an increased rectilinearity R,

$$R_{[LS]} = 1 - \left(\frac{\lambda_1 + \lambda_2}{2 \cdot \lambda_0}\right)_{[LS-w,\,LS]}. \qquad \text{(IV.2-32)}$$

The three Eigen-values $\lambda_0, \lambda_1, \lambda_2$ (λ_0 being the largest) are obtained by a Principal Component Analysis over a 1s long window. Although the S-phase is linear polarized, its arrival is often accompanied by a decrease in rectilinearity. This is because the polarization direction changes from the less polarized P-coda to the S-phase. While the S-phases slowly increase in amplitude, the superposition of the perpendicular polarized phases often causes a drop in the observed total rectilinearity. To account for this effect, not only a high rectilinearity but also a change in the

rectilinearity

$$\left|\left(\Delta_t R\right)\right| \tag{IV.2-33}$$

is considered as trigger condition, which may be positive or negative. Δ_t Is again computed as in IV.2-32.

Amplitude contrast. The S-phase is usually characterized by an increase in amplitude. This increase is significant for distant events, where P- and S-phase are separated, but may be weak with local events, where P-and S-phase overlap. The amplitude criterion for potential S-phase arrivals is therefore reckoned to be fulfilled if the contrast is at least unity

$$A_C \geq 1\ , \tag{IV.2-34}$$

although the optimal threshold for the analyzed data set is slightly higher (~ 1.2). This criterion could be further optimized to reduce early drop-outs by a time-dependent threshold, which increases as time elapses from the P-arrival. The amplitude contrast is defined as

$$A_{C,[LS]} = \left[\max_{LS-w2<s<LS}\left|V_{[s]}\right|\right] \Big/ \left[\max_{LS-w2-w1<s<LS-w2}\left|V_{[s]}\right|\right]$$
$$\text{with} \tag{IV.2-35}$$
$$V_{[s]} = \sqrt{X_{[s]}^2 + Y_{[s]}^2 + Z_{[s]}^2}\ ,$$

comparing the peak amplitude between two sliding windows. A shorter second window (w2) (approx. one tenth of the 1s long first window) allows again for an early and close-to-real-time recognition of the increasing amplitude.

Dominant Frequency. The fourth quality used is the S-phase's lower dominant frequency f_0. At the S-phase arrival we therefore expect to observe a decrease, yielding a negative frequency change rate

$$\Delta_t f_0 < 0\ . \tag{IV.2-36}$$

As positive and negative rates are observed at all times, also a high impulsiveness of the frequency change

$$\left|\Delta_t \Delta_t f_0\right| , \tag{IV.2-37}$$

is required, constituting the last criterion for in this detector. To estimate the impulsiveness localized around the evaluated time, the second application of Δ_t uses a much shorter length for the first window (*w1*) than usual (~ 0.2s long). The dominant frequency observed on the horizontal traces will be approximated by

$$f_{0,[LS]} = \frac{sr}{4\mathrm{w}} \cdot \frac{\sum_{s=LS-w}^{LS} \sqrt{\left(X_{[s]} - X_{[s-1]}\right)^2 + \left(Y_{[s]} - Y_{[s-1]}\right)^2}}{\max_{LS-w<s<LS} \sqrt{X_{[s]}^2 + Y_{[s]}^2}} , \tag{IV.2-38}$$

measuring the arc length of the signal, which is proportional to the signal's dominant frequency. The quotient of the sample rate (*sr*) and window length in samples (*w*) normalizes the duration. The factor 4 relates to the fundamental arc length of a 1Hz sine with unit amplitude. Since the arc length also scales with the amplitude, we further normalize by the highest amplitude observed in the 1s interval. Due to this correction, the approximation decreases for both, a signal whose frequency decreases as also for a signal whose amplitude increases. Both are characteristic for the S-phase arrival.

The optimal thresholds were determined by a multi-dimensional grid-search to

$$\begin{aligned} &\theta_E = 0 \\ &\theta_{dE} = 0.15 \ \tfrac{1}{s} \\ &\theta_R = 0.3, \ \theta_{dR} = 0.11 \ \tfrac{1}{s} \\ &\theta_A = 1.2 \\ &\theta_{df} = -0.02 \ \tfrac{1}{s^2}, \ \theta_{ddf} = 0.14 \ \tfrac{1}{s^3} \end{aligned} \quad . \tag{IV.2-39}$$

Figure IV.2:10 shows typical timelines of the individual parameters and the resulting S-potential.

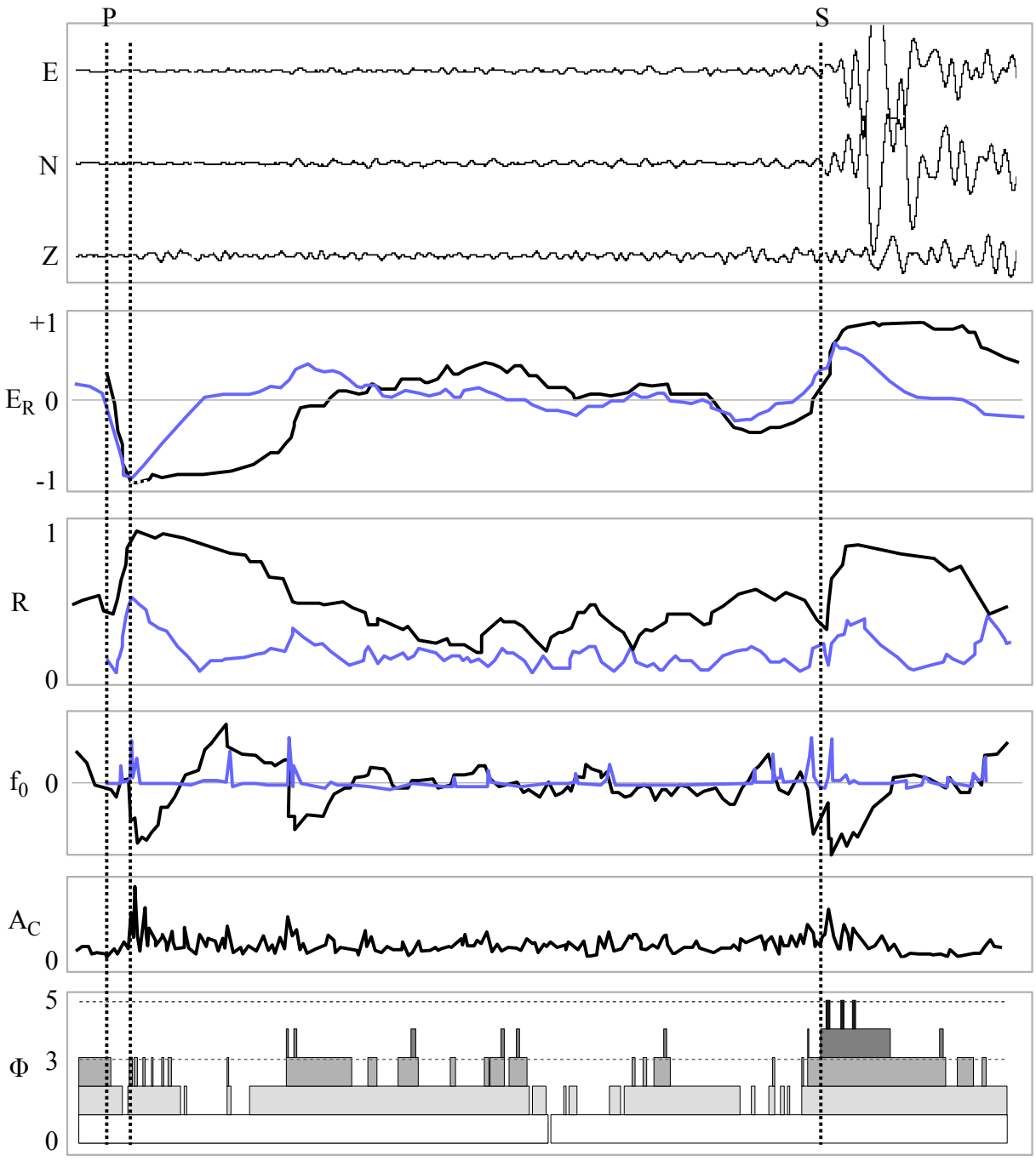


Fig. IV.2:10: Typical evolution of parameters and S-potential (bottom panel) between P-and S-phase arrival (vertical dashed line). The second vertical dashed line indicates the end of the data interval being used to rotate the traces into their Eigen-space ($|V^{\parallel}|$, $V_1^{\perp}$ and $V_2^{\perp}$). The top panel displays the three components of the recorded seismicity. The next four panels display polarization, rectilinearity, dominant frequency change and amplitude. Blue curves indicate the respective change rates. After rising from -1, the polarization fluctuates around 0 in the coda and approaches +1 at the S-arrival, accompanied by a positive growth rate. The rectilinearity shows high absolute values and change rates at both phase arrivals. The S-phase shows the expected drop in dominant frequency and two distinct peaks in the impulsiveness. The amplitude contrast is slightly raised at the arrival. The S-potential reaches its drop-out threshold (top horizontal dashed line) at the beginning of the S-phase arrival. The algorithm switches to on-hold (pausing the reporting of times) when reaching the on-hold threshold (lower horizontal dashed line), e.g. several times in the P-coda and just before the S-arrival. Since the algorithm is still on-hold when the drop-out is reached, the latest timestamp concerning the absence of S-phase energy is reported right before the visual S-phase arrival.

The use of these four parameters and their time derivatives provides assurance concerning the absence of S-phases in real-time (fig. IV.2:5) and thereby allows to assess a lower bound for the S-P interval and (via the P- and S velocity model) of the hypocentral distance. This information in turn can be used to obtain real-time estimates of the lower bound of the event magnitude, e.g. using the peak displacement of the early P-phase. (The peak displacement of the first 2 seconds following P-arrival correlates well with the event magnitude, saturating at about $M \simeq 6.5$ (Lancieri and Zollo, 2008).) For larger magnitudes, the longer rupture process requires larger time windows for proper estimates. Early estimates can therefore generally only provide a lower bound for the magnitude, as the rupture and received signal amplitudes may still grow with time. This estimate requires the knowledge of the hypocentral distance. NYAS provides minimum hypocentral distances, and supports therefore true minimum magnitude estimates at a given point in time, often already before a second P-phase detection occurs. This allows an efficient warning when the minimum magnitude exceeds the common warning threshold of $M_L = 5$ (given the confidence that the signal is indeed seismic, which can be gained by arrays). The NYAS' growing lower bound for the hypocentral distance complements the shrinking maximum bound obtained from the NYAD cell. As the growth of the NYAS sphere freezes as soon as possible S-signatures are registered ("on-hold"), it is more reliable than real time S-picks and S-P constraints, as only a lower bound is predicted and the drop-out condition may be set very low. Taking up for instance the initial example that solely relied on the polarization parameter (fig. IV.2:6 - IV.2:9) - even for a low drop-out condition of $E^R > 0$ (i.e. as much energy on the traces perpendicular to the P-trace as on the P-trace itself), significant information is added to the system at barely any risk. With this drop-out condition, prediction would have stopped in figures (IV.2:6 - IV.2:9) with a minimum hypocentral distance of 7km for a true hypocentral distance of 14km (fig. IV.2:6), 14km of true 47km (fig.IV.2:8), 16km of true 53km (fig. IV.2:10) and 33km of true 101km (fig. IV.2:9). As the magnitude depends on the logarithm of the distance ($1.6 \cdot \log_{10}(d)$) these early estimates would have caused an underestimation of less then 1 magnitude unit. This shows that NYAS can be formulated in a safe way that allows to add valuable information without adding the risk of late picks. In contrast to NYAD cells, the Not-Yet-Arrived-S (NYAS) constraint is based on single (independent) stations. It is therefore not affected by inter-station latencies and can be applied independent of the network at any given station, e.g. allowing the application in on-site-systems. Such systems may be established using small-aperture-arrays, providing location by back-azimuth and NYAS, slowness estimates for event characterization, inter-array relative amplitudes as discriminator between anthropogenic and seismic events (via the

distance), and advanced S-phase picking based on beam-steering and variations in slowness. These real-time small-aperture networks are currently developed and evaluated within the DeadSeaNet project (Wust-Bloch et. al., 2016).

IV.3 **Sequential Real-time Constraining of Hypocenter Locations**

After having derived and analyzed the different constraints (NYAD,BA,EDT,S-P) individually in the context of a real-time analysis, these can be now used combined to constrain hypocenters in real-time. The use of these (mostly independent) constraints adds robustness in case of ambiguities and is therefore well suited for an early location that is based on the few data given by the onsets of one to five stations. Evaluating latency corrected NYAD and NYAS regions, back azimuth beams, hyperbolic EDTs and S-P constraints, it integrates the search for the point of highest likelihood in space-time as well as its uncertainties. Evaluating uncertainties in picking (e.g. by multi-pick analysis) and velocity model allows to yield robust congruent regions, identifying the possible hypocenter region, by superposing the individual constraints. This allows to obtain reliable minimum and maximum distance estimates, contrasting the widely used approaches in realtime location schemes today that only identify the most probable epicentral distance by minimizing residua sums or evaluating likelihood functions (e.g. based on hyperbolic constraints) (Satriano et al., 2011). Having minimum and maximum bounds for the distance allows for a real-time assessment of lower and upper bounds of the event magnitude at a given time. Especially the minimum distance, yielding a lower limit in magnitude, is of special importance as the magnitude can only *grow* with time (e.g. for a still growing fault plane). The minimum magnitude rising over a given threshold (e.g. M_L 5) may therefore be used as a warning trigger. The focal depth is commonly fixed (to an expected value based on the catalog) until the fourth P-phase detection, when first depth estimates become available. Fig. (IV.3:1) shows an example for a typical real-time location procedure, using a shallow event scenario in the Dead Sea Region with the DeadSeaNet network configuration: The detection of a first P-onset initiates the location procedure by computing a NYAD cell, whose volume gradually shrinks with time, combined with a NYAS sphere growing with time (fig. IV.3:1a). Until the time of second detection this distance interval narrows. Only fractions of seconds after the first detection, a first back-azimuth beam can be formed to intersect NYAD edges, thus often reducing the maximal distance (fig. IV.3:1b).

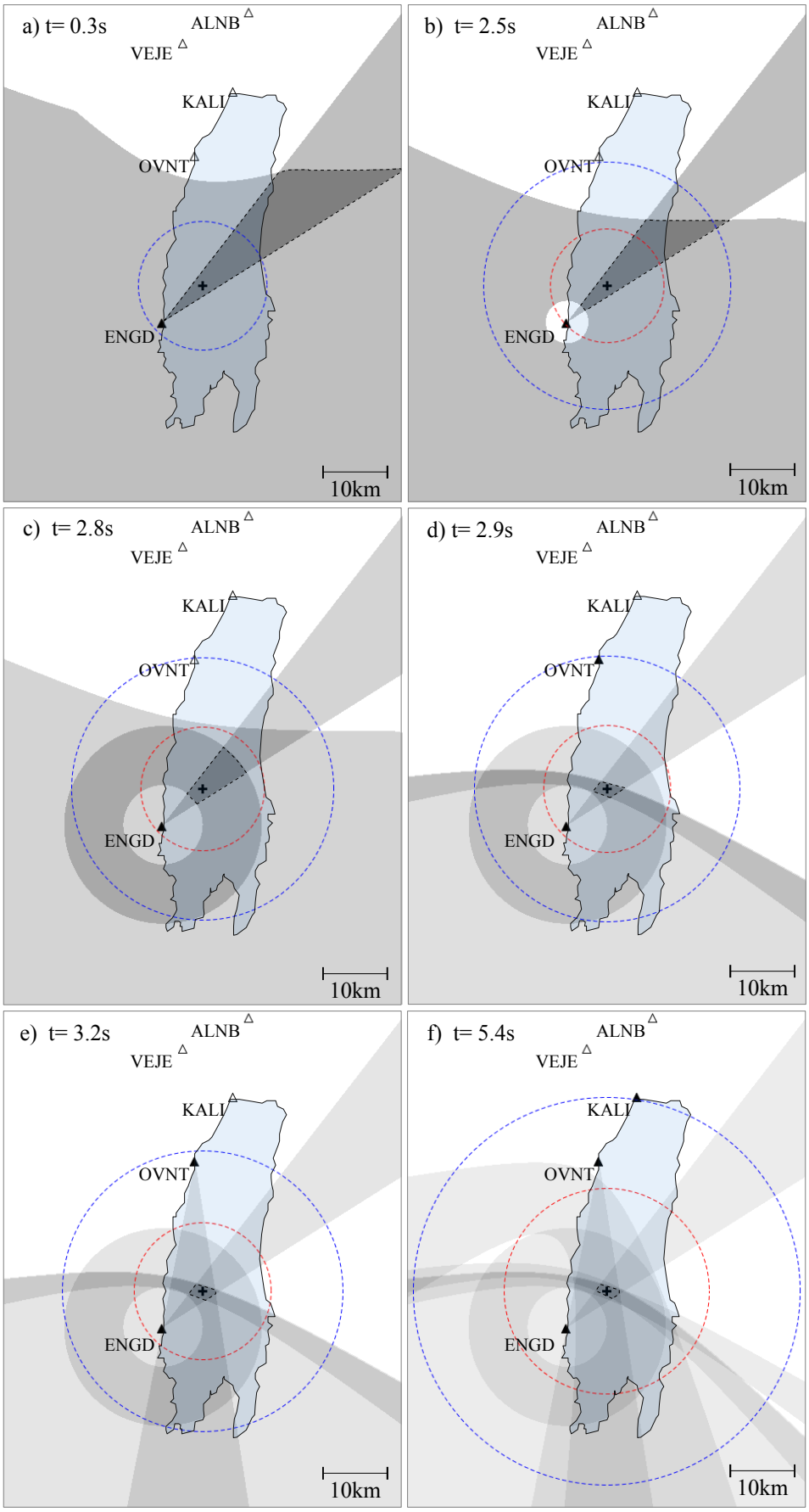


Fig. IV.3:1: Evolving location (dashed region) by superposition of constraints.

a) t=0.3s. A third of a second after first detection at ENGD, the first back azimuth limits the maximum epicentral distance in conjunction with the large NYAD region. The blue dashed circle indicates the p-wave front, propagating from the epicenter (bold black cross).

b) t=2.5s. Until the S-arrival at ENGD, the not-arrival of the S-phase spans out a growing sphere around ENGD, effectively constraining a minimum hypocentral distance. The shrinking NYAD cell further reduced the maximum distance.

c) t=2.8s. The arrived S-phase constrains a circular region around ENGD, defines an updated minimum and maximum distance, as well as maximum hypocentral depth. The intersection with the back azimuth defines a small region, enclosing the hypocenter. Also, the maximum hypocentral depth is fixed.

d) t-2.9s. The arrival of the p-wave at the second station (OVNT) creates a hyperbolic location constraint (EDT) between both stations, which in conjunction with S-P and back azimuth constrains a precise patch, containing the hypocenter.

e) t=3.2s. Shortly after, the second back azimuth obtained at OVNT stabilizes the location.

f) t=5.4s. The p-wave arriving at the third station KALI generates 2 additional EDTs, further constraining the hypocenter location.

After 2-3 seconds, first magnitude estimates based on peak velocity resp. displacement, are provided. Since a three second window contains enough spectral content to determine the magnitude up to M_L 6.5, a minimum magnitude (depending on the minimum hypocentral distance) may already be determined (and possibly exceed the warning level) at this time. Once the first S-phase arrives and can be detected with a sufficiently high confidence level, the hypocentral distance interval is fixed to a narrow interval (fig. IV.3:1c). In combination with the already determined back azimuth beam, the hypocenter location would be limited to a small region. However, often the P-phase is registered at a second (or even third or fourth) station before the S-phase reaches the first station and not every S-detection reaches a confidence level high enough to be used in automated schemes. With a detection at a second station the location is constrained to lie within an EDT region within the bounds of the further shrinking NYAD cell and growing NYAS sphere (fig. IV.3:1d). The first back-azimuth had already further constrained this region. Often, the location is sufficiently constrained by EDT and back-azimuth that NYAD cell information becomes redundant. Fractions of seconds later, the second back-azimuth beam further stabilizes estimations of minimal and maximal epicentral distance (fig. IV.3:1e). With three P-detections, three hyperbolic constraints exist, which intersect in (at least) one common region (fig. IV.3:1f). Inside the network this region is usually precise, outside the network it may be extensive. In both cases, multiple ambiguous congruent regions may exist at this stage for the hyperbolic system, but are usually resolved by the back-azimuth information.

Two real-time scenarios of California events are presented in the following chapter, demonstrating the performance gain that is obtained by harnessing the various constraints (especially back-azimuth), over a standard P-phase based location.

V **Application Examples**

This chapter will provide several examples of the application of distinct constraints for earthquake location, demonstrating the different techniques and measures, derived in the previous chapters. The first example will describe the analysis of a weak event that was recorded at the Voelkersen Gas field, Nov 22nd 2012. Here the the uncertainty driven analysis will be contrasted with the classical approach, further demonstrating the importance of 3D-models and proper uncertainty measures. In the second example an anthropogenic event will be used to demonstrate the analysis for highly ambiguous phase arrivals. While in the first two examples the data of up to twelve stations is evaluated, the third example will focus on a scenario working with a minimal amount of given information in the field of real time location for earthquake early warning.

V.1 **Voelkersen Gas Field Event**

On Nov 22nd, 2012 an event of Magnitude M_L2.9 was detected in the Voelkersen Gas field. The official report (Bischoff et al., 2013) located the event to 52,97° North ± 1,6 km, 9,212° East ± 2,5 km in a focal depth of 4,9 km ± 1,6 km using HYPOSAT (Schweitzer, 2001) with a 1D-layer-model. The gas field consists of a sandstone layer in 4.8km depth, from which $1.2 \cdot 10^9 m^3$ gas are extracted every year. The continual decrease in pore-pressure within the reservoir leads to stress changes which may induce seismic events. According to the report, the event seems to be located in the depth of the gas field sandstone layer. Since 2008 three other events have been located in the vicinity of the gas field with Magnitudes between M_L1,9 and M_L2.8 for which the focal depths were determined to 5-10km, 4 ± 5 km and the third being undetermined. The following analysis serves to demonstrate the concept of distinct constraints, side by side with residua estimates and classical jackknifing, demonstrating the impact that stations at different depths, different models and uncertainty measures exert on the solution. It is further meant to visualize the uncertainties in location that we usually overlook. The report (Bischoff et al., 2013) was based on a 1D-layer model, which is derived from Dahm et al. (2007). Here, we will take this model as starting point and will finally constrain the solution using a now given 3D-velocity model.

V.1.1 **Data**

The used dataset consists of 12 stations of velocity and accelerometer sensors, on which P- and S-phase onsets are manually picked. While p-phases can be determined with high accuracy, the S-phases are often masked. After rough p-picking, a pre-location is performed.

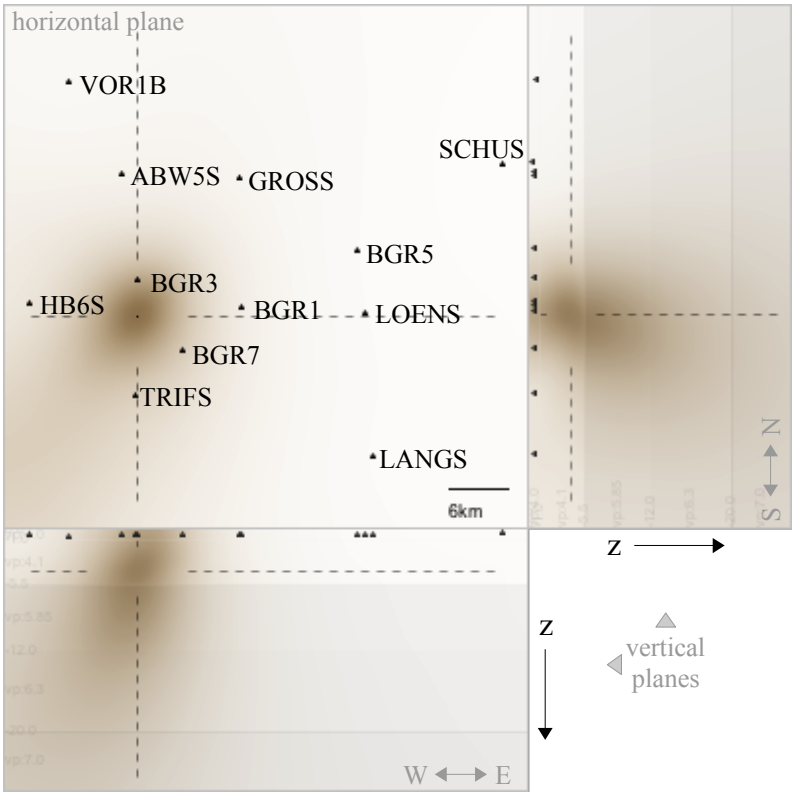


Fig. V.1.1:1: Pre-location. RMS plot. A pre-location based on preliminary P-picks allows to sort stations by epicentral distance for better phase identification by visual correlation..

This allows to sort the the stations by their (approximate) epicentral distance, providing the analyst the possibility to follow the phases through the expanding wave field, yielding a better phase identification. The epicentral distances range between 6km and 43km. Several traces (e.g GROSS, fig. V.1.1:3 marked by dashed box) show a clear bi-modal wave-package in the S-phase. Tracking this signal pattern in traces where the S-phase can hardly be identified (eg. BGR3), provides also there for a rough identification. In some of those cases, however, we need to settle with a rather large picking-interval than with a pick of a specific onset (e.g. TRIFS).

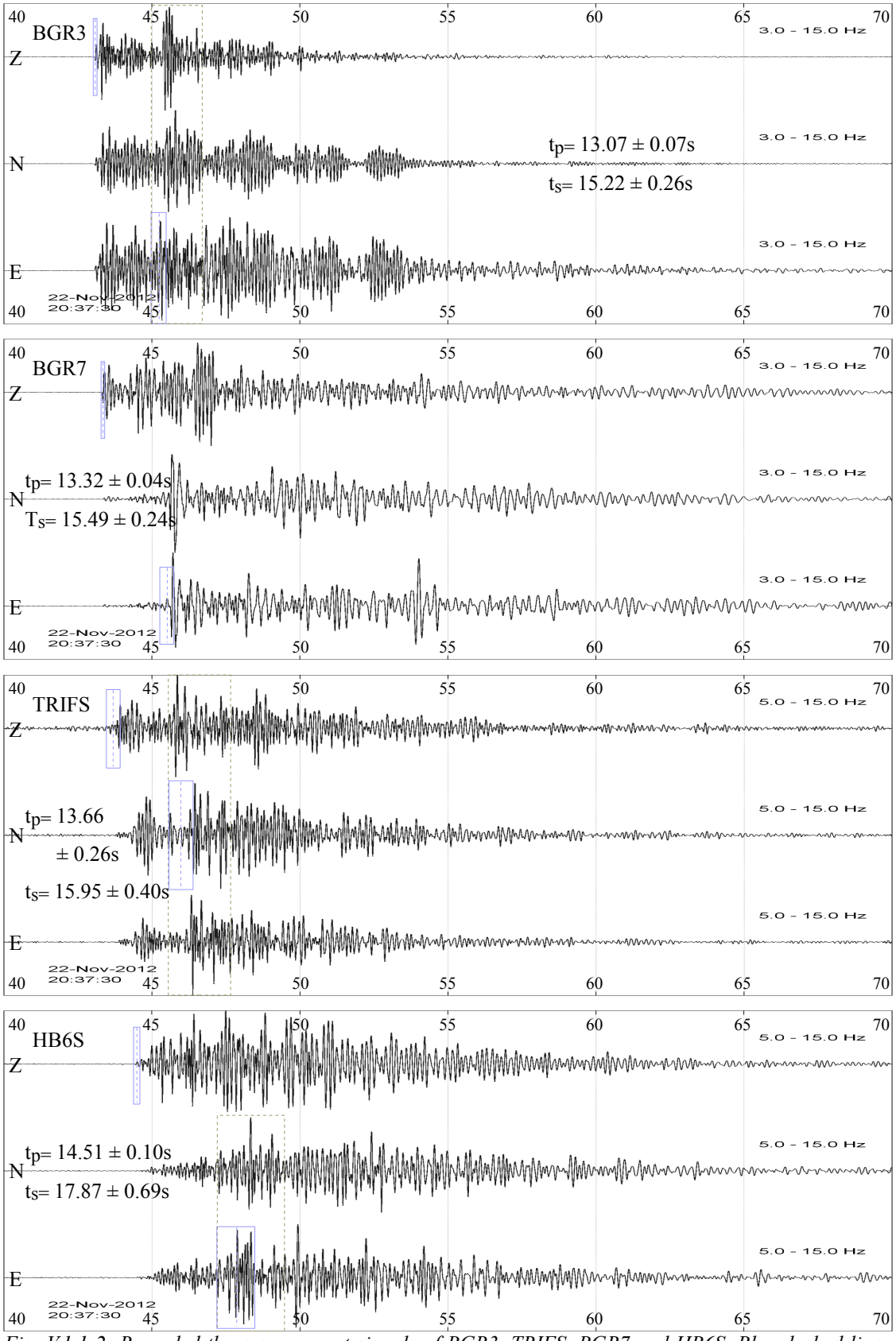


Fig. V.1.1:2: Recorded three-component signals of BGR3, TRIFS, BGR7 and HB6S. Blue dashed lines mark the phase onset, resp. the center of the uncertainty interval (blue box). Dashed boxes mark the pattern of the S-phase wave package.Time axis is in seconds after 20:37:30.

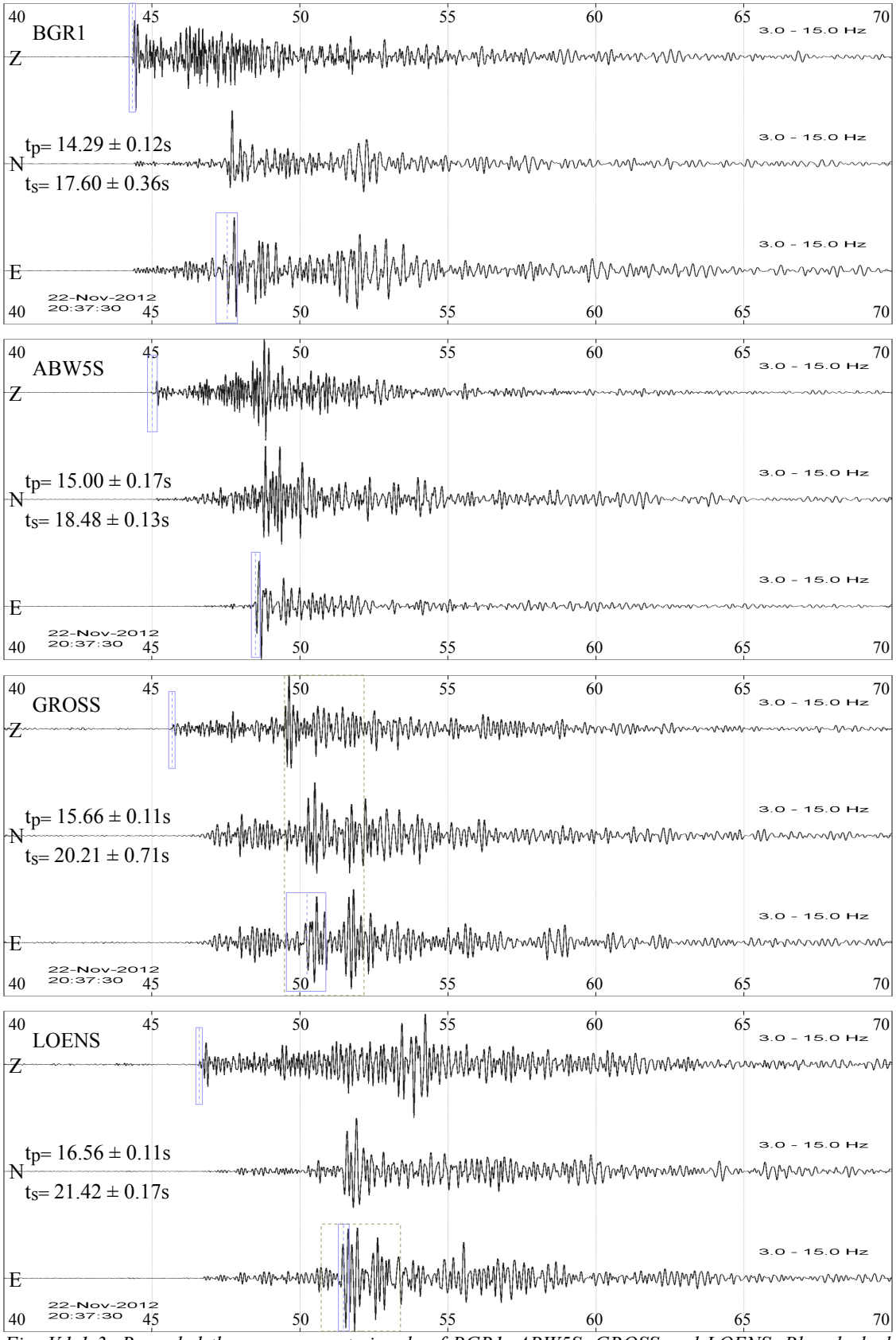


Fig. V.1.1:3: Recorded three-component signals of BGR1, ABW5S, GROSS and LOENS. Blue dashed lines mark the phase onset, resp. the center of the uncertainty interval (blue box). Dashed boxes mark the pattern of the S-phase wave package. Time axis is in seconds after 20:37:30.

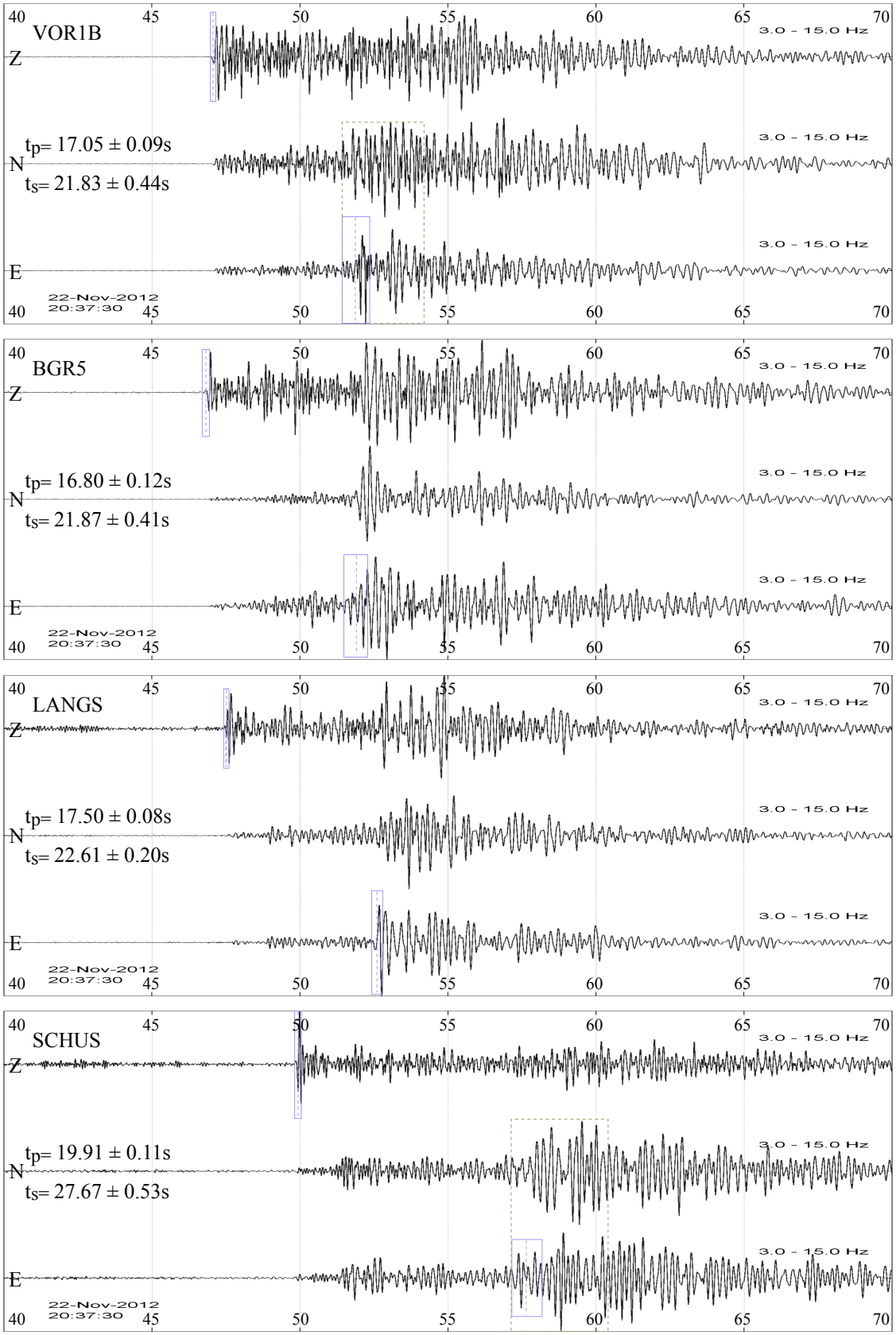


Fig. V.1.1:4: Recorded three-component signals of VOR1B, BGR5, LANGS and SCHUS. Blue dashed lines mark the phase onset, resp. the center of the uncertainty interval (blue box). Dashed boxes mark the pattern of the S-phase wave package. Time axis is in seconds after 20:37:30.

As the system is well-overdetermined, it is sufficient to treat picking-uncertainties as intervals rather than Level-of-confidence curves. As the stations are part of different networks, we will make the assumption that all digitizers are synchronized and that the error (due to e.g. unsynchronized drifting data loggers) is well below the picking uncertainty. Practically this would be evaluated by analysis of teleseimic signals, allowing the time shifts be corrected.

Station ID	lat	lon	elevation	P-onset	S-onset	dP [s]	dS [s]
ABW5S	53.113699	9.223998	23	20:38:15.00	20:38:18.48	0.17	0.13
GROSS	53.109322	9.399226	20	20:38:15.66	20:38:20.21	0.11	0.71
HB6S	52.999569	9.087641	28	20:38:14.51	20:38:17.87	0.10	0.69
LANGS	52.860865	9.591874	39	20:38:17.50	20:38:22.61	0.08	0.20
LOENS	52.987822	9.582618	60	20:38:16.57	20:38:21.44	0.11	0.17
SCHUS	53.119762	9.787329	87	20:38:19.91	20:38:27.67	0.11	0.53
TRIFS	52.916618	9.243451	20	20:38:13.66	20:38:16.00	0.26	0.10
VOR1B	53.196064	9.147802	-173	20:38:17.05	20:38:21.83	0.09	0.44
BGR1	52.99475	9.39985	[0]	20:38:14.29	20:38:17.60	0.12	0.36
BGR3	53.01995	9.24733	[0]	20:38:13.07	20:38:15.29	0.07	0.17
BGR5	53.04473	9.57024	[0]	20:38:16.80	20:38:21.87	0.12	0.41
BGR7	52.95723	9.3143	[0]	20:38:13.32	20:38:15.49	0.04	0.24

Table V.1.1.1: Station data; dP and dS specify the picking uncertainties for P- and S-phases. Vor1B is a borehole station. Depths for BGR stations were undefined in the data set. Based on topography their elevation would range between 41m-63m. Due to the large station-to-station distances the depth error is insignificant for the analysis.

V.1.2 **P-P based Location**

Classical graphical Jackknifing would now search for the region of highest line density (fig V.1.2:1a), disregarding the timing uncertainties. Even so, using all given data, the epicenter can be roughly identified. Now available vertical cross sections also reveal the agreement over depth, indicating the focal depth, but a quantitative description of the hypocenter region in this manner is difficult. A close look reveals that not all constraints agree with the solution: Choosing a small subset of data given in the onsets of BGR1, BGR3, BGR5, LOENS and SCHUS for better overview, shows that the point of maximal line density stands in discrepancy to several constraints. Re-evaluating the discrepant constraints concerning their phase onsets shows that the P-onsets are non-ambiguous. In classical Jackknifing the analyst would now vary the onsets of discrepant constraints to estimate their impact on the constraint shape and possibly bring the constraints into agreement. This sensitivity analysis can now simply be bypassed by

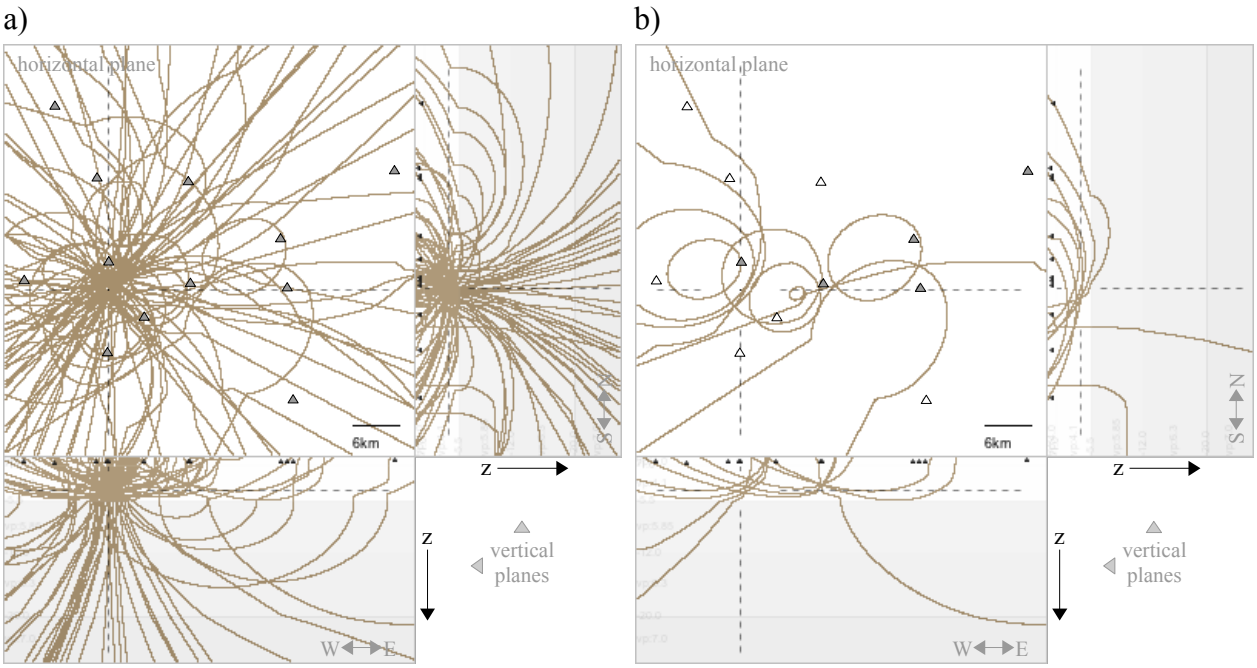


Fig. V.1.2:1: Cross-sections of exact hyperbolic constraints of a) all P-phases b) the subset of BGR1, BGR3, BGR5, LOENS and SCHUS. The point of highest line density in (a) stands in discrepancy to several individual constraints (b).

applying the assigned picking uncertainties of fig. V.1.1:2 - V.1.1:4:

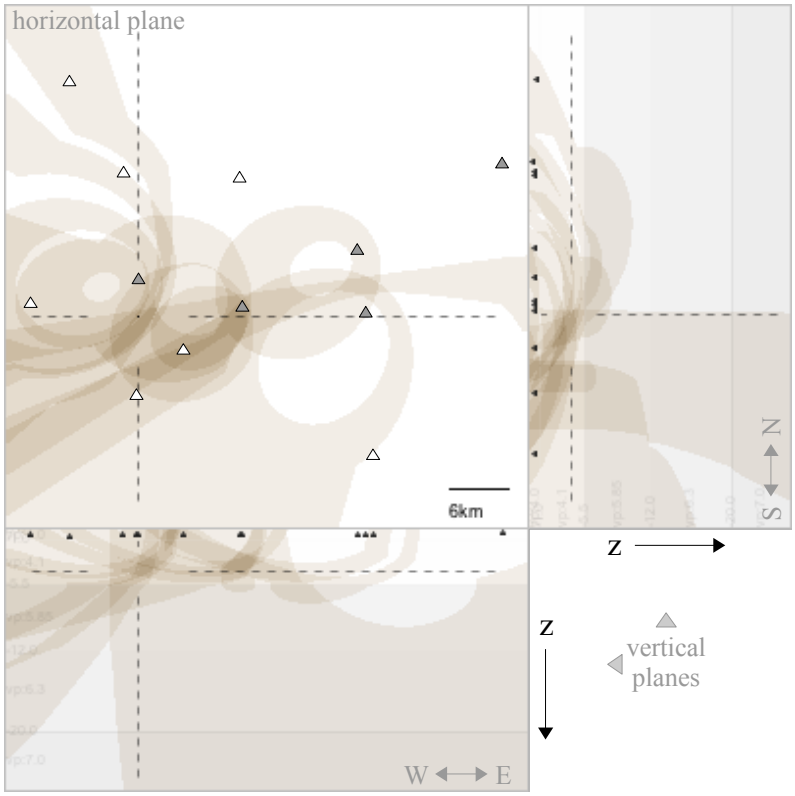


Fig. V.1.2:2: Applying picking uncertainties widens the constraints. Although this increases overlap and agreement of constraints, no region of common agreement is found. The color intensity scales with the amount of agreeing constraints.

This widens the constraints to cover all regions that would have been covered by varying all onset picks in all combinations within the assigned uncertainty windows. Although this creates here regions of increased constraint agreement, no feasible region can be found. With properly assigned picking uncertainties and the station locations being properly known, the error should be found in the used velocity model. Most inconsistent constraints can be linked to the most distant North-Eastern station SCHUS, which indeed indicates inconsistent model assumptions. The next step is therefore to include velocity model uncertainties. Although more detailed information is available, we will now apply a simple globally constant model uncertainty: For the selected subset of stations, first congruent regions appear with a 1% uncertainty in velocity (fig. V.1.2:3a). To obtain a point for which the constraints of all 12 stations agree, a 3% uncertainty is required (fig. V.1.2:3b, only the subset of stations is shown). The congruent zone now describes all the locations that are possible by variation of onsets and velocity model. Looking only at the subset of stations, this region is extensive (unlimited) (fig. V.1.2:3b).

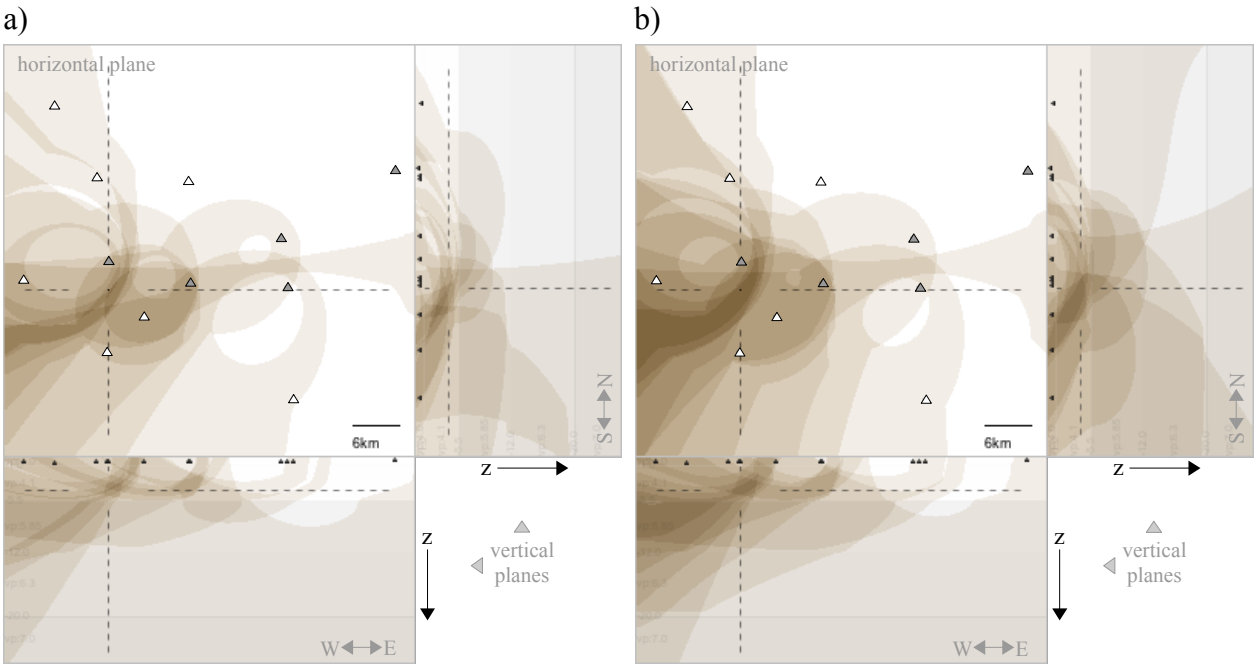


Fig. V.1.2:3: Constraints with picking- and model uncertainties. a) 1% model uncertainty b) 3% model uncertainty. For (a) small clusters of congruent agreement appear. To satisfy the constraint conditions of all twelve stations a 3% velocity uncertainty is required. For the subset of stations this yields an extensive zone of congruency.

Adding the constraints of stations BGR7 and VOR1B reduces the congruent region (fig. V.1.2:4) significantly. Yet, the current subset of solutions is located on one side of the location, which poses an ill-conditioned problem (as given for an event right outside the network edge): The solution is robust in azimuth but not in distance.

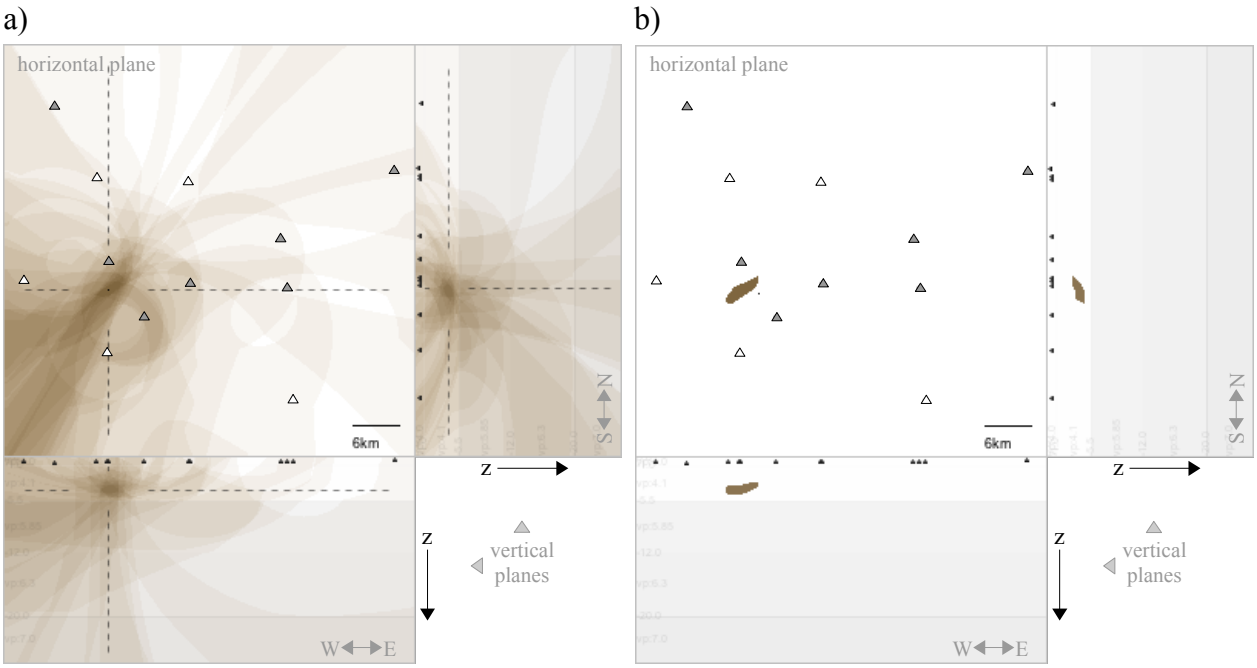


Fig. V.1.2:4: Added stations BGR7 and VOR1B. a) cross sections of constraints. b) projection of congruent zone. The depth is well constrained.

The congruent region (in fig. V.1.2:4b) shows that the solution is already well constrained in depth. This is due to the refractor at 5.5km depth where the velocity jumps from 4.1km/s to 5.85km/s, which limits certain hyperboloids (e.g. P^{BGR1}-P^{SCHUS}) to the space above. Using the complete set of constraints the mapping takes over the character of a discrete PDF with a localized maximum region of constant likelihood (fig. V.1.2:5a). The congruent zone is a compact sharp region. However, due to the lack of additional data, the uncertainty in the velocity model is unknown. Until now, a homogeneous uncertainty in velocity of 3% was assumed. This leads to coherent agreement of all hyperbolic constraints, resembling the point of maximum likelihood for location, yet, it does not provide a reliable extent of the hypocenter region (i.e. location uncertainty). The model uncertainty might be well larger which would lead to larger regions of constraint agreement. Indeed, strong perturbations exist in the first layer boundary, which resembles a geological fold. Due to the fold-structure, several regions are assigned the wrong (layer) velocity by the 1D-model. Increasing the uniform velocity uncertainty from 3% to 10% leaves the solution well constrained horizontally but extends the congruent depth interval to [-1.6, -9.5]km (fig. V.1.2:6).

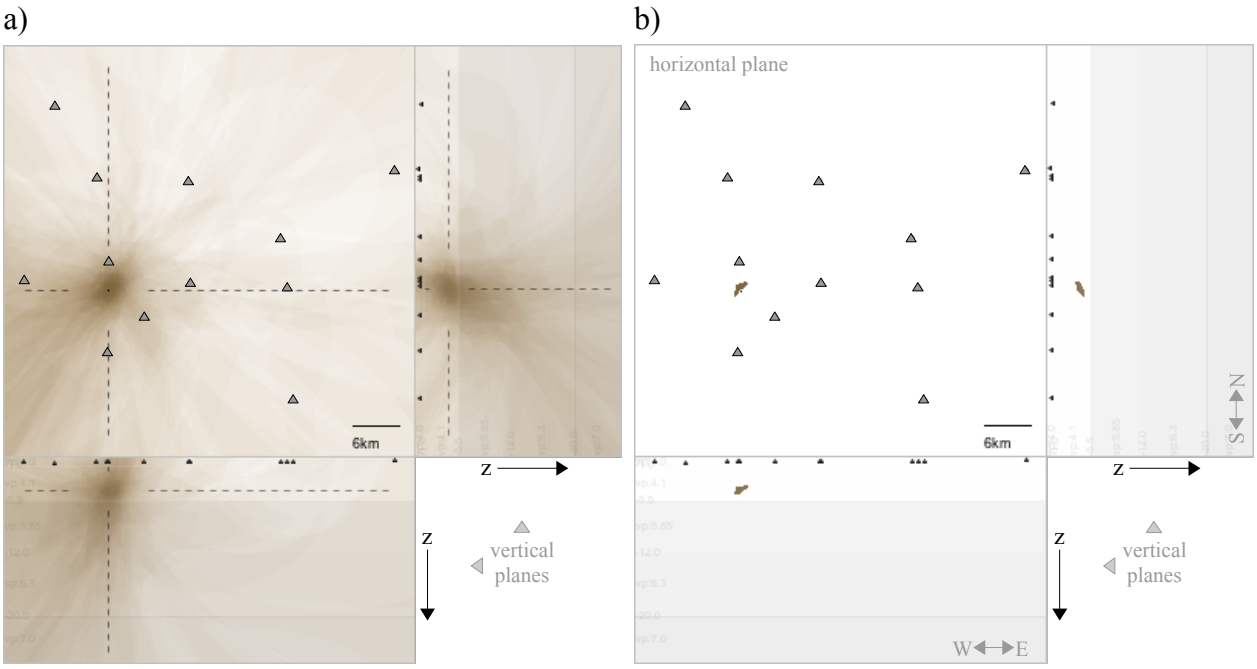


Fig. V.1.2:5: Solution for the complete set stations for 3% variation in velocity model. Similar to the first RMS solution (fig. V.1.1:1) the a) region of similar probability appears vertically stretched, but horizontally well constrained. b) the congruent zone limits the solution to a small sharp region.

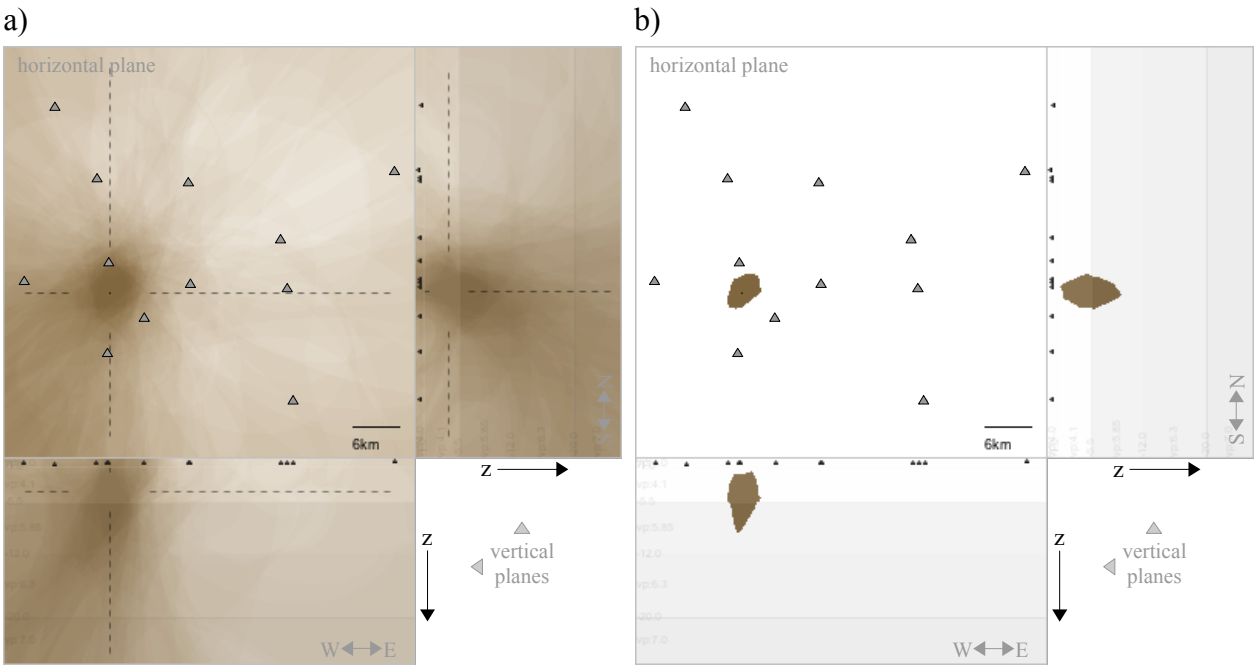


Fig. V.1.2:6: Solution for the complete set of stations for 10% variation in velocity model. a) Cross-section of constraints b) projection of larger, but constrained congruent region.

However, due to the high contrast of velocities at this border (jumping from 4.1km/s to 5.85km/s at the Zechstein-base), this leads to velocity errors of even up to 30%. Increasing the model uncertainty uniformly to this amount, would leave the solution hardly constrained as the example with *only* a 20% uniform uncertainty (fig. V.1.2:7) shows,

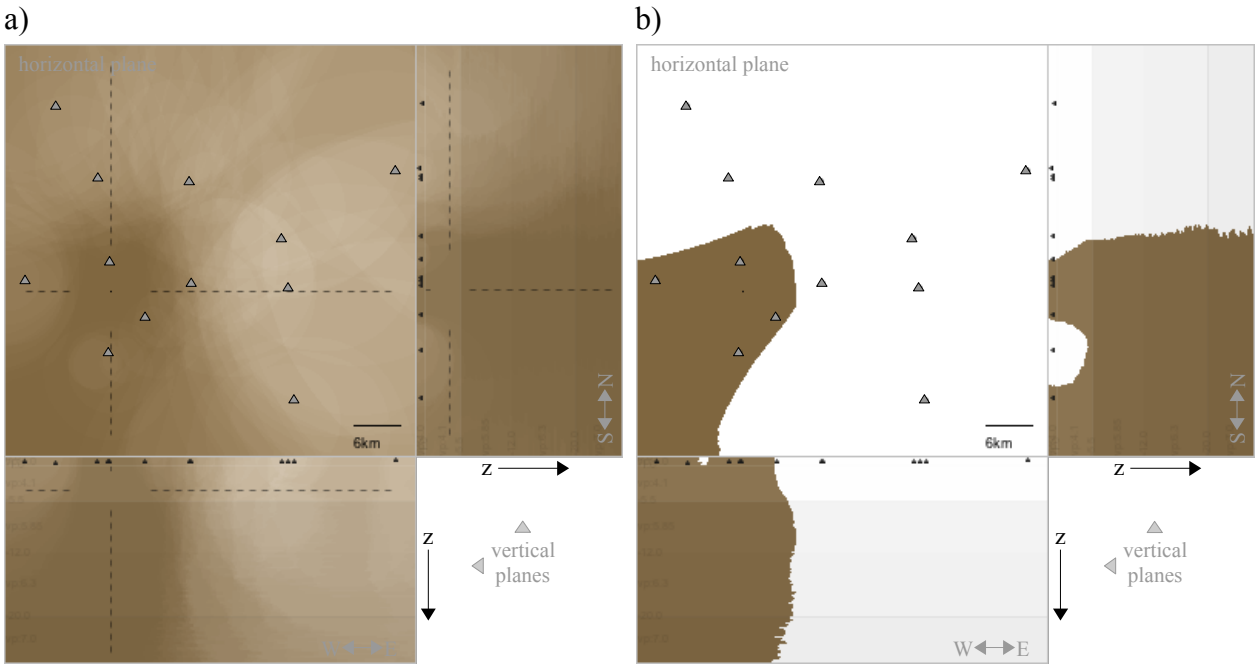


Fig. V.1.2:7: Solution for the complete set stations for a uniform 20% variation in velocity model. a) Cross-section of constraints b) projection of larger, but constrained congruent region.

providing no useful information. Always, when additional information is available, general uncertainties can be reduced and the precision increased. By increasing the precision of the model description, we can increase the precision of the location: Since we are aware of the fact that mainly the layer-boundary at -5.5km depth is affected by this strong perturbation (the Zechstein-base reaching up to -1.3km depth in a fold structure), we may assign the uncertainty to the depth of the layer boundary, rather than assigning it to the velocities in general. Additionally, the uncertainty in velocity is increased to 5% in the perturbation zone while it is left at 3% for the rest of the model. This description yields a well constrained location, showing a hypocentral depth interval ranging from surface to -6.5km (fig. V.1.2:8). This region encloses the point that will be identified as the most likely location when we use a 3D model that was made available by the industry.

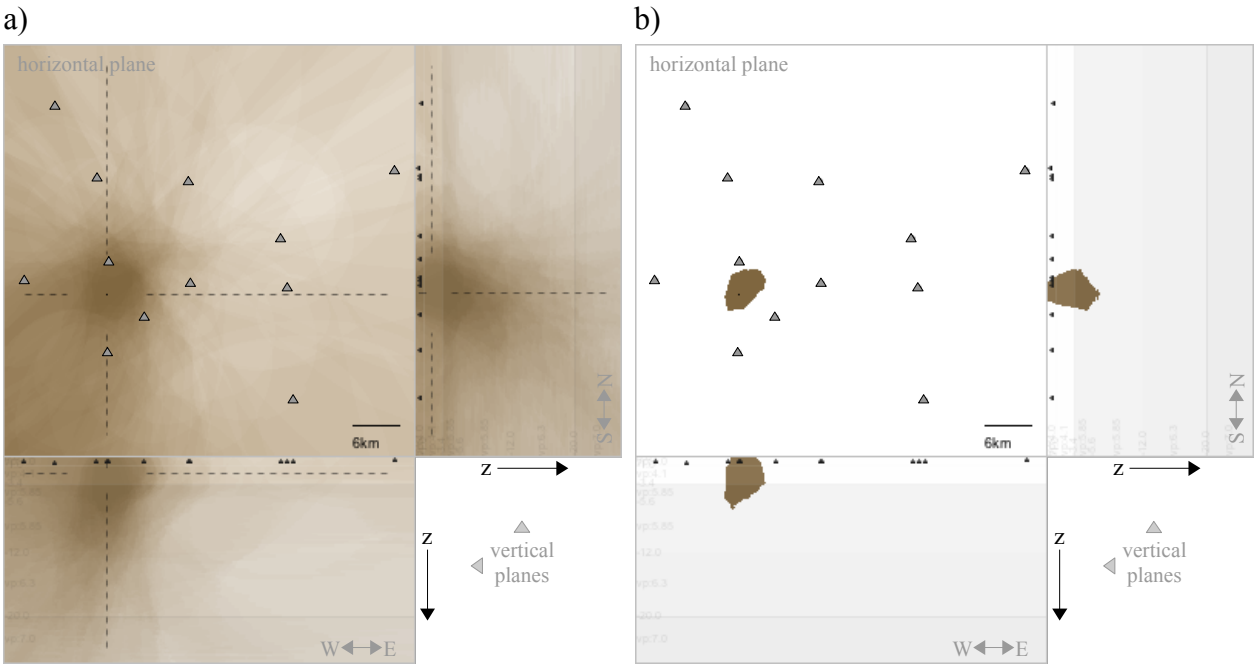


Fig. V.1.2:8: Solution for a specific variation in velocity model. The 5.5km layer depth is varied between this depth and the depth of -1.3km. In this region the velocity uncertainty is increased to 5%, the rest of the model holds a general uncertainty of 3%. The vertical extent of the coherent zone reaches from surface to -6.5km depth. The lateral a) Cross-section of constraints b) projection congruent region.

With this 3D-model for P-velocities, 100% constraint congruency is already achieved with a sub percent model uncertainty. The model error for P-velocities is expected to be less than 2%, yielding the solution displayed in fig. V.1.2:9.

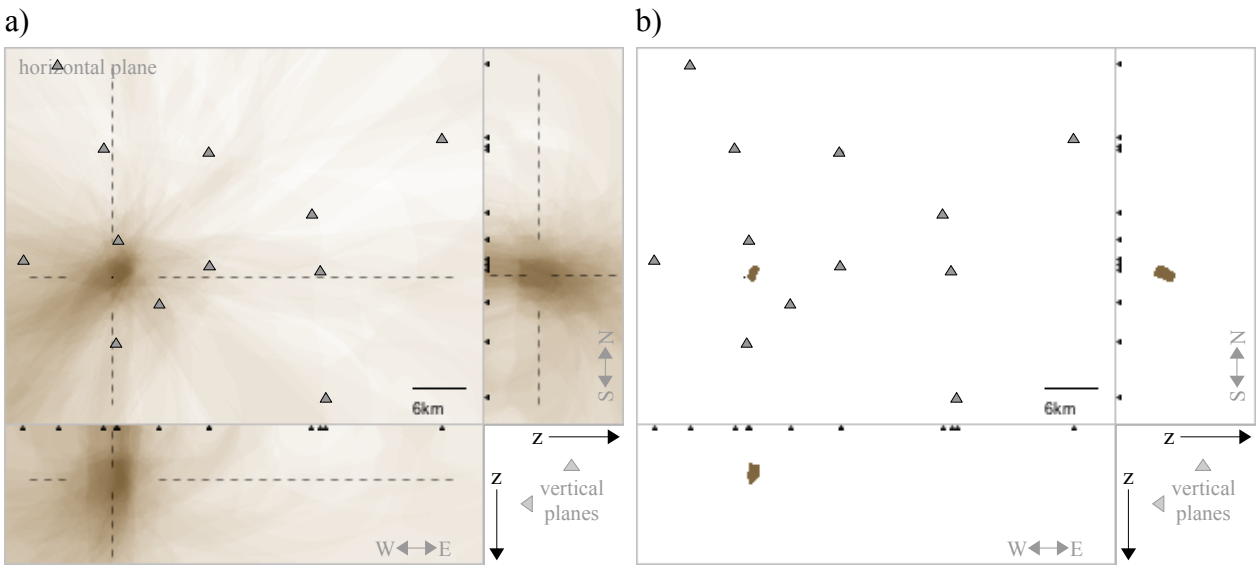


Fig. V.1.2:9: hyperbolic constraints using the 3D velocity model with 2% uncertainty. a) Cross-section of constraints b) projection of the congruent region.

To visualize the accuracy of the 3D-model, the subset of stations used in fig. (V.1.2:1), BGR1, BGR3, BGR5, LOENS and SCHUS is plotted disregarding velocity uncertainties (fig. V.1.2:10a)

and disregarding every uncertainty (picking&model) (fig. V.1.2:10b), respectively. While no agreement was found for the layered model, here the constraints form a significant congruent region only regarding picking uncertainties. Further disregarding picking-uncertainties (fig. V.1.2:10b), *most* constraints still intersect in one point, but *all* constraints lie within a circle of small radius (dashed black circle). The model and the location can therefore be trusted to be of good quality.

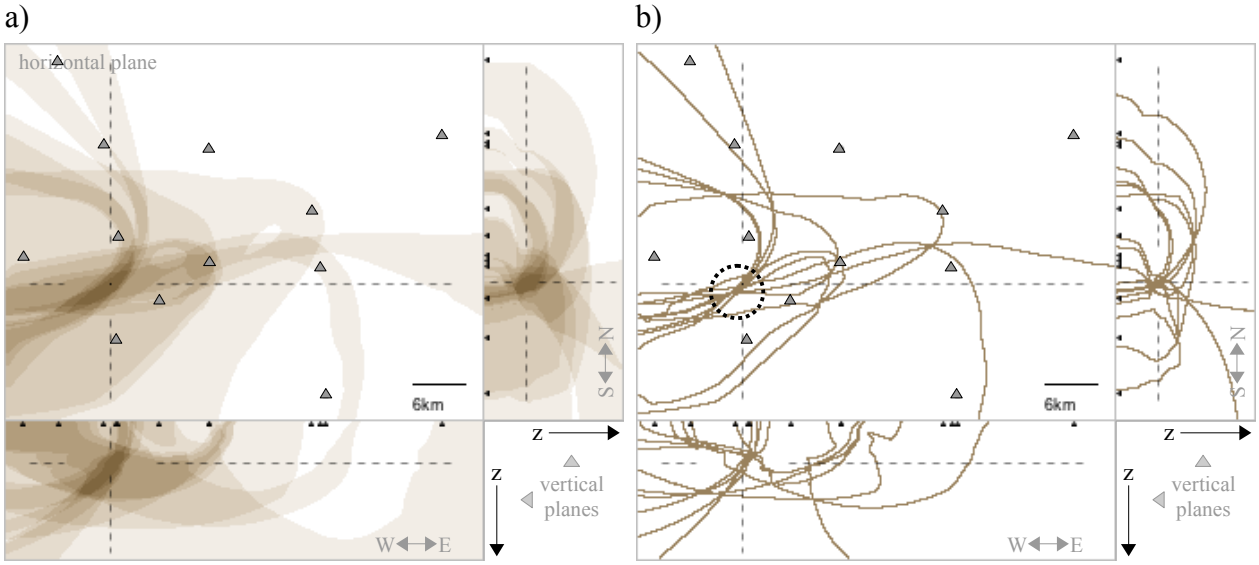


Fig. V.1.2:10: Subset BGR1, BGR3, BGR5, LOENS and SCHUS in the 3D-model a) disregarding velocity model uncertainties and b) disregarding every uncertainty. Most constraints intersect in a single point. All constraints run through the black dashed circle.

This example shows how we can use simple 1D models with simple uncertainty proxies (fig. V.1.2:8) to obtain reliable hypocentral bounds, yielding the confidence and accuracy that otherwise only much more detailed model information can provide (fig. V.1.2:10). To complete the picture, we will add the available S-phase information in the following chapter.

V.1.3 **Additional S-P Constraints**

Adding S-P constraints, we will continue to use the 3D-model. Since S-velocities were not determined in the geophysical prospecting, a higher uncertainty of 3% will be applied to them. The S-velocity-model will be based on the the P-velocity model over a fixed vp/vs-ratio. The most likely ratio will be estimated using a Wadati-diagram respecting the picking uncertainties. Using that $Tt_p = t_p - t_0$, the vp/vs-ratio can be derived from equation I.2.3-6 (chapter II.1.2.1) and be

expressed as

$$\frac{v_p}{v_s} = 1 + \frac{t_s - t_p}{t_p - t_0} \,. \qquad \text{(V.1.3-1)}$$

Averaged values from the Wadati-diagram yield a vp/vs of 1.74 for a regression curve respecting all uncertainty intervals (fig. V.1.3:1, graph A). A more likely regression, focussing on closer stations, yields a higher ratio of 1.81 (fig. V.1.3:1, graph B), which will be used for the location.

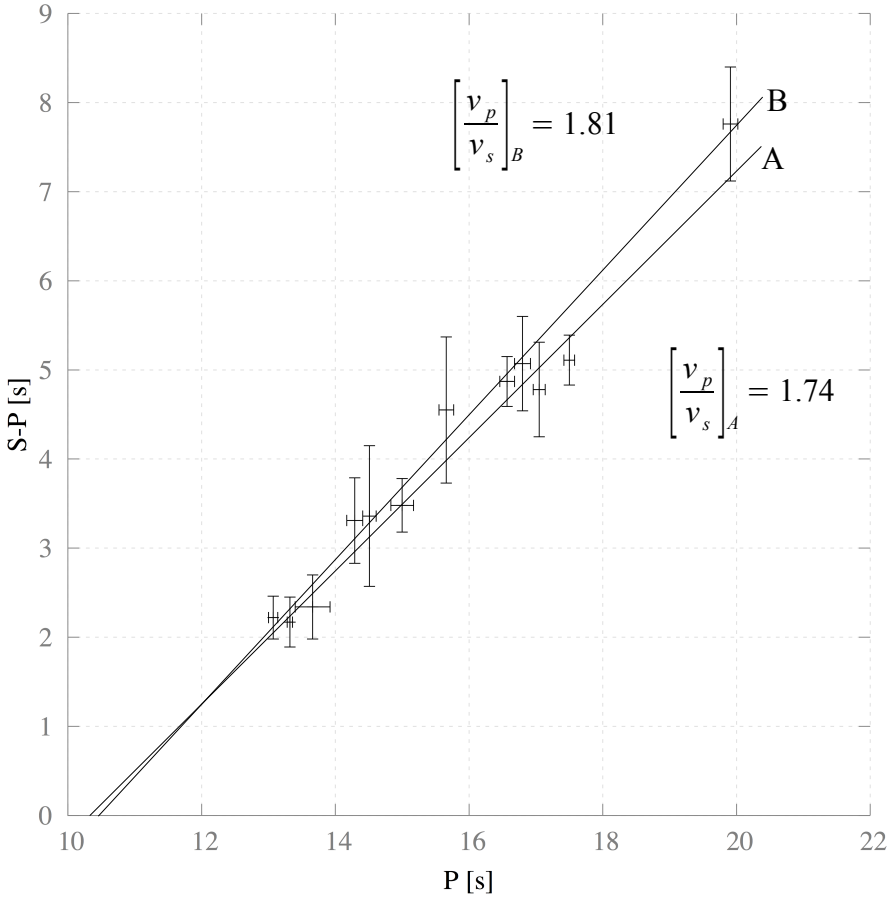


Fig. V.1.3:1: Wadati-diagram. Data is displayed indicating its picking uncertainties dP and d(S-P)=dS+dP according to table (V.1.1.1). Regression line A respects all dSP uncertainty intervals and yields a vp/vs-ratio of 1.74. A more likely regression is shown in line B, putting the focus on closer stations, yielding a vp/vs-ratio of 1.81.

As no fixed model for the S-velocity exists, we can interactively change the value of the vp/vs-ratio around the value of 1.81 to observe its impact on the solution and to check for other possible maxima in likelihood. The solution for S-P information based on this ratio of 1.81, is shown in the following figure (fig. V.1.3:2). The 3% uncertainty proves to be sufficient to achieve S-P

congruency, centered around the solution of the P-P-constraints. For smaller vp/vs-ratios the S-P congruent zone moves South-West-ward and descends in depth, in discrepancy to the precise P-P constraints. For higher vp/vs-ratios up to 1.92 the solution centers well around the P-P based location, moving North-East-ward, rising in depth while shrinking the possible congruent zone and depth interval to the one determined by the P-P solution. Above that value the congruent zone splits in several zones before it disappears. This behavior supports the most likely vp/vs-ratio derived from the Wadati-diagram. For comparison, using the layered model and uncertainty free constraints, BGR1, BGR7 and TRIFS would have required an extremely high vp/vs-ratio of 2.15 to bring the corresponding S-P constraints into agreement.

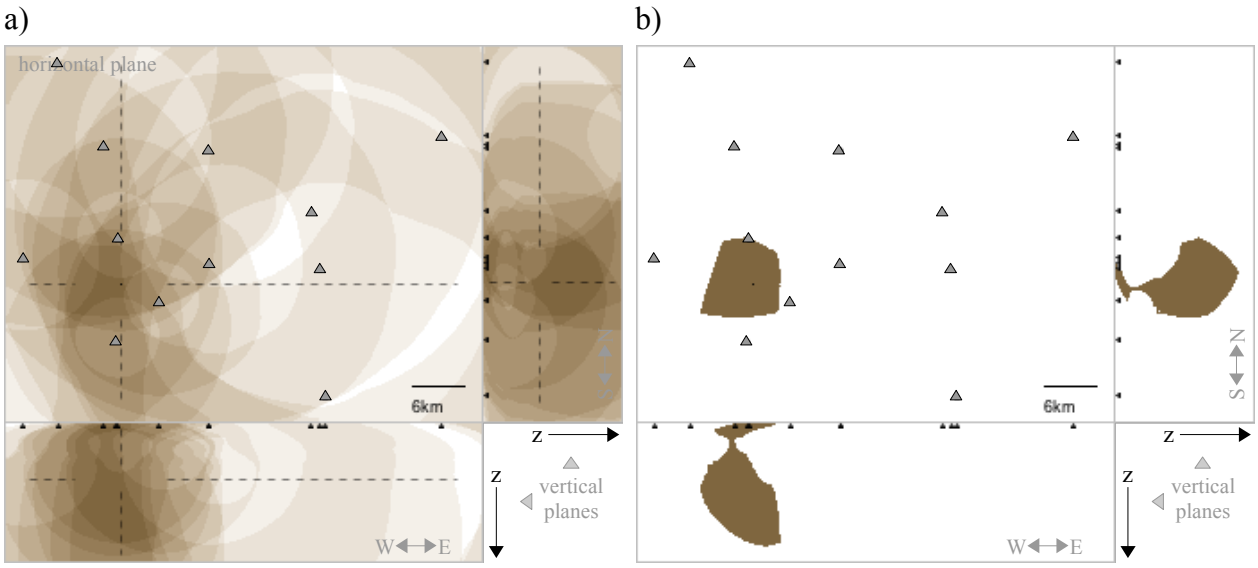


Fig. V.1.3:2: S-P constraints regarding picking and 3% S-wave velocity uncertainties for a vp/vs-ratio of 1.81. A large but well constrained congruent region is formed a) cross-sections of constraints b) projection of coherent zone.

The derived ratio of 1.81 leads to a stable congruency between both constraint groups over a range of vp/vs-ratios. The larger uncertainties in S-picking and S-velocities causes the congruent S-P region to be significantly larger than the one solely based on P-onsets. Fig. V.1.3:3 shows the overlap of both regions. While varying the vp/vs-ratio, both congruent zones never fully overlap. The zone of common congruency (for both S-P and P-P) corresponds to the Southern part of the congruent region of the P-P constraints (fig.V.1.3:3b, S-N Panel).

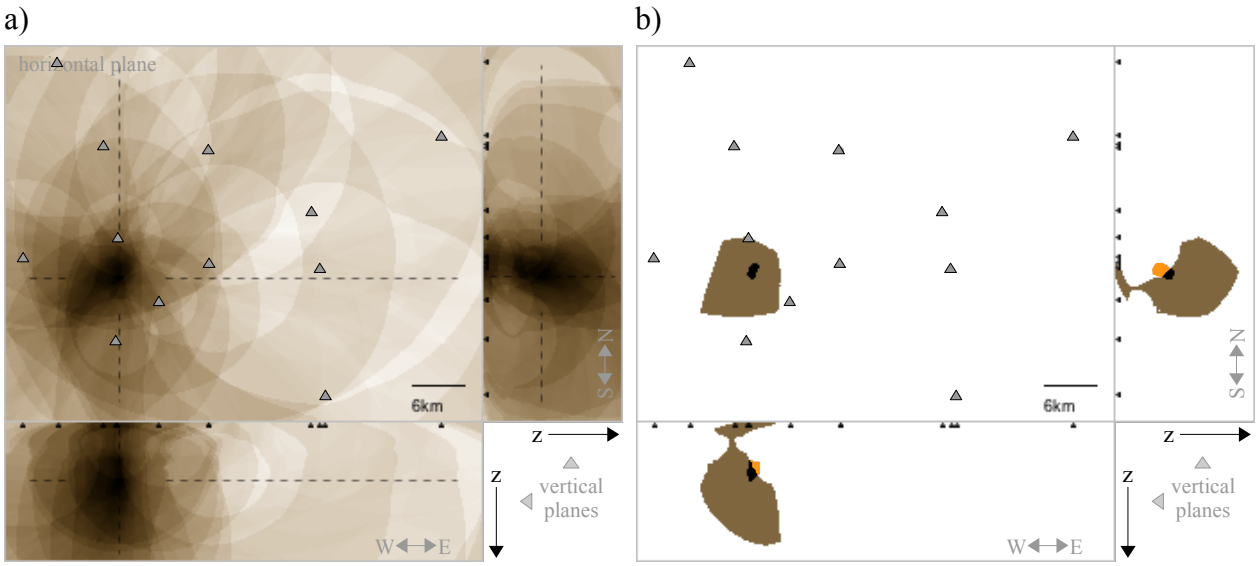


Fig. V.1.3:3: S-P and P-P constraints with 2% P-wave velocity and 3% S-wave velocity uncertainties. a) cross-sections of constraints b) projection of congruent zone. P-P coherency is marked orange for visibility purposes. The overlap region of both constraint groups is black.

For the most likely vp/vs-ratios, the intersection of both zones further corresponds to the deeper end of the P-P region (fig.V.1.3:3b, S-N Panel). The point most trusted lies at [52.979115, 9.248010] in a depth of -6.3km. Figure (V.1.3:4a) shows the overlap of the congruent regions at this point. An increase of model uncertainties of P to 3% and S to 4% displays the stability of this solution (V.1.3:4b).

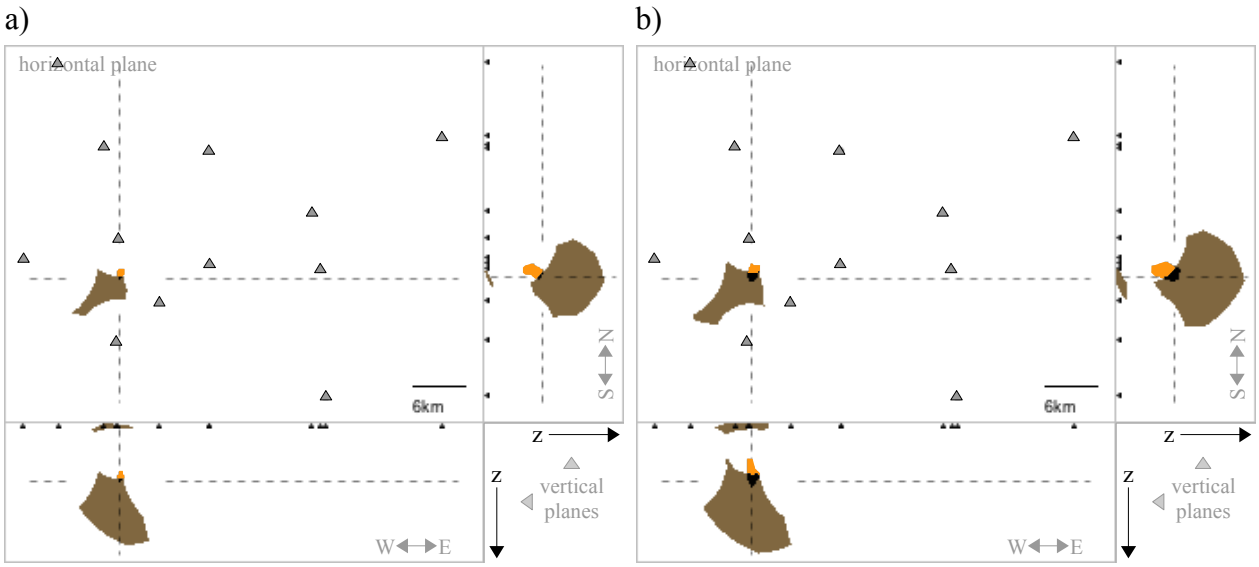


Fig. V.1.3:4: Overlap of coherent regions at the chosen hypocenter. a) model uncertainties of 2% (P) and 3%(S). b) Test for solution stability and solution uncertainties by increasing constraint model uncertainties to 3%(P) and 4% (S).P-P coherency is marked orange for visibility purposes. The overlap region of both constraint groups is black.

The congruent region for P-P constraints reaches from -4.4km to -6,7km for a 2% P-velocity uncertainty. The congruent region for combined S-P and P-P constraints assuming a 3% S-velocity uncertainty reaches from -6.1 to 6.7 km depth. This solution represents the collection of points that satisfies both constraint groups, marking all locations that are possible within the uncertainty margins. However, while a significant region of the P-P region intersects with the S-P region, this is not the case vice-versa. While the point of highest probabilities for the P-P constraints lies close to that intersection region, it lies south-west and deeper for the S-P information. This signifies that the S-P information based on the velocity model assumptions is not very trustworthy. Higher vp/vs-ratios like 1.87 would lead to a full coverage with the congruent P-P region. Likewise would a S-model variation of 6% do: Similar to the case of the P-based location in the layered model, the given S-uncertainties do not allow to further constrain the P-based location in a reliable way, but only to stabilize the solution. The impact of the S-P information to constrain the hypocenter would have been significantly larger if the event was situated “outside” the network. More precise model information could possibly be obtained by assigning material dependent vp/vs ratios to the layers described in the 3D-model using core hole data. However, especially within the reservoir the ratio would still vary by several percent, dependent on pore pressure changes. Here, we will therefore base the event interpretation solely on the P-P region. This yields a most probable hypocenter at 52.979° North, 9.248° East, -6.3km depth and a reliable hypocenter interval [52.977°, 52.990°] North, [9.242°, 9.257°] East, [-6.7km, -4,4km] depth. This corresponds to horizontal spans of 1.5km (N), 1.1km (E) and a vertical span of 2.3km (z).

The official report based on the 1D model (Bischoff et al., 2013) located the event to 52,973° North ± 1,6 km, 9,212° East ± 2,5 km in a focal depth of 4,9 km ± 1,6 km, 2.5km west of the location determined here. It places the location dominantly above the modeled Zechstein-base (-5.5km depth). The location determined by the use of the 3D model including the location uncertainties determined by the distinct constraints places the hypocenter well below the Zechstein-base, which actually lies at -4km depth in this area. This seems to be supported by the wave forms, since no significant splitting in P-onset due to refracted and direct travel paths can be observed. For an event just above the base (as reported), the time difference ($P_{ref} - P_{direct}$) would go up to approx. 2.5s for the stations in largest distance.

V.2 **Probabilistic Multi-Pick Analysis**

The following location example pertains to the class of on- or near surface events of anthropogenic nature, typically given in impact events (e.g. airplane crashes, structure collapses), im- or explosions. These weak events (M<2) usually only generate one usable phase, and can often be identified by the nonexistence of the typical spectral energy signature consisting of an impulsive high frequent P-onset, followed by S-energy of lower frequency. If no arrays are given, location will be predominantly based on hyperbolic location. Being generated near surface, the highest observed amplitudes are carried by surface waves, while phases of weaker induced body waves may not be visible. Surface (Rayleigh-Lamb) waves hold velocities below the shear velocity and are slightly affected by the velocity model. This type of "Ground Roll" surface wave travels with approximately 92% of the near surface shear wave velocity (Xia et al., 1999)

$$v_{surface} = 0.92 \cdot v_{shear} \,. \tag{V.2-1}$$

Due to the lower velocity, location is less affected by picks off the true onset as it would be the case for P-phase based location. Additionally, the amplitude of surface waves decays slower with distance ($A \sim r^{-0.5}$) than for body waves. However, due to these events' low magnitude, they are often only seen on few stations and might be masked by local events or disguised by local noise bursts. Correlated phases are usually identified by their signal pattern, showing similarity in waveform or spectral signature. For low SNR conditions this signature may be not identifiable at every station, and ambiguities might exist in the phase onset determination. However, phases do not only show their correlation by signature but are linked by their common source. Being able to account for velocity model (and picking) uncertainties, the set(s) of truly corresponding phases and the most probable location may be found by searching for (all) the combination(s) of phases that can be mapped to a common source - in a single step using a multi-pick-analysis (chapter II.3.4): The robust velocity model is hereby used to "filter out" (most) improper combinations (usually by arrival-time differences, too large to be consistent) and is "associating" the set(s) of reasonable phases. The treatment of model uncertainties is crucial for this analysis, assuring that corresponding phases are not filtered out because of model inaccuracies. Depending on phase pattern, station distribution and velocity model, the result may appear as one single zone or as a cluster of distributed congruent zones. When several zones are shown, then these depict a macro uncertainty

in location due to the ambiguity in phase picking. This cluster may be caused by an interference of several simultaneous events, or the improper combination of uncorrelated phases in general, that yield congruent solutions 'just by chance'. With several simultaneous events, the congruent region will consist of all events' hypocenters alongside possible pseudo-locations, that are caused by incorrect combinations of phases. A forward modelling of travel times from the congruent region identifies the set of related phases and allows the analyst to re-evaluate their signal pattern concerning similarity and plausibility. [If pseudo-locations exist, the forward modelling from one of such would identify a pseudo-associated set of phases. These can, however, often be ruled out due to inconsistent signatures, amplitudes, etc... .]

We will now pick up the example given at the beginning of this thesis (chapter I.1), in which the true phase (that originated from the unknown source) can not be clearly identified on all traces. This event is the third in a sequence of signals (fig. V.2:1).

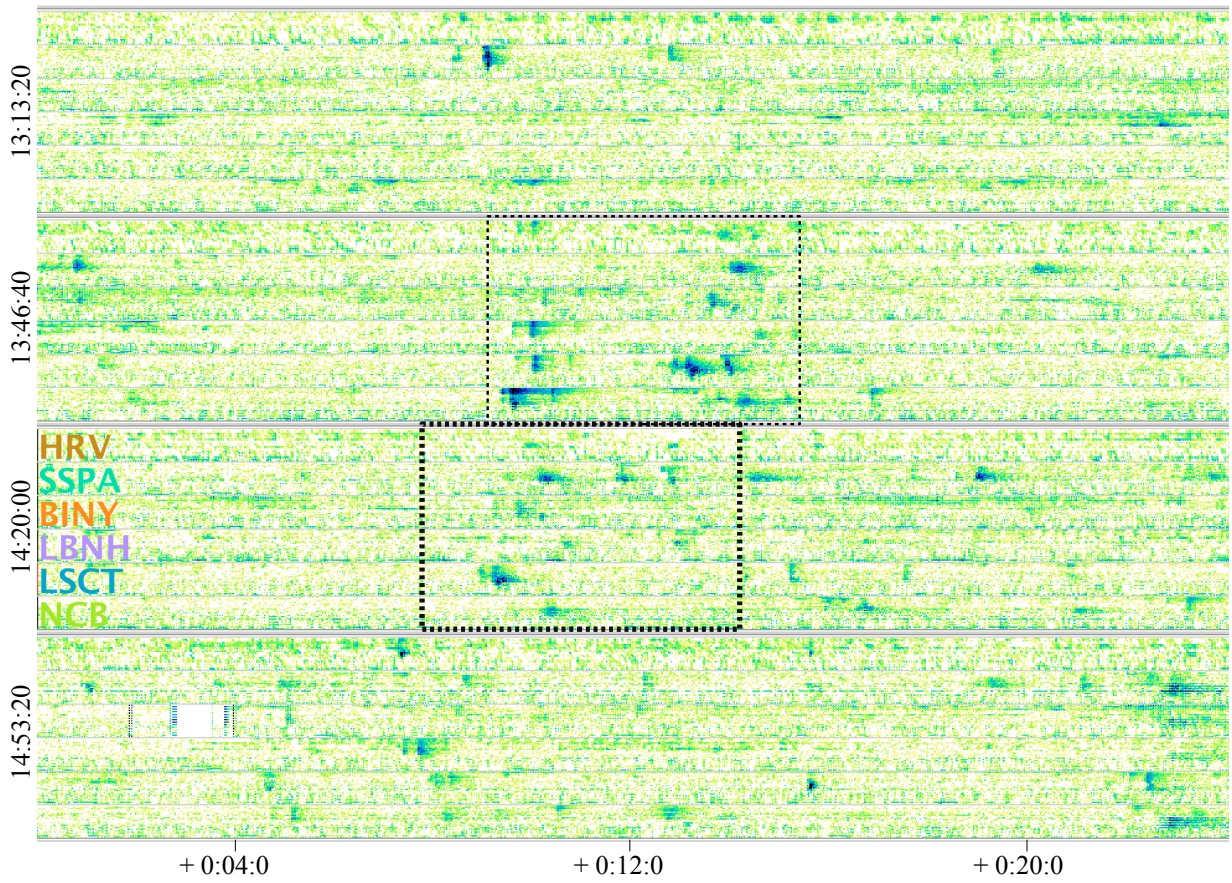


Fig. V.2:1: Sonogram (noise adapted spectral density maps) plot of the three component stations HRV, SSPA, BINY, LBNH, LSCT and NCB of the IU and US network. Two windows (dashed boxes) with significant events but multiple possible phase onsets are marked. The bold window (bottom) marks the analyzed event. Significant signal energy is found between 1 and 15Hz.

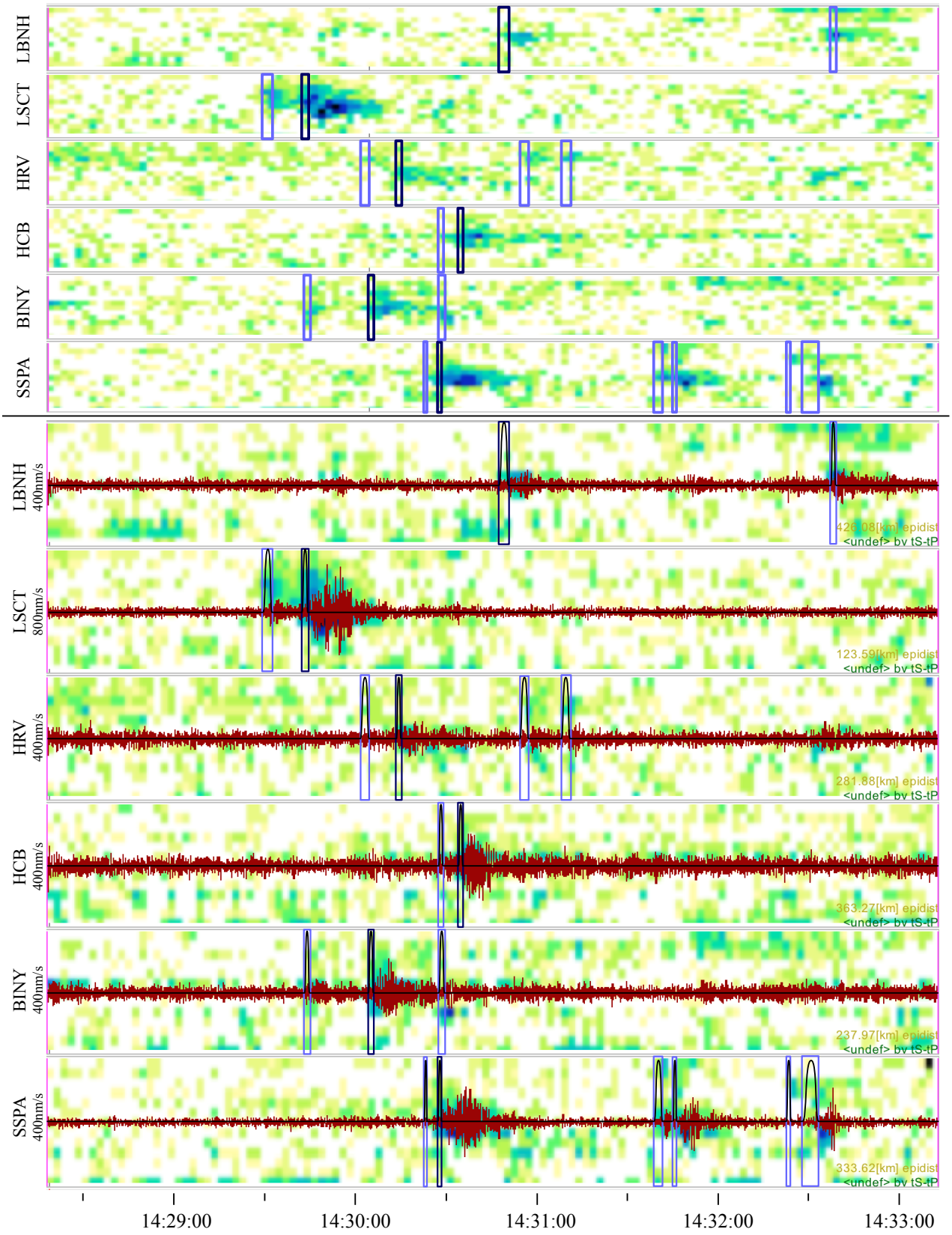


Fig. V.2:2: Sonogram (top) and Seismogram (bottom) traces of the event window (fig. V.2:1) with selected phases (blue boxes indicate onset uncertainties, top) and probability curves (black solid lines, bottom), extending over all phase candidates. Between the candidates the LOC curve is "muted", falling to zero.

We will use this probabilistic constraint approach to evaluate the observed phase pattern on those six stations in a multi-pick analysis (fig. V.2:2). Here, the various ambiguous phase onsets would have yielded 576 trial combinations (although many being not likely) but are instead elegantly analyzed using the probabilistic constraint approach, displaying a combined solution for all physically (or model based) possible phase combinations.

The sonogram traces allow a comparison of spectral signatures and a rough picking. Station LSCT shows two clear onsets, which can still be identified on station HRV, some 165 km away. Stations in larger distance (HCB and SSPA) show similar signatures, but make the discrimination between both phase onsets difficult. Additionally, station SSPA shows three similar signatures, where two of them seem to stem from local events. At station BINY the correlation becomes difficult as three different phase onsets with different signatures exist.

According to chapter II.3.4 the level-of-confidence (LOC) curve is now raised over all phase-candidates per station. The LOC curve only differs from zero in the vicinity of the phase-candidates, and can be interpreted as being “muted” in between. For every constraint being formed between two stations, several bands of hyperbolic “modes” are now “muted” according to the “muted” regions in the cross-correlated LOC curves. Likewise can the visible bands be described as “exitated modes”, based on the non-zero regions of the LOC-curves. The following figure (V.2:3) shows the hyperbolic bands created between the phases of stations SSPA and LBNH. Based on the S-velocity model for the Appalachian and Grenville Province (Viegas et al., 2010) the propagation velocity for the surface waves is estimated as $v_{surface} = (3.5 \pm 0.07)\,\mathrm{km/s}$.

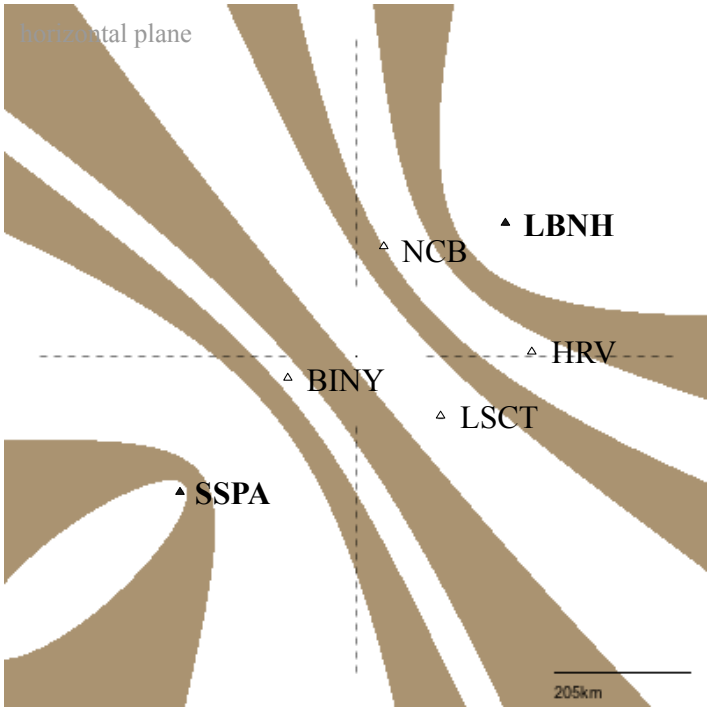


Fig. V.2:3: Hyperbolic modes excitated for the multiple phases of stations SSPA and LBNH.

This information hardly constrains the location and requires additional stations to disqualify most of the proposed hyperbolic bands: Using all stations, the modes corresponding to the set of truly correlated phases lead to a congruent region south of the network in the resulting map (fig. V.2:4), surrounded by side lobes of lower likelihood generated by other phase combinations.

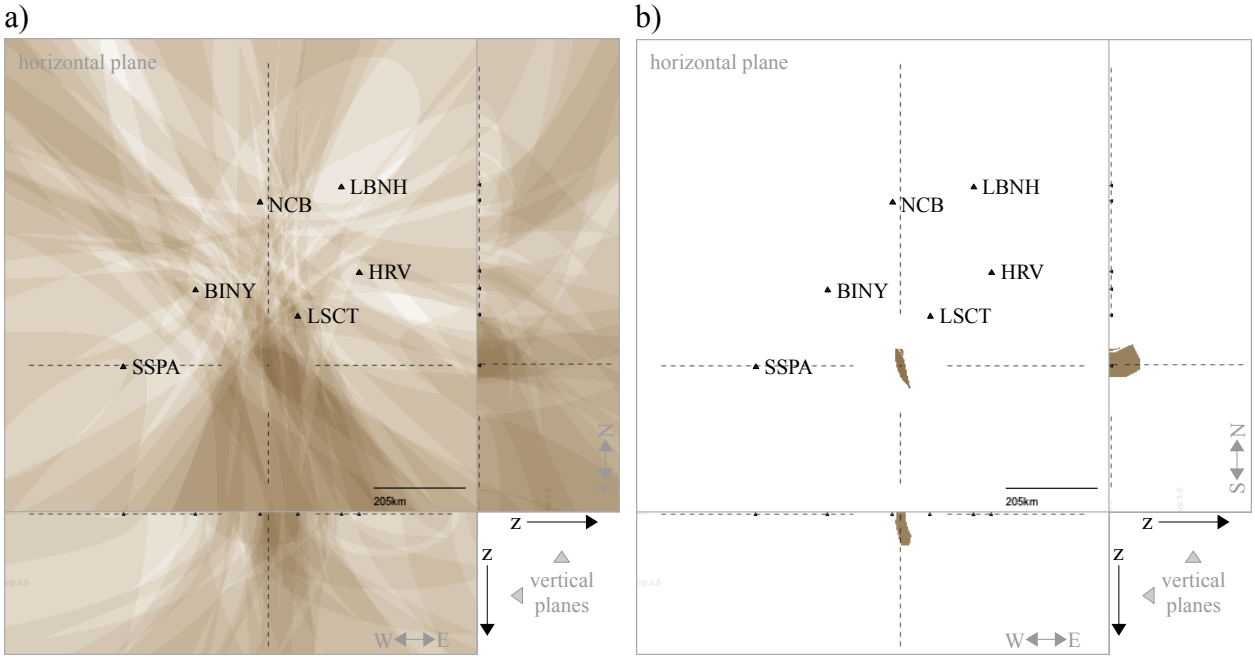


Fig. V.2:4: a) Multi-phase likelihood map using all six stations showing a region of increased congruity south of the sub-network and b) only showing the congruent zone for which all constraints agree.

The multi-pick analysis therefore displays the solution spread in regard of ambiguous phases and supports the identification of the most plausible phase-set and location. It is hereby important that the true phase is among the chosen ones as otherwise the true location itself would only appear as side lobe in the likelihood map. Additional constraints (e.g. particle motion back-azimuth) could be applied to discriminate the true solution in case of multiple equivalent maxima. However, in this example only one confined region exists south of the network. Figure V.2:5 shows the stability of the solution (constrained region) over the model assumption (mean surface wave velocity): While a velocity of 3.5km/s was chosen for the analysis based on the regional S-velocity, runs for different velocities show that the congruent region is stable for mean surface velocities above 3.4km/s, up to 3.8km/s. For slower velocities the solution moves into depth, for faster velocities it would have ceased to exist.

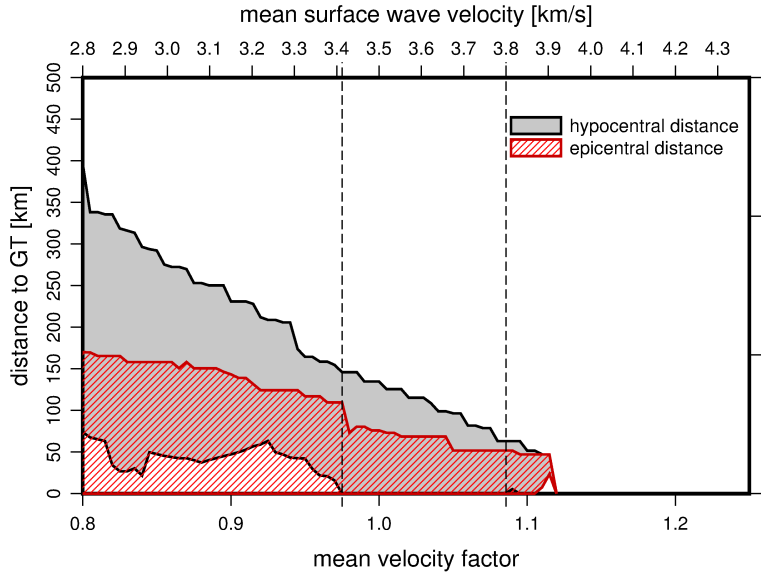


Fig. V.2:5: Robustness of the constrained region over the chosen surface wave velocity. The graph displays the constrained region's epicentral- and hypocentral distance interval to the ground truth (GT) location: A graph touching the X-axis means that the congruent region encloses the ground truth location. A mean velocity of 3.5km/s (mean velocity factor 1.0) was chosen for the analysis based on the regional S-wave velocity. Based on the hypocentral distance graph, mean velocities between 3.4km/s and 3.8km/s yield correct location estimates.

The correlated set of phase onsets originating from the determined location can be identified via a forward modelling from the constrained congruent region, yielding the set of arrival-times (dark blue boxes in fig. V.2:2). Using only these identified phases yields fig. V.2:6.

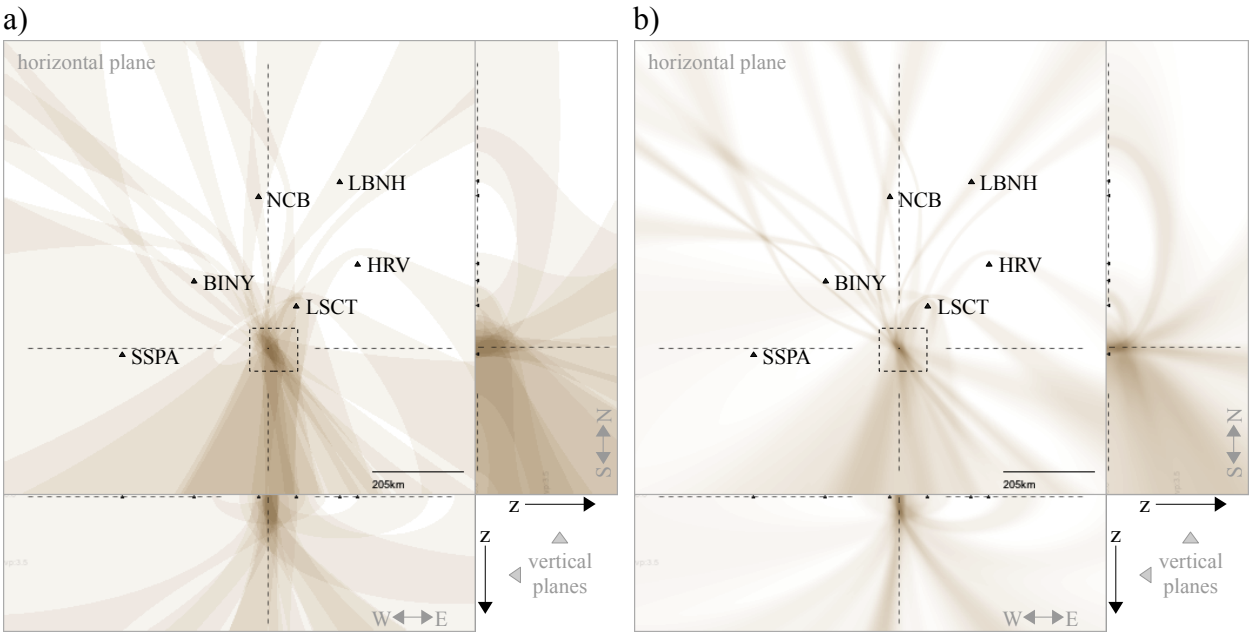


Fig. V.2:6: Constraints for the selected set of phases a) using flat weights describing the congruent zone and b) in probabilistic representation using the assigned level-of-confidence curves. The dashed box defines the zoom window for fig. V.2:7 over the region of increased congruity.

A zoom into the zone of higher congruity yields figure V.2:7. The point of highest likelihood on surface (marked as +) lies in 3.9km distance to the point marked by the red star in the congruent zone. The signals displayed in fig. V.2:1 were recorded on September 11th, 2001 and are witnesses of the catastrophic and tragic event of the collapse of the first and second tower of the World Trade Center. Its location is marked by the red star.

a) b)

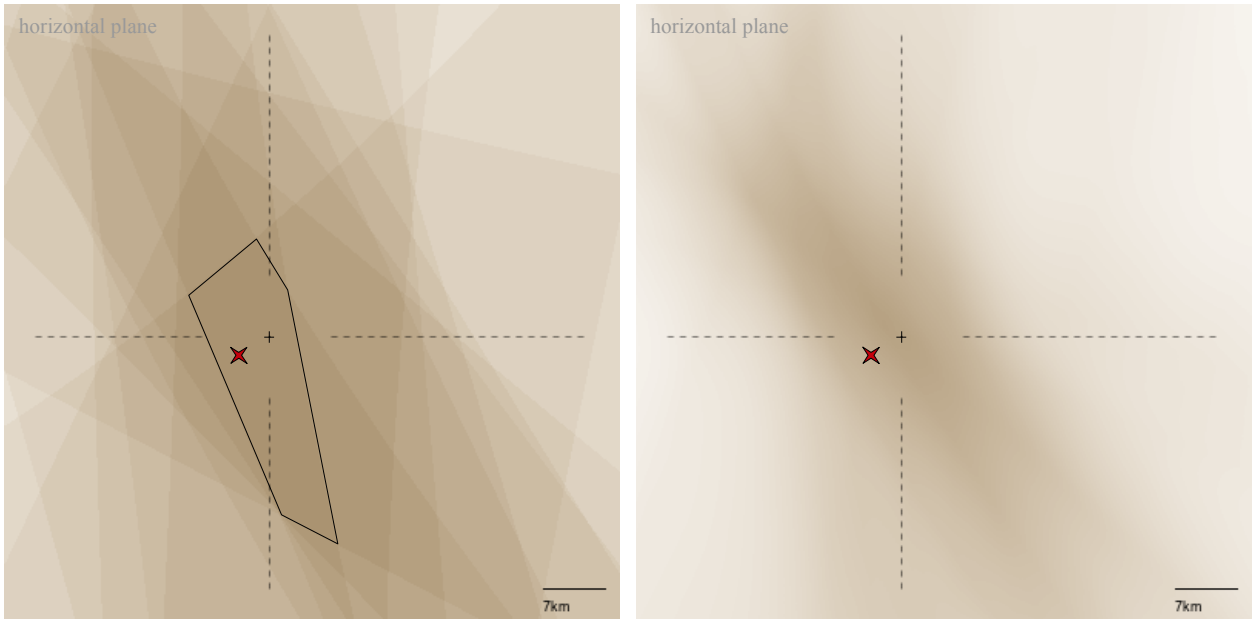


Fig. V.2:7: Zoom into region of highest coherence of fig (V.2:6). a) Coherent region of constraint agreement. The region that fulfills all constraint demands is marked by the polyhedron (solid black line). The point of highest probability on surface (panel (b)) is marked by a +. The red star marks the location of the World Trade Center Twin Towers.

An unofficial report based on the seismic records of the Palisades station (PAL), 34km North of the WTC, estimates the magnitude of the second collapse to M_L2.3. However, the data of that station was not made available to IRIS over this time frame and was therefore not used in the analysis. The closest station to the event given was LSCT with 126 km epicentral distance. While the signal could be clearly observed at PAL with a Peak-to-Peak SNR over 50, LSCT, which showed the strongest signal of the data set used, only held a SNR of 5. Although the SNR at station HRV ranges between only 0.75 and 1.07, the event signature and phase onset is clearly identifiable in the sonogram plot. This displays the strength of sonogram detection and probabilistic location in the context of sparse and doubtful data.

V.3 **Real-Time Location**

In this last example, we will use the concept of distinct constraints for real-time locations in the context of earthquake early warning (EEW). In this case the P-phase onsets are clear due to the high event magnitudes and can be detected automatically. But contrary to earlier examples, here the location procedure begins with the first available P-phase arrival and is hence prone to ambiguities and larger uncertainties. Incorporating model uncertainties, the concept of distinct constraints provides for the robustness needed in EEW, providing trustworthy estimates. The real-time location of two events will be presented with focus on back-azimuth information, which is not yet used in today's EEW algorithms but will allow to yield robuster estimates with fewer detections - and further allows to resolve ambiguities in the early location estimate. The first event will demonstrate a sparse network scenario being situated at the network edge (fig. V.3:1 (1)). The other event is situated inside the network (fig. V.3:1 (2)), in a region of higher station density.

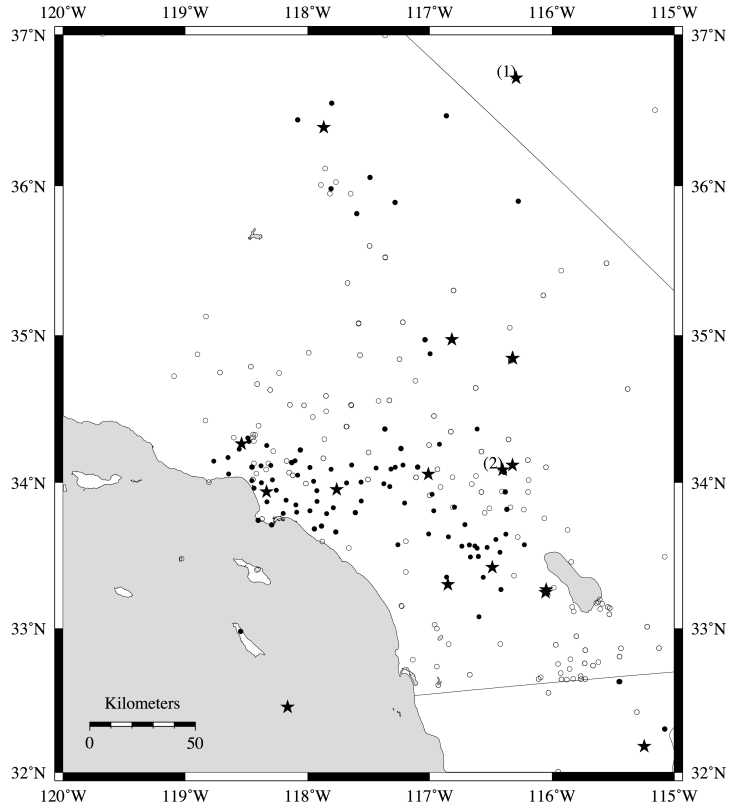


Fig. V.3:1 Location map (Southern California) with 17 (3.2 < M < 5.7) earthquakes (stars) selected for the study of real-time locations focused on the back-azimuth estimate obtained from particle motion. White circles indicate CISN stations that triggered and black circles show CISN stations that provided 3 component data used for the analysis. (1) and (2) mark the two events that are shown in this chapter, demonstrating a real-time location.

The aim of early-warning is to provide reliable warning in the shortest time possible, decreasing the blind-zone and providing enough time for public warning and triggering automatic mechanisms (as stoping trains etc...). P-phase arrival based location estimates become useful with three detections (if the focal depth can be roughly estimated based on the catalog). Depending on station density and distribution, this amount of detections may only become available with time and will, even then, often produce ambiguous results (e.g.fig. II.4:1). Back-azimuth information (BA), on the other side, can be obtained within fractions of seconds, either by arrays or particle motion estimates. In future, small-array-networks may deliver fast and precise back azimuth information. However, these are usually not given today and particle motion estimates must fill the gap. It was shown in chapter IV.1 that such estimates can be obtained reliably within fractions of seconds in most cases. Back azimuth information can therefore be used almost instantaneously to constrain the epicenter region and resolve ambiguities in the early location estimate (chapter III.2). Automatic detection of the later arriving S-phase is based on changes in polarization and signal frequency content (e.g. Nakamura, 1988). Its detection, however, is less trustworthy and is therefore neglected in these examples. In real-time location, the location problem is (initially) under-determined. The focus therefore doesn't lie on a point of highest likelihood but on the region of congruency that provides reliable location estimates, in particular estimates of the minimum and maximum distance to the first stations to compute reliable magnitude estimates (Sadeh et al., 2014). Since the final magnitude of large earthquakes is still increasing with time (while the rupture still grows) it may not always be determined within a few seconds. For this reason, especially the minimum distance is of interest, which allows to determine a true lower bound for the magnitude: The minimum magnitude cannot decrease with time and it rising over a certain threshold (commonly M_L5) may be used as reliable early alarm trigger. The concept of NYAS (chapter IV.2) can be used to reliably constrain the lower bound of the hypocentral distance based on the absence of the S-phase and hence do so before the S-phase arrives. For earthquake early warning, a-priori information like the average hypocentral depth of the region is used to fix the focal depth for the early estimates. This concept will be followed for this example. In praxis, however, the constraints should rather be evaluated over a given depth interval.

In the following, the application of distinct constraints will be demonstrated for the real-time analysis of two California earthquakes (marked as (1) and (2) in fig.V.3:1), focussing on the stability and gain in time that particular back-azimuth estimates provide for the early real-time location.

V.3.1 **Off-Network Event**

The event of Magnitude M_w4.6 occurred June 14th, 2002 in an area of low station density, just outside the CISN network. The first detection occurred at station FUR 9.5s after the event, in 59km epicentral distance. Figure V.3.1:1 displays the evolving location.

> [The plus marks the epicenter from which the wave front (dashed black circle) propagates. Detected stations are marked solid black. The hypocenter region is given in the congruent region (indicated by the darkest shading). The minimum and maximum distances are indicated in each panel. Values in parentheses indicate results without the use of back azimuth beams. The procedure was performed twice, for latency variations of 4 [resp. 5] seconds to demonstrate the large impact that these values have on the formation of the NYAD cell. Both values lie below the actual values based on transport and package size of most networks today. Values pertaining the 5s variation are noted in square brackets.]

The location procedure is initiated by the arrival of the first P-phase at station FUR. The NYAD cell provides the first location estimates, which extends far outside the network with a maximal epicentral distance of 302 [500] km (figure V.3.1:1a). A third of a second after the first detection the first back azimuth beam is computed, allowing a reduction of the maximum distance from 269 [450] km to 172 [190] km (figure V.3.1:1b). While time progresses and the wave front expands, the NYAD cell shrinks in size and constrains a maximum distance of 111km just fractions of seconds before the P-phase arrives at the second station (figure V.3.1:1c). With the second detection, 5.6s after the first arrival, a hyperbolic EDT surface is formed by the difference of first and second arrival time (figure V.3.1:1d). In itself, it constrains a minimum distance of 21 [22] km and a maximum of 107 [120] km in conjunction with the NYAD cell. The back-azimuth beam further increases the minimum distance to 44 [42] km. The second back azimuth beam is established within 0.17 seconds and stabilizes the solution (figure V.3.1:1e). The third detection occurs 7.1s after the first detection and constrains the epicentral distance to between 46 [42] km and 63 [62] km (figure V.3.1:1f). The stations of the first three detections are aligned nearly linearly and yield two ambiguous solutions - a scenario quite common for detections at the network edge. Back-azimuth and NYAD information are both able to disqualify one of the solutions. NYAD information, however, is stronger affected by network latencies. Particle motion, on the other hand, may at times produce estimates that are 180° off the true direction due to a non-effective flipping condition. This problem can be bypassed by locating by axis rather than by direction (additional back-facing beam, blue). From the second station on, the true direction is usually identified by the intersection of both azimuth-axes'.

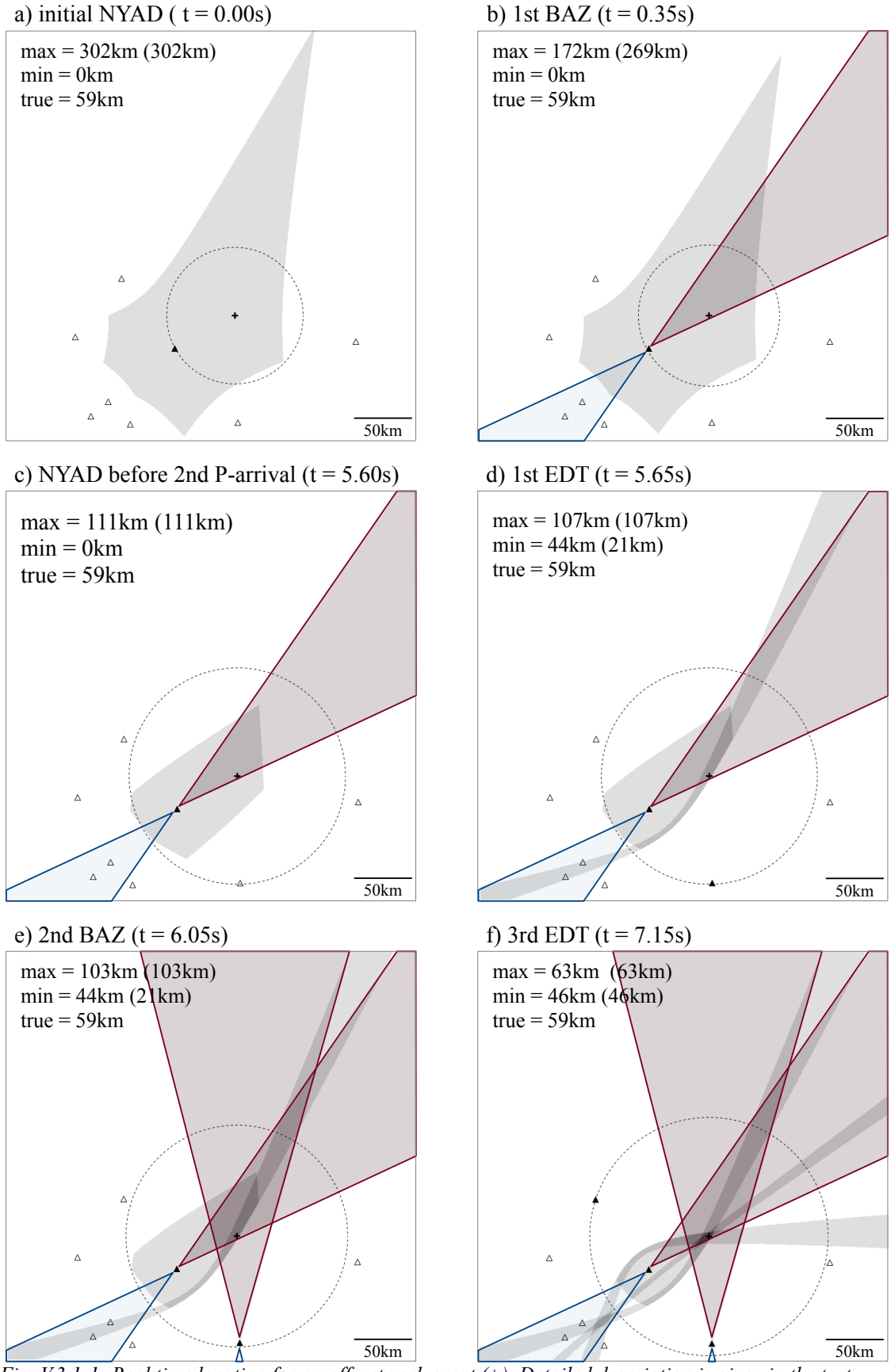


Fig. V.3.1:1: Real time location for an off-network event (+). Detailed description is given in the text.

Figure (V.3.1:2) shows four snapshots of the congruent region for the 5s latency variation. The larger latency yields significantly larger NYAD regions and therefore larger maximum distances.

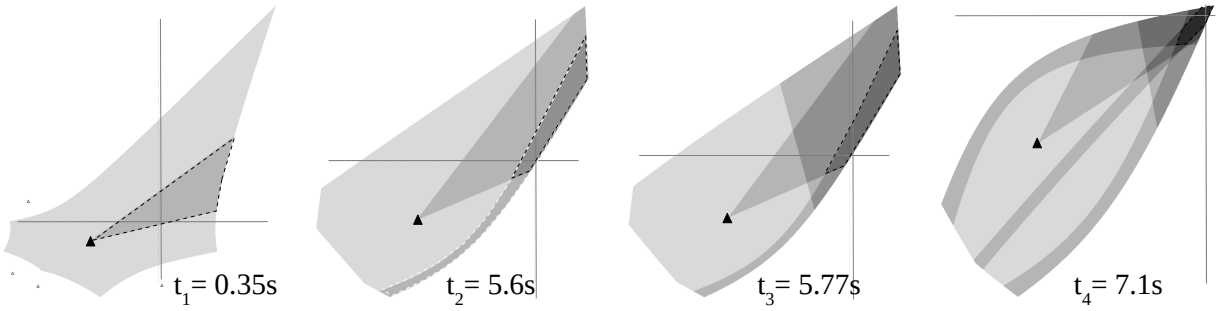


Fig. V.3.1:2: Evolving location constraints. The possible area containing the epicenter lies within the dashed polygon, which is defined by (t_1): the area of the (infinite) back-azimuth beam being cropped by the NYAD cell; (t_2): the intersection of EDT surface (white dashed line) and back-azimuth beam. (t_3): the intersection of EDT surface (right edge of NYAD cell) and two back-azimuth beams and (t_4): the intersection of three EDT surfaces and two back-azimuth beams.

Figure (V.3.1:3) displays the improvement gained by including back azimuth information into the location process. The time marks correspond to the snapshots of fig. (V.3.1:2). The gain is quantified comparing the congruent zones using and neglecting BA information. The size of the congruent zone is expressed by its moment of area (MoA) (chapter III.1) with reference point at the epicenter. The evolution of the epicentral distance interval (fig. V.3.1:3) shows that the largest improvement is gained in the time between the first and second detection, where the BAZ beam reduced the NYAD's hypocentral region, which in turn also lowered the maximum

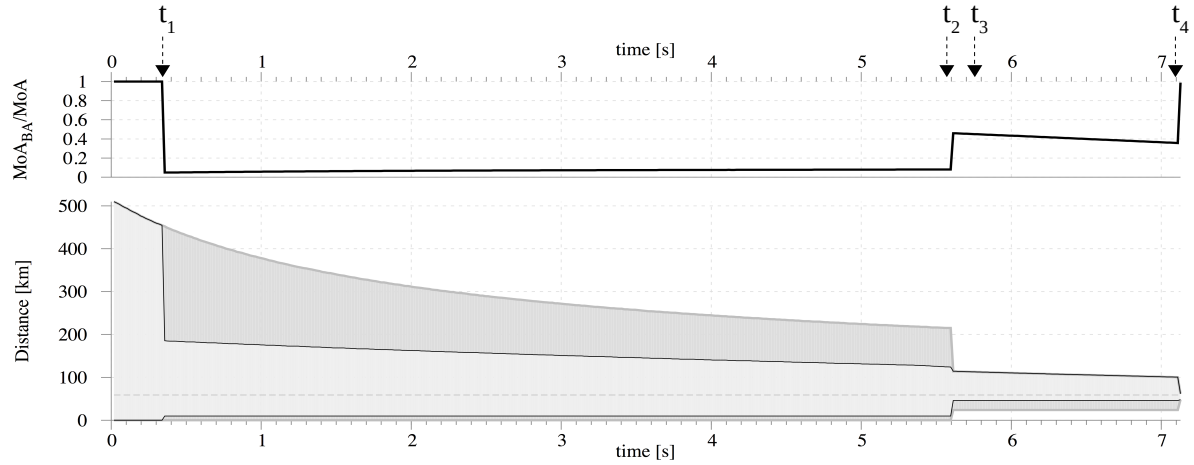


Fig. V.3.1:3: Time evolution of estimated min and max distances and ratio of moment of area, comparing realtime location using (MoA_{BA}) and neglecting (MoA) back-azimuth information. Dark grey indicates the possible epicentral distance interval disregarding back-azimuth information. Light grey marks the evolution using back-azimuth information.

hypocentral distance. The additional use of NYAS would have provided a minimum distance that would have grown to 44km already before the arrival of the second trigger (at $t = 5.5\text{s} < t_2$).

V.3.2 **Intra-Network Example**

The event of Magnitude M_w4.73 took place Aug 15th, 1992 in the Mojave Desert, in only 2km epicentral distance to station BRCC. Despite the given high station density, the initial NYAD cell is large due to package lengths of several seconds. This results in large variations in total latency, which cause the cell to even encompass near by stations (fig. V.3.2:1a). Here again, latency variations of 4 [resp. 5] seconds are used. The first back azimuth information becomes available 0.08 seconds after the first detection. This reduces the constrained NYAD region almost instantaneously to the small area which is being defined by the BA-beam within this cell (fig. V.3.2:1b). The maximum distance of 90 [183] km based on NYAD information only is hereby instantly reduced to 46 (50) km. Just before the second detection, the NYAD cell has significantly reduced its size, limiting the BA-beam to a maximum distance of 23 (36) km (fig. V.3.2:1c). The second detection occurs 2.3 seconds after the first. The generated EDT surface constrains the area to the Northern part of the previous NYAD cell. Due to the event being very close to the station, the EDT is extremely bent. The uncertainties in the velocity model and the given station-event geometry cause the strongly bent EDT to be extremely wide, keeping it from constraining the maximum distance any further (fig. V.3.2:1d). Within 0.1 seconds the second BA-beam is computed, also stabilizing but not further constraining the location (fig. V.3.2:1e). With the third detection 2.95 seconds after the first arrival, the back azimuth beam is further constrained by one of the three generated EDT surfaces to a compact area with a maximal epicentral distance of 7 [8] km and a minimal epicentral distance of 1.7km, for a true epicentral distance of 2km (fig. V.3.2:1f).
A location based solely on the three hyperboloids and NYAD constraints would at this stage still only be poorly constrained. Neglecting BA information would have resulted in significantly larger areas with a smaller minimal distance of 1.2 km and maximum distances of 19.2 [27] km (compared to 7 [8] km with BA information). Additional later P-arrivals would have been required to narrow down this region. This example shows that back-azimuth information is capable of speeding up the location procedure also inside networks where higher station densities are given: The further the event is situated from the network's Voronoi cell corners (which corresponds to events being closer to one or two stations than they are to the third), the higher is the potential to speed up the location procedure by back-azimuth information.

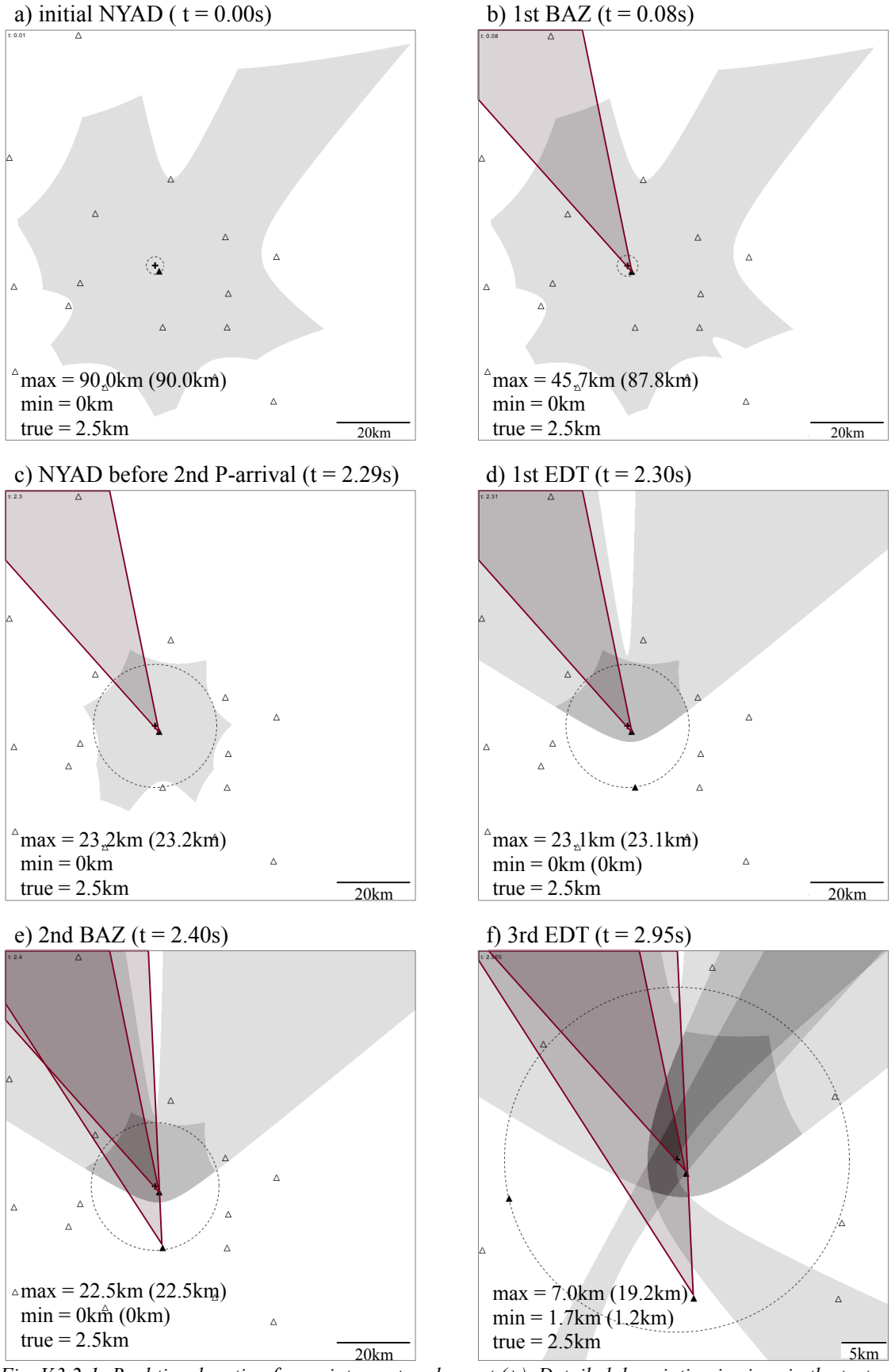


Fig. V.3.2:1: Real time location for an intra-network event (+). Detailed description is given in the text.

Figure (V.3.2:2) shows snapshots of the congruent region for the 5s latency variation. The larger latency yields significantly larger NYAD regions and therefore larger maximum distances.

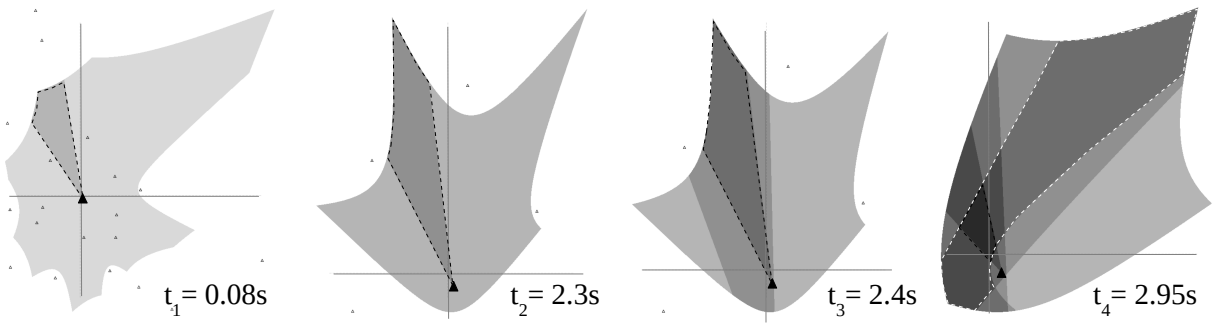


Fig. V.3.2:2: Evolving location constraint. (t_1): Large NYAD cell due to large latency variance. (t_2): The area of the strongly curved EDT surface extends over the Northern bounds of the NYAD cell. (t_4): Due to the first EDT surface being large, the epicenter is constrained only poorly (white dashed line) even using three available EDT surfaces. Using back-azimuth information the area is further constrained (black dashed line).

The figure (V.3.2:3) shows the time evolution of the constrained region, expressed as ratio between the moment of area (MoA_{BA}) using BA information and the moment neglecting it (MoA). The minimum and maximum distances are plotted in fig. V.3.2:3. The MoA ratio jumps down at the end of the time evolution (t_4) where one would expect it to rise (see fig. V.3.1:3). This is caused by the station-event geometry. In this example the event is very close to the first station which leaves the congruent region only poorly constrained when BA information is neglected.

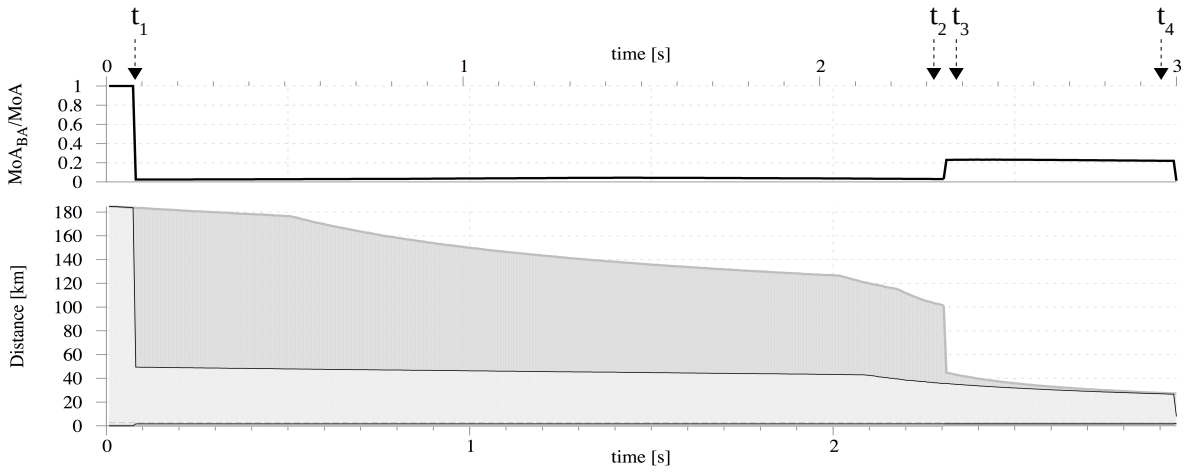


Fig. V.3.2:3: Time evolution of estimated min and max distances and moment of area ratio, comparing realtime location regarding (MoA_{BA}) (light grey) and neglecting (MoA) (dark grey) back-azimuth information.

Both examples show the stability and time gain that back-azimuth estimates may bring to real-time location schemes. This study used a worst case scenario for the NYAD region, with all stations lagging behind the first triggered station with maximum latency. This is obviously not the standard as latency varies with each station as well as over time: Some station's data will be delayed, while other stations data will already have arrived. Some data flow will stall, and will later be received faster than standard recording speed. This will cause very dynamic NYAD cell boundaries and would render the actual NYAD cell-size smaller than here displayed. However, considering that the actual variation in today's networks' latencies may significantly exceed the used values (Brown et al., 2011), the displayed NYAD cell-size may be reasonable. Since the impact of the latency variance on the NYAD cell-size is large (compare the values of the 5s and 4s scenarios) - a reduction of the variance by even a single second through e.g. smaller data packets will significantly improve (reduce) the maximum distance estimates. Ideally, data transport is switched from discrete data packets to continuous streams to minimize quantization latencies, as currently being implemented in the DeadSeaNet small-aperture array network. Chapter IV.1 showed, that back azimuth estimates can be obtained reliably from the particle motion of three-component sensors for large magnitude events, independent of the epicentral distance. These estimates are sooner available than most other constraints and allow to resolve the ambiguities that may occur with P-P constraints in early location, allowing to stabilize and speed up the location procedure. With only two detections, three independent constraints are obtained (two BA- and one EDT constraint). This provides the redundancy needed to assess the trustworthiness of the location estimate. Even with three detections, a standard P-arrival-time based location would provide only the same number of constraints, those however, being mutually dependent (which impedes the assessment of the solution's trustworthiness), often yielding ambiguous solutions. Especially in sparse networks, at network edges, for off-network events, or when events are closer to the first than to later stations, significant improvement can be gained by the incorporation of back-azimuth information. Coupled with NYAS information, which provides growing minimum hypocentral distances from the time of first detection, this becomes the real-time equivalent of classical array location combining azimuth and distant information. Together with the incorporation of model uncertainties for the travel-time dependent constraints, the methodology of distinct constraints provides for early and trustworthy location estimates for earthquake early warning.

VI **Summary and discussion**

The objective of this work lied in a methodological development that would provide for a more accurate single event location in the context of sparse and uncertain data, advancing the analysis to the complexity of the location problem demanded by recent applications, and accounting for the uncertainties, often hidden in location. Especially the impact of wrongly chosen velocity models remains un-evaluated in most location approaches, today. This makes the interpretation of single event catalogs difficult, as the listed error bounds only give a rough idea of the uncertainties, but do not reflect the true error margins (as they only pertain to the model having been used). The hypocenter estimate may actually be significantly displaced to the true location - outside of the assumed error margins. In the location with sparse and doubtful data this problem is amplified and may easily lead to the misinterpretation of events. Not only is the pre-requisite of a Gaussian-based analysis not given (which makes RMS schemes error-prone), but due to often only poorly-constrained problems, location uncertainties are larger and errors due to mis-picked phases and wrongly chosen models have a significantly larger impact. Frequently appearing location ambiguities additionally complicate the hypocenter determination.

To provide for accurate location estimates, use was made of the Jackknife to dissect datasets into under-determined subgroups, which yielded a pair-wise evaluation of phase arrivals (or triple groups for array beams, respectively). Analyzing the two-member groups over the difference between both arrival-times, eliminated the time dimension from the location problem and lead to a purely spatial problem of three dimensions that provided for a direct search for the hypocenter. In it, the hypocenter region is constrained by each arrival-time difference-group, identifying all points whose respective travel-time difference confirms the observed arrival-time differences. In this constraint based approach, discrepancies between the constraints can be traced back to individual phase data; Their impact on the solution can be evaluated and a plausible and trustworthy database be established, based on which a global outlier-resistant location estimate is recomposed.

To this fundament, a non-gaussian description of picking-uncertainties was added as well as a general formulation translating model uncertainties into uncertainties in travel-time, to cover errors in the model assumptions. Mapping these time-based constraints in their uncertainties to space, corresponds to a sensitivity analysis performed within the location procedure, identifying every possible location that can be brought into agreement with the observed arrival-time data&model in

their uncertainties. This provides for robust hypocenter regions, even when stations are sparse and data few.

Level-of-confidence functions were introduced to weigh the arrival-time (i.e. picking) interval. This in turn puts weight on regions in space, leading to a most likely hypocenter location. Together with reconstructed, approximated probability distributions for the model uncertainty, a fully probabilistic analysis is formed. This doesn't only serve to identify the most likely location, but is an important instrument indicating the trustworthiness of the congruent (or feasible) region: Almost any (wrongly picked) phase could be brought into mutual agreement with other data, when uncertainties are assigned unrealistically high: At the point when uncertainties are just high enough to yield constraint agreement, the congruent zone is extremely small for a well-conditioned problem, indicating a high precision. This precision is obviously only apparent. Towards the constraint edges the likelihoods are very low. A probabilistic analysis identifies this solution of apparent precision as unlikely. It is therefore the product of likelihood and precision of the feasible region that quantifies the solution quality.

Raising the Level-of-confidence function over ambiguous phase candidates introduced the concept of multi-pick evaluation, which provides accurate hypocenter regions by excitating (e.g. hyperbolic) constraint modes in space, identifying the possible regions that correlate with all reasonable phase combinations. As most arrival-time differences of phase-candidate combinations are not consistent in itself nor with the model, they do not yield feasible regions, in which case precise hypocenter regions are estimated for highly ambiguous input data.

Several aspects required a change in the numerical approach computing the constraints, to provide for all given demands in real-time. These were: the proper treatment of picking and model uncertainties; The requirement to use velocity models and station distributions of any complexity; and the need to “open” the solution space via vertical cuts for the analyst to comprehend the increasingly complex constraint behavior. These demands could be satisfied by decoupling constraint algorithms and travel-time (model) calculations (which where combined for earlier semi-analytical approaches computing exact constraints). On the model side, fast interpolating travel-time calculation schemes were developed for grids, rendering layered models. The portation of Podvin&Lecomte's wave-front travel-time code (1991) enabled the use of 3D models for distinct constraints, providing for more accurate predictions whenever this information is available. Global travel-time scaling factors allow to modify models per station in real-time by scaling the cached

travel-time field and its uncertainties. This allows to study the impact that e.g. variations in one station's vp/vs-ratio have on the solution. Fast uncertainty approximation schemes were developed that support a real-time analysis: For uncertain models, computed via a Monte-Carlo-approach that varies structure and velocity, minimal and maximal travel times are extracted to approximate and reconstruct a real time applicable travel-time likelihood distribution. Alternatively, in a (slower) post processing step (with a significantly larger memory consumption), the full density function may be evaluated and be used for location. To describe the effect of model uncertainties more accurately, model-ensembles may likewise be used for post-processing: Keeping the model variation synchronized among all constraints, allows ensembles to yield more precise locations and more accurate location uncertainty estimates. All uncertainty intervals are formulated as discrete probability mass functions, allowing the incorporation of any given likelihood distribution, to provide for the proper treatment in the context of few and/or doubtful data. For the direct search, constraints are evaluated over grid cells rather than grid points, to ensure the identification of sharp hypocenter regions, independent of the grid's resolution. A varying resolution (the grid cell size increasing with distance) is used to extend the grid to identify possible ambiguous solutions in distance without higher computational cost. As the spatial uncertainties grow likewise with distance, this keeps the (discretization based) precision comparable over the whole grid's extent.

The analysis of ambiguities in multilateration revealed a zonation scheme that allows to identify locations prone to showing ambiguous solutions. This provides a simple measure for evaluating the stability and trustworthiness of locations, e.g to estimate whether existing catalog locations are trustworthy. A precision mapping allows to estimate the expected precision of hypocenter regions for a given network layout and the used constraint classes. In combination, these aspects support for better network layout planing.

By having incorporated the uncertainties in picking and model (which, mapped to space, may be interpreted as a performed sensitivity analysis), provided for automated schemes identifying the most likely solutions as well as the corresponding feasible region using the parameter of constraint congruency (which may be further used to e.g. identify the most likely phase onsets in a multi-pick analysis). This real-time capable global search scheme provides robust location estimates based on few data, providing e.g. for accurate predictions that are crucial in earthquake early warning systems. At the early stage, arrival-time based location commonly yields ambiguous solutions. Back-azimuth information was shown to be available early on and to be capable of resolving these ambiguities, by disqualifying either one of them. As small-array networks are not the norm today, a

robust real-time particle-motion back-azimuth estimator was developed for three-component single stations (accurate for signals with SNR>10). This opens the door for rapid location in earthquake early warning, today. The concept of Not-Yet-Arrived-Data was extended to the delaying of the S-phase (NYAS), alongside a real-time S-phase detector that provides true lower bounds for the S-P interval. This enables the early estimation of lower bounds for the magnitude and with that a possible reduction of the blind-zone and an increase in available warning-time. As NYAS (contrary to NYAD) does not rely on other network stations, it provides for the use in independent on-site systems based on small-aperture arrays, currently being evaluated in the DeadSeaNet project.

So far, the concept of phase-difference-based constraints has been applied to constrain the solution based on the phase-differences of S and P-onset [S-P], direct and refracted P-onset [P_{ref}-P_{dir}] at one station, the difference of P-, S- or surface phase arrivals at two stations and can be applied to every phase that the travel time solvers support. For the small-array based analysis, stacking might reveal reflected phases that may be included in future. Another useful type of location constraint (although not a precise one) may be given in amplitude ratios. The ratio of the amplitudes of a phase recorded at two stations, yields similar to S-P constraints a spherical constraint (but non-concentric). This requires site-corrections and is further impeded by the focal mechanism and radiation pattern of shear-failure events, which complicates the amplitude comparison between different stations. This constraint will therefore hold large uncertainties, rendering the location less precise. However, becoming available with only two detections it may complement the determination of the distance estimate in on-site systems, before S-information becomes available. Amplitude modelling is complex and requires wavefront construction algorithms (Vinje et al., 1993). However, approximating amplitude estimates can also indicate, whether first arrivals might fall below noise level, and may be prone to unintentional neglection. Another future requirement will probably lie in the implementation of anisotropic model solvers (e.g. Sadri and Riahi, 2010).

The first application example visualized the extent of uncertainties in location due to unknowns in the velocity model. Although we may properly include uncertainties into our location codes, it remains difficult to quantify the unknown. How much our simplified used models differ from the true velocity field, only becomes quantifiable with many events or by geophysical prospecting that provides uncertainty measures for the velocity field inverted (i.e. the model). As long as there is no exact knowledge of the model (which includes all its static and dynamic parameters), one can only default to the use of security parameters as is custom in engineering. As Lomax (2009) indicated, new formulations have to be developed to communicate location errors. Today's methods that do

not only resolve the point of smallest RMS but the more complex solution environment, demand a more complex form for reporting location “errors”, or better, location “uncertainties”. With increasing computational power we are able to increase the dimensionality of the location problem, which we once reduced in complexity until our few data could provide for “overdetermined” systems (using fixed model estimates). Performing a higher dimensional analysis (i.e. varying models) causes also the complexity of the solution and its uncertainty measures to increase. One way to report the location uncertainty could follow the suggestion of the first chapter, in reporting the bounding polyhedron around the feasible hypocenter region.

I hope that this work has added to the robustness and accuracy of single event locations. The presented ideas leave much room for future development - and although this last chapter concludes this thesis, the last chapter on this topic is surely long not yet written.

Notes of thanks and acknowledgment

I would like to express my deepest thanks and highest respect to my advisor, Manfred Joswig, who raised me in one of the most interesting location approaches out there. His example of “thinking out of the (black-)box”, perseverance and never underestimating the power of the human mind in analysis as guiding and final instance, as the given freedom to develop ideas outside the norm have contributed much to this work. I enjoyed and will never forget our adventure in Lapland and will miss our evening discussions in your office that made me perceive the institute much more like home than a work place. I am very grateful for having gotten to know Ruedi Widmer-Schnidrig, in whom i always had an example of a true and excellent seismologist before me, who would always take the time to bring the depths of seismology closer to us. I would like to thank Martin Haege, Dieter Kurrle and Marco Walter who tutored me in geophysics and Nanoseismic Monitoring from the beginning, always displaying an admirable example of “how it's done”. Hillel Gilles Wust-Bloch and Alon Ziv had much impact on the developed ideas for real time location estimates in earthquake early warning. I am looking forward to our future work! I thank my colleges Sabrina, Benny, Rolf and Uwe, who inspired me and made the work together very enjoyable. I would like to thank my co-examiners Edi Kissling and David Steinberg whose expertise I highly value, who were willing to take the time reading through this long document and enrich this work through their experience and insights. Acknowledgment truly has to be given to all those scientists that laid the foundation for this work, many of them being mentioned in the references.

Obviously, not much of this work would have been possible without the support of my wonderful wife Anna. Her comfort and patience when ideas didn't turn into results too quickly, and her love for a husband who is absorbed by the beauty of rocks and mathematical formulas is beyond price. One thing i know with*(out un)*certainty - I love you beyond measure!

My wonderful son Joel - together we located a miaowing cat! - I thank you for always reminding me to take breaks to play and build tree houses. My cute little Hadassah - what would that time have been without you! My deepest gratitude is toward my parents who supported me all the way, to my father who always reminded me not to become too theoretical as physicist. I think, geophysics was an excellent choice.

Above all, i thank Elohim, in whom all things are possible.

References

Bantchev, B., (1995), Pseudo-ellipse, Mathematics & Informatics Quarterly, Vol.5, No.2 (May 1995), pp.65-71.

Berry, M. J., West, G. F., (2013), Reflected and Head Wave Amplitudes in a Medium of Several Layers. In The Earth Beneath the Continents (eds J. S. Steinhart and T. J. Smith). doi:10.1029/GM010p0464

Bischoff, M., Boennemann C., Fritz J., Gestermann N., Plenefisch T., (2013), Untersuchungsergebnisse zum Erdbeben bei Voelkersen (Landkreis Verden) am 22.11.2012. Seismologische Auswertung. Landesamt für Bergbau, Energie und Geologie (LBEG), Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), retrieved from www.lbeg.niedersachsen.de/download/80075/Erdbeben_bei_Voelkersen_Landkreis_Verden_am_22.11.2012.pdf

Bormann, P. (Ed.) (2012). New Manual of Seismological Observatory Practice (NMSOP-2), IASPEI, GFZ German Research Centre for Geosciences, Potsdam; http://nmsop.gfz-potsdam.de;
DOI: 10.2312/GFZ.NMSOP-2 urn:nbn:de:kobv:b103-NMSOP-2.

Bregman Y., Bartal Y,. Ben Horin Y., Leonard G., (unpublished), Back Azimuth Error for Tripartite Micro Arrays under Plane Wave Assumption, retrieved from http://www.soreq.gov.il/pdf/pirsumim/pirsimimfull/1484.pdf.

Brown, H. M., Allen R. M., Hellweg M., Khainovski O., Neuhauser D., Souf A., (2011), Development of the ElarmS methodology for earthquake early warning: Realtime application in California and offline testing in Japan, Soil Dynam. Earthquake Eng. 31, 188-200.

Cansi, Y. (1995) An automatic seismic event processing for detection and location: The P.M.C.C. Method, GEOPHYSICAL RESEARCH LETTERS, VOL. 22, NO. 9, PAGES 1021-1024, MAY 1, 1995

Capon, J., (1969), "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE* , vol.57, no.8, pp.1408,1418, Aug. 1969, doi: 10.1109/PROC.1969.7278

Cete, A., (1977), Seismic Source Location in the Ruhr District, Proceedings First Conference on Acoustic Emission/Microseismic Activity in Geological Structure and Materials, The Pennsylvania State University, June 1975, Trans Tech Publications, Clausthal-Zellerfeld, Germany, pp. 231-241

Chan, Y. T., K. C. Ho, (1994), “A simple and efficient estimator for hyperbolic location,” IEEE Trans. Signal Process., vol. 42, no. 8, pp. 1905–1915.

Chew, V. (1966) Confidence, prediction and tolerance regions for multivariate normal distribution. Journal of the American Statistical Association, 61, 605–617.

Cua, G., Heaton T., (2007), The virtual seismologist (VS) method: a Bayesian approach to earthquake early warning, in Earthquake Early Warning Systems, P. Gasparini, G. Manfredi and J. Zschau (Editors), Springer-Verlag, Berlin, 85-132.

Dahm, T., Krüger, F., Stammler, K., Klinge, K., Kind, R., Wylegalla, K., Grasso, J.-R., (2007), The 2004 Mw 4.4 Rotenburg, Northern Germany, earthquake and its possible relationship with gas recovery. Bulletin of the Seismological Society of America, 97, S. 691-704.

Dahlen, F.A., Hung, S.-H., Nolet, G., (2000), "Fréchet kernels for finite-frequency travel times - I. Theory", Geophys. J. Int., 141, pp. 157-174.

Deichmann, N., Garcia-Fernandez M., (1992), Rupture geometry from high-precision relative hypocenter locations of microearthquake clusters, Geophys.J.Int., 110,501-517.

Dubois, D., Fargier, H. & Prade, H. Appl Intell, (1996), Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty, October 1996, Volume 6, Issue 4, pp 287–309, https://doi.org/10.1007/BF00132735

Duncan, P.M., (2005), Is there a future for passive seismic?, First Break, volume 23, June 2005, p.111-115

Eisermann A., (2008), Berechnung der Wellenpropagation durch ein geneigtes Schichtenmodell zur Lokalisierung von Erdbeben (Diploma thesis), Retrieved from http://www.geophys.uni-stuttgart.de/~ase/Berechnung_der_Wellenpropagation_durch_ein_geneigtes_Schichtenmodell_zur_Lokalisierung_von_Erdbeben.pdf

Eisermann A., Ziv A., Wust-Bloch G.H., (2015), Real-Time Back Azimuth for Earthquake Early Warning, Bulletin of the Seismological Society of America, Vol. 105-4, doi: 10.1785/0120140298

Geiger, L., (1910), Herbsetimmung bei Erdbeben aus den Ankunfzeiten, K. Gessell. Wiss. Goett. 4, 331-349.

Geiger, L., (1912), Probability method for the determination of earthquake epicenters from the arrival time only, Bull. St. Louis Univ. 8, 60-71.

Green, P. E., Frosch R. A., Romney C. F., (1965), Principles of an experimental large aperture seismic

array (LASA), Proc. IEEE 53, 1821-1833.

Guitton, A. W.W. Symes, (2003), Robust inversion of seismic data using the Huber norm, GEOPHYSICS, VOL. 68, NO. 4 (JULY-AUGUST 2003); P. 1310–1319, 10 FIGS. 10.1190/1.1598124

Horiuchi, S., Negishi, H., Abe, K., Kamimura, A., and Fujinawa, Y. (2005), An Automatic Processing System for Broadcasting Earthquake Alarms, Bulletin of Seismological Society of America , 95 , 708-718.

Husen, S., E. Kissling, N. Deichmann, S. Wiemer, D. Giardini, and M. Baer (2003), Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland, J. Geophys. Res., 108, doi: 10.1029/2002JB001778.

Husen, S., Hardebeck J.L., (2010), Earthquake location accuracy, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-55815573. Available at http://www.corssa.org.

Joswig, M., (1990), Pattern recognition for earthquake detection, Bull. Seism. Soc. Am., 80, 170-186.

Joswig, M. (1999). Automated Processing of seismograms by SparseNet. *Seismological Research Letters*, *70*(6), 705-711.

Joswig, M., (2000), Automated Event Location by Seismic Arrays and Recent Methods for Enhancement, Advances in Seismic Event Location, pp 205-230, DOI 10.1007/978-94-015-9536-0_8

Joswig, M., (2006), Nanoseismic Monitoring: Method and First Applications. Retrieved from *www.geophys.uni-stuttgart.de/lehre/summerschool/third/tutorial.pdf.*

Joswig, M., (2008), Nanoseismic Monitoring fills the gap between microseismic networks and passive seismic. First Break, 26(6), 121-128

Kao H., Shan S.-J., (2004), The Source-Scanning Algorithm: mapping the distribution of seismic sources in time and space, GJI, Volume 157, Issue 2, 1 May 2004, Pages 589–594, https://doi.org/10.1111/j.1365-246X.2004.02276.x

Kissling, E., (1988), Geotomography with local earthquake data, Rev. of Geophysics

Kissling, E., (1993), Deep structure of the Alps:what do we really know?, Phys. Earth Planet. Int., 79,87-112.

Kissling, E., W. L. Ellsworth, D. Eberhart-Phillips, U. Kradolfer: Initial reference models in local earthquake tomography, J. Geophys. Res., 99, 19,635-19,646, 1994.

Kissling, S. Solarino, M. Cattaneo, (1995), Improved seismic velocity reference model from local earthquake data in Northwestern Italy, Terra Nova, Volume 7, Issue 5, pages 528–534, September 1995, DOI:10.1111/j.1365-3121.1995.tb00554.x

Klein, R.W., (1978), Hypocenter Location Program HYPOINVERSE Part 1: Users Guide to Versions 1,2,3, and 4. U.S. Geol. Surv. OFR 78-694, 113 pp.

Küperkoch, L., Meier, T., Brüstle, A., Lee, J., Friederich, W., EGELADOS working group, (2012), Automated determination of S-phase arrival times using autoregressive prediction: application to local and regional distances, Geophysical Journal International, doi: 10.1111/j.1365-246X.2011.05292.x.

Kuyuk, H.S., Allen R.M., (2014), Designing a Network-Based Earthquake Early Warning Algorithm for California: ElarmS-2 Bull. Seismo. Soc. Am., 104, 162-173, doi:10.1785/0120130146

Kvaerna, T., Doornbos, D., (1986), An integrated approach to slowness analysis with arrays and tree component stations. Norsar semiannual technical summary, 1 october 1985 - 31 march 1986, NORSAR, Kjeller, Norway. Scientific Report No. 2-85/86.

Labák P., M. Joswig, L. Fojtíková, P. Dewez, F. Guendel, (2004), Detection capability of 3-component seismic stations and tripartite mini-arrays: CTBT monitoring of artificial nanoevents with M<0, EGU 2004

Lahr, J.C. (1989), HYPOELLIPSE: A computer program for determining local earthquake hypocentral parameters, magnitude and first motion pattern (Y2K compliant version), U.S. Geol. Surv. Open-File Rep., 92 pp.

Lancieri, M. and A. Zollo, (2008), A Bayesian approach to the real-time estimation of magnitude from the early P and S wave displacement peaks, JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, B12302, doi:10.1029/2007JB005386

Lay T, T.C. Wallace, (1995), Modern Global Seismology, Academic Press, London, p.217..231

Leonard M., (2000), Comparison of Manual and Automatic Onset Time Picking, *Bulletin of the Seismological Society of America*, December 2000, v. 90, p. 1384-1390

Lin, K. and Sanford A. R., (2001), Improving Regional Earthquake Locations Using a Modified G Matrix and Fuzzy Logic, Bulletin of the Seismological Society of America, 91, 1, pp. 82–93, February 2001.

Larmat, C., Montagner J-P., Fink M., Capdeville Y., Tourin A., Clevede E., (2006), Time-reversal imaging of seismic sources and application to the great Sumatra

earthquake, GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L19312, doi:10.1029/2006GL026336

Lee, W.H.K., and J.C. Lahr (1975), HYPO71 (revised): A computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes, U.S. Geol. Surv. Open-File Rep., 64 pp.

Leonard M., (2000), Comparison of Manual and Automatic Onset Time Picking, *Bulletin of the Seismological Society of America*, December 2000, v. 90, p. 1384-1390.

Lienert, B.R., Berg E., and Frazer L.N., (1986), HYPOCENTER: An Earthquake Location Method Using Centered, Scaled and Adaptively Damped Least Squares. Bull. Seism. Soc. Am., v.76, pp. 771-783.

Lior I.,Ziv A., Madariaga R., (2015), P-Wave Attenuation with Implications for Earthquake Early Warning, Bulletin of the Seismological Society of America, Vol. 106 no. 1 13-22 DOI: 10.1785/0120150087

Lockman, A. B., Allen R. M., (2005), Single-station earthquake characterization for early warning, Bull. Seism. Soc. Am. 95, 2029–2039.

Lomax, A., Virieux J., Volant P., Berge C., (2000), Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, in Advances in Seismic Event Location Thurber, C.H., and N. Rabinowitz (eds.), Kluwer, Amsterdam, 101-134.

Lomax, A., Michelini A., Curtis A., (2009), Earthquake Location, Direct, Global-Search Methods, in Encyclopedia of Complexitzy and System Science, Part 5, Meyers, R. A. (ed.), Springer, New York, pp. 2449-2473, doi:10.1007/978-0-387-30440-3.

Lomax, A., (2011), NLLoc - Non-linear, earthquake location program, retrieved from http://alomax.free.fr/nlloc/soft6.00/NLLoc.html

Lomnitz C., (2006), Three Theorems of Earthquake Location, Bulletin of the Seismological Society of America, Vol. 96, No. 1, pp. 306–312, February 2006, doi: 10.1785/0120050039

Menke, W., and D. Schaff (2004), Absolute earthquake locations with differential data, Bull. Seism. Soc. Am., 94, 2254-2264.

Miller, R.G., (1974), The Jackknife – A Review, Biometrika, Vol. 61, No.1. (Apr., 1974), pp. 1-15.

Milne, J., (1886), Earthquakes and Other Earth Movements, Appelton, New York.

Moser, T.J., van Eck T., Nolet, G., (1992), Hypocenter determination in strongly heterogeneous earth models using the shortest path method., J. Geophys. Res., 97, 6563-6572.

Nakamura, Y,. (1988), On the urgent earthquake detection and alarm system (UrEDAS), in Proc. of the 9th World Conference on Earthquake Engineering, 2–9 August 1988, Tokyo–Kyoto, 673-678.

Niu F., Li J., (2001), Component azimuths of the CEArray stations estimated from P-wave particle motion, Earthquake Science, February 2011, Volume 24, Issue 1, p. 3-13.

Noda, S., Yamamoto S., Sato S., Iwata N., Korenaga M., Ashiya K., (2012), Improvement of back-azimuth estimation in real-time by using single station record, Earth Planets Space 64, 305-308.

Pavlis, G.L. (1986), Appraising earthquake hypocenter location errors - a complete, practical approach for single-event locations, Bull. Seism. Soc. Am., 76, 1699-1717.

Podvin, P. and Lecomte, I., (1991), Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools., Geophys. J. Int., 105, 271-284.

Prieto, G. A., D. J. Thomas, F. L. Vernon, P. M. Shearer and R. L. Parker, (2007). Confidence intervals for earthquake source parameters, Geophys. J. Int 168, 1227-1234. doi:10.1111/j.1365-246X.2006.03257.x

Pujol J., (2004), Earthquake Location Tutorial: Graphical Approach and Approximate Epicentral Location Techniques, Seismological Research Letters, Volume 75 (1), p. 63-74

Quenouille, M.H., (1949), Approximate tests of correlation in time-series. J.R. Statist.Soc. B 11, 68-84, [§1].

Quenouille, M.H., (1956), Notes on bias in estimation. Biometrika 43, 353-60, [§§1, 2.1].

Rabinowitz, N. Steinberg D.M., (1988), Effect of single arrival time on the hypocentral focal depth determination. *Geophysical Journal* 94, 567-570.

Rabinowitz, N. Steinberg D.M., (1990), Optimal configuration of a seismographic network – a statistical approach. *Bulletin of the Seismological Society of America* 80, 187-196.

Ruff, L.J, 2001, How to Locate Epicenters, Seismological Research Letters, Volume 72 (2), p. 197

Rydelek P. And Pujol, J., (2004), Real-time seismic warning with a 2-station subarray. *Bull. Seism. Soc. Am.* 94, 1546-1550.

Sadeh, M., Ziv A., Wust-Bloch H,. (2014)., Real-time

magnitude proxies for earthquake early warning in Israel, Geophys. J. Int.

Sadri, M., Riahi M.A., (2010), Ray tracing and amplitude calculation in anisotropic layered media, Geophys. J. Int. (2010) 180, 1170–1180 doi: 10.1111/j.1365-246X.2009.04464.x

Satriano, C., Lomax A., Zollo A., (2008), Real-time evolutionary earthquake location of seismic early warning, Bull. Seism. Soc. Am. 98,1482–1494.

Satriano, C., Elia L., Martino C., Lancieri M., Zollo A., Iannaccone G., (2011), PRESTo, the earthquake early warning system for Southern Italy: Concepts, capabilities and future perspectives, Soil Dynam. Earthq. Eng. 31, no. 2, 137–153, doi: 10.1016/j.soildyn.2010.06.008.

Schweitzer, J. (1997): HYPOSAT – a new routine to locate seismic events. NORSAR Scientific Report 1-97/98, 94-102, NORSAR, Kjeller, Norway, November 1997.

Schweitzer, J. (2001): HYPOSAT – an enhanced routine to locate seismic events. Pure and Applied Geophysics 158, 277-289.

Shearer, P.M., (1997), Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence., J. Geophys. Res., 102, 8269-8283.

Sick, B., Walter, M., Joswig, M. (2012). Visual Event Screening of Continuous Seismic Data by Supersonograms. Recent Advances in Nuclear Explosion Monitoring Vol. 2 - Pure and Applied Geophysics, PAGEOPH. doi:10.1007/s00024-012-0618-x

Spiesberger, J.L., (2001), Hyperbolic location errors due to insufficient numbers of receivers, 2001 Acoustical Society of America, DOI: 10.1121/1.1373442

Steinberg, D.M., Rabinowitz N., (2003), Optimal seismic monitoring for event location with application to On Site Inspection of the Comprehensive Nuclear Test Ban Treaty. *Metrika*, **58**, 31-57.

Swanson, P.L., Estey L.H., Boler F.M., Billington S., (1992), Accuracy and Precision of Microseismic Event Locations in Rock Burst Research Studies, Report of Investigations 9395, UNITED STATES DEPARTMENT OF THE INTERIOR, BUREAU OF MINES.

Tarantola, A. and Valette, B., (1982), Inverse problems = quest for information., J. Geophys., 50, 159-170.

Viegas, G.M., Baise L.G., Abercrombie R.E., (2010), Regional Wave Propagation in New England and New York, Bulletin of the Seismological Society of America, Vol. 100, No. 5A, pp. 2196–2218, October 2010, doi: 10.1785/0120090223

Vinje, V., Iversen E., Gjoystdal H., (1993), Traveltime and amplitude estimation using wavefront construction, GEOPHYSICS, VOL. 58, NO. 8 (AUGUST 1993); P. 1157-1166

Waldhauser, F., (2001), hypoDD - A program to compute double-difference hypocenter locations. Technical report, U.S. Geol. Survey, Menlo Park, CA.

Walter, M., U. Niethammer, S. Rothmund, M. Joswig, (2009), Joint analysis of the Super-Sauze (French Alps) mudslide by nanoseimic monitoring and UAV-based remote sensing, First Break, 27(8), p.75-82

Wust-Bloch, G.H., Joswig M., (2006), Pre-collapse identification of sinkholes in unconsolidated media at Dead Sea area by `nanoseismic monitoring' (graphical jackknife location of weak sources by few, low-SNR records). Geophysical Journal International. 01/2006; 167(3):1220-1232. DOI: 10.1111/j.1365-246X.2006.03083.x

Wust-Bloch G. H., Ziv A., Eisermann A. S., Al-Dabbeek J., Al-Zoubi A., (2016), DeadSeaNet: Cross-border array of mini-arrays for EEWS, Proceedings of the ECGS & ESC/EAEE Joint Workshop: Earthquake and Induced Multi-Risk Early Warning and Rapid Response, Volume 31, P.87

Xia, J., Miller R.D., Park C.B., (1999), Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves, GEOPHYSICS, VOL. 64, NO. 3 (MAY-JUNE 1999); P. 691

Xiaochun. X., Sahni S., Rao N. S. V., (2008), On basic properties of localization using distance-difference measurements, *11th Int.Inform. Fusion Conf.*, pp. 1-8