Proceedings of the 10th International Conference on Ecological Informatics

Translating Ecological Data into Knowledge
and Decisions in a Rapidly Changing World.

24-28 September 2018, Jena, Germany

PROCEEDINGS

10TH INTERNATIONAL CONFERENCE ON ECOLOGICAL INFORMATICS

Translating Ecological Data into Knowledge and Decisions in a Rapidly Changing World

24-28 September 2018,
Jena, Germany

Edited by-

Jitendra Gaikad[1,2], Birgitta König-Ries[1,2], and Friedrich Recknagel[3]

[1] Friedrich Schiller University Jena, Germany; [2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany; [3] University of Adelaide, Australia

Disclaimer

The ideas and opinions expressed in this document are those of the authors and do not necessarily represents the views of the Friedrich Schiller University, iDiv, University of Adelaide and Journal of Ecological Informatics. The authors are fully responsible for the submitted material (abstracts) in this document and should be contacted directly for further information as preferred.

The proceedings should be cited as:

Gaikwad, J., König-Ries, B., & Recknagel, F. (Eds). Proceedings of the ‘10th International Conference on Ecological Informatics: translating ecological data into knowledge and decisions in a rapidly changing world’, Jena, Germany, 24-28 September, 2018.

Preface

It is with great pleasure to welcome you to the 10th International Conference on Ecological Informatics hosted by the Friedrich Schiller University, Jena, Germany.

The Conference Proceedings are an impressive display of the current scope of Ecological Informatics as schematically represented in Figure. 1.

Whilst Data Management, Analysis, Synthesis and Forecasting have been lasting popular themes over the past nine biannual ICEI conferences, ICEI 2018 addresses distinctively novel developments in Data Acquisition enabled by cutting edge *in situ* and remote sensing technology.

The here presented ICEI 2018 program captures well current trends and challenges of Ecological Informatics towards: • regional, continental and global sharing of ecological data, • thorough integration of complementing monitoring technologies including DNA-barcoding, • sophisticated pattern recognition by deep learning, • advanced exploration of valuable information in 'big data' by means of machine learning and process modelling, • decision-informing solutions for biodiversity conservation and sustainable ecosystem management in light of global changes.

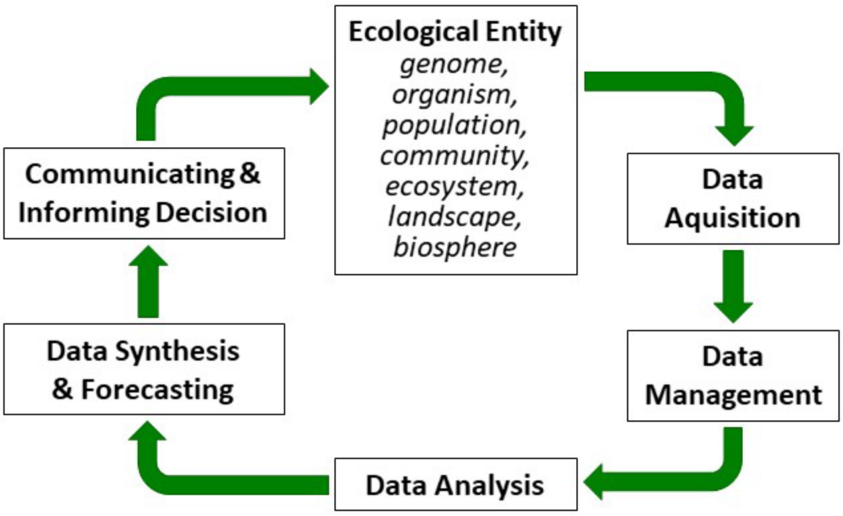


Figure. 1: Scope of Ecological Informatics

I herewith wish all delegates an inspiring week of science, communication and socialising.

Adelaide, 5th September 2018

Friedrich Recknagel

Friedrich Recknagel
Scientific Program Chair ICEI 2018
University of Adelaide
AUSTRALIA

Table of Content

Session R1.1

ECOLOGICAL MONITORING BY CAMERA, THERMAL AND ACOUSTIC IMAGES

Ecological monitoring by camera, thermal and acoustic images

Marie Roch[1]

[1]San Diego State University, United States

Corresponding author(s) e-mail: mroch@mail.sdsu.edu

ABSTRACT:

Camera traps, thermal infrared videos and soundscapes are non-invasive monitoring techniques of elusive animals without significantly sacrificing analytical accuracy. These methods reduce field hours for estimating demographic parameters, inventory species and migration patterns. Automatic classification systems of camera, thermal and acoustic images allow large datasets to be analyzed over short timescales and yield valuable information for natural resource decision-making.

This session welcomes papers on all aspects of automatic acquisition, classification and interpretation of image-based data of biodiversity.

KEYWORDS: Camera traps, Acoustic images, Image-based data, Thermal infrared videos

Keeping the Human in the Loop: Towards Automatic Visual Monitoring in Biodiversity Research

***Joachim Denzler*[1,2], *Christoph Käding*[1], *Clemens-Alexander Brust*[1]**

[1]*Comouter Vision Group, FSU Jena, Germany*; [2]*Michael Stifel Center Jena for Data-driven and Simulation Science, Jena, Germany*;

Corresponding author(s) e-mail: joachim.denzler@uni-jena.de

ABSTRACT:

More and more methods in the area of biodiversity research grounds upon new opportunities arising from modern sensing devices that in principle make it possible to continuously record sensor data from the environment. However, these opportunities allow easy recording of huge amount of data, while its evaluation is difficult, if not impossible due to the enormous effort of manual inspection by the researchers. At the same time, we observe impressive results in computer vision and machine learning that are based on two major developments: firstly, the increased performance of hardware together with the advent of powerful graphical processing units applied in scientific computing. Secondly, the huge amount of, in part, annotated image data provided by today's generation of Facebook and Twitter users that are available easily over databases (e.g., Flickr) and/or search engines. However, for biodiversity applications appropriate data bases of annotated images are still missing.

In this presentation we discuss already available methods from computer vision and machine learning together with upcoming challenges in automatic monitoring in biodiversity research. We argue that the key element towards success of any automatic method is the possibility to keep the human in the loop - either for correcting errors and improving the system's quality over time, for providing annotation data at moderate effort, or for acceptance and validation reasons. Thus, we summarize already existing techniques from active and life-long learning together with the enormous developments in automatic visual recognition during the past years. In addition, to allow detection of the unexpected such an automatic system must be capable to find anomalies or novel events in the data.

We discuss a generic framework for automatic monitoring in biodiversity research which is the result of collaboration between computer scientists and ecologists of the past years. The key ingredients of such a framework are initial, generic classifier, for example, powerful deep learning architectures, active learning to reduce costly annotation effort by experts, fine-grained recognition to differentiate between visually very similar species, and efficient incremental update of the classifier's model over time. For most of these challenges, we present initial solutions in sample applications. The results comprise the automatic evaluation of images from camera traps, attribute estimation for species, as well as monitoring in-situ data in environmental science. Overall, we like to demonstrate the potentials and open issues in bringing together computer scientists and ecologist to open new research directions for either area.

KEYWORDS: computer vision, machine learning, automatic visual monitoring, camera trapping

Automating biological monitoring on the Northern Andes of South America: combining biology and machine learning for conservation

***Juan M. Daza*[1], *Claudia Isaza*[1], *Carol Bedoya*[1,2], *Estefany Cano*[1], *Diana C. Duque*[1], *William E. Gómez*[1], *Camilo Sánchez-Giraldo*[1]**

[1]*Universidad de Antioquia, Colombia*; [2]*University of Canterbury, New Zealand*

Corresponding author(s) e-mail: juanm.daza@udea.edu.co

ABSTRACT:

The northern Andes in South America harbors one of the most diverse biological diversity on the planet. Yet, it is one of the most threatened regions as a result of habitat fragmentation, invasive species, agriculture and cattle grazing, and global climate change. It is therefore critical to implement robust conservation strategies and effective monitoring plans. In the region, biological monitoring relies on traditional methods such as direct observation and capture. These methods are expensive and require a large effort specially for rare species. As an alternative, automated passive bioacoustics allow to obtain large amounts of data both in time and space and in comparison, with traditional methods at low cost. The main challenge in passive monitoring is to handle and analyze these rivers of information in order to obtain meaningful results from acoustic data. We have implemented a passive bioacoustic monitoring since 2012 on the northern Andes in Colombia, a highly diverse region in the Neotropics. Our goals are two fold: first, we want to develop analytical strategies to process large amounts of sound files and second we are interested in answering biological questions from individuals to the landscape. As a result of this monitoring, we have developed a machine learning algorithm based on syllable recognition to automatically identify frog species (*Ecol. Inf.* 24: 200-209). We also have developed an algorithm to estimate the amount of rain from acoustic recordings (*Ecol. Ind.* 75:95-100). We have answered biological questions ranging from acoustic niche partitioning, interaction of traditional community indices with acoustic indices, and association of acoustic indices with landscape features. Now, we are using passive monitoring to fit complex occupation models and to determine assembling rules in anuran communities. In addition, we are assessing acoustic indices aiming to develop tools with more functions for soundscape analysis (*Ecol. Inf.* 45:16-25). We show how the continuous feedback between biologists and engineers will spike the implementation and analysis of passive monitoring in imperiled tropical hotspots.

KEYWORDS: Bioacoustics, Colombia, Conservation, Neotropics, Passive monitoring.

Understanding the Relationship between Soundscape and Landscape Features in a Tropical Andean Environment

Camilo Sánchez-Giraldo**[1, 4]**, Estefany Cano**[1]**, William E. Gómez**[2]**, Diana C. Duque**[2]**,
***Claudia Isaza**[2]**, Carol Bedoya**[3]**, Juan M. Daza**[1]*

[1]*Grupo Herpetológico de Antioquia, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia*; [2]*SISTEMIC, Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia*; [3] *Department of Biological Sciences, University of Canterbury, Christchurch, New Zealand*

Corresponding author(s) e-mail: camilo.sanchez@udea.edu.co

ABSTRACT:

The acoustic component of the landscape (soundscape) has been suggested as an indicator of landscape conditions as it has been related to physical, biological and anthropogenic features of the local environment. Despite of acoustic indices have been used as measurements of soundscape complexity, their relationship with landscape features has shown to be highly variable and not direct. This relationship has not been examined on the tropical Andes then it is not clear if soundscape can be a reliable indicator of landscape conditions. In this study, we assess the relationship between 14 acoustic indices and landscape conditions in an area on the northern Andes of Colombia. This region is highly fragmented and characterized for a highly diverse community of animals and plants thus representing an excellent opportunity to test such relationship. Soundscape recordings were obtained from 31 randomly selected sites surveyed among May and July 2017. A Song Meter SM4 device was deployed at each site for five consecutive days, and programmed to collect 1-min recordings every 15 minutes for a total of 95 samples per day. Recordings were obtained as monaural 16 bits and at a sampling rate of 22.05 kHz. Out of 14 indices, we calculated 9 non-correlated acoustic indices for each 1-min recording and selected the maximum value per hour to estimate the hourly average over the five days at each site. Thirteen landscape features were derived from satellite images and metrics describing vegetation, fragmentation, water availability, terrain, and soil attributes at 100 m radius around each site. Indices did not present a similar pattern of variation with respect to landscape conditions. Acoustic evenness (AE) and temporal entropy (TE) indices were related to changes in terrain conditions, while acoustic diversity index (ADI) was associated to fragmentation. The bioacoustic index (BI), acoustic complexity index (ACI), and the number of peaks (NP) were positively related with water availability. Our results suggest that in Andean environments specific acoustic indices could be connected to particular aspects of landscape.

KEYWORDS: Acoustic Indices, Soundscape, Landscape, Northern Colombian Andes

The Diversity of Heath Flowering Phenology– Revealing Fine Scale Patterns of Heterogeneity by High Resolution Drone Cameras

***Carsten Neumann*[1], *Robert Behling*[1], *Sibylle Itzerott*[1], *Gabriele Weiss*[2], *Matthias Wichmann*[3], *Jörg Müller*[3]**

[1] *Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Germany*; [2]*ecostrat GmbH Berlin, Germany*; [3] *Heinz Sielmann Foundation, Germany*

Corresponding author(s) e-mail: carsten.neumann@gfz-potsdam.de

ABSTRACT:

Unmanned aerial vehicles (UAV) open up new perspectives for the repetitive spatial monitoring of vegetation stands and possibly even for fine scale analyses of individual plants. The potentials of recent UAV camera systems for integrating high spatial resolution (< 10 cm per image pixel) and area wide mapping are undergoing research. In particular, robust algorithms are needed for a spatially explicit characterization of individual plants, their structural composition and related functional traits. Such information can be used to examine ecological interdependencies that determine habitat establishment, and in general, to evaluate the habitat quality for nature conservation purposes [1].

In our study we developed a UAV-based methodological procedure to analyse open heathland areas that are protected in the European Natura 2000 network, and therefore, need to be managed to preserve a favourable conservation status [2]. We mapped the flowering phase of regenerated dwarf shrub heath (*Calluna vulgaris*) stands 2 years after burning and of unmanaged old stands using a standard RGB drone camera. From the UAV imaging point cloud a digital orthophoto and a digital surface model was generated with 2 cm pixel resolution. A method was developed that extracts every individual plant on the basis of RGB-colour value classification, texture filtering and local maxima estimation on a normalized digital surface model. For every plant the maximum plant height as well as the RGB colour distribution could be extracted and related to field surveys of plant traits. On that basis, a colour model was constructed that statistically predicts the phenological heath plant status with regard to fractional cover of flowers, fruits, vegetative shoots and senescent plant components.

Our study reveals that the flowering phase of *Calluna vulgaris* is spatially much more divers than expected. Colour models from simple RGB imagery show that phases of completely vegetative, withered, fruit and full flowering plants spatially coexist in heterogeneous alternations of individual plants, in fact, independent of growth age after burning (height < 40cm). Only unmanaged old stands (height > 40 cm) exhibit significantly lower fractions of fruit and flowers while senescence is increased. However, the high phenological diversity after fire management shows that driving factors of shoot extension and flower productions, thus, are triggered by additional mechanisms that are not affected by habitat management practice. UAV-based imagery can help to observe such effects for a more detailed evaluation of management effects.

KEYWORDS: unmanned aerial vehicles, habitat management, dwarf shrub heath, phenology, Calluna vulgaris

REFERENCES:

1. Neumann C, Weiss G, Schmidtlein S, Itzerott S, Lausch A, Doktor D, Brell M., 2015. Gradient-based assessment of habitat quality for spectral ecosystem monitoring. Remote Sensing. 7(3), 2871-98.

2. Habitat Directive, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Union. 206, 7-50.

Using automated species identification in passive acoustic recording to test the acoustic niche partition hypothesis in Neotropical frogs

Estefany Cano[1,3], Carol Bedoya[2], Claudia Isaza[2], Juan M. Daza[1]

[1] *Grupo Herpetológico de Antioquia, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia.*
[2] SISTEMIC, Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia.

Corresponding author: estefany.cano@udea.edu.co

ABSTRACT:

Technological advances in acoustic recorders and development of automatic identification of acoustic signals have boosted the use of passive acoustic monitoring to address questions at different levels of ecological organization (organisms, population, community, landscape). We used an acoustic data set of 24752 sound files, obtained through passive acoustic monitoring collected during 8 months in a neotropical forest on the northern Andes of Colombia to determine if frog species in a community show acoustic niche partition. To stablish of presence of a frog species and obtain spectro-temporal parameters and acoustic activity patterns, we implemented an automatic species identification method based on a fuzzy classifier and Frequency Cepstral Coefficients. We compared acoustic activity patterns and spectral-temporal characteristics among species to determine the extent of acoustic niche partition. The validation test for the automatic identification method was done with approximately 100 calls of 6 frog species detected, we obtained accuracies between 70% and 93%. We found differences in both spectro-temporal characteristics and activity patterns among all detected frog species. Our results show that the use of Frequency Cepstral Coefficients has high accuracy in the identification of advertisements calls of frogs and show that methodology based on a fuzzy classifier allow infer patterns in acoustic activity patterns and spectral characteristics in long temporal scale. Additionally, our results suggest that frog species in an Andean community partition their acoustic niche in time and spectral axes.

KEYWORDS: Amphibian communication, Colombia, Machine Learning, South America,

Automated recognition of people and identification of animal species in camera trap images

***Laura Hoebeke*[1], *Michiel Stock*[1], *Stijn Van Hoey*[2], *Jim Casaer*[2], *Bernard De Baets*[1]**

[1]*Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium*;
[2]*Research Institute for Nature and Forest, Belgium*

Corresponding author(s) e-mail: laura.hoebeke@ugent.be

ABSTRACT:

Camera traps are increasingly being used in wildlife monitoring. The great advantage of camera traps in comparison with other sampling methods is that very accurate data can be collected without the animal being collared or tagged nor the researcher being present. However, such camera trapping frameworks produce high volumes of pictures which often need to be reviewed manually. Convolutional neural networks can be used to automate this labour intensive process.

In our work, we use existing manually labelled images from a camera trap study conducted by the Research Institute for Nature and Forest in collaboration with Hasselt University (Belgium) to train a convolutional neural network for identifying animal species. Images were annotated using the camera trap application Agouti (www.agouti.eu). In this way images can be automatically labelled or the network can be incorporated into annotation applications to provide a suggestion to the users and as such speed up the annotation process.

In addition to conveying the presence or absence of species, the images may contain other useful information, for example animal attributes and behaviour. Therefore, getting help from wildlife enthusiasts via citizen science may be desirable to review the large amounts of data. However, since cameras are mounted in public nature reserves, there always exists the risk that passers-by have triggered the camera traps. For privacy reasons, images showing people cannot be made public. Removing these images from the dataset can be automated by training the network to recognise people in addition to identifying animals species, before the data can be made available to volunteers.

KEYWORDS: image recognition, classification, camera traps

A method for automatic creation of a vegetation map using high-resolution aerial photographs of unmanned aerial vehicles

***Masatoshi Denda*[1], *Sayuri Yoshioka*[2, 3], *Yoshiko Maeda*[2]**

[1] *Public Works Research Institute, Japan*; [2] *Kokusai Kogyo Co. LTD, Japan*

Corresponding author(s) e-mail: denda@pwri.go.jp

ABSTRACT:

Woodland overgrowth caused by the reduced flood disturbance and succession resets are critical technical issues to ensure effective river management in Japan. Woodland overgrowth can increase the risk of flooding and decrease the bio-diversity. Japanese river managers must control the woodland overgrowth and maintain the condition of vegetation. To maintain the woodland in a good condition, Japanese river managers must monitor the state of vegetation dynamics. Currently, Japanese river management uses vegetation maps to monitor the state of vegetation dynamics. Although vegetation maps are effective in monitoring the state of vegetation dynamics, creating vegetation maps will incur high cost.

Unmanned aerial vehicles(UAVs) exhibit strong potential to monitor the status of vegetation conditions. As an example of strong potential, we consider the potential of high-resolution aerial photographs. High-resolution aerial photographs provide detailed information about the vegetation growth conditions such as surface solid material and vegetation invasion on the river terrace. The high-resolution aerial photographs aid in understanding the vegetation dynamics in a limited area. If we are attempting to monitor the condition of vegetation dynamics across a large area, analysis of the images will enhance the quantitative monitoring of large spatial heterogeneous areas that are located within the river environment.Based on this background, we conducted image analysis using the high-resolution aerial photographs and applied decision tree analysis to the image analysis results (objects) after which we verified whether it is possible to create an automatic vegetation map. First, we apply object-based segmentation methods, which group together individual image pixels to objects with common characteristics using the image spectrum information (RGB information and brightness) and shapes. Second, we applied the machine learning methods (such as decision tree analysis) to the objects along with cross validation methods.We classified the objects into training data (supervised data) and validation data. After applying machine learning based on the training data, the accuracies of the machine learning methods were evaluated.From the results, we successfully created some approximate automatic vegetation maps. The automatic vegetation map represents the main part (dominating part) of the landscape (water area, gravel bed condition, annual vegetation community) and monitors the investigation and initial succession of the woodlands. The precision of the automatically created vegetation map is approximately 65%, and 35% of the errors are observed to be concentrated among minor objects. These results are produced by minor segmentation, and we hypnotize that reconsidering the initial step will improve the precision.

KEYWORDS: Unmanned aerial vehicles, objects-based clustering, decision tree analysis, vegetation map

Complimenting long-term bird monitoring observations with acoustic sensors and camera traps: best of both worlds

***Elizabeth Znidersic*[1], *Michael Towsey*[3], *David M. Watson*[1], *Kelly W. Roy*[2], *Sarah E. Darling*[4], *Anthony Truskinger*[3], *Paul Roe*[3]**

[1]Institute for Land, Water and Society, Charles Sturt University, Elizabeth Mitchell Drive, Albury 2640, NSW, Australia;[2]Oak Ridge National Laboratory. Environmental Sciences Division, Oak Ridge. Tennessee 37831, USA; [3]QUT Ecoacoustics Research Group, Science and Engineering Faculty, Queensland University of Technology, Australia; [4]Oak Ridge National Laboratory, Facilities and Operations Directorate, Oak Ridge, Tennessee 37831, USA

Corresponding author(s) e-mail: eznidersic@csu.edu.au

ABSTRACT:

In this era of dynamic environmental change, monitoring is becoming even more critical as species experience escalating pressures associated with reduced habitat availability, climate disruptions and invasive species. Although long-term monitoring is central to the assessment of spatial and temporal movements in species richness and abundance, this labour intensive approach is being increasingly challenged to maximize efficiency and cost effectiveness. We investigated the application of two recently developed technological approaches to biodiversity monitoring and compared them with long-term occurrence data for marsh bird species at the Oak Ridge Reservation, Tennessee. Oak Ridge National Laboratory staff have collected bird data for over 67 years (1950–2017), yielding observational records of 232 species. During May–June 2017, we deployed remote camera traps and autonomous acoustic sensors for a period of 35 days. The resultant accumulation of acoustic data necessitated an automated method to assist in the analysis and interpretation. We used a novel ‘soundscape’ technique, long-duration false-colour spectrograms to detect specific species. The acoustic sensors confirmed the presence of the Least Bittern (*Ixobrychus exilis*), a species previously only detected on 3 occasions at the Oak Ridge Reservation. In addition, a Purple Gallinule (*Porphyrio martinicus*) was ‘captured’ in a series of camera trap images, the first Roane County and Oak Ridge Reservation record of this species. Rather than replacing conventional monitoring approaches, these recently-developed methods are a useful complement, especially for furtive species and inhospitable habitats. As well as generating independently verifiable records, acoustic sensors and camera-traps provide novel means to interact with a wide range of stakeholders, generating engaging sounds and images to personalize a data-driven narrative.

KEYWORDS: Monitoring methods, Acoustic monitoring, Camera trap, long-duration false-colour spectrogram, soundscape, marsh bird

Optimisation of video monitoring of fish for reef assessment and management

Stijn Bruneel[1]***, Amber Schoeters***[1]***, Rafael Bermudez***[2]***, Peter Goethals***[1]

[1]*Department of Animal Science and Aquatic Ecology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium*; [2]*Facultad de Ingeniería Marítima, Ciencias Biológicas, Oceánicas y Recursos Naturales, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador*

Corresponding author(s) e-mail: Stijn.Bruneel@ugent.be

ABSTRACT:

Underwater video recordings provide a panoramic window into aquatic ecosystems. Camera technologies are rapidly advancing and become increasingly popular in ecological studies. However, to allow optimal use of these techniques, scientifically underbuilt methodologies are key. To this end, during a sampling campaign in 2017, reef fish were monitored via different setups of GoPro cameras at 10 sites in Santa Cruz and Floreana islands of the Galapagos Archipelago. The objective of this study was to determine the effect of scale, tides, camera setups, duration, repetitions as well as (number of) persons involved in snorkeling. Based on numerical analysis, an optimal approach was developed in order to maximize the reliability, repeatability and efficiency of fish monitoring.

KEYWORDS: Video monitoring, Reef fish, Method development, Galapagos

Annotating Species Trait Images with Absolute Size Information Using Mobile Devices

***Martin Hofmann*[1]*, Marco Seeland*[1]*, Patrick Mäder*[1]**

[1]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: marco.seeland@tu-ilmenau.de

ABSTRACT:

The projection of a real world scenery to a planar image sensor inherits the loss of information about the 3D structure as well as the absolute dimensions of the scene. For image analysis and object classification tasks, however, absolute size information can make results more accurate. Especially capturing, mapping, and analyzing species traits depends on the availability of accurate trait information. For morphological traits, relevant information is mostly constituted by geometric measurements of species characters such as leaf area, plant height, or wing width. Capturing such information requires cumbersome measurements. An alternative is taking images of the species, which can be analyzed later in a lab. However, the creation of size annotated image datasets is effort-intensive and typically requires measurement equipment not available to public image contributors. In this talk, we propose an effective annotation method that utilizes the camera within smart mobile devices to capture the missing size information along with the image. The approach builds on the fact that with a camera, calibrated to a specific object distance, lengths can be measured in the object's plane. We use the camera's minimum focus distance as calibration distance and propose an adaptive feature matching process for precise computation of the scale change between two images facilitating measurements on larger object distances. Eventually, the measured object is segmented and its size information is annotated for later analysis. We report on a user study showing that humans are able to retrieve the calibration distance with a low variance. We also demonstrate that the proposed approach facilitates a measurement accuracy comparable to manual measurement with a ruler and outperforms state-of-the-art alternative methods in terms of accuracy and repeatability. Consequently, the proposed method allows in-situ size annotation of traits in images without the need for additional equipment or an artificial reference object in the scene. Ultimately, we discuss the automated extraction and analysis of a set of morphological traits from size-annotated images of plant flowers and leaves.

KEYWORDS: trait annotation, metric scale, size measurement, image acquisition, mobile device

Session R2.1

UNDERSTANDING SPECIES DISTRIBUTION, POPULATION DYNAMICS AND PHENOLOGY BY MACHINE LEARNING

Understanding Species Distribution, population Dynamics and Phenology by Machine Learning

***Antonino Staiano*[1], *Friedrich Recknagel*[2], *Alberto Lanzoni*[3]**

[1]*University of Naples Parthenope, Italy;* [2] *The University of Adelaide, AUS*; [3]*University of Bologna, Italy*

Corresponding author(s) e-mail: antonino.staiano@uniparthenope.it

ABSTRACT:

Inferential modelling by machine learning techniques allows efficient and user-friendly analysis and synthesis of highly complex ecological data. Methods like random forest, quantile regression forest, Maxent, GARP have been successfully applied for species distribution modelling of extensive spatial data, resulting in species response curves that describe a species' response to a given habitat condition and illustrate specific habitat requirements for the species. Unsupervised artificial neural networks, regression trees and canonical correspondence analyses allow to ordinate and classify complex ecological data. Applications of support vector machines, supervised artificial neural networks and evolutionary algorithms to large spatial and temporal data allow to predict population dynamics and reveal phenology.

This session welcomes papers on all aspects of inferential modelling of ecological data by means of novel machine learning techniques.

KEYWORDS: Machine Learning, Random Forest, Quantile Regression Forest, Species Distribution, Population Dynamics, Maxent, GARP, Inferential Modelling, Hybrid Evolutionary Algorithm, Neural Network

Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data

Patrick Schratz[1], Jannes Muenchow[1], Jakob Richter[2], Eugenia Iturritxa[3], Alexander Brenning[1]

[1]*FSU Jena, Germany*; [2]*TU Dortmund, Germany;* [3]*NEIKER, Spain*

Corresponding author(s) e-mail: patrick.schratz@uni-jena.de

ABSTRACT:

Machine-learning algorithms have gained popularity in recent years in the field of ecological modeling due to their promising results in predictive performance of classification problems. While the application of such algorithms has been highly simplified in the last years due to their well-documented integration in commonly used statistical programming languages such as R, there are several practical challenges in the field of ecological modeling related to unbiased performance estimation, optimization of algorithms using hyperparameter tuning and spatial autocorrelation. We address these issues in the comparison of several widely used machine-learning algorithms such as Boosted Regression Trees (BRT), kNearest Neighbor (WKNN), Random Forest (RF) and Support Vector Machine (SVM) to traditional parametric algorithms such as logistic regression (GLM) and semi-parametric ones like Generalized Additive Models (GAM). Different nested cross-validation methods including hyperparameter tuning methods are used to evaluate model performances with the aim to receive bias-reduced performance estimates. As a case study the spatial distribution of forest disease (Diplodia sapinea) in the Basque Country in Spain is investigated using common environmental variables such as temperature, precipitation, soil or lithology as predictors.

Results show that GAM and Random Forest (RF) (mean AUROC estimates 0.708 and 0.699) outperform all other methods in predictive accuracy. The effect of hyperparameter tuning saturates at around 50 iterations for this data set. The AUROC differences between the bias-reduced (spatial cross-validation) and overoptimistic (non-spatial cross-validation) performance estimates of the GAM and RF are 0.167 (24%) and 0.213 (30%), respectively. It is recommended to also use spatial partitioning for cross-validation hyperparameter tuning of spatial data. The models developed in this study enhance the detection of Diplodia sapinea in the Basque Country compared to previous studies.

KEYWORDS: spatial modeling, machine learning, model selection, hyperparameter tuning, spatial cross-validation

Statistically reinforced machine learning for nonlinear interactions of factors and hierarchically nested spatial patterns

***Masahiro Ryo*[1, 2], *Matthias C. Rillig*[1, 2]**

[1]*Freie Universität Berlin, Germany*;
[2]*Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Germany*

Corresponding author(s) e-mail: masahiroryo@gmail.com

ABSTRACT:

A data-driven approach employing machine learning holds promise for applications in ecological informatics. Machine learning can automatically find unexpected relationships between variables and will help generate hypotheses. Despite recognizing such an advantage, many ecologists may feel uncomfortable with fully relying on a data-driven approach that is quite different from the hypothesis-driven approach (i.e. frequentist statistics). With several examples from our ongoing projects, we discuss how to apply machine learning in ecology in terms of both advantages and caveats. First, we introduce an emerging technique that blends machine learning and frequentist statistics: Statistically-reinforced machine learning. This modeling approach has potential to discover nonlinear and interaction relationships, relying on statistical significance without the requirement that the user specifies a priori which variables interact. We show an example where machine learning discovered the interactions of an elevational gradient, biogeoclimatic class, and forest coverage as the most important interactions from > 40000 possible three-way interactions for explaining freshwater biodiversity patterns across Switzerland. Second, we also introduce how to model spatial patterns that are hierarchically nested across multiple scales using machine learning, considering spatial autocorrelation that is largely overlooked.

KEYWORDS: Machine learning, variable interaction, biodiversity, spatial distribution, hypothesis generation

Reconstruction and Recognition of Spatial Patterns from Sparse Data in the Problem of Biological Invasion

Natalia Petrovskaya

University of Birmingham, UK

Corresponding author(s) e-mail: n.b.petrovskaya@bham.ac.uk

ABSTRACT:

In the modern world of big data there are still a plenty of ecological applications where data available for analysis are extremely sparse because of financial, labour, and other reasons. Accurate reconstruction of spatial patterns from sparse data remains a challenging problem where the results of reconstruction may heavily depend on the sampling protocol. One example is a problem of biological invasion where distinguishing between a patchy spatial density pattern and a continuous front spatial density pattern is crucial for monitoring and control of the invasive species. From the pattern recognition viewpoint, a continuous front density distribution can be classified as a single object, while a `no front' patchy invasion presents a collection of separate objects in a spatial domain. Sensitivity of such classification to the definition of a monitoring protocol remains an open question and will be discussed in the talk. Two basic properties of the monitoring protocol (i.e. the cutoff density value and the number of sampling locations) will be investigated and it will be shown how their variation affects reconstruction of spatial density patterns and recognition of the invasion type.

KEYWORDS: biological invasion, spatial pattern, monitoring protocol, pattern recognition

Modeling Green Peach Aphid populations exposed to elicitors inducing plant resistance on peach

***Alberto Lanzoni*[1], *Francesco Camastra*[2], *Angelo Ciaramella*[2], *Antonino Staiano*[2], *Giovanni Burgio*[1]**

[1] *Dipartimento di Scienze e Tecnologie Agro-Alimentari – Entomologia, Alma Mater Studiorum Università di Bologna, Italy;* [2] *Dipartimento di Scienze e Tecnologie, Università di Napoli Parthenope, Italy*

Corresponding authors e-mail: alberto.lanzoni2@unibo.it; antonino.staiano@uniparthenope.it

ABSTRACT:

Matrix Population Models (MPMs) are not commonly used to simulate arthropod population dynamics with applications to pest control assessment in agricultural context. However, an increasing body of studies are prompting the finding of optimization techniques to reduce uncertainty in matrix parameters estimation. Indeed, uncertainty in parameters estimates may lead to significant management implications. Here we present a case study where MPMs are used for assessing the efficacy of treatment with elicitors inducing plant resistance against pathogen, such as laminarin, for the control of the Green Peach Aphid (*Myzus persicae* Sulzer) populations on peach. Such demographic approach could be particularly suitable to study this kind of compounds, which are mainly characterized by causing sub-lethal effects rather than acute mortality. An artificially assembled system [1] was arranged since it is well suited to follow the fate and behavior of a population exposed to elicitors activating chemical defense in plant. The obtained data, consisting of population time series, were used to generate a stage-classified projection matrix. The general model used to simulate population dynamics consists of a matrix containing *i*) survival probabilities (the probability of growing and moving to the next stage and the probability of surviving and remaining in the same stage), and *ii*) fecundities of the population. Most of the used methods for estimating the parameter values of stage-classified models rely on following cohorts of identified individuals [2]. However, in this study the observed data consisted of a time-series of population vectors where individuals are not distinguished. The relationship between the observed data and the values of the matrix parameters that produced the series involves an estimation process called *inverse problem*. Since all demographic analyses rely on how much the estimated parameters of the matrix are able to represent population dynamics, a Genetic algorithm for inverse parameter estimation was used in order to find a better model fit for the observed stage class distributions. These results were compared to those obtained by the quadratic programming method [3] used for determining the set of parameters that minimizes the residual between the collected data and the model output.

KEYWORDS: Elicitors, Laminarin, Matrix population model, Genetic algorithm, Inverse problem

REFERENCES:

1. Macfadyen, S., Banks, J.E., Stark, J.D., Davies, A.P., 2014. Using semifield studies to examine the effects of pesticides on mobile terrestrial invertebrates. Annu. Rev. Entomol. 59, 383-404.
2. Caswell, H., 2001. Matrix population models, second ed. Sinauer Associates Inc., Massachusetts.
3. Wood, S.N., 1994. Obtaining birth and mortality patterns from structured population trajectories. Ecol. Monogr. 64, 23-44.

ACKNOWLEDGMENT: This work was supported by Gruppo Nazionale per il Calcolo Scientifico (GNCS-INdAM).

Testing the strengths of relationships between otter populations, fish and macroinvertebrate communities as well as habitat conditions across three Korean rivers by inferential modelling based on the hybrid evolutionary algorithm HEA

Sungwon Hong[1], Friedrich Recknagel[2], Tae-Soo Chon[1, 3], Gea-Jae Joo[1]

[1]*Pusan National University, RK*; [2]*University of Adelaide, Australia*; [3]*Ecology and Future Research Association, RK*

Corresponding author(s) e-mail: nanhsw@pusan.ac.kr

ABSTRACT:

The Eurasian otter *Lutra lutra* is semi-aquatic carnivore and apex predator in aquatic systems. Since Korean government has implemented comprehensive clean water plans (1989 ~ 1997), improved aquatic food-web structure sustains otter population recovery. In this study, using hybrid evolutionary algorithm (HEA), we would demonstrate the influential food chains to the otter populations. We used 22 chains of the food-web structures (11 water qualities measurements (BOD, NH_3N, NO_3N, TN, PO_4P, TP, water temperature, dissolved oxygen (mg/l), pH, conductivity, and turbidity), Diatom, chlorophyll a, five Macro-Benthic Invertebrates categories (Mollusca, Anthropoda, Annelida, Nematomorpha, and Platyhelminthes), and four fish categories (benthivore, herbivore, planktivore, and piscivores)). According to the 22 chains, we investigated spraint densities (no. spraint per 600m) as otter population indices at 250 sites in Nakdong River basin (NR, 2014-2016, three years), 92 sites in Youngsan River basins (YR, 2016), and 83 sites in Seumjin River basin (SR, 2016). In NR of 2014, otter populations seemingly affected by both micro-invertebrates (r^2 = 0.32) and fish densities (r^2 = 0.31). However, overall of NR (2014-16), otter populations have been more precisely expected by fish densities (r^2 = 0.41), other than water quality measurements (r^2 = 0.37), and macro-invertebrates (r^2 = 0.32). In Seumjin and Youngsan River basins, otter populations were more explained by macro-invertebrate (r^2 = 0.40), than fish (r^2 = 0.26) and water qualities measurements (r^2 = 0.28). Different river basins and years showed the different thresholds of different food chains. We concluded that otter population status could result in different sensitivity of chain of food-web structures.

KEYWORDS: *Lutra lutra*, food-web structure, model ensemble, hybrid evolutionary algorighm (HEA), South Korea

Community and Population Abundance Patterns in Benthic Macroinvertebrates in Streams Unravelled by Species Abundance Distribution and Machine Learning

***Tae-Soo Chon*[1,2], *Gyu-Suk Kwak*[2], *Yong-Hyuck Jang*[1], *Jaehan Choi*[3], *and Joo-Baek Leem*[1]**

[1]Ecology and Future Research Association, Busan (Pusan) 46228 Republic of Korea; [2]Division of Biological Sciences (Prof. Emer.), Pusan National University, Busan 46241 Republic of Korea; [3]Department of Physics, Inha University, Incheon 22212 Republic of Korea

Corresponding author e-mail: tschon.chon@gmail.com

ABSTRACT:

Stream communities/populations reveal complex ecological phenomena in space and time in responding to variable sources of natural and anthropogenic disturbances in watershed areas. Benthic macroinvertebrate communities were collected by the Surber sampler in streams across different levels of pollution in Korea. Community and population abundance patterns were analysed by species abundance distributions (SADs) and machine learning including self-organizing map (SOM). Ecological responses to natural (e.g., seasonality) and anthropogenic (e.g., pollution) environmental variability were illustrated at community (e.g., species richness) and population (e.g., sudden increase) levels. Community persistency was overall observed in SADs in least polluted sites except the disturbing periods (e.g., flooding), consisting of a small number of dominant species, intermediately abundant species and rare species. In contrast community/population structure was highly variable in response to different degrees of pollution including eruption of tolerant populations and changes in SAD patterns. The community and environmental variables were correspondingly differentiated across different levels of disturbances according to the SOM. Suitability of SAD and machine learning was demonstrated to address changes in ecological integrity and water quality in stream ecosystems.

KEYWORDS: Community structure, Water quality, Ecological integrity, Aquatic ecosystems, Environmental impact

Modelling urban bird breeding sites with a random forest classifier using indicators of spatial heterogeneity in plant communities derived from earth observation data

***Thilo Wellmann*[1], *Angela Lausch*[2], *Sebastian Scheuerl*[1], *Dagmar Haase*[1, 2]**

[1]*Humboldt Universität zu Berlin, Germany*; [2]*Helmholtz Centre for Environmental Research – UFZ, Germany*

Corresponding author e-mail: thilo.wellmann@hu-berlin.de

ABSTRACT:

Birds strongly respond to vegetation structure and composition, yet typical habitat models based on earth observation (EO) data use pre-classified data such as land use state classes for the habitat modelling. Since this neglects factors of internal spatial composition of the land use classes, we propose a new scheme of deriving multiple continuous indicators of urban vegetation heterogeneity using high-resolution earth observation datasets. The deployed concepts encompass spectral trait variations for the quantification of vegetation heterogeneity as well as subpixel vegetation fractions for the determination of the density of vegetation. Both indicators are derived from RapidEye data, thus featuring a continuous resolution of 5 meters. Using these indicators of plant heterogeneity and quantity as predictors, we can model the breeding bird habitats with a random forest machine learning classifier for our case study Leipzig while exclusively using one input dataset. Separate models are trained for the breeding ranges of 60 urban bird species (including 10 on the German red list), featuring medium to high accuracies (54–87%). Analysing similarities between models regarding variable importance of the single predictors allows species groups to be discriminated based on their preferences and dependencies regarding the amount of vegetation on the one hand, and its structure and heterogeneity on the other. The combination of continuous high-resolution EO data paired with a machine learning technique creates novel and very detailed insights into the ecology of the urban avifauna opening up possibilities of analysing and optimising different greenspace management schemes or future urban developments concerning overall bird species diversity or single species under threat of local extinction.

KEYWORDS: Habitat model, spatial heterogeneity, Random forest, Earth observation, urban avifauna

A machine learning approach to the assessment of the vulnerability of *Posidonia oceanica* meadows

***Elena Catucci*, Michele Scardi**

University of Rome Tor Vergata, Italy

Corresponding author(s) e-mail: catucci.elena@gmail.com

ABSTRACT:

In this study, we adopted a modelling approach to assess the vulnerability of *Posidonia oceanica* meadows, the most widespread seagrass in Mediterranean Sea. *P. oceanica* has a crucial ecological role all over the basin. In fact, this seagrass is a habitat-forming species that can extend from the surface to 45 m depth, forming meadows. These meadows rank among the most valuable ecosystem in the Mediterranean Sea, in term of the services they provide. However, in areas where alterations of environmental conditions happened, regression of the meadows may occur. Despite it is one of the main targets of conservation actions all over the basin, *P. oceanica* is declining at alarming rate, especially due to the anthropogenic impacts. Thereby, there is a urgent need to study the effects of environmental factors that could affect its ecological status.

We used a Random Forest for developing a Habitat Suitability Model (HSM) for *P. oceanica* in the Italian seas. The use of HSMs has been especially promoted to support ecosystem assessment and conservation planning, since they allow to better understand both the habitat requirements and the potential distribution of species.

Since the spatial distribution of meadows in Italian seas is already known, we used the HSM predictions to evaluate the suitability of the habitat for *P. oceanica* at large spatial scale and, consequently, we assessed the vulnerability of the meadows. Particularly, our occurrence data included both areas were *P. oceanica* was known as living and regressed meadows. After the RF training, we validated the model using an independent test set and we evaluated the performance using both ROC curve and K statistic. The results showed that the HSM presented a quite good level of accuracy. Thus, we carried out a spatial analysis of the HSM predictions in relation to the actual ecological status of *P. oceanica*. The results showed that in areas where living meadows actually occurred, high habitat suitability predictions were significantly more frequent. On the contrary, where regressed meadows were actually observed, predictions indicated low habitat suitability for *P. oceanica*.

This study stressed that modeling can effectively support the assessment of ecosystem status as well as conservation actions.

KEYWORDS: habitat suitability, random forest, *Posidonia oceanica*, assessment, conservation

Supervised learning methods to predict species interactions based on traits and phylogeny

***Michiel Stock*[1], *Bernard De Baets*[1]**

[1]*Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium*

Corresponding author e-mail: michiel.stock@ugent.be

ABSTRACT:

Species interaction networks - pollination networks, host-phage networks, food webs and the like - are key tools to study community ecosystems. Biologists enjoy working with networks as they provide a sound mathematical description of a system and come equipped with a large toolkit to analyze various properties such as stability, diversity or dynamics. Species interaction networks can be obtained experimentally or by field observations. Modern techniques such as DNA barcoding and camera traps, coupled with large databases, contribute further to the popularity of networks in ecology.

In practice however, a collected network rarely contains all *in sito* interactions, as this would require an unfeasible large sampling effort. Species distributions are also subject to changes, for example due to climate change, which leads to new potential interactions. It is of great importance to be able to predict such interactions, for example to anticipate the effect of exotic species in an ecosystem.

In our work, we study how to use supervised machine learning tools to be able to predict new species interactions. Based on an observed network, we learn a function that takes as inputs the description of two species (e.g. traits, phylogenetic similarity or a morphological description) and predicts whether these two species are likely to interact or not. This framework for *pairwise learning* is based on kernels and similar methods have been highly successful for predicting molecular networks and for recommender systems, as used by companies such as Netflix and Amazon. We have shown that these methods can detect missing interactions in many different types of species interaction networks. A large focus of our work is on how the accuracy of these models can be estimated realistically. Our methods are available in an R package called **xnet**, making them easy to use for ecology researchers.

KEYWORDS: Species interaction networks, Pairwise learning, Machine learning

Overall and site-specific response of the macroinvertebrate community of Swan Coastal Plain Wetlands (West Australia) to water quality gradients revealed by GF and HEA

Jawairia Sultana[1], ***Friedrich Recknagel***[1], ***Jennifer A. Davis***[2] ***and Bruce C. Chessman***[3]

[1]*University of Adelaide, Australia;* [2]*Charles Darwin University, Australia;* [3] *University of New South Wales, Australia*

Corresponding author(s) e-mail: friedrich.recknagel@adelaide.edu.au

ABSTRACT:

The Swan Coastal Plain is situated along the Western Australian seaboard and accommodates a large number of permanently, seasonally and episodically flooded wetlands. Many of these wetlands are affected by eutrophication and hydrological changes. A systematic monitoring program has been conducted between 1989 and 1990 to assess the environmental status of 41 selected wetlands based on measurements of 19 physical-chemical attributes and the collection of 253 macroinvertebrate taxa samples (Davis et al. 1993). This study analysed 35 wetlands with consistent data collected in Nov 1989 and Nov 1990 by means of Gradient Forest (GF) and the Hybrid Evolutionary Algorithm (HEA). Whilst GF allows identifying macroinvertebrate taxa with the "strongest overall response" to gradients in "important" physical-chemical attributes, HEA allows to model population dynamics of the taxa depending on "important" attributes identified by GF along all 35 wetlands. HEA models are represented by IF-THEN-ELSE rules whereby IF-conditions disclose attribute thresholds that indicate changes in the species abundance across the wetlands. GF suggested different ranking of the attributes EC, TN and DIP for both years as well as different taxa assemblages for same attributes in 1989 and 1990. Since results for merged data were also different, only the year-by-year specific results have been taken into account. When inferential models have been built for the 4 species that responded "strongest" to EC, DIP and TN in 1989 and 1990 by HEA, the threshold conditions fall in the range of overall gradients of these attributes discovered by GF. GF and HEA proved to be complementary tools for identifying overall attribute gradients and species- and site-specific thresholds in complex ecological data sets.

KEYWORDS: Bio-indication, Wetlands, Macroinvertebrates, Gradients, GF, HEA

REFERENCES:

Davis, J.A., Rosich, R.S., Bradley, J.S., Growns, J.E., Schmidt, L.G. and F. Cheal, 1993. Wetlands of the Swan Coastal Plain. Vol. 6: Wetland classification on the basis of water quality and invertebrate community data. Water Authority of Western Australia.

Causal relationships of *Cylindrospermopsis* dynamics with water temperature and N/P-ratios: a meta-analysis across lakes with different climate based on inferential modelling by HEA

Friedrich Recknagel*[1]*, Tamar Zohary*[2]*, Ilia Ostrovsky*[2]*, Jacqueline Rücker*[3]*,
***Philip Orr*[4]*, Christina Castello Branco*[5]*, Brigitte Nixdorf*[3] *and Ricardo Tezini*[6]**

[1]University of Adelaide, Australia; [2]Israel Oceanographic and Limnological Research, Israel
[3] Brandenburg University of Technology, Germany; [4]Griffith University, Australia; [5]Federal University of the State of Rio de Janeiro, Brazil; [6]University of Brasilia, Brazil

Corresponding author(s) e-mail: friedrich.recknagel@adelaide.edu.au

ABSTRACT:

The cyanobacterium *Cylindrospermopsis raciborskii* is globally distributed by taking advantage of sophisticated adaptation mechanisms. Akinetes that germinate only when water or sediment temperatures reach 22–23°C enable *C. raciborskii* to survive the cold months in temperate waters. Heterocyst cells allow *C. raciborskii* to grow in N-limited waters by performing N_2-fixation. High uptake affinity and storage capacity for phosphorus make *C. raciborskii* competitive in lakes with seasonally pulsing P-concentrations. In addition, buoyancy by gas vesicles allows *C. raciborskii* to access horizons with optimum light and nutrient conditions within the water column. This study analyses limnological data patterns of four meso- to eutrophic lakes with different environmental and climate conditions that have a history of *C. raciborskii* blooms: the temperate Langer See (Germany), the Mediterranean Lake Kinneret (Israel), the sub-tropical Lake Wivenhoe (Australia) and the tropical Lake Paranoa (Brazil). The meta-analysis is based on inferential models for *C. raciborskii* solely driven by either water temperature (WT) or N/P-ratios (N/P) for six selected years of data from each lake that include three 'high abundance' and three 'low abundance' years. The models discovered by the hybrid evolutionary algorithm HEA achieved average coefficients of determination $r^2>0.5$, and disclosed lake-specific WT- and N/P-thresholds indicating timing and magnitudes of bloom events of *C. raciborskii*.As outcome of the meta-analysis, following hypotheses will be will addressed: (1) *C. raciborskii* is overwintering in lakes with warmer climates and therefore less reliant on germination of akinetes. (2) *C. raciborskii* reaches highest abundances during episodes of warmest water temperatures in lakes across different climates. (3) Growth of *C. raciborskii* is thriving fastest in lakes across different climates during episodes of N-limitation by performing N2 fixation by heterocyst cells. (4) *C. raciborskii* is coping with P-limiting conditions by utilising its mechanisms for uptake and storage of inorganic as well as organic P.

KEYWORDS: Cyanobacteria bloom, *Cylindrospermopsis raciborskii*, Lakes, Climate, Meta-Analysis, Inferential models, HEA

Remote Sensing based Estimation of Forest Biophysical Variables using Machine Learning Algorithm

Ritika Srinet, Subrata Nandy, N.R. Patel

Indian Institute of Remote Sensing, Indian Space Research Organisation, Dept. of Space, Govt. of India
Dehradun 248001, India

Corresponding author(s) e-mail: ritikasrinet@gmail.com; subrato.nandy@gmail.com

ABSTRACT:

Leaf Area Index (LAI), Fraction of Intercepted Photosynthetically Active Radiation (fIPAR) and forest Aboveground Biomass (AGB) are important regulatory parameters for several functions of the forest canopy. An accurate information about the spatial variability of these biophysical variables is vital to capture the variability in estimates of gross primary productivity, carbon exchange and microclimate in terrestrial ecosystems. The present study aims at developing predictive models for generating spatial distribution of LAI, fIPAR and AGB by integrating remote sensing imagery and field data using random forest (RF) regression algorithm. The study was carried out in a tropical moist deciduous forest of Uttarakhand, India. Various spectral and texture variables were derived using Sentinel-2 data of 10 April 2017. In-situ measurements of LAI, incident Photosynthetically Active Radiation (PAR) above canopy (I_o), below canopy (I), and diameter at breast height (dbh) were taken. fIPAR and AGB were calculated. RF regression algorithm was used to optimize the variables to select the best predictor variables. Three models, using only spectral variables, only texture variables and both spectral and texture variables were tested. For all three biophysical variables, the models using both spectral and texture variables gave better results. The best predictor variables were used to map the spatial distribution of LAI, fIPAR and AGB. On validation, the models were able to predict LAI with R^2=0.83, %RMSE = 13.25%, fIPAR with R^2=0.87, %RMSE = 13.24%, and AGB with R^2=0.85, %RMSE = 12.17%. The estimated biophysical parameters showed high interdependence (LAI-fIPAR R^2= 0.71, LAI-AGB R^2=0.75 and fIPAR-AGB R^2= 0.74). The results showed that RF can be effectively applied to predict the spatial distribution of forest biophysical variables like LAI, fIPAR and AGB with adequate accuracy.

KEYWORDS: Leaf area index, Fraction of intercepted photosynthetically active radiation, Aboveground Biomass, Remote sensing, Random forest

A mixed model approach to modelling global habitat suitability and invasion risk of the American bullfrog

***Desiree Andersen*[1]*, Yikweon Jang*[1]**

[1] *Ewha Womans University, Republic of Korea.*

Corresponding author(s) e-mail: desireeka93@hotmail.com, jangy@ewha.ac.kr

ABSTRACT:

American bullfrogs (*Lithobates catesbeianus*), which are native to the eastern United States, have been widely introduced to other parts of the world through food and pet trade. These include parts of North, Central, and South America, Western Europe, and parts of Asia. In many of these regions, they have become invasive by predating or outcompeting native animals. Because of the potentially damaging impact on ecosystems where this species is not native, it is pertinent to delineate through habitat suitability models where it is most likely to find suitable habitat outside of its native range in order to effectively prevent future outbreaks. Here, we use presence points available through the Global Biodiversity Information Facility (GBIF) along with mixed method habitat suitability modelling to determine areas of highest suitability and areas most at risk of invasion. We first ran an ecological niche factor analysis (ENFA) on a global grid of presence and pseudo-absence points using five uncorrelated bioclimatic factors (Bio1: annual temperature; Bio2: mean diurnal range; Bio12: annual precipitation; Bio14: precipitation of driest month; Bio15: precipitation seasonality). The original and ENFA modelled presence-pseudoabsence points were then fed separately into random forest algorithms using the same five bioclimatic variables. The random forest result of original data was considerably more conservative than the result from the ENFA modelled points, which is more similar to previous MAXENT models of this species found in the literature; however, the result of the original data indicates areas where the species is already present and therefore presents more risk than what is predicted by ENFA modelling. The two results were subsequently averaged to indicate areas at greatest risk of invasion; points with a value > 0.5 were considered risk areas. Our final model showed that the highest risk areas are located in Western North American, South America, Western Europe, parts of Africa, Japan, Southeast Asia, Western Australia, and New Zealand. Within these areas, the American bullfrog already occurs at high levels in Western North America and moderate levels in South America, Western Europe, and Japan; the species also has limited occurrences in Southeast Asia.

KEYWORDS: American bullfrog, invasive species, habitat suitability models, ecological niche factor analysis, random forest

Dynamics of Four Cyanobacteria in the Nakdong River, South Korea over 24 years (1993-2016) Patternized by an Artificial Neural Network

***Hyo Gyeom Kim*[1]*, Sungwon Hong*[1]*, Dong-Kyun Kim*[2]*, Gea-Jae Joo*[1]**

[1]*Pusan National University, South Korea*; [2]*University of Toronto, Canada*

Corresponding author(s) e-mail: gjjoo@pusan.ac.kr

ABSTRACT:

Cyanobacterial communities are affected by both biotic and abiotic factors, which makes it difficult to predict the timing and magnitude of their occurrences. Ecological or physiological traits of each species have been delineated, but changes in environmental conditions contribute to define their community dynamics. In this study, we used a self-organizing map (SOM, an artificial neural network algorithm) to identify patterns in cyanobacterial assemblages and environmental factors that affect them. Water quality was checked weekly or biweekly over a 24-year period, and meteo-hydrological data in the lower reach of the Nakdong River were used to forecast the abundance of four cyanobacterial genera: Anabaena, Aphanizomenon, Oscillatoria, and Microcystis. Classification of the SOM output neurons into six clusters that grouped the exemplar with similar community assemblages allowed to determine the major abiotic factors. Highest abundance of Anabaena sp. and Oscillatoria sp. were classified over multiple clusters, while Microcystis sp. were not assigned to a specific cluster. Microcystis sp., which are the most dominant cyanobacteria in the Nakdong River, are less sensitive to environmental factors and have competence to coexist with other species. Unlike the other species, Aphanizomenon sp. showed a higher abundance in cluster five, where water temperature and total phosphorus concentration were relatively low and transparency was highest. Increased water quantity due to the anthropogenic modification (2009–2011) on the riverine area and nutrient management policy implemented from 2011 decreased the phosphorus concentration. This change might be attributed to the re-emergence of Aphanizomenon sp., having low temperature tolerance, in 2014. Therefore, our study suggests that the shifting of the N:P ratio and abiotic factors influence the distribution of subdominant cyanobacterial groups.

KEYWORDS: Artificial neural networks, Self-organizing map, Cyanobacteria, Temporal species dynamics

Integrating context-based recommendation with deep CNN image classification for on-site plant species identification

Hans Christian Wittich[1], ***David Boho[1]***, ***Patrick Mäder[1]***

[1]*TU Ilmenau, Germany*

Corresponding author e-mail: hans-christian.wittich@tu-ilmenau.de

ABSTRACT:

Accurate plant species identification is essential for many scenarios in botanical research and conservation of biodiversity. Since a main obstacle is the large number of possible candidate species to consider, assistance through automatic identification techniques is highly desirable. On one side, photos of plant organs taken by users in the field can effectively be used in machine learning-based image classification, predicting the most likely matching taxa. At the same time, metadata on the user's spatio-temporal context usually goes unused despite its potential to be considered as an additional aspect to augment and improve prediction quality.

We develop a recommender system utilizing a user's context to predict a list of plant taxa most likely to be observed at a given geographical location and time. Using a data-driven approach, we integrate knowledge on plant observations, species distribution maps, phenology and environmental geodata in order to calculate contextual recommendations on a local scale. The resulting model facilitates fine-grained ranking of plant taxa expected to occur in close proximity to a user in the field.

Focusing on the territory of Germany with a list of the most common wild flowering plant taxa we are presented with a 2.8k class problem. Using a NASNet deep convolutional neural network trained on 860k taxon-labelled plant images we can presently achieve a 82% top-1 prediction accuracy. For a recommender system the combination of biogeographical information, phenology and habitat suitability models is showing viable results, being able to reduce the list of candidate taxa on average more than threefold with a recall of 25% for the top 20 list positions, 50% for the first 70 and a 90% recall for the full recommended list, based on contextual metadata alone. We show how prediction performance can be improved by merging context-based recommendations with deep learning image classification.

KEYWORDS: Plant Species Identification, Deep Learning, Recommender System, Spatio-Temporal Context

Session S1.1

EARTH OBSERVATION FOR ECOSYSTEM ANALYSIS AND DECISION MAKING

Earth Observation for ecosystem analysis and decision making

***Klaus Joehnk*[1], *Janet M. Anstee*[2]**

[1]*CSIRO Land and Water, Australia*; [2]*CSIRO Ocean and Atmosphere, Australia*

Corresponding author(s) e-mail: Klaus.Joehnk@csiro.au; Janet.Anstee@csiro.au

ABSTRACT:

Earth observation data is acquired from satellites, airborne platforms, and in-situ measurements on land and in water. Combined with modelling tools this yields detailed insights into ecosystem functioning from small to large scales and allows for improved prediction algorithms, which then provides decision makers with reliable and up-to-date information.

In this special session we want to address the challenges and opportunities of the use of Earth observation data in modelling and forecasting of ecosystem trends. Applications of remote sensing in water quality modelling, biodiversity, invasive species, vegetation mapping, primary productivity etc. are discussed.

KEYWORDS: Earth observation, remote sensing, water quality, bio-optics

Forest biodiversity estimated from the space: testing the Spectral Variation Hypothesis comparing Landsat 8 and Sentinel 2 using a multi-temporal Rao Q

Michele Torresani*[1]*, Duccio Rocchini[2, 3, 4], Ruth Sonnenschein[5], Marc Zebisch[5], Giustino Tonon[1]

[1]University of Bolzano/Bozen, Italy; [2]Department of Biodiversity and Molecular Ecology, Fondazione E.Mach, Italy; [3]Center Agriculture Food Environment, University of Trento, Italy; [4]Centre for Integrative Biology, University of Trento, Italy; [5] Earth Observation Institute, European Research Academy, Italy

Corresponding author e-mail: michele.torresani@natec.unibz.it

ABSTRACT:

Forests cover about 30 percent of the earth surface, they are the most biodiverse terrestrial ecosystems and they are at the base of many ecological processes and services. The loss of forest biodiversity makes in risk the benefits that the humans derived from theme. The assessment of biodiversity is therefore an important and essential goal to achieve, that however can result difficult, time consuming and expensive if estimated through field data. Through the remote sensing it is possible to estimate in a more objectively way the species diversity, using limited resources, covering broad surfaces with high quality and standardized data. One of the method to estimate biodiversity from remote sensing data is through the Spectral Variation Hypothesis (SVH) , which states that the higher the spectral variation of an image, the higher the environmental heterogeneity and the species diversity of that area. The SVH has been tested using different indexes and measures; recently in literature, the Rao's Q index, applied to remote sensing data has been theoretically tested as a new and innovative spectral variation measure. In this paper for the first time, the SVH through the Rao's Q index has been tested with an NDVI time series derived from the Sentinel 2 (with a spatial resolution of 10m) and Landsat 8 satellites (spatial resolution of 30m) and correlated with data of species diversity (through Shannon's H) collected in forest. The results showed that the Rao's Q is a grateful spectral variation index. For both the sensors, the correlation with the field data had the same tendency as the NDVI trend, reaching the highest value of correlation (through the coefficient of determination R^2) in June, when the NDVI was at its peak. In this case the correlation reached a value of R^2=0.61 for the Sentinel 2 and of R^2=0.45 for the Landsat 8, showing that the SVH is scale and sensor dependent. The SVH tested with optical images through the Rao's Q index showed grateful and promising results in alpine forests and could lead to as much good results with other remote sensing data or in other ecosystems.

KEYWORDS: Forest biodiversity, Spectral Variation Hypothesis, Rao's Q, NDVI, Sentinel 2, Landsat 8

Gross Primary Production and spring onset linked by spatio-temporal data analysis

***Emma Izquierdo-Verdiguier*[1, 3], *Alvaro Moreno-Martinez*[2], *Raul Zurita-Milla*[3], *Gustau Camps-Valls*[1], *Steve Running*[2]**

[1]*Universitat de València, Spain*; [2]*University of Montana, USA*; [3]*University of Twente, the Netherlands*

Corresponding author(s) e-mail: emma.izquierdo@uv.es

ABSTRACT:

The Earth is a complex networked spatio-temporal system, and we want to monitor and understand the complex changes that happen therein. One of current well-known changes is the climate change and, phenology is one of the best indicators of it. The phenological products provide information of biological events in space and time. New phenological products based on weather data are available and provide consistent information on the times recurrent biological events. Earth observation satellites can also be used to study vegetation dynamics over large areas. Despite the potential complementarity of these information sources, relatively few studies have examined their links in detail. Studying the relationship between them and understanding the changes that happen between them are necessary to better describe the variations produced by the impacts of climate change onto the vegetation over time.

This work explores the spatio-temporal data relations between Gross Primary Production (GPP) and a suite of temperature based spring onset indices based on correlation and clustering. The analysis is done at 1 km spatial resolution for the period 2000 to 2015, in the contiguous US. We used the MODIS annual cumulative GPP and the leaf and bloom indices derived from the extended spring indices (SI-x) models. These indices were obtained using Daily Surface minimum and maximum temperature data (Daymet).

Preliminary results confirm the correlation between the GPP and SI-x products in the central region of the USA, whereas both the Western and Eastern coasts are mainly anti-correlated (i.e. later spring onset leads to less cumulative GPP). In addition, the Leaf index presents lower correlations than the Bloom index, indicating that larger periods are needed to see significant direct effects. Regarding clustering preliminary results, the regionalization is dependent to the latitude at east of the USA. These results help to develop an understanding of the links among the green-wave components for monitoring the impacts of inter-annual and decadal climatic variability on vegetation dynamics.

KEYWORDS: GPP, Extended spring indices, Correlation, Clustering, Cloud computing, Big data

Monitoring the spread of invasive plant species in Germany – how many species can we possibly detect by remote sensing and what data do we need?

Sandra Skowronek[1]***, Stefanie Stenzel***[1,]***, Hannes Feilhauer***[2]

[1]*German Federal Agency for Nature Conservation, DE*; [2] *Friedrich-Alexander-University of Erlangen-Nürnberg, DE*

Corresponding author(s) e-mail: Sandra.Skowronek@BfN.de

ABSTRACT:

Combining remote sensing and field data allows for the detection of some invasive alien plant species with an adequate accuracy. Especially the use of satellite data for larger areas or UAS (unmanned aerial system) data for smaller sites may provide alternatives to classical field mapping approaches. A main advantage is that satellite or UAS data is potentially more cost-efficient then the use of for example hyperspectral data, which was frequently applied in research on the detectability of invasive species in the past. This study discusses the possibilities and limitations of remote sensing to contribute to the detection of invasive alien plant species in Germany. Taking into account previous studies on the topic, we estimate the potential for a successful detection of relevant invasive plant species in Germany. Main criteria to determine the potential for detection are the species characteristics (size, detectable traits, habitat) as well as their similarity to other native species.

For 19 of the 42 species examined, the use of remote sensing data is most probably successful, mainly for larger species and species with characteristic features such as colorful flowers or leaves. For another 10 species the detection might eventually be feasible. For about 13 species, especially hydrophytes living below the water surface and other species lacking any characteristic features, the detection is currently not possible.

We can conclude that remote sensing remote may offer efficient solutions for a small or large scale monitoring of certain invasive plant species or to control the management success and thus support decision-making. In general, more research is needed to develop cost-efficient and user-friendly solutions.

KEYWORDS: invasive alien species, mapping, monitoring, aerial photography, satellites, UAV

Spatial Patterns of Leaf Mass per Area of Wetland Vegetation under Water Stress Analyzed with Imaging Spectroscopy

Hannes Feilhauer[1], Thomas Schmid[2], Ulrike Faude[3], Salvador Sánchez-Carrillo[4], Santos Cirujano[5]

[1]*FAU Erlangen-Nürnberg, Germany*; [2]*CIEMAT, Madrid, Spain*; [3]*Freelancer, Spardorf, Germany;* [4]*NCN-CSIC, Madrid, Spain;* [5]*RJB-CSIC, Madrid, Spain*

Corresponding author(s) e-mail: hannes.feilhauer@fau.de

ABSTRACT:

Plant and community traits of wetland vegetation show a high intra-specific plasticity, originating from the high variability of environmental conditions. Remote sensing approaches promise to be able to retrieve some of these traits and their plasticity from the spectral reflectance signal of the canopy. In the present study, we evaluate a remote-sensing based approach for an analysis of spatial patterns of leaf mass per area (LMA), a key trait for ecosystem functioning and good negative correlate of potential growth rate. The test was conducted in Las Tablas de Daimiel, a National Park in Central Spain. This wetland was affected by a long-term drought, which introduced pronounced trait plasticity as part of the adaptation mechanisms of the vegetation to reduced water availability as well as a decrease in photosynthetic activity. Imaging spectroscopy (HyMap) data of the wetland were acquired in 2009 at peak drought intensity. At the same time, a field campaign was conducted. We applied an inversion of the PROSAIL model on these data to map the LMA distribution across the wetland. PROSAIL is a radiative transfer model that simulates the physical principles of light absorption and scattering in a vegetation canopy. The inversion enables the retrieval of trait information from the spectral signal. Furthermore, we assessed trends in photosynthetic activity and changing species composition across the wetland by analyzing time series of the normalized difference vegetation index (NDVI) as determined from various multispectral sensors. The mapped LMA values were analyzed within and between stands of different species and communities along a gradient of changing photosynthetic activity and species composition.

LMA values retrieved for stands of species with high photosynthetic activity at peak drought intensity closely met values reported in trait data bases. The observed intra-specific LMA variability is in line with the expected plasticity of this trait along a moisture gradient that is reflected in a change in photosynthetic activity and species composition. We thus conclude that remote sensing approaches provide sufficient detail to trace the LMA-response of wetland vegetation to long-term drought stress.

KEYWORDS: Remote sensing, Plant traits, Drought, Radiative transfer, Random forest

A constrained depth-resolved artificial neural network model of marine phytoplankton primary production

F. Mattei*, *M. Scardi

Experimental Ecology and Aquaculture Laboratory, Department of Biology, University of Rome "Tor Vergata", Rome, Italy.

Corresponding author e-mail: francesco.mattei90@yahoo.it

ABSTRACT:

Marine phytoplankton primary production is a process of paramount importance not only in biological oceanography, but also in a wider perspective, due to its relationship with oceanic food webs, energy fluxes, carbon cycle and Earth's climate.

As field measurements of this process are both expensive and time consuming, indirect approaches, which can estimate primary production from remotely sensed imagery are the only viable solution.

We developed a depth-resolved model of marine phytoplankton primary production using an Artificial Neural Network, namely a three-layer perceptron trained with the Error Back-Propagation algorithm.

Despite numerous variables could be useful to estimate primary production, we chose to use predictive variables that can be acquired by remote sensing in order to enhance the practical value of the model. Indeed, using exclusively this type of predictors in combination with a depth-resolved approach allows to expand the two-dimensional view from satellite images to the estimated three-dimensional distribution of phytoplankton primary production.

Since the vertically integrated values of this process are the basis for any connection to other levels of the pelagic food web, it is worth noting that, once integrated, the primary production estimates of this depth-resolved model are more accurate than those obtained from a similar vertically integrated approach.

We also tried to improve the accuracy of the primary production estimates using constraints during the training procedure. Those constraints were based on theoretical knowledge of the marine photosynthesis process. Accordingly, the training phase has been modified in order to add penalty terms to the solutions which were not compliant with the constraints. For instance, one of the constraints acts as a selection tool for the shape of the modelled production profile.

The above-mentioned approach not only enhanced the ecological soundness of the artificial neural network predictions. In fact, the constrained version of the model also explained a larger share of variance than the original one.

KEYWORDS: Phytoplankton primary production, Artificial Neural Networks, Depth-resolved model, constrained training, remote sensing

Exploring environmental variables based on ecotopes derived by remote sensing

Julien Radoux[1], Brieuc François[1], Thomas De Maet[1], Quentin Vandersteen[1], Elie Khalil[1], Pierre Defourny[1]

[1]*Université catholique de Louvain, Belgium*

Corresponding author e-mail: Julien.radoux@uclouvain.be

ABSTRACT:

LifewatchWB contributes to the European Research Infrastructure Consortium for biodiversity and ecosystem research (Lifewatch) by providing online tools for the visualization of environmental variables. It consists in two Web portals (www.uclouvain.be/lifewatch) providing data about i) weekly anomalies of the dynamic land cover properties (snow cover and vegetation greenness) and ii) high resolution characterization of ecotopes. The topic of this abstract is the ecotope visualization interface.

Ecotopes are the smallest ecologically functional units. They can be mapped by intersecting a large number of thematic layers (soil type, land cover, topographic types). However, this creates a lot of polygons with an exponentially increasing number of categorical values combinations. The visualization and analysis of those polygons is therefore difficult. On the other hand, the LifewatchWB ecotopes consist in irregular polygons derived from geographic object-based image analysis (grouping adjacent pixels of similar properties to create a partition of irregular polygons). Each polygon is then characterized using a set of continuous fields: quantitative variables (climate, elevation, slope, artificial light, contextual variables…) are either interpolated or averaged at the level of the ecotopes depending on the resolution of the input layers; categorical variables (soil types, land cover types…) are provided as proportions inside the ecotopes. The ecotope database is an open data layer currently available for the Walloon Region (Belgium). It consists in 1.2 million polygons with more than 80 quantitative fields. Those fields cannot be visualized together, therefore several visualization tools are provided on the interface: 1) each variable can be visualized individually in grey level, 2) any set of three variables can be visualized as a color composite and 3) queries based on potentially all variables can provide binary outputs. In addition, an online tool is available to extract the values of the ecotope characteristics in a set of point locations for further analysis in other softwares. With its polygon-based structure that is homogeneous with respect to the land cover and the topography, ecotopes provides meaningful landscape units with a large set of precomputed variables. Those variables are easy to visualize in a WebGIS and have been combined with species observations to run habitat and biotope models.

KEYWORDS: WebGIS, ecotopes, GEOBIA, integrated data

Estimating Gross Primary Productivity in Crops with Satellite Data, Radiative Transfer Modeling and Machine Learning

***Aleksandra Wolanin*[1], *Luis Guanter*[1], *Gustau Camps-Valls*[2], *Luis Gómez-Chova*[2], *Gonzalo Mateo García*[2], *Christiaan van der Tol*[3], *Yongguang Zhang*[4]**

[1]*GFZ German Research Centre for Geosciences, Germany*; [2]*University of Valencia, Spain*;
[3]*University of Twente, the Netherlands;* [4]*Nanjing University, China*

Corresponding author(s) e-mail: ola@gfz-potsdam.de

ABSTRACT:

Monitoring spatio-temporal changes in terrestrial gross primary productivity (GPP) of crops is key for estimating, understanding and predicting global carbon fluxes. Satellite remote sensing has been widely applied in the last decades to monitor agricultural resources, and the amount and quality of remote sensing data continuously increase. Since recently, and partly due the European Copernicus Programme, an unprecedented amount of open access data suitable for agriculture observations is now available. Benefiting from recent developments in satellite remote sensing technology, great advances in machine learning and advancements in our understanding of photosynthetic processes leading to increasingly complex and detailed photosynthesis models, we developed a hybrid approach to model GPP using satellite reflectance data by combining radiative transfer modeling and machine learning (ML).

We have combined process-based model SCOPE with ML algorithms to estimate GPP of C3 crops using a variety of satellite data (Sentinel-2, Landsat and MODIS) and ancillary meteorological information. We link reflectance and meteorological data directly with crop GPP, bypassing the need of retrieving the set of input vegetation parameters needed to represent photosynthesis in an intermediate step, while still accounting for the complex processes of the original model.

Several ML models, trained with the simulated data, were tested and validated using flux tower data. First, we tested our approach using Sentinel-2 data, which provide high frequency of observation, high spatial resolution of 20 m and multiple bands including red edge. Our final neural network model was able to estimate GPP at the tested flux towers with r2 of 0.92 and RMSE of 1.38 gC m^{-2} d-1. Our model successfully estimated GPP across a variety of C3 crop types and environmental conditions, including periods of no vegetation, even tough it did not use any additional local information from the site. Since our learning approach is fast and efficient in the test phase and, at the same time, is based on a process-based model (and not on local empirical relationships), it can be applied globally. Furthermore, the simulated training dataset can be easily adapted to band settings of different instruments, assuring thus consistency among many sensors. However, such a global application requires high computational power and therefore we applied our approach to Landsat and MODIS data using Google Earth Engine (GEE) platform that provides cloud computing resources for processing large geospatial datasets. The results were validated using the FLUXNET2015 Dataset.

KEYWORDS: Remote sensing, Crops, Gross Primary Productivity (GPP), Machine Learning, Neural Networks, Google Earth Engine (GEE)

Estimating Grassland Biomass - Potentials and Limitations of Point Cloud Analysis

***Thomas Moeckel*[1], *Michael Wachendorf*[1]**

[1]*Kassel University, Germany*

Corresponding author(s) e-mail: thmoeck@uni-kassel.de

ABSTRACT:

Quantifying above ground biomass of grasslands is important information for grassland management and the understanding of ecological processes in grassland habitats. Often, allometric relationships between grassland height and biomass are used for biomass estimation. While these methods may be used in intensively used grassland with a homogenous canopy surface, in heterogenous grasslands it is not possible to repeat these measurements on larger areas. Recent technological advances in active and passive remote sensing offering new opportunities for estimations of grassland biomass. Many studies using remote sensing data for biomass estimation are based on the analysis of optical remote sensing sensors and are situated in forests and agricultural crops. Small temporal and spatially heterogenous grasslands were often neglected due to their complex vegetation structure. Just recently, point cloud data based either on terrestrial laser measurements (TLS) or on photogrammetric image analysis (SfM) approaches were investigated for their potential of biomass estimation in grasslands. The focus of this talk will be on evaluating the potential of TLS and SfM derived point clouds in deriving biomass estimation of grasslands with very different land use intensities. Both approaches show promising results for predicting grassland biomass (R^2 ranging from 0.48 to 0.79 and from 0.35 to 0.81 for TLS and SfM respectively). TLS always performs better, which could be explained by the higher point densities and thus higher information content about the vegetation structure. However, under consideration of price and expert knowledge UAV based point clouds also produce satisfying results. Another aspect of the talk will be the comparison of performance aspects (e.g. computing time) of different point cloud analysis strategies. It can be shown that two scans of the same location from different aspects already provide detailed information about biomass and additional scans only lead to an unnecessary increase in data volume while maintaining consistent prediction quality. Various analysis methods will be test for extracting information from the point clouds. Here methods based on canopy surface height show the best prediction performance for biomass. Concluding, it is possible to say that both TLS and SfM-based point clouds have a good potential for deriving biomass information of grasslands, independent of land use intensity. However, to derive final conclusions the stability of the statistical relationships needs to be test over several growing periods. For the future, also the fusion of point cloud information with spectral information should be tested, as better biomass prediction models can be expected from this.

KEYWORDS: Biomass, Grassland, Point cloud, Terrestrial laser scanner, UAV RGB

Integration of near-surface and satellite observations for algal bloom detection

***Janet M. Anstee*[1], *Klaus Joehnk*[2], *Phillip Ford*[1], *Tim Malthus*[1], *Elizabeth Botha*[1], *Marit van Oostende*[1], *Eric Lehmann*[3], *Xavier Ho*[3], *Stephen Gensemer*[4]**

[1]CSIRO Oceans and Atmosphere, Australia; [2]CSIRO Land and Water, Australia; [3]CSIRO Data61, Australia; [4]CSIRO Data61, Australia; [5]CSIRO Manufacturing, Australia.

Corresponding author(s) e-mail: janet.anstee@csiro.au

ABSTRACT:

Retrieval of water quality information from satellite imagery can provide resource managers with an improved understanding into the spatial variability of the water body. In light of the increasing availability of 'analysis ready data' (ARD) satellite imagery in open datacubes*, either on cloud-based services or on high performance computing environments, development of operational monitoring systems is becoming feasible.

Near-surface sensors can assist in more rapid and widespread algal bloom monitoring at a much higher temporal resolution. Remote sensing imagery, whilst cost effective, may not be optimal in terms of spatial or spectral resolution and can be greatly enhanced with the integration of near-surface observations. We describe pathways to use field-based near-surface sensors to calibrate and validate satellite remote sensing. These methods allow early detection of algal blooms and assist in the early warning for management intervention.

We have designed and deployed several low-cost, near-surface sensors at several inland water sites around eastern Australia. The data is transferred using mobile networks where it is processed into spectral information. From this data and coincident field bio-optical measurements, we have developed algorithms for quantitative estimation of blue-green algal-specific pigments (phycocyanin) and chlorophyll concentrations. We have tested these algorithms for detection using a number of existing satellite sensors and report on results here.

These methods have applied next-generation monitoring technology and when combined with hydrologic modelling will provide aquatic observations and forecasts. These will lead to improved management preparedness to respond to environmental challenges, e.g., a harmful algal blooms.

KEYWORDS: Open Datacube, algal bloom detection, monitoring

REFERENCES:

1. Open Data Cube. 2018. "OpenDataCube". opendatacube. https://www.opendatacube.org/.

A neural network approach to infer the 3D chlorophyll-a field from remote sensing observations

***Michela Sammartino*[1, 2], *Salvatore Marullo*[3], *Rosalia Santoleri*[1], *Michele Scardi*[2]**

[1]*National Research Council - Institute of Atmospheric Sciences and Climate (CNR-ISAC), Rome, Italy*; [2]*Department of Biology, University of Rome -Tor Vergata, Rome, Italy*; [3]*ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Frascati, Italy*

Corresponding author e-mail: michela.sammartino@artov.isac.cnr.it

ABSTRACT:

Phytoplankton is one of the most important marine bioindicators to be monitored in their time and space dynamics. In this context, chlorophyll-a (Chla) represents one of the most common proxies for phytoplankton biomass.

At surface, the use of satellite sensors has allowed to acquire data on phytoplankton variability at high spatial and temporal resolution. On the contrary, the vertical Chla pattern remains less explored. In fact, the characterization of Chla in subsurface layers is still restricted to the collection of in situ samples and field activities, which are time consuming and require a significant economic effort.

A big challenge is represented by the possibility to exploit high resolution satellite imagery to obtain information also about the vertical distribution of water column properties. In this study, we propose a combined use of satellite observations and Machine Learning techniques to realize the extension of the Chla field from 2D to 3D.

Here, a multilayer perceptron trained with an error back-propagation algorithm (BPN) is used to infer the Chla vertical profile from surface data only. Satellite estimates of Chla and Sea Surface Temperature (SST) are used as co-predictors for the Chla vertical reconstruction. The network is trained and validated using a large in situ dataset of temperature and Chla profiles collected in the Mediterranean Sea from 1998 to 2015. Our results are comparable to similar attempts, but highlight promising features that might prove useful to tackle some of the practical difficulties that characterize studies on the vertical Chla variability. In fact, given the synopticity of the satellite data, this method represents the first attempt at creating a working alternative to classical models and at overcoming the discontinuous nature of in situ sampling. The potentiality of neural networks in generalizing complex scenarios, without a priori assumptions, is a very useful feature in the monitoring of Chla into the oceans; however, the prediction capability of such neural networks is strictly depending on training dataset features, that deeply influence the network's performance.

KEYWORDS: Chlorophyll-*a*, Mediterranean Sea, Machine Learning, Vertical profile, Satellite.

Towards Understanding the Trends of Informal Harvesting of Sand Forest in Maputaland, South Africa

Ryan Nel[1], Kevin Mearns[1], Maarten Jordaan[1]

[1]*University of South Africa, South Africa*

Corresponding author(s) e-mail: 48160407@mylife.unisa.ac.za; Mearnkf@unisa.ac.za

ABSTRACT:

Indigenous forests and savannah provide numerous benefits for rural communities and are utilised as a source of firewood, building material and for woodcraft production. Currently, there is insufficient information on the magnitude of human pressure affecting one such important forest community, namely Sand Forest, particularly in communal areas. Sand Forest is regarded as being critically endangered and is considered to hold various endemic species. The fragmented patch occurrence of this rare and valuable forest type, combined with the lack of necessary knowledge and prior interest in its management, has resulted in the Sand Forest being subjected to uncontrolled utilisation within communal areas.

The temporal monitoring of the spatial structures of forest areas, such as Sand Forest, within landscapes has been recommended in order to detect and model deteriorating trends in the forest structures and functioning. Remote sensing is critical in the generation of data that enables the identification and quantification of degraded and deforested areas. This study aims to contribute towards understanding the effects that could emerge from trends of informal Sand Forest wood harvesting, quantified through a spatial-temporal analysis. Quantifying the impact of a declining canopy closure resulting from selective wood harvesting required the use of remote sensing techniques and procedures that could potentially account for this effect.

In addition, the study envisaged predicting the future changes of Sand Forest that will take place as a result of continued informal wood harvesting. The ability of trajectory analysis to predict potential changes based on observed and quantified trends provides a new dynamic to conservation and management strategies. In understanding where and how much Sand Forest will be lost in the forthcoming years, more appropriate and accurate recommendations on conservation and management can be made. Furthermore, priority areas can be more readily identified for both conservation and for management intervention.

KEYWORDS: Sand Forest, Informal wood harvesting, temporal monitoring, Trajectory analysis

Challenges and opportunities of Earth observation for the prediction of water quality in inland waters

***Klaus Joehnk*[1], *Janet M. Anstee*[2]**

[1]*CSIRO Land and Water, Australia*; [2]*CSIRO Ocean and Atmosphere*

Corresponding author(s) e-mail: Klaus.joehnk@csiro.au

ABSTRACT:

Water quality in lakes and river systems has deteriorated worldwide due to intensification in land use and associated nutrient loading or changes in natural flow regimes. The most obvious impacts are increase in the frequency of harmful algal blooms caused by potentially toxic cyanobacteria and fish kills due to hypoxia. Other problems are not immediately visible or have indirect impacts like contamination by metals and pathogens, or vector borne diseases depending on wetting and increased temperature. To reduce health and economic risks posed by such water quality issues, there is an increased need for early warning systems. While Earth observation of inland aquatic systems can give an account of historic conditions and current state, integrating hydrodynamic and hydrologic modelling tools with predictive capabilities allow for timely intervention and optimised management options.

On a local scale Earth observation can be used to drive hydrodynamic simulations for short term prediction of harmful algal blooms in specific water bodies allowing for early warning and providing operating strategies for risk minimisation for, e.g., water treatment plants or reservoirs (case studies shown here). Combined with local hyperspectral sensors it is even possible to discriminate cyanobacteria species based on their pigments and thus infer potential toxicity. A generalisation of these methods on a regional or continental scale not only yields an early warning account for a larger region, e.g. state wide, but can yield a risk estimation based on weather forecast. In combination with hydrologic modelling tools EO is applied in ecological impact studies of flood inundation, e.g., the generation of hypoxic conditions in lowland rivers, or the spread of a carp virus for pest eradication in a large basin.

Although there is a large spectrum of water quality issues where EO can lead to better insight, spatial and temporal resolution of satellite sensors limits their application. Other techniques of remote sensing are necessary to fill these gaps.

KEYWORDS: harmful algal blooms, hypoxia, flood inundation, early warning, hydrodynamic modelling

Mapping of invasive plant species with Sentinel-1 and -2 data calibrated with UAV-based training data

***Teja Kattenborn*[1], *Javier Lopatin*[1], *Michael Förster*[2], *Fabian Fassnacht*[1]**

[1]*Institite of Geography and Geoecology (IFGG), Karlsruhe Indtitute of Technology (KIT), Germany*
[2]*Geoinformation in Environmental Planning Lab, Technische Universität Berlin, Germany*

Corresponding author(s) e-mail: teja.kattenborn@kit.edu

ABSTRACT:

Up-to-date maps on depicting the current status of plant invasions are highly valuable for efficient mitigation measures. In this context, remote sensing has been reported to be a useful tool for continuous mapping of invasive species over larger spatial extents. So far an important restriction for the development of flexible, operational approaches was the limited availability of cost-free and spatially highly-resolved datasets in many parts of the world. Since 2014, ESA´s Sentinel-1 and 2 satellites provide cost-free EO data with global coverage, relatively high temporal and spatial resolution. This data is ascribed a high potential for differentiating plant species. One important perquisite to operationally use such data for an operational mapping of invasive species is an efficient collection of reference data to train and validate mapping procedures applied to Sentinel data. We hypothesize that data collected from unmannead aerial vehicles (UAV) can be an efficient alternative of traditional field surveys. Applying UAV data instead of GNSS-coded (global navigation satellite system) field data comprises several advantages: (1) more area can be mapped in a given time frame, (2) increased area accessibility, (3) the UAV data share the bird's eye perspective of the satellite data and are hence directly compatible, and (4) the extraction of target species may be possible with automatized classification algorithms. We hence developed and tested a workflow for three invasive woody plant species in southern Chile which firstly includes an automatic extraction of spatially continuous species cover from UAV imagery by combining a sparse set of photo-interpreted presence points of the target species with a Maximum Entropy (MaxEnt) classifier. Secondly, we use this species cover data as reference to train a random forest model with multitemporal Sentinel observations to predict the canopy cover of the three target species over large areas. Our results show that the three invasive species considered can be mapped with very high accuracy as there is a very high agreement between UAV and Sentinel predictions ($R2 > 0.9$, $NRMSE < 10\%$). We thus conclude that the proposed methodology can be used as a blueprint for operational monitoring of invasive plant species.

KEYWORDS: Unmanned aerial vehicle, Invasive species, random forest, MaxEnt

Interactive tool for real-time delivery of remote sensing based vegetation maps and support of botanical data collection

Adam Kania[1]

[1]*definity, Poland*

Corresponding author(s) e-mail: kania@definity.pl

ABSTRACT:

The process of classification and creating vegetation maps based on botanical reference and remote sensing data typically takes many steps and some time.

Thanks to recent advances in speed and processing power of modern computing resources and through careful selection of processing algorithms, the whole workflow of creating vegetation mapping products can be performed in interactive times on modern workstations. Some algorithms and techniques used in a new, interactive module of Vegetation Classification Studio, that enable interactive work with vegetation maps will be presented.

Vegetation Classification Studio is a modern software tool implementing whole end-to-end workflow of creating vegetation maps – starting from botanical data pre-processing, through learning a machine learning model, validation and generation of quality assessment reports and metrics, prediction of the model on the whole study site and creating final raster map visualizations – in multiple formats. The new Remote Sensing Lab add-on is a browser-based interactive user interface developed with the aim of giving researcher interactive response times from applying new settings to presenting a set of reports and raster map visualizations of the whole study area.

Thanks to careful engineering of the rendering process, optimizations in the processing pipeline, with prediction applied at multiple resolutions and zoom levels, with dynamic rendering synchronized with current viewport, multiple levels of caching at different stages of the pipeline – the aim to facilitate realtime, interactive feedback to the researcher using just single modern workstation has been achieved.

This ability to deliver classification results fast has been integrated into an ongoing effort aimed at optimizing targeted and efficient collection of botanical reference data – especially in time- or budget-constrained regimes – to best use all available (and often limited) resources (remote sensing and pre-existing and recently-collected botanical reference) to obtain best possible vegetation maps.

To support that, both unsupervised (tSNE dimensionality reduction algorithm based) and supervised (classification) approaches are used to give an early insight into the whole variability of study area. The aim is to assist the experts in the field to better understand the terrain in the process of vegetation mapping, facilitate planning of reference data collection, and support fast orientation in unknown terrain in order to optimize reference data collection.

KEYWORDS: Classification, Vegetation maps, Fuzzy classification, Visualization, Interactive tool

Modelling tree species diversities of the Afromontane forest ecosystem with satellite remote sensing and macro-ecological data

Ralph Adewoye[1], [2]Christian Huttich, [3]Hazel Chapman.

[1]Remote Sensing/ Tropical Biodiversity Observation Group. Forestry Research Institute of Nigeria; [2]Remote Sensing Department, University of Würzburg; [3]Nigerian Montane Forest Project. School of Biological Sciences, University of Canterbury. Christchurch, New Zealand.

Corresponding authors. e-mail: adewoye.ralph@frin.gov.ng; christian.huettich@uni-wuerzburg.de

ABSTRACT:

Montane forests situated in the afro tropical region (Afromontane forests) are on the list of the world's most threatened ecosystem. These ecosystems are highly diverse and adjudged as repositories of genetic diversities. Information on the biodiversity of such an important area is a prerequisite for effective conservation and management strategy. Ecologists have relied on the traditional method of field survey to quantify biodiversity of large area, which often is time consuming, costly and dependent on expert knowledge. This has led to the conclusion that field measurements represent estimates rather than absolutes. Information on landscape biodiversity can be optimized through use of ecological proxies. This research examines the application of Spectral Variation Hypothesis (SVA) in an Afromontane forest ecosystem using features derived from high and medium resolution images combined with macro ecological data to predict tree species distribution. Alpha diversity (α) of tree species ≥ 10cm were calculated from in situ data obtained from survey of two study sites. The Object Based Image Analysis (OBIA) was adopted for the tree species distribution modelling. Spectral and textural metrics from QuickBird image were computed with the segmentation algorithm. While the macro ecological parameters (slope, Aspects, Elevation and solar energy radiation/ annum) were derived from 30m ASTER DEM. The relationships between diversity and spectral, textural features derived from the two images and the macro-ecological parameters were assessed with random forest algorithm. Elevation (r=0.75), slope (r=0.56) and aspects (r=0.34) were the determinant of tree species distribution in the study area. While spectral and textural features significantly contributed to the enhancement of the alpha diversity model in QuickBird image. QuickBird spectral and textural heterogeneity showed a significant correlation with species richness (r=0.87) and (r=0.53) respectively. The empirical models developed can be used to predict landscape-level species density in the Afromontane forests of Nigeria and the adjourning Cameron highlands.

KEYWORDS: Afromontane, Spectral Variation Hypothesis, Macro ecology, OBIA, Random Forest Algorithm.

Connectivity and Synchronisation of Lake Ecosystems in Space and Time - CONNECT

Stella A Berger[1], Sabine Wollrab[1], Jens C Nejstgaard[1], Hans-Peter Grossart[1], Gabriel Singer[1], Franz Hölker[1], Andreas Jechow[1], Jürgen Fischer[2], Thomas Ruhtz[2], Peter Gege[3], Torsten Sachs[4], Matthias Labrenz[5], Gunnar Lischeid[6], Rüdiger Röttgers[7], Thomas Schneider[8]

[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin/Stechlin, Germany; [2]Free University of Berlin (FUB), Germany; [3]German Aerospace Center (DLR), Oberpfaffenhofen, Germany; [4]German Research Centre for Geosciences (GFZ), Potsdam, Germany; [5]Leibniz-Institute for Baltic Research (IOW), Warnemünde, Germany; [6]Leibniz-Centre for Agricultural Research (ZALF), Müncheberg, Germany; [7]Helmholtz Centre Geesthacht (HZG), Germany; [8]Technical University Munich (TUM), Germany

Corresponding authors e-mail: Berger@igb-berlin.de, Wollrab@igb-berlin.de

ABSTRACT:

Within the project CONNECT we are establishing a collaborative network between experts in remote sensing (RS) and freshwater ecology to study connectivity and coherence of lake ecosystems in a regional context at unprecedented temporal and spatial resolution. The overall aim is to understand the yet unexplained variation in phytoplankton dynamics among river-connected German lowland lakes, many of which are presently classified as in poor to bad ecological status. These lakes often face a high risk of eutrophication, mass development of harmful algal blooms, and high production of greenhouse gases. We suggest if measured on adequate temporal and spatial scales much of the among-lake variation in phytoplankton dynamics to be explained by the strength of hydrological lake-to-lake and lake-to-catchment connectivity as modulated by lake depth and mixing regime. This may have profound implications for the maximum intensity, spatial range and regional-scale magnitude of eutrophication impacts. We will use (i) a large-scale experimental manipulation of lake connectivity, and (ii) an observational field campaign contrasting deep and shallow river-connected lakes, to challenge this research frontier by an innovative combination of automatic high- frequency *in situ* measurements with state of the art near-to-far RS technology. Climate change is expected to alter the hydrology, and thus the connectivity of lake-river systems. However, it is also predicted to increase extreme weather events leading to an increased input of nutrients as well as colored dissolved organic matter (cDOM). By providing data of high spatio-temporal coverage, CONNECT will provide basic high quality data to better understand mechanisms of eutrophication at the local and regional scale. Our data, thus, provide a valuable basis to improve current management of such river-connected lake ecosystems under future climate scenarios. To reach this ambitious goal, the project will (i) build a cross-disciplinary collaborative network of excellence, (ii) develop a mechanistic understanding of lake ecosystem functioning at local and regional scale, (iii) improve future environmental monitoring and interpretation of available data from inland waters, and (iv) support more effective integrated management of river-connected lakes to mitigate eutrophication impacts.

KEYWORDS: river-lake chains, mesocosms, remote sensing, phytoplankton dynamics, green house gases

Session S1.2

ANALYSIS OF ECOACOUSTIC RECORDINGS: DETECTION, SEGMENTATION AND CLASSIFICATION

Analysis of ecoacoustic recordings: detection, segmentation and classification

Jérôme Sueur[1], Dan Stowell[2]

[1]*Muséum national d'Histoire naturelle-CNRS, France*; [2]*Tuffs University, USA*; [3]*Queen Mary University of London, UK*

Corresponding author(s) e-mail: jerome.sueur@mnhn.fr; dan.stowell@qmul.ac.uk

ABSTRACT:

Ecoacoustics is a newly emerged discipline that aims at tackling ecological research questions through the lens of sound analysis [1,2]. Ecoacoustics covers several questions in marine, freshwater and terrestrial environments dealing with biodiversity monitoring, population ecology, community ecology and landscape ecology. One of the key approaches of ecoacoustics consists in identifying sounds of ecological importance in environmental recordings that were collected in an unattended way by automatic recorders. This search task is made difficult by the occurrence of background noise due to human activities, the co-occurrence of several sounds of interest, the degradation of the sounds of interest related to their propagation in the environment, a high-degree of variability of the sounds of interest, a large amount of data, and a lack of reference archives [3]. Solutions including computer processes are currently in development to try to get around these difficulties. This session will be the occasion to report and share new techniques involving signal analysis, machine learning, deep learning and high dimension statistics for advances in detection, segmentation, supervised and unsupervised classification of sound events.

KEYWORDS: sound, monitoring, signal analysis, automatic identification

REFERENCES:

1. Towsey, M. Parsons, S., Sueur, J. 2014. Ecology and acoustics at a large scale. *Ecological Informatics,* 21, 1-3.

2. Sueur, J., Farina, A. 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound Biosemiotics 26, 493-502.

3. Stowell, D. 2017. Computational bioacoustic scene analysis, in Virtanen, T.; Plumbley, M. D. & Ellis, D. (Eds.), Computational analysis of sound scenes and events. Springer, pp. 303-333.

AUREAS: a tool for recognition of anuran vocalizations

William E. Gómez[1], ***Claudia V. Isaza***[1], ***Sergio Gómez***[1], ***Juan M. Daza***[1], ***Carol Bedoya***[2]

[1]*Univesidad de Antioquia, Colombia*; [2]*University of Canterbury, New Zealand*

Corresponding author(s) e-mail: westeban.gomez@udea.edu.co

ABSTRACT:

Implementing techniques that facilitate and automatize the species monitoring is a task that can be carried out in several ways. One of them is bioacoustics analysis, which focuses on analyzing the soundscapes of a large number of recordings. One of the bioacoustics monitoring methods is species call recognition, which can be tackled commonly using four stages: cleaning, segmentation, feature extraction, and classification. Depending on the methodology several stages can be omitted and additional stages can also be considered, e.g., feature selection. In order to propose a monitoring system, the research groups SISTEMIC and Grupo Herpetologico de Antioquia (GHA) of the Universidad de Antioquia have implemented techniques focused on anuran calls recognition, specifically fuzzy clustering algorithms. In the segmentation stage, only the anuran vocalization segments were extracted from the spectrogram, ignoring the remaining frequencies. This procedure allows removing noise that can be found at other frequencies. Then, for the feature extraction stage, descriptors based on the MFCC (Mel Frequency Cepstral Coefficients) were computed, but with the difference that Mel scale was removed and the coefficients were computed only from the segmentation of the call. For the classification stage, an unsupervised algorithm called LAMDA (Learning Algorithm for Multivariate Data Analysis) was used to classify the segmented calls. LAMDA allows creating new classes that were not included in the learning process but were presented in the recognition step. The complete methodology was used to create the software AUREAS, which was tested using a database of 7 anuran species (1712 segments, including segments where there are no calls) from the northern Andes of Colombia. This software can identify these anuran calls with a F1-score of 0.88, allowing us to obtain activity patterns of the anuran species, which is a useful tool to monitor these species at the time. Now, the software is being modified to be able to recognize avian calls that have generally more inter-species variation. Therefore, more features based on other approaches (Wavelet-based, linear predictive codes, perceptual linear prediction) were included in order to abstract different attributes of the calls. To identify the most informative features, a feature selection stage was included, which used wrapper and filter strategies. In this conference, the methods implemented in the software and the different study cases to identify anuran and birds calls will be presented. We will illustrate the main challenges that are still required to solve issues in segmentation, feature extraction, and classification stages.

KEYWORDS: Animal calls, Classification, Monitoring, Colombia's reservoir

Content Description of Very-long-duration Recordings of the Environment

Michael Towsey[1]***, Aniek Roelofs***[2]***, Yvonne Phillips***[1]***, Anthony Truskinger***[1]***, Paul Roe***[1]

[1]*Queensland University of Technology, Australia*; [2]*University of Amsterdam, Netherlands*

Corresponding author e-mail: m.towsey@qut.edu.au

ABSTRACT:

Long-duration sound recordings are an established technique to monitor terrestrial ecosystems. Acoustic sensing has several advantages over personal field-surveys, but a disadvantage is that technological advances enable collection of much more audio than can be listened to. Machine learning methods can identify individual species, but these are time-consuming to build and if the species of interest is absent, nothing is revealed about recording content. Visual methods have also been developed to interrogate long-duration recordings but ultimately, interpretation of acoustic recordings must be ground-truthed by listening to the actual sound. However, the ear is constrained to listen in real-time. Even if one listens to 10 hours of one-minute segments, selected randomly from one year of recording, this represents only a 0.11% sample of the data.

For this study, we recorded 13 months of continuous audio in natural Australian woodland. We divided the audio into one-minute segments, which yields a content description at one-minute resolution. The feature set representing each segment consists of summary and/or spectral acoustic indices. Our objective in this investigation is two-fold: (1) to maximise content description of a very-long-duration recording while keeping listening to manageable levels; and (2) to determine how content description is influenced by the choice of acoustic features and other variables.

We begin by clustering the acoustic feature vectors using the k-means algorithm. Given sufficient clusters (k = 60), each cluster can be interpreted as a discrete acoustic state within the year-long soundscape. We describe four findings:

1. Listening to the medoid minute of each cluster (the minute whose feature vector is closest to the cluster centroid) yields a similar content description to that obtained by listening to a random sample of ten minutes from each cluster. This represents a ten-fold reduction in listening effort.

2. Although k-means is known to produce different clustering outcomes depending on cluster initialisation, we find that content description is little affected by different runs of k-means.

3. Different feature vectors yield a slightly different content description depending on which acoustic events have been ‘targeted’ by the selected features.

4. Training a Hidden Markov Model on the year-long cluster sequence helps to identify the underlying acoustic communities and can be used to obtain a more fine-grained labelling of sound-sources of interest.

KEYWORDS: Acoustic indices, Soundscape ecology, Eco-acoustics, Acoustic communities

What male humpback whale song chorusing can and cannot tell us about their ecology: strengths and limitations of passive acoustic monitoring of a vocally active baleen whale

***Anke Kügler*[1], *Marc O. Lammers*[2]**

[1]*Hawaii Institute of Marine Biology, University of Hawaii at Manoa, USA;* [2]*Hawaiian Islands Humpback Whale National Marine Sanctuary, USA*

Corresponding author(s) e-mail: akuegler@hawaii.edu

ABSTRACT:

Passive acoustic monitoring (PAM) with autonomous bottom-moored recorders is widely used to study cetacean occurrence, distribution, and behaviors, as it is not subject to factors that limit other traditional observation methods (e.g. vessel, land and aerial-based surveys) such as inclement weather, sighting conditions, the remoteness of study sites, etc. PAM projects range widely from presence/absence studies to addressing questions about noise exposure.

On the Hawaiian breeding ground, an estimated 12,000 humpback whales migrate annually from their high-latitude feeding grounds in Alaska to mate and give birth. While on the breeding grounds males produce an elaborate acoustic display known as song. Multiple males usually sing concurrently resulting in an ongoing chorus of song throughout the season. This song, primarily produced at peak frequencies under 2 kHz, can be captured using PAM, thus providing a powerful tool to monitor the humpback whale population and to study the actual song itself and its ecological importance. Calculating root-mean-square sound pressure levels (RMS SPLs in dB re 1 µPa) in one-octave bands using data from long-term PAM recordings made off Maui, we can compare the low frequency acoustic energy (0-1.56 kHz) produced by singing males across time and sites. Over the season, energy levels start increasing in December, peak in February and March, before dropping in late March through April. This mirrors the whales' migratory pattern and indicates that male chorusing can be used as a proxy for relative whale abundance. The uses for PAM can be varied, including monitoring the relative abundance at the same site over time and comparing occurrence among multiple recording locations to understand spatial patterns of habitat use.

However, on the breeding grounds and in contrast to males, females and mother-calf pairs have been shown to be acoustically cryptic and any PAM study on humpback whales will predominantly capture singing males. Furthermore, some observed spatial and temporal acoustic patterns can be ambiguous. The relationship between recordings from singing males and whale abundance is still poorly understood and previous attempts to correlate acoustic recordings with abundance have proved challenging. We propose that linking acoustics with other survey methods, particularly land and vessel-based observations, will greatly reduce the ambiguity typical of each individual method. Continued efforts to relate trends in acoustic levels to actual whale numbers and to understand the influence of recording strategies are planned.

KEYWORDS: Humpback Whale, Song, Passive Acoustic Monitoring, Ecology, Multi-method

Improving acoustic monitoring of biodiversity using deep learning-based source separation algorithms

***Mao-Ning Tuanmu*[1], *Tzu-Hao Lin*[2], *Joe Chun-Chia Huang*[1], *Yu Tsao*[3], *Chia-Yun Lee*[1]**

[1]*Biodiversity Research Center, Academia Sinica, Taiwan*; [2]*Department of Marine Biodiversity Research, Japan Agency for Marine-Earth Science and Technology, Japan*; [3] *Research Center for Information Technology Innovation, Academia Sinica, Taiwan*

Corresponding author(s) e-mail: mntuanmu@gate.sinica.edu.tw

ABSTRACT:

Passive acoustic monitoring of the environment has been suggested as an effective tool for investigating the dynamics of biodiversity across spatial and temporal scales. Recent development in automatic recorders has allowed environmental acoustic data to be collected in an unattended way for a long duration. However, one of the major challenges for acoustic monitoring is to identify sounds of target taxa in recordings which usually contain undesired signals from non-target sources. In addition, high variation in the characteristics of target sounds, co-occurrence of sounds from multiple target taxa, and a lack of reference data make it even more difficult to separate acoustic signals from different sources. To overcome this issue, we developed an unsupervised source separation algorithm based on a multi-layer (deep) non-negative matrix factorization (NMF). Using reference echolocation calls of 13 bat species, we evaluated the performance of the multi-layer NMF in separating species-specific calls. Results showed that the multi-layer NMF, especially when being pre-trained with reference calls, outperformed the conventional supervised single-layer NMF. We also evaluated the performance of the multi-layer NMF in identifying different types of bat calls in recordings collected in the field. We found comparable performance in call types identification between the multi-layer NMF and human observers. These results suggest that the proposed multi-layer NMF approach can be used to effectively separate acoustic signals of different taxa from long-duration field recordings in an unsupervised manner. The approach can thus improve the applicability of passive acoustic monitoring as a tool to investigate the responses of biodiversity to the changing environment.

KEYWORDS: Bat echolocation calls, Blind source separation, Multi-layer non-negative matrix factorization, Passive acoustic monitoring

Acoustic sensor networks and machine learning: scalable ecological data to advance evidence-based conservation

***Matthew W. McKown*, David J. Klein**

[1]*Conservation Metrics, USA*

Corresponding author(s) e-mail: matthew.mckown@conservationmetrics.com

ABSTRACT:

Evidence-based frameworks have helped to transform decisions in medicine, education, agriculture, and international development. Conservation has lagged other fields in embracing the data-driven revolution, largely because of the difficulty and expense of collecting ecological data over large areas and long time-scales. The decline of forest cover, species loss, and impacts of climate change make it imperative that we develop better tools to measure ecological change and conservation outcomes. Passive acoustic sensors that expand survey effort and machine learning techniques that automate data-processing, are one approach for collecting robust and cost-effective ecological metrics at the required scale. We present data from three case studies where we have developed and applied Deep Learning models to analyze passive acoustic survey data to 1) detect rare and elusive species, 2) estimate population trends from call rates, and 3) quantify wildlife impacts in the built environment and test potential mitigation measures. Our approach to acoustic analysis has contributed rigorous metrics to inform decisions for more than 100 conservation monitoring projects around the globe over the last six years. We discuss our efforts to develop transparent and efficient acoustic analysis workflows, our experience on the relative strengths and limitations of acoustic approaches, and the potential for combining soundscape indices with detection/classification approaches focused on target-species.

KEYWORDS: passive acoustic monitoring, Deep Learning, evidence-based conservation

Information retrieval from marine soundscape by using machine learning-based source separation

Tzu-Hao Lin [1], ***Tomonari Akamatsu*** [2], ***Yu Tsao*** [3], ***Katsunori Fujikura*** [1]

[1] *Department of Marine Biodiversity Research, Japan Agency for Marine-Earth Science and Technology, Japan*; [2] *National Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency, Japan*; [3] *Research Center for Information Technology Innovation, Academia Sinica, Taiwan*

Corresponding author(s) e-mail: schonkopf@gmail.com

ABSTRACT:

In remote sensing of the marine ecosystem, visual information retrieval is limited by the low visibility in the ocean environment. Marine soundscape has been considered as an acoustic sensing platform of the marine ecosystem in recent years. By listening to environmental sounds, biological sounds, and human-made noises, it is possible to acoustically identify various geophysical events, soniferous marine animals, and anthropogenic activities. However, the sound detection and classification remain a challenging task due to the lack of underwater audio recognition database and the simultaneous interference of multiple sound sources. To facilitate the analysis of marine soundscape, we have employed information retrieval techniques based on non-negative matrix factorization (NMF) to separate different sound sources with unique spectral-temporal patterns in an unsupervised approach. NMF is a self-learning algorithm which decomposes an input matrix into a spectral feature matrix and a temporal encoding matrix. Therefore, we can stack two or more layers of NMF to learn the spectral-temporal modulation of *k* sound sources without any learning database [1]. In this presentation, we will demonstrate the application of NMF in the separation of simultaneous sound sources appeared on a long-term spectrogram. In shallow water soundscape, the relative change of fish chorus can be effectively quantified even in periods with strong mooring noise [2]. In deep-sea soundscape, cetacean vocalizations, an unknown biological chorus, environmental sounds, and systematic noises can be efficiently separated [3]. In addition, we can use the features learned in procedures of blind source separation as the prior information for supervised source separation. The self-adaptation mechanism during iterative learning can help search the similar sound source from other acoustic dataset contains unknown noise types. Our results suggest that the NMF-based source separation can facilitate the analysis of the soundscape variability and the establishment of audio recognition database. Therefore, it will be feasible to investigate the acoustic interactions among geophysical events, soniferous marine animals, and anthropogenic activities from long-duration underwater recordings.

KEYWORDS: Source separation, soundscape information retrieval, ecosystem sensing, marine soundscape

REFERENCES:

1. Lin, T.-H., Fang, S.-H., Tsao. Y. 2017. Improving biodiversity assessment via unsupervised separation of biological sounds from long-duration recordings. Sci Rep, 7: 4547.
2. Lin, T.-H., Tsao. Y., Akamatsu, T. 2018. Comparison of passive acoustic soniferous fish monitoring with supervised and unsupervised approaches. J. Acoust. Soc. Am. Express Letters, 143: EL278.
3. Lin, T.-H., Tsao. Y. 2018. Listening to the deep: Exploring marine soundscape variability by information retrieval techniques. OCEANS'18 MTS/IEEE Kobe / Techno-Ocean 2018, in press.

A Novel Set of Acoustic Features for the Categorization of Stridulatory Sounds in Beetles

***Carol Bedoya*[1], *Eckehard Brockerhoff*[1, 2], *Michael Hayes*[3], *Richard Hofstetter*[4], *Daniel Miller*[5], *Ximena Nelson*[1]**

[1]*School of Biological Sciences, University of Canterbury, NZ*; [2]*SCION (New Zealand Forest Research Institute), NZ*; [3]*Department of Electrical and Computer Engineering, University of Canterbury, NZ*; [4]*School of Forestry, Northern Arizona University, USA*; [5]*USDA, Forest Service, USA*

Corresponding author e-mail: carol.bedoya@pg.canterbury.ac.nz

ABSTRACT:

Scolytinae and Platypodinae (Coleoptera, Curculionidae), also known as bark and ambrosia beetles, are two closely-related subfamilies of weevils with sound production capabilities. Both subfamilies contain ca. 7400 species (~ 250 genera) and are infamous due to the symbiotic relationship of some species with tree-killing fungi. These beetles spend most of their life history inside plant tissues, which has yielded a complex variety of mating systems associated with specific acoustic communicatory interactions and stridulatory mechanisms. Despite being some of the few taxa with properly-developed acoustic communication in a medium other than air or water, bark and ambrosia beetle acoustics remains one of the most understudied areas in ecoacoustics. We used sounds produced by bark and ambrosia beetles to develop a new set of acoustic parameters for the identification and categorization of stridulations. Our goal was to generate a group of attributes able to extract information from the non-syntactic acoustic communication commonly found in insects. To achieve this, we implemented a series of non-linear transformations in the time-frequency representation of each stridulatory call, thus mapping the intrinsic spectro-temporal information to a 2-dimensional feature space where inter-specific call variability was more discernible. The proposed approach allows the generation of taxonomically-based hierarchies, and the accurate discrimination of species with similar stridulatory sounds. Some applications that can be derived from this study include non-invasive monitoring methods for ecological studies inside tree logs, acoustic sexing and identification protocols for species with stridulatory-based communication, and techniques for automatic detection of insects in border biosecurity scenarios.

KEYWORDS: Bark Beetles, Ambrosia Beetles, Acoustic Detection, Animal Identification, Stridulation

Noise robust 2D bird localization via sound using microphone arrays

Daniel Gabriel[1], *Ryosuke Kojima[2], Kotaro Hoshiba[3], Katsutoshi Itoyama[1], Kenji Nishida[1], Kazuhiro Nakadai[1, 4]*

[1]*Tokyo Institute of Technology, Japan*; [2]*Kyoto University, Japan*; [3]*Kanagawa University, Japan;*
[4]*Honda Research Institute Japan Co., Ltd.*

Corresponding author e-mail: gabriel@ra.sc.e.titech.ac.jp

ABSTRACT:

Birds in the wild are difficult to localize, because their sizes tend to be small, they move swiftly, and they are often visually occluded. However, their location information is crucial for ethological studies on birds' behaviour. Recently, automating the process has been studied as a hot topic, where spatial sensors and sensor networks are commonly used. To avoid the visual occlusion problem, many studies focus on acoustic signal processing by applying microphone arrays and perform 1D azimuth localization through bird songs. In this study, we perform 2D sound source localization in the Cartesian coordinates using azimuths from multiple microphone arrays. To estimate the exact bird's location, we calculate the intersection points of these azimuth lines. Although this approach is simple and easy to be implemented, it has two main issues. One is that even small noise interference in azimuth values results in corrupting the localization data. This leads to a problem, where the intersection points between the azimuth lines do not intersect in one point for a single bird, but in several points. This proves difficulty in estimating the exact location of each bird. Especially in a far-field application, even small noise corruption leads to large localization errors. The other issue is that in the bird's natural habitat, elements such as leaves, grass and rivers are natural noise sources. It is difficult to extract the bird songs in such a noisy environment. We propose an algorithm involving statistic methods, sound feature analysis and machine learning. Based on this approach, a noise robust bird localization system has been established. We have performed numerous simulations to further understand the limitations of the system. Based on the results we have also derived the system's design guidelines, describing how the results change depending on the number of microphone arrays, signal-to-noise ratio, bird's distance from the devices, array's transfer function, type of the singing bird and specific parameter settings used in the algorithms. Such detailed guidelines support interested researchers in creating a similar system, which can contribute to ethological researches.

KEYWORDS: Auditory scene analysis, localization, robot audition

Fine-scale observations of spatiotemporal dynamics and vocalization type of birdsongs using microphone arrays and unsupervised feature mapping

***Reiji Suzuki*[1], *Shinji Sumitani*[1], *Naoaki Chiba*[1], *Shiho Matsubayashi*[2], *Takaya Arita*[1], *Kazuhiro Nakadai*[3, 4], *Hiroshi G. Okuno*[5]**

[1]*Nagoya University, Japan*; [2]*Osaka University, Japan*; [3]*Tokyo Institute of Technology, Japan*;
[4]*Honda Research Institute Japan Co., Ltd., Japan*; [5]*Waseda University, Japan*

Corresponding author(s) e-mail: reiji@nagoya-u.jp

ABSTRACT:

In ecoacoustics, natural sounds can be both the tool (for the indirect measurement of biodiversity or habitat quality) and the subject (to understand the properties of sound and its function in the environment) of ecological research [1]. We believe that deploying a robotic monitoring system into a field is a promising approach to contribute to both cases. We are developing a portable system, HARKBird, to localize birdsongs, which automatically extracts sound sources and its direction of arrival (DOA) [2]. HARKBird consists of a laptop PC with an open source software for robot audition HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [3] combined with low-cost microphone arrays. We have shown the existence of temporal overlap avoidance in the singing behaviors of some forest species and successful spatial localization of song posts of the great reed warblers by using multiple microphone arrays [1, 4]. However, it took much cost to classify many localized sounds to quantitatively analyze their behavioral patterns in detail. We introduce two example playback experiments and their analysis using HARKBird and unsupervised feature mapping for classification of localized sound sources. One is to observe how playback patterns can affect vocalization and spatial movement of an individual of Japanese bush-warbler (Horornis diphone), by estimating the DOA of his songs. We used a deep autoencoder to classify separated sound sources in an interactive way after recordings, and could roughly discriminate songs of the target individual from other sound sources. We quantitatively observed that he tended to sing a specific type of songs less frequently and move actively during which conspecific songs were replayed from a loudspeaker. The other playback experiment attempted to perform a real-time 2D localization and offline classification of songs of Spotted Towhee (Pipilo maculatus) using the dimension reduction algorithm t-SNE and a clustering algorithm DBSCAN. We successfully estimated the position of a loudspeaker replaying conspecific songs and classify three song and call types of the individual semi-automatically with high accuracy. We also found significant differences in the spatial distributions of these vocalization types. Results imply that these fine-scale data, in time and space, can be utilized for not only analyzing behavioral patterns of individuals, but also estimating the spatial structure and the acoustic quality of the habitat space. This work was supported in part by JSPS/MEXT KAKENHI: JP16K00294, JP17H06841, JP18K11467, and JP17H06383 in #4903.

KEYWORDS: sound source localization, birdsongs, HARK, deep autoencoder, t-SNE

REFERENCES:

1. Farina, A., Gage, S.H., 2017. Ecoacoustics: A new science, in: Farina, A. and Gage, S. H. (Eds.), Ecoacoustics: The Ecological Role of Sounds, Wiley, pp. 1-11.

2. Suzuki, R., Matsubayashi, S., Nakadai, K., Okuno, H.G., 2017. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. Journal of Robotics and Mechatronics. 27(1), 213-223.

3. Nakadai, K., Okuno, H.G., Mizumoto, T., 2017. Development, deployment and applications of robot audition open source software HARK. Journal of Robotics and Mechatronics, 27(1), 16-25.

4. Suzuki, R., Matsubayashi, S., Saito, F., Murate, T., Masuda, T., Yamamoto, K., Kojima, R., Nakadai, K. and Okuno, H. G., 2018. A spatiotemporal analysis of acoustic interactions between great reed warblers (*Acrocephalus arundinaceus*) using microphone arrays and robot audition software HARK. Ecology and Evolution, 8(1), 812-825.

Articulating citizen science, automatic classification and free web services for long-term acoustic monitoring: examples from bat monitoring schemes in France and UK

***Bas Yves**[1,2]**, Newson Stuart**[3]**, Barré Kévin**[1]**, Julien Jean-François**[1]** and Kerbiriou Christian**[1]*

[1]*CESCO-MNHN, France*; [2]*CEFE-CNRS, France*; [3]*BTO, UK*

Corresponding author(s) e-mail: yves.bas@mnhn.fr

ABSTRACT:

Monitoring biodiversity over large spatial and temporal scales is crucial for assessing the impact of global changes and environmental mitigation measures. Bats often have high conservation prioritisation owing to their trophic position, habitat associations and threat level, and many have dedicated management plans. However, poor knowledge of species' ecology, identification issues and surveying challenges mean that large-scale monitoring to produce required distribution and abundance information is less developed than for some other taxa. Exciting possibilities applicable to professional and citizen science are offered by new recording techniques and methods of semi-automated species recognition based on sound detection. Static detectors deployed to record bats throughout whole nights have been recommended for standardised acoustic monitoring but until recently cost and lack of software to support the analyses of such data has prohibited wide uptake. Such monitoring schemes have recently been deployed in both Britain and France allowing the fast and standardized collection of millions of bat records together with very interesting data on non-targeted taxa such as bush-crickets. Such data management led us to develop generic and open tools: (1) the Tadarida software toolbox providing a generic detection and classification of sound events, and (2) an open dedicated web portal (www.vigiechiro.herokuapp.com) to allow participants to manage and upload their data, then being processed trough Tadarida to get a quick feedback on the content of the data.

We demonstrate how such data can accurately describe pronounced ecological patterns for numerous species at different scales: spatial variation in activity as a proxy for relative abundance, habitat selection and phenology of seasonal and nocturnal activity. If maintained in the long term, such schemes will also greatly improve estimates of species temporal trends and hence the assessment of conservation priorities.

The feedback produced by these two monitoring schemes allows us the opportunity to provide recommendations for the sustainability of long-term acoustic monitoring of bats. These include a database that is adaptively managed to allow all raw data to be re-analysed every time automatic identification makes significant progress, while keeping the link with expert validation to ensure consistency in the semi-automated process. More importantly, there are real benefits of developing long-term acoustic monitoring within a collaborative framework. Specifically, (1) for collaboration among bat scientists for the collection of reference sound data, because diversity and quantity of the reference library remains a limiting factor for automatic identification, and (2) for work on bats to consider the wider acoustic monitoring of other species groups by working with other zoologists to share resources and costs.

KEYWORDS: Acoustic monitoring, citizen science, multi-taxa, population trends

Session S1.3

COMPUTER VISION IN ENVIRONMENTAL SCIENCES

Computer vision in environmental sciences

Shinji Fukuda[1], Jeffrey Tuhtan[2]

[1]Tokyo University of Agriculture and Technology, Japan; [2]Tallinn University of Technology, Estonia

Corresponding author(s) e-mail: shinji-f@cc.tuat.ac.jp; jeffrey.tuhtan@ttu.ee

ABSTRACT:

Image-based methods are at the forefront of artificial intelligence applications. This special session provides a forum for researchers and professionals using image-based methods to study species, population, biodiversity and the abiotic environment. The topics of this special session include:

-UAV imagery

-GIS/orthoimagery

-Video tracking/motion estimation

-Object recognition and classification

-High speed imaging

-Multispectral remote sensing

KEYWORDS: image analysis, machine learning, artificial intelligence, classification/regression

NAIRA a tool to automatic mammals genera identification in Camera Trapping Pictures

Claudia Isaza[1], Luis Pulido[1], Angelica Diaz-Pulido[2, 3]

[1]SISTEMIC, Engineering Faculty, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia; [2]Instituto de Investigación de Recursos Biológicos, Alexander von Humboldt, Calle 28A # 15-09, Bogotá, Colombia; [3]Grupo de Ecología y Evolución de vertebrados, Instituto de Biología, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

Corresponding author(s) e-mail: victoria.isaza@udea.edu.co;fernando.pulido@udea.edu.co;adiaz@humboldt.org.co

ABSTRACT:

Camera traps are an efficient tool for detecting terrestrial mammals and birds. Vast amounts of presence data and information about distribution and size of mammals and birds have been generated in a very short time using this approach. The first step into a monitoring study, using camera trapping is to analyse the photographs. However, processing times, when it is done manually, can take longer because of the large amount of data collected. Pre-selecting relevant pictures using an automatic system and automatically identify the animals is an alternative to reduce the analysis time.

The challenges to automatic identify the mammal´s genus from camera trap photos are: few examples of some genera (unbalanced classes problem), variation in light levels, constant changes in the scene, animal partially occluded, blurred photographs and other variations resulting from the natural dynamics of the ecosystem. Until now there has not been a computational tool to help in the specific task of labelling the animal´s genus. Thus, it is necessary to design new tools that automate the processing of these photographs.

Here we introduce a new version of the software NAIRA. This software uses Machine Learning algorithms as an alternative to automatically labelling mammal genera. The photographs are classified into pictures with animals and without animals using a fuzzy classifier, after that the images are segmented to extract the area with animal and a second classification distinguishes between birds and mammals using an Artificial Neural Network (ANN). Finally NAIRA identifies the genera in the detected mammals using Support Vector Machine (SVM) and Bag of Words (BoW). This version of software includes a proposal to identify the level of incertitude in the machine decision.

The results over a database (Andean, Caribbean and Pacific regions in Colombia) with 70780 photos and strong unbalance between classes are promising; it was possible to automatically identify photos of animals and to differentiate among birds and 20 mammal genera (average accuracy 95%). The incertitude analyse is usefully to prevent a misclassification when the algorithm does not have a high level of certitude. In this case, only the 4,45% of pictures (with high classification incertitude level) has to be analysed by an human expert.

The attendees will know in detail the functions of the software to recognize the advantages of the processing of this type of images with NAIRA.

KEYWORDS: Camera traps, Image Processing, Support Vector Machine, Artificial neural networks, Taxonomic Rank

Assessment of permanent grasslands in Latvia using spectral remote sensing techniques

Dainis Jakovels, Agris Brauns, Jevgenijs Filipovs, Juris Taskovs, Ruta Abaja

Institute for Environmental Solutions, "Lidlauks", Priekuli Parish/County, LV-4101, Latvia

Corresponding e-mail: dainis.jakovels@videsinstituts.lv

ABSTRACT:

Permanent grasslands (meadows and pastures) are the most common agricultural land use type covering 34% (0.65 million hectares) of agricultural land in Latvia. The Common Agriculture policy (CAP) stipulates that the EU Member States have to designate permanent grasslands, ensure that farmers do not convert or plough them and that the ratio of permanent grasslands to the total agricultural area does not decrease by more than 5% in order to receive support payments. However, semi-natural grassland habitats require appropriate management activities to ensure their long-term conservation. The European Commission report (2015) required by the Birds and Habitats directives concludes that 'grasslands and wetlands have the highest proportion of habitats with an unfavourable-bad and deteriorating status' in the EU, while the midterm review of EU biodiversity strategy 2010-2020 highlighted that grassland habitat change presents a high risk to biodiversity. Latvia's rural development programme (2014-2020) has identified only 47 thousand hectares of biologically valuable grasslands. These grasslands are semi-natural meadows and pastures that include species and habitat types of EU importance. 70-90% of EU importance grassland habitats in Natura 2000 sites were in poor condition in Latvia during 2012. There is a clear interest from a number of end-users (e.g. the Nature Conservation Agency, the Rural Support Service,) for grassland mapping and management practice monitoring solutions.

In order to prevent loss of high nature value grasslands and increase sustainability of semi-natural grassland management, the Integrated Planning tool was developed in frames of LIFE+ project "Integrated planning tool to ensure viability of grasslands" (LIFE Viva Grass ENV/LT/00018). Spectral remote sensing technique was used for preparation of necessary inputs for the tool from Cesis Municipality in Latvia - mapping of grasslands, detection of overgrowth with shrubs/trees and spread of invasive species (Sosnowsky's hogweed) as well as assessment of grass biomass. High spatial and spectral resolution of hyperspectral airborne data obtained with flying laboratory ARSENAL was complemented with temporal dimension of Sentinel-2 satellite data in order to achieve the best result reaching classification accuracy >90%.

KEYWORDS: spectral image analysis, grasslands, Sosnowsky's hogweed, hyperspectral airborne data, multispectral Sentinel-2 data

Exploiting Taxonomic Relations in Image-based Plant Species Classification

***Marco Seeland*[1], *David Boho*[1], *Patrick Mäder*[1]**

[1]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: marco.seeland@tu-ilmenau.de

ABSTRACT:

Modern plant taxonomy describes phylogenetic relationships of taxa based on their morphological and genetic similarities. Hence, taxonomical relations reflect very different levels of visual resemblance: from close overall resemblance over similarity of only specific morphological characters to similarities mainly expressed on the molecular level, hence almost no visual resemblance. It is an open research question to which extent phylogenetic relations are reflected by visual similarities. Whereas previous studies on automated plant identification from images focused solely on the species level, we investigated classification at higher taxonomic levels, i.e., at genus and family level. We show that visually observable characters are indeed present at higher taxonomic levels and that they can be learned from natural images by state-of-the-art machine learning techniques. This allows for high classification accuracy which increases with taxonomic level and even facilitates the taxonomic identification of non-trained species. Targeting further improvement in classification accuracy, we investigated different strategies of exploiting taxonomic relations during classifier training and inference.

KEYWORDS: Plant Identification, Machine Learning, Image Classification, Deep Learning, Convolutional Neural Networks

Computer vision applications using multispectral UAS imagery: comparing pixel and object-based methods for automatic classification of river landscapes

Jeffrey A. Tuhtan*[1], Philipp Thumser[2], Christian Haas[2]

[1]*Tallinn University of Technology, Estonia*; [2]*I AM HYDRO, Germany*

Corresponding author e-mail: jetuht@ttu.ee

ABSTRACT:

The use of unmanned aerial system (UAS) imagery in environmental sciences has rapidly increased due to the ease of use and low cost [1]. In general, the ecological informatics community widely uses spatial data analyses, especially via GIS. A growing number of applications now include UAS imagery, which can provide centimetre scale multispectral data [2]. Processing this information requires the application of image-based remote sensing techniques. Specifically, the differences between unsupervised and supervised image classification methods are discussed, with a focus on pixel and object-based computational methods [3]. Examples relevant to ecological studies are presented using multispectral imagery collected of river landscapes to illustrate how UAS data can be used to classify complex spatial features such as vegetation and submerged regions of different depths, including turbulent flows and complex lighting and shade conditions.

KEYWORDS: UAS, Multispectral imagery, Classification, River, Remote sensing

REFERENCES:

1. Arif, M.S.M., Gülch, E., Tuhtan, J.A., Thumser, P., Haas, C., 2016. An investigation of image processing techniques for substrate classification based on dominant grain size using RGB images from UAV. Int. J. Remote Sens. 0, 1–23. https://doi.org/10.1080/01431161.2016.1249309

2. Hugenholtz, C.H., Whitehead, K., Brown, O.W., Barchyn, T.E., Moorman, B.J., LeClair, A., Riddell, K., Hamilton, T., 2013. Geomorphological mapping with a small unmanned aircraft system (sUAS): Feature detection and accuracy assessment of a photogrammetrically-derived digital terrain model. Geomorphology 194, 16–24. https://doi.org/10.1016/j.geomorph.2013.03.023

3. Black, M., Carbonneau, P., Church, M., Warburton, J., 2014. Mapping sub-pixel fluvial grain sizes with hyperspatial imagery. Sedimentology 61, 691–711. https://doi.org/10.1111/sed.12072

Tracking swimming Lefua echigonia to assess the impact of crayfish introduction

***Shinji Fukuda*[1], *Jeffrey Tuhtan*[2]**

[1]*Tokyo University of Agriculture and Technology, Japan*; [2]*Tallinn University of Technology, Estonia*

Corresponding author(s) e-mail: shinji-f@cc.tuat.ac.jp

ABSTRACT:

Lefua echigonia is an endangered freshwater species in Japan, whose habitats are decreasing due to urbanization and modernization of agriculture. In addition, invasion by non-native crayfish, *Procambarus clarkii*, is known to be a threat to *L. echigonia*. Whereas we observed *P. clarkii* at *L. echigonia* habitats such as large gravels, little is known about the impact of *P. clarkii* invasions on *L. echigonia*. In this study, we conducted a laboratory experiment using a 25-cm cylinder to observe swimming behaviour of *L. echigonia* with varying number of *P. clarkii* individuals and the presence/absence of a rectangular obstacle at the center. Swimming trajectories were obtained from a series of images taken 30 frames per second, based on which swimming speed, acceleration, and distance between individuals were calculated for the assessment of behavioural changes with *P. clarkii* introduction. No feed was given for both species during experiment. While both species moved along the wall during experiment under no obstacle condition, *L. echigonia* was more active than *P. clarkia*. Behavioural changes were observed when the obstacle was placed in the center. Foraging by *P. clarkii* was expected but no such a behaviour was observed even under the condition where more *P. clarkii* individuals were put together. Whereas no foraging behaviour was observed, some *L. echigonia* individuals stayed around the brick which is similar to *P. clarkia*. This suggests potential habitat overlaps between the two species. Further study can reveal species competition under more complex, dynamic instream conditions for which image analysis with high-speed/high-resolution cameras can be used as a tool for reliable assessment.

KEYWORDS: Tracking, Swimming behaviour, Species competition, Image analysis

Trends in machine learning for plant species identification

Jana Wäldchen[1]*, Michael Rzanny*[1]*, Marco Seeland*[2]*, Patrick Mäder*[2]

[1]*Max Planck Institute for Biogeochemistry, Germany*; [2]*Technische Universität Ilmenau, Germany*

Corresponding author(s) e-mail: jwald@bgc-jena.mpg.de

ABSTRACT:

Current rates of species loss triggered numerous attempts to protect and conserve biodiversity. Species conservation, however, requires species identification skills, a competence obtained by intensive training and experience. Field researchers, land managers, educators, civil servants, and the interested public would greatly benefit from accessible up-to-date tools automating the process of species identification. Also professional taxonomists are asking today for more efficient methods to meet identification requirements. The availability of smart portable devices, digital cameras as well as the mass digitization of natural history collections led to an explosion of available image data of living organism stored in online databases. This rapid increase in biological image data in combination with modern machine learning methods, such as deep learning, is a key development that offers new opportunities towards automated species identification. The idea of automated species identification is approaching reality. We review the technical status quo on computer vision approaches for plant species identification, introduce publicly available applications and benchmark datasets for image based plant species identification, highlight the main research challenges to overcome in providing applicable tools and conclude with a discussion of open and future research thrusts.

KEYWORDS: automated plant species identification, machine learning, deep learning, computer vision

Deep Learning for Cracking the Leaf Code

Dimitri Belousow[1]***, Georg Graser***[1]***, Marco Seeland***[1]***, Patrick Mäder***[1]

[1]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: dimitri.belousow@tu-ilmenau.de, marco.seeland@tu-ilmenau.de

ABSTRACT:

The complexity and variation of angiosperm leaf shape and venation impeded botanists and taxonomists to identify more decisive traits allowing to increase the relevance of the leaf architecture for taxonomic classification. It was demonstrated in a recent publication, that a computer vision algorithm based on gradient histograms and trained on chemically cleared leaves successfully learns and visualizes leaf features allowing to classify species into botanical groups above species level. Making use of state-of-the-art machine learning and computer vision methods, we show that deep learning approaches further increase the classification accuracy. Moreover, we evaluate gradient-weighted guided backpropagation for visualizing the neural activations relevant for classification in order to highlight the location of novel and decisive leaf characters.

KEYWORDS: leaf architecture, leaf venation and shape, computer vision, deep learning, guided backpropagation, higher level classification

Session S1.4

INTEGRATING DATA FOR ANALYSIS – HOW FAR ARE WE?

Integrating data for analysis – how far are we?

Corinna Gries[1]***, Margaret O'Brien***[2]***, Kristin Vanderbilt***[3]***, Colin Smith***[1]

[1]*University of Wisconsin, Madison, USA*; [2]*University of California, Santa Barbara, USA,* [3]*Florida International University, Miami, USA*

Corresponding author(s) e-mail: cgries@wisc.edu

ABSTRACT:

This session will discuss advances and tools for the second step of the data life cycle in ecological research. The first step being either field data collection and/or data discovery, the second step has been called the 'janitorial' step, but is mostly known as data cleaning, data harmonization, or data integration. It poses major data management challenges and is frequently a time consuming process with estimates of up to 80% of the data analysis. The reasons for this large proportion of effort range from those that cannot be addressed with technological solutions and are rooted in sampling methods to those that are related to data organization and semantics that may be addressed with developing technologies. This session will explore the progress that is being made toward reducing the effort needed for pre-analysis data harmonization. Encouraged are: (1) reports on data integration projects spanning the range of employing and advancing semantics, ontology, linked data, specific tools, workflow systems, and standards developments, (2) considerations of an approach's promise for a high return on the investment and/or whether it will it significantly improve documentation of data manipulations, (3) experiences and discussions focusing on comparing effectiveness in reducing time spent in data integration, (4) technological gaps and shortcomings.

KEYWORDS: Data science, Data reuse, Data integration, semantics, Linked data

Whip: Human and machine-readable specifications for data

***Stijn Van Hoey*[1]*, Peter Desmet*[1]**

[1]*Research Institute for Nature and Forest (INBO), Belgium*

Corresponding author(s) e-mail: stijn.vanhoey@inbo.be

ABSTRACT:

Different tools and technologies are available to clean and harmonize data. Independent of the tool used, the ability to assess the quality of a data set and identify potential errors is crucial for harmonization efforts. The necessity becomes even more apparent in the context of data publication, (re)use and aggregation.

Documentation and guidelines about the data requirements provide guidance in this process and enable to communicate what to expect from the data, but are mostly intended for humans only. To facilitate the harmonization process, we propose the usage of a specification file, describing the constraints to which the data should comply. Its syntax is human- and machine-readable, so it can be used to communicate expected data quality/conformity and to validate data automatically. The scope of the set of specifications can be specific to a dataset, researcher or research community, which allows bottom-up and top-down adoption. As an example, we apply the specifications to verify data mapped to the biodiversity information standard Darwin Core.

In this talk, we will present "whip", a proposed syntax and format to express data specifications. Whip allows to define column-based constraints for tabular (tidy) data with a number of rules. We will also demonstrate a software application (called "pywhip") to validate data sets using these specifications. We hope it will trigger a discussion on how to express data specifications and communicate data quality expectations.

KEYWORDS: Data Harmonization, Data Quality, Documentation, Specifications

sPlot – the global vegetation-plot database

Francesco Maria Sabatini[1,2], **Milan Chytrý**[3], **Jürgen Dengler**[1,4,5], **Florian Jansen**[6], **Borja Jiménez-Alfaro**[7], **Valério D. Pillar**[8], **Helge Bruelheide**[1,2] **& the sPlot Consortium**[1]

[1]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany; [2]Geobotany and Botanical Garden, Institute of Biology, Martin Luther University, Halle, Germany; [3]Department of Botany and Zoology, Masaryk University, Brno, Czech Republic; [4]Vegetation Ecology Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland; [6]University of Bayreuth, Bayreuth, Germany; [5]Faculty of Agriculture and Environment, University of Rostock, Germany; [7]Research Unit of Biodiversity (UO, CSIC, PA). Oviedo University, Campus de Mieres Edificio de Investigación – 5ª planta C/ Gonzalo Gutiérrez Quirós s/n. 33600 Mieres, Spain; [8]Department of Ecology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Corresponding author(s) e-mail: Francesco.sabatini@botanik.uni-halle.de

ABSTRACT:

Vegetation-plot databases contain biodiversity data on presence and relative abundance of plants co-occurring in the same community. Compared to databases based on occurrence records of individual species aggregated at the level of grid cells, vegetation-plot databases have the advantage of providing information on species relative cover, co-occurrences, and to provide more reliable information on true absences. Although large collections of plant community data are now available at national to regional level, they are rarely accessible at continental or global extents, as their compilation is technically and conceptually challenging, due to different data formats and taxonomical nomenclatures used.

Here we present the sPlot database, which merges and standardizes data contributed by more than 100 regional, national and continental databases, and contains records from 1,121,244 vegetation plots, for a total of 23,586,216 plant species entries with relative cover or abundance. All plots are georeferenced, although with varying precision, their size varies from less than 1 m^2 to 25 ha, and span from year 1885 to 2015. The vegetation-plot data are stored in a SQLite database, managed with TURBOVEG v3 software, and further processed in R for data integration and analysis.

In order to make sPlot suitable for the exploration of global patterns in taxonomic, functional and phylogenetic diversity at the plant community level we performed three steps. 1) We standardized the species lists of the different databases in sPlot through the construction of a taxonomic backbone using existing databases on accepted plant species names. 2) We calculated functional attributes of each plot (community-weighted means and variances of traits) using gap-filled data from the global plant trait database TRY. 3) We generated a phylogeny for 50,167 out of the 54,519 vascular plant species occurring in sPlot. Finally, in addition to the information provided by the data owners, we retrieved for each plot information on environmental conditions (i.e. climate, soil) and the biogeographic context (i.e. biomes) from external sources.

sPlot provides a unique, integrated global repository of data that would otherwise be fragmented in unconnected and structurally inconsistent databases at national or regional level. We believe that sPlot can be the basis for a new generation of studies, not only to address fundamental ecological questions related to

plant diversity patterns or community assembly rules, but also to foster further development and testing of macroecological theories and as an information baseline for refining interdisciplinary conservation studies in a human-dominated, changing world.

Further information: https://www.idiv.de/en/sdiv/working_groups/wg_pool/splot.html

KEYWORDS: Database, global, plant diversity, Vegetation

Integration and dissemination of aquatic biodiversity and ecosystem services data for case studies focusing on ecosystem-based management

***Aaike De Wever*[1], *Astrid Schmidt-Kloiber*[2], *Juan Arevalo*[3], *Declan Dunne*[4], *Ana Luisa Barbosa*[3], *Koen Martens*[1], *Alejandro Iglesias Campos*[3]**

[1]*Royal Belgian Institute of Natural Sciences, Belgium*; [2]*University of Natural Resources and Life Sciences, Vienna, Austria*; [3]*Intergovernmental Oceanographic Commission of United Nations Educational, Scientific and Cultural Organization, France; [4]University College Cork, Ireland*

Corresponding author(s) e-mail: aaike.dewever@naturalsciences.be

ABSTRACT:

Aquatic ecosystems –from marine and coastal to freshwater– are rich in biodiversity and home to a diverse array of species and habitats, providing numerous economic and societal benefits to the European population. Many of these valuable ecosystems are at risk of being irreversibly damaged by human activities and pressures, including pollution, contamination, hydromorphological alterations, invasive species, overfishing and climate change. These pressures threaten the sustainability of these ecosystems, their provision of ecosystem services and ultimately human well-being.

AQUACROSS (Knowledge, Assessment, and Management for AQUAtic Biodiversity and Ecosystem Services aCROSS EU policies – http://aquacross.eu) seeks to advance the application of ecosystem-based management for aquatic ecosystems in an effort to support the timely achievement of the EU 2020 Biodiversity Strategy and other international conservation targets. In this regard, AQUACROSS aims to develop and test an assessment framework through a series of cases studies which considers the full array of interactions within aquatic ecosystems, including human activities.

Following the Horizon 2020 Open Research Data Pilot, AQUACROSS addresses the challenge of bringing together newly generated as well as existing data used in the framework of the case studies, while at the same time supporting the project partners in terms of data integration and harmonisation. Through a lightweight CKAN-based information platform, we aim on one hand to support project partners in terms of discovery and data access (in interoperable formats), while on the other hand we offer operational support to open up raw and processed data for use in other contexts and disseminating these data and results. The latter is facilitated by the capabilities of the CKAN software which includes on-the-fly analysis and visualisation tools and enables easy data access through external software and tools such as R, QGis and Phyton. During this presentation we will report on the main lessons learned during this data integration exercise and focus on selected case study examples.

KEYWORDS: Aquatic biodiversity data, Freshwater, Marine, Coastal, Ecosystem services data, Data integration, Information platform, Open Data

The Netherlands Biodiversity Data Services and the R package nbaR: Automated workflows for biodiversity data analysis

***Hannes Hettling*, Maarten Schermer, Rutger Vos, Daphne Duin**

Naturalis Biodiversity Center, Leiden, The Netherlands

Corresponding authors e-mail: hannes.hettling@naturalis.nl, daphne.duin@naturalis.nl

ABSTRACT:

The value of data present in natural history collections for research in biodiversity, ecology and evolution cannot be overstated. Naturalis Biodiversity Center of the Netherlands, home to one of the largest natural history collections in the world, launched a large-scale digitisation project resulting in the registration of more than 38 million specimen objects, many of them annotated with descriptive metadata, such as geographic coordinates or multimedia content. Other resources hosted at Naturalis include species occurrence records and comprehensive taxonomic checklists, such as the Catalogue of Life. As our institution strongly believes in the Open Science paradigm, we seek to make our data available to the global biodiversity research community, enhancing data analysis workflows, as for example (i) the modelling of present, past and future species distributions using specimen occurrence data, (ii) time calibration of (molecular) phylogenies using dated specimen occurrences, (iii) taxonomic name resolution or (iv) image data mining. To this end, we developed the Netherlands Biodiversity Data services [1], providing centralized access to biodiversity data via state of the art, open access interfaces and a mechanism to assign persistent identifiers to all records. Data are retrieved from heterogeneous sources and harmonized into a document store that complies with international data standards such as ABCD (Access to Biological Collection Data [2]). Employing the Elasticsearch engine, our infrastructure features complex query options, near real-time queries, and scaling possibilities to secure foreseen data growth. Focusing on availability and accessibility, the services were designed as a versatile, low-level REST API to allow the use of our data in a broad variety of applications and services. For programmatic access to our data services, we developed client libraries for several programming languages. Here we present the R package 'nbaR' [3], a client especially targeted to an audience of biodiversity researchers. The R programming language has found wide acceptance in this field over the past years and our package facilitates convenient means to connect our data resources to existing tools for statistical modelling and analysis. The abstraction layer introduced by the client lets the user formulate even complex queries in a convenient manner, thereby lowering the access threshold to our data services. We will demonstrate the potential and benefits of services and R client by integrating nbaR with state-of-the art packages for species distribution modelling and time calibration of phylogenetic trees into a single analysis workflow.

KEYWORDS: biodiversity data, automated access, API, automated analysis workflows, R package

REFERENCES:

1. Netherlands Biodiversity Data services – User documentation. http://docs.biodiversitydata.nl (accessed 17 May 2018).

2. Access to Biological Collections Data task group. 2007. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/115 (accessed 17 May 2018).

3. nbaR GitHub repository. https://github.com/naturalis/nbaR (accessed 17 May 2018).

Integrating data and analysis: On bridging data publishers and computational environments

***Markus Stocker*[1,2], *Uwe Schindler*[2], *Robert Huber*[2]**

[1]*TIB Leibniz Information Centre for Science and Technology, Welfengarten 1 B, 30167 Hannover, Germany*;
[2]*PANGAEA Data Publisher for Earth & Environmental Science, MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Str. 8, 28359 Bremen, Germany*

Corresponding author(s) e-mail: markus.stocker@tib.eu

ABSTRACT:

Prior to analysing data, researchers today need to perform the ‘janitorial’ step of the data life cycle. This step involves cleaning, harmonizing, or integrating data and typically relies on loading data from one or multiple sources into a computational environment and one of its native data structures. Loading data consumes only a small fraction of the estimated 80% of time consumed by the ‘janitorial’ step overall in data analysis. Yet, it is baffling how much effort it can take to load data into a native data structure of a computational environment.

What could arguably be as straightforward as providing a DOI to a specialized function that returns the corresponding data (and metadata) represented in a data structure native to the computational environment in reality generally amounts to resolving the DOI using a browser, navigating a landing page to identify data and metadata, download data to a file, and ultimately load the data from the file using one of several specialized functions that read data in one of many file formats. The matter is further complicated by Web APIs that - while easing access and download - generally require prior knowledge for how to retrieve data. Such knowledge needs to be encoded in programming code using the computational environment of choice. Surely the required pieces of technology exist to directly access data given a DOI and negotiate content between data provider and consumer so that the computational environment can automatically load data into a native data structure. Yet we still have some way to go before the subtask of loading data into a computational environment is truly easy.

Using PANGAEA as a data publisher and a couple of other data sources, and Jupyter as a computational environment, in this talk we highlight the problem and delineate a solution. Specifically, we will demonstrate how, given a DOI name, PANGAEA data can be automatically loaded into a Python Data Analysis Library (pandas) DataFrame with a mere call of a specialized function. We will also discuss some of the challenges and implications of performing such operation on Linked Data. While the prototype does not do justice to the complexity of generalizing the implementation over heterogeneous data sources and data types, we argue the talk contributes to improving how a minor but necessary subtask of the data life cycle may be executed in computational environments, and thus contribute to seamless integration of data and analysis.

KEYWORDS: Computational Environments, Data Publishers, Data Analysis, Integration, Linked Data

Development of a reef fish community evaluation system for the Galapagos islands

***Peter Goethals*[1], *Heleen Raat*[1], *Stijn Bruneel*[1] *Rafael Bermudez*[2], *Marie Anne Eurie Forio*[1]**

[1]*Department of Animal Science and Aquatic Ecology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium*; [2]*Facultad de Ingeniería Marítima, Ciencias Biológicas, Oceánicas y Recursos Naturales, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador*

Corresponding author(s) e-mail: Peter.Goethals@ugent.be

ABSTRACT:

Reef fish in several islands of the Galapagos are affected via habitat deterioration, fisheries and water quality changes. Sound data acquisition and processing techniques to quantify the impact of anthropogenic pressures are therefore of paramount importance. In this study we developed an indicator system to assess the quality of fish communities. The evaluation is based on both taxonomic and trait-based approaches and allowed us to study ecosystem functioning and the role of certain species and groups in their habitats. The data were collected during a sampling campaign in 2017, during which reef fish were monitored via video recording of diver transects.

KEYWORDS: fish indices, trait-based approaches, bio-monitoring, reef systems, video monitoring

Harmonizing long-tail ecological data sets for synthesis

***Corinna Gries*[1], *Margaret O'Brien*[2], *Colin Smith*[1]**

[1]*University of Wisconsin, Madison, USA*; [2]*University of California, Santa Barbara, USA*

Corresponding author(s) e-mail: cgries@wisc.edu, margaret.obrien@ucsb.edu

ABSTRACT:

With the almost forty year existence of the Long-term Ecological Research (LTER) Network and other networks worldwide, a diverse array of long-term observational and experimental data is becoming increasingly available. A number of data repositories are making the data accessible and the accompanying detailed metadata allow meaningful reuse, repurpose and integration with other data. However, in synthesis research the largest time investment is still in discovering, cleaning and combining primary datasets until all data are completely understood and converted to a similar format. There are two approaches to achieving this data regularity: a) to prescribe the format before data collection starts, or b) to convert primary data into a flexible standard format for reuse. Prescribed formats have rarely been successful due to a wide range of ecosystems, original research questions and varying sampling methods. Hence, we took the second approach: define a flexible intermediate data model, and convert primary data. In the context of the Environmental Data Initiative's data repository, this allows to maintain the original data format, which is most convenient for answering the original research questions, add a conversion script in R to reformat the data into the intermediary format and make this available to synthesis research as Level 1 datasets. This pre-harmonization step may be accomplished by data managers because the dataset still contains all original information without any aggregation or science question specific decisions for data omission or cleaning. Although the data are still distributed into distinct datasets, they can easily be discovered and converted into other formats. We'll be briefly introducing the ecocomDP, an intermediary format for long-term community observation data to initiate the discussion of collaboration with other groups working in this space and how to extend this working model to other ecological data types.

KEYWORDS: Data harmonization, Data reuse, Long-term

BBN models as trade-off tools for ecosystem services

***Marie Anne Eurie Forio*[1]*, Helena Ryckebusch*[1]*, Wout Van Echelpoel*[1]*, Peter Goethals*[1]**

[1]*Ghent University, Belgium*

Corresponding author(s) email: peter.goethals@ugent.be

ABSTRACT:

Human activities (e.g. agriculture and industry) have been increasing within the river basins. These agricultural and industrial impacts affect the water quality in the basin and the related ecosystem and ecosystem services. We aim to identify the link between water management options and ecosystem services, and how the ecosystem services change when certain management actions are applied. Thus, a case study in the Guayas River basin (Ecuador) is presented. A trade-off tool was developed to assess ecosystem services. A model was developed with both biophysical and management variables as input to calculate the effect on multiple ecosystem services (food production, fresh water availability and recreational value). With these three ecosystem services, a trade-off tool was constructed with the use of Bayesian Belief Networks (BBNs). Based on the model outcome, it can be stated that in general, the changes in management leading to an increase in food production, made the fresh water availability and recreational value go down. With further research on the influences of different variables on ecosystem functions and services, this BBN model can become a valuable trade-off tool to be implemented in both water and environmental management and decision making.

KEYWORDS: Ecosystem services, Food production, Biodiversity, Water quality

Leveraging Cloud Computing and IoT to Improve Research Solutions for Ecological Modelling

Fabiana Santana[1]

[1]*University of Canberra, Australia*

Corresponding author(s) e-mails: Fabiana.Santana@gmail.com

ABSTRACT:

Developing strategies for conservation and sustainable use of the environment and natural resources is mandatory nowadays, as the effects of global change are affecting populations worldwide and economic crises are threatening even well-known pro-environmental societies. Leveraging powerful technological resources such as the ones provided by cloud computing and internet of things (IoT) may help to build cost-effective solutions to keep improving research outcomes in ecological modelling, even under a financially-challenging scenario. Cloud computing allows the on-demand delivery of configurable computing resources (e.g., networks, servers, storage, applications and services). Different combinations of resources can be delivered under different models, being IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service), and SaaS (Software-as-a-Service) the most common ones. Public clouds such as Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform offer cloud services based on the pay-as-you-go pricing model. If properly used, they allow the provision of secure, reliable, and performance-efficient solutions which are also cost-optimised. The pay-as-you-go model reduces considerably the upfront investments, so a researcher may be able to develop demos and proofs-of-concept with little funding. Community clouds and virtual laboratories represent technological-efficient and cost-effective solutions to support collective research. For example, the Australian National Collaborative Research Infrastructure Strategy (NCRIS) supports both the Biodiversity and Climate Change Virtual Lab (http://bccvl.org.au/) and the EcoCloud (http://www.ecocloud.org.au/) projects, which provide different cloud-based solutions for ecological modelling. IoT can be seen as a gigantic network of connected devices (called things). The things contain embedded technology to collect data (such as sensors and cameras) and/or interact with the external environment, including people and species. Examples of things would be mobiles, drones, robots, coffee makers, washing machines, and wearable devices. Intelligent things are things that are not only interconnected but also deliver the power of Artificial Intelligence (AI)-enabled systems. This work describes an architectural solution to connect things to a cloud-based platform, leveraging cloud computing and IoT to improve the current processes and state-of-art research in Ecological Modelling. The things, in this case, are robots, drones, sensors, and cameras that are being used to collect species data in the field. The data collected is used as input to an AI-cloud-based system, which remotely controls the things in the field, turning them into intelligent things to improve their data collection capability. Challenges in data acquisition, communication, security, interoperability, integration, big data, performance, latency, and AI-algorithms are discussed, as well as the main pillars to successfully deploy cloud-based IoT solutions.

KEYWORDS: Cloud computing, Internet of Things, Ecological modelling, Ecological monitoring, Data acquisition and integration.

Session S1.5

THE GERMAN FEDERATION FOR BIOLOGICAL DATA (GFBio) - FROM DATA AQUISITION TO ANALYSIS

The German Federation for Biological Data (GFBio) - from data acquisition to analysis

***Michael Diepenbroek*[1], *Birgitta Koenig-Ries*[2], *Anton Güntsch*[3]**

[1]*MARUM, Germany*; [2] *FSU Jena, Germany*; [3]*Freie Universität Berlin, Germany*

Corresponding author(s) e-mail: mdiepenbroek@pangaea.de

ABSTRACT:

The German Federation for Biological Data (GFBio) aims to set up a sustainable, service oriented, national data infrastructure facilitating data sharing and stimulating data intensive science in the fields of biological and environmental research. GFBio follows a holistic approach including technical, organizational, cultural, and policy aspects. The development of the infrastructure is essentially based on the collective experience and expertise of leading researchers from multiple disciplines as well as on a network of complementary and professional data facilities in the biological and environmental sciences communities, including PANGAEA, major German natural history collection data repositories, and selected facilities from the molecular biology research community.

The workshop will address a broad range of aspects related to biological research data management, archiving, interoperability, and analysis.

KEYWORDS: data acquisition, data archiving, data discovery, terminologies, data integration, data visualization, data analysis

The Development of a Data Recommender System for Improving the Discovery of Environmental and Biological Scientific Datasets

***Anusuriya Devaraju*[1], *Uwe Schindler*[2], *Michael Diepenbroek*[2], *Robert Huber*[2], *Jens Klump*[1], *Markus Stocker*[3]**

[1]*CSIRO Mineral Resources, Australia*; [2]*MARUM, University of Bremen, Germany*; [3]*German National Library of Science and Technology (TIB), Germany*

Corresponding author(s) e-mail: anusuriya.devaraju@googlemail.com

ABSTRACT:

In biological and environmental science, there has been good progress in setting up several data repositories to provide greater access to scientific data. However, users cannot realize their values if they cannot quickly locate the datasets required for their scientific research and applications. Recent studies on data retrieval practices [1,2] have revealed that current data portals lack effective data discovery solutions. Text or keyword search provided by the data portals matches user queries and data descriptions to rank the relevant datasets. This type of search depends on how well the data owners described the datasets, or how the users expressed their information needs. It may yield either empty results or too many almost identical datasets. Additionally, users need to be familiar with the structure and terminology of the portal to obtain meaningful results. The text-based search may produce top-ranked search results, which may be retrieved from the same data collection, sharing common attributes. As a result, users are unlikely to discover novel datasets. Therefore, we need an innovative data discovery solution that complements the existing search tools on the portals to deliver relevant and new datasets to users. To address this challenge, we developed a data recommender system for scientific datasets. We describe the system in the context of the PANGAEA data portal. PANGAEA is a data publisher for Earth and Environmental Sciences and hosts more than 370,000 datasets with more than 12 billion measurements from various disciplines. It uses Elasticsearch to index and to support full-text search on the datasets. We present the design and development of the recommender system and describe how a data search engine can be used to build a scalable data recommender system. The data recommender system uses the metadata of datasets and the user interactions extracted from the data server logs to deliver two types of recommendations, i.e., 'similar datasets' and 'users who are interested in this item are also interested in..'. In addition to PANGAEA, we provide some insights on how we may apply the system in the context of other scientific data portals such as the German Federation for Biological Data (GFBio). Building a data recommender on top of Elasticsearch enhances the scalability and maintainability of the recommender system. Our work is an essential contribution towards developing a real-world recommender system for improving scientific dataset discovery.

KEYWORDS: Recommender System, Open Scientific Data, Digital Library, PANGAEA, Data Discovery

REFERENCES:

1. Kathleen Gregory, Paul T. Groth, Helena Cousijn, Andrea Scharnhorst, and Sally Wyatt. 2017. Searching data: A review of observational data retrieval practices. CoRR, http://arxiv.org/abs/1707.06937.
2. Dagmar Kern and Brigitte Mathiak. 2015. Are there any differences in data set retrieval compared to well-known literature retrieval? Springer International Publishing, Cham. 197–208. https://doi.org/10.1007/978-3-319-24592-8_15

The EDIT Platform for Cybertaxonomy - an integrated software environment for biodiversity research data management

Anton Güntsch[1]***, Andreas Müller***[1]***, Patrick Plitzner***[1]***, Tilo Henning***[1]***, Norbert Kilian***[1]***, David Fichtmüller***[1]***, Maren Gleisberg***[1]***, Naouel Karam***[2]***, Claudia Müller-Birn***[1]***, Walter G. Berendsohn***[1]

[1]*Freie Universität Berlin, Germany*; [2]*Fraunhofer-Institut für offene Kommunikationssysteme FOKUS, Germany*

Corresponding author e-mail: a.guentsch@bgbm.org

ABSTRACT:

The Platform for Cybertaxonomy [1], developed as part of the EU Network of Excellence EDIT (European Distributed Institute of Taxonomy), is an open-source software framework covering the full breadth of the taxonomic workflow, from fieldwork to publication [2]. It provides a number of tools for full, customized access to taxonomic data, editing and management, and collaborative team work.

At the core of the platform is the Common Data Model [3], offering a comprehensive information model covering all relevant data domains: names and classifications, descriptive data (morphological and molecular), media, geographic information, literature, specimens, persons, and external resources [4]. The model adheres to community standards developed by the Biodiversity Information Standards organization TDWG [5]. Apart from its role as a software suite supporting the taxonomic workflow, the platform is a powerful information broker for a broad range of taxonomic data providing solid and open interfaces including a Java programmer's library and a CDM Rest Service Layer.

In the context of the DFG-funded "Additivity" project ("Achieving additivity of structured taxonomic character data by persistently linking them to preserved individual specimens", DFG project number 310530378), we are developing components for capturing and processing formal descriptions of specimens as well as algorithms for aggregating data from individual specimens in order to compute species-level descriptions [6]. Well-defined and agreed descriptive vocabularies referring to structures, characters and character states are instrumental in ensuring the consistency and comparability of measurements. This will be addressed with a new EDIT Platform module for specifying vocabularies based on existing ontologies for descriptive data. To ensure that these vocabularies can be re-used in different contexts, we are planning an interface to the Terminology Service developed by the German Federation for Biological Data (GFBio) [7]. The Terminology Service provides a semantic standards aware and harmonised access point for distributed or locally stored ontologies required for biodiversity research data management, archiving and publication processes [8].

The interface will work with a new OWL export function of the CDM library, which provides EDIT Platform vocabularies in a format that can be read by the import module of the Terminology Service. In addition, the EDIT Platform will be equipped with the ability to import semantic concepts from the Terminology Service using its API and keeping a persistent link to the original concept. With an active pipeline between the EDIT Platform and the GFBio Terminology Service, terminologies originating from the taxonomic research process can be re-used in different research contexts as well as for the semantic annotation and integration of existing research data processed by the GFBio archiving and data publication infrastructure.

KEYWORDS: taxonomic computing, descriptive data, terminology, inference

REFERENCES:

1. EDIT Platform for Cybertaxonomy. http://www.cybertaxonomy.org (accessed 17 May 2018).

2. Ciardelli, P., Kelbert, P., Kohlbecker, A., Hoffmann, N., Güntsch, A. & Berendsohn, W. G., 2009. The EDIT Platform for Cybertaxonomy and the Taxonomic Workflow: Selected Components, in: Fischer, S., Maehle, E., Reischuk, R. (Eds.): INFORMATIK 2009 – Im Focus das Leben. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings 154. Köllen Verlag, Bonn, pp. 28;625-638.

3. Müller, A., Berendsohn, W. G., Kohlbecker, A., Güntsch, A., Plitzner, P. & Luther, K., 2017. A Comprehensive and Standards-Aware Common Data Model (CDM) for Taxonomic Research. Proceedings of TDWG 1: e20367. https://doi.org/10.3897/tdwgproceedings.1.20367.

4. EDIT Common Data Model. https://dev.e-taxonomy.eu/redmine/projects/edit/wiki/CommonDataModel (accessed 17 May 2018).

5. Biodiversity Information Standards TDWG. http://www.tdwg.org/ (accessed 17 May 2018).

6. Henning T., Plitzner P., Güntsch A., Berendsohn W. G., Müller A. & Kilian N., 2018. Building compatible and dynamic character matrices – Current and future use of specimen-based character data. Bot. Lett. https://doi.org/10.1080/23818107.2018.1452791.

7. Diepenbroek, M., Glöckner, F., Grobe, P., Güntsch, A., Huber, R., König-Ries, B., Kostadinov, I., Nieschulze, J., Seeger, B.; Tolksdorf, R. & Triebel, D., 2014. Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform: The German Federation for the Curation of Biological Data (GFBio), in: Plödereder, E., Grunske, L., Schneider, E., Ull, D. (Eds.): Informatik 2014 – Big Data Komplexität meistern. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings 232. Köllen Verlag, Bonn, pp. 1711-1724.

8. Karam, N., Müller-Birn, C., Gleisberg, M., Fichtmüller, D., Tolksdorf, R., & Güntsch, A., 2016. A Terminology Service Supporting Semantic Annotation, Integration, Discovery and Analysis of Interdisciplinary Research Data. Datenbank-Spektrum, 16(3), 195–205. https://doi.org/10.1007/s13222-016-0231-8.

Publishing and analysing biodiversity data rapidly, repeatably and FAIR-ly for agile policy relevant results

Quentin Groom[1]***, Sonia Vanderhoeven***[2]***, Lien Reyserhove***[3]***, Damiano Oldoni***[3]***, Tim Adriaens***[3] ***and Peter Desmet***[3]

[1]*Meise Botanic Garden, Belgium*; [2]*Belgian Biodiversity Platform, Belgium*; [3]*Research Institute for Nature and Forest (INBO), Belgium*

Corresponding author(s) e-mail: quentin.groom@plantentuinmeise.be

ABSTRACT:

Informing policy on invasive species requires rapid mobilization of biodiversity data from many sources and converting these raw data into simple metrics and reliable information. Yet the data are collected by a wide variety of actors, both professional and amateur, and they are often divided by political and language barriers. Belgium is typical in that sense, we struggle with the fragmented data sources and restrictions on data mobility, while trying to answer the policy needs at both national and local levels. In 2017 we launched the TrIAS project that aims to resolve some of these problems. We envisage a future where data are rapidly mobilized, the spread of exotic species is regularly monitored, future risks are evaluated and potential impacts assessed [1]. In many ways we have similar aims to GFBio, though we may have found different solutions to the same problems. TrIAS is building workflows that openly publish species information and primary biodiversity data, then harvest those and other data, to create indicators, predictive models and policy support documentation. TrIAS is a consortium of 12 Belgian institutions together with another 9 stakeholder organizations. We aim to address terrestrial, freshwater and marine environments. The organizations involved include those involved in citizen science, pure and applied research, and wildlife management. TrIAS is an open science project; all the software are shared under an MIT Licence; all the data are shared under a CC0 licence waiver and all the documentation is shared under Creative Commons licenses. The hope is that this approach will contribute to the post-project sustainability, because the data and software can all be reused as a whole or in part, either in Belgium, or anywhere else it is needed. Such reuse need not be confined to alien species monitoring, but there is also a need for repeatability and rapid mobilization in other fields, such as red-listing of conservation worthy species.

KEYWORDS: invasive species, rapid data publication, sustainability, biodiversity monitoring, cost-effective, Global Biodiversity Information Facility, R tools, controlled vocabularies, policy relevant

REFERENCES:

1. Vanderhoeven, S., Adriaens, T., Desmet, P., Strubbe, D., Backeljau, T., Barbier, Y., Brosens, D., Cigar, J., Coupremanne, M., De Troch, R., Eggermont, H., Heughebaert, A., Hostens, K., Huybrechts, P., Jacquemart, A., Lens, L., Monty, A., Paquet, J., Prévot, C., Robertson, T., Termonia, P., Van De Kerchove, R., Van Hoey, G., Van Schaeybroeck, B., Vercayie, D., Verleye, T., Welby, S., Groom, Q. 2017. Tracking Invasive Alien Species (TrIAS): Building a data-driven framework to inform policy. Research Ideas and Outcomes. 3, e13414. https://doi.org/10.3897/rio.3.e13414

A flexible *Diversity Workbench* tool to publish biodiversity data from SQL database networks through platforms like GFBio

***Markus Weiss*[1], *Tanja Weibulat*[1], *Stefan Seifert*[1], *Juan Carlos Monje*[2], *Marcel Ruff*[1], *Dieter Neubacher*[1], *Wolfgang Reichert*[1] & *Dagmar Triebel*[1]**

[1]SNSB IT Center and Botanische Staatssammlung München, Staatliche Naturwissenschaftliche Sammlungen Bayerns (SNSB), Germany; [2]Staatliches Museum für Naturkunde Stuttgart (SMNS), Germany

Corresponding author(s) e-mail: triebel@bsm.mwn.de

ABSTRACT:

The *Diversity Workbench* (DWB, www.diversityworkbench.net) is a suite of MS SQL databases and data processing tools designed for the management of research data in biology, ecology and geosciences. The software apart from the underlying DBMS is open source. The complete software package is freely available. The DWB with ten data domain-specific databases, one generic database and some independent data processing tools is the result of 20 years software development. The DWB is suitable for single researchers, research groups of any size but also appropriate to set up networks for long-term data repositories and data centers. To operate automated transfer of bio- and geodiversity data for publication from those in-house networks, a new DWB tool was implemented. The core functions include the filtering and transformation of data and metadata from selected in-house data collections stored in productive master SQL databases. The tool is designed for use by database administrators and scientific data curators.

It fulfils three major steps for each single data collection:

- term, taxa and metadata assignment with parallel data export and creation of a not publicly accessible first-level cache MS SQL database independent from the DWB master database network; unifying data from data domain-specific DWB source databases inside of a (institutional) firewall
- re-organisation of data, filtering according later data package assignment and creation of a second-level publicly accessible PostgreSQL database
- creation of a publicly accessible data package properly formatted for data harvesting tools of web portals and for data mapping and provision software like the BIOCASe Provider Software.

With this DWB tool for guiding data publication some major challenges in bio- and geodiversity research are addressed: a) The data filtering, transformation and publication can be done periodically and is realized without data change and information loss in the linked in-house master databases which might be curated in the long-run. b) A data expert and data scientist is able to handle this transformation tool and organize data publication with minor involvement of a database administrator. c) The data packages are configured for publication according to the individual requests of data producers who often ask for anonymization of certain persons, intend to withhold single data units, set embargos and have to blur geographic coordinates. d) With the automated data transfer for publication, the tool guarantees a reproducible path from the original source to the presentation on a platform.

The data centers SMNS and SNSB are using the tool to guide their data publication through the GBIF global biodiversity data network (https://www.gbif.org/) and through the GFBio platform (https://www.gfbio.org/). Furthermore, it is used to create specifically formatted, publicly accessible cache databases with filtered and aggregated content for thematically focused information portals like the Botanischer Informationsknoten Bayern (http://daten.bayernflora.de/de/index.php). Summarizing, the new DWB tool is able to support a wide range of data transfer and transformation tasks for data publication from DWB networks. It is included in the published software versions of *DiversityCollection* and described in its manual. Future planning includes the extension of the tool to involve the data filtering and transformation from *DiversityDescriptions* as a generic DWB source database.

KEYWORDS: automated data transfer, biodiversity data, data filtering, data publication, GFBio data centers

Managing and publishing fungal community barcoding data by use of the process-oriented schema MOD-CO and a GFBio data publication pipeline

Janno Harjes[1], Dagmar Triebel[2], Tanja Weibulat[2], Anton Link[2], Gerhard Rambold[1]

[1]Dept. of Mycology, University of Bayreuth, Germany; [2]IT Center, Staatliche Naturwissenschaftliche Sammlungen Bayerns, München, Germany

Corresponding author(s) e-mail: triebel@bsm.mwn.de

ABSTRACT:

The need to fulfil FAIR guiding principles for data management and publication [1] directly affects researchers, i.e., data producers as well as data managers. Data management has to be set up well already at an early stage of the data life cycle. This is demonstrated by a best practice work- and dataflow '*Fungal community barcoding data'*, which has been established as side product in the context of the project 'GBOL 2 Mycology', German Barcode of Life initiative (https://www.bolgermany.de/). The work- and dataflow was set up by applying the newly published MOD-CO schema, Version 1.0 which has been implemented as an instance of the database application DiversityDescriptions for data management, and for making data compliant to GFBio infrastructure for data archiving and publication.

The comprehensive conceptual schema MOD-CO for 'Meta-Omics Data of Collection Objects' Version 1.0 was published as Linked Open Data representation in spring 2018 [2]. The process-oriented schema describes operations and object properties along the work- and dataflow from gathering environmental samples, to the various transformation, transaction, and measurement steps in the laboratory up to sample and data publication and archiving. By supporting various kinds of relationships, the MOD-CO schema allows for the concatenation of individual records of the operational steps along a workflow. The MOD-CO descriptor structure in version 1.0 comprises 653 descriptors (concepts) and 1,810 predefined descriptor states, organised in 37 concept collections. The published version 1.0 is available as various schema representations of identical content (https://www.mod-co.net/wiki/Schema_Representations).

This schema has been implemented as data structure in the relational database DiversityDescriptions (DWB-DD) (https://diversityworkbench.net/Portal/DiversityDescriptions), a generic component of the Diversity Workbench environment (https://diversityworkbench.net). DWB-DD is considered being appropriate to be applied as a LIMS (Laboratory Information Management System) and ELN (Electronic Laboratory Notebook) for organising *Fungal community barcoding data*' and similar data collections in molecular laboratories. Its data export interface provides guidance to generate data and metadata in the formats CSV and XML, the latter following the SDD metadata schema with involvement of extensions by metadata elements from EML and ABCD standards; for community standards see: https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_relevant_for_the_collection_data_domain_within_the_GFBio_network. The research data themselves are organised according to the MOD-CO data schema.

The data package of the work- and dataflow '*Fungal community barcoding data'* is going to be submitted to GFBio after having been checked for GFBio compliance and to be published under a creative common license. Suggestions for standardized citation will be provided, a DOI assigned, and long-term data archiving ensured.

KEYWORDS: DiversityDescriptions, German Barcode of Life (GBOL), German Federation for Biological Data (GFBio), MOD-CO conceptual schema, use case for community barcoding data

REFERENCES:

1. Wilkinson, M.D. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. – Sci. Data 3: 160018. DOI: 10.1038/sdata.2016.18.

2. Rambold, G., Yilmaz, P., Harjes, J., Link, A., Glöckner, F.O., Triebel, D. 2018. MOD-CO schema – a conceptual schema for processing sample data in meta'omics research (version 1.0). http://mod-co.net/wiki/MOD-CO_Schema_Reference.

Unfolding existing Data Publication Practice in Research Data Workflows in the Biological and Environmental Sciences – First Results from a Survey

Felicitas Löffler[1], ***Tina Astor***[2], ***Claudia Müller-Birn***[3]

[1]*Heinz-Nixdorf-Professur für verteilte Informationssysteme, Friedrich-Schiller-Universität Jena, Germany*; [2]*Georg-August-Universität Göttingen, Germany*; [3]*Freie Universität Berlin, Germany*

Corresponding author(s) e-mail: felicitas.loeffler@uni-jena.de

ABSTRACT:

In recent years, data publication workflows get more and more attention [1,2]. In order to obtain FAIR data [3], reviewers, data curators and other stakeholders have realized that not only the submitted data matter but also the underlying process to create that data within existing research practice. A better understanding of existing data publication practices in research workflows will help service providers such as data repositories (Pangaea [4], ENA [5], GenBank [6]) to support their users with more appropriate services and tools when submitting data, and otherwise, will sustain the role of data repositories in research practice. Such improved coordination will minimize the workload of researchers and data curators and will facilitate the review process of all stakeholders with respect to reproducibility. Furthermore, well-documented data publication workflows may improve data retrieval and finally data reuse in a long run.

One obstacle towards comprehensible and properly described research workflows is the fact that data publication workflows in the life sciences are hard to define. Scholars have their very individual disciplinary background, research skills and experiences. In some domains such as biodiversity, scholars work from several weeks to years to collect and analyze often heterogeneous data from various sources, such as collections, environmental or molecular data repositories. Thus, reconstructing their work process after the project is finalized is very difficult if not impossible. However, our goal is to reveal the state of the art on how scholars manage their data in their research practices. We are in the process of setting up a survey whose general structure is organized according to the GFBio Data Lifecycle [7]. The results will allow us to reveal typical data practices workflows that can be used to evaluate the suitability of existing data repository portals, such as GFBio [8].

In our talk, we present the first insights of the survey.

KEYWORDS: data publication workflows, data practices, biological and environmental data, green life sciences, biodiversity

REFERENCES:

1. Dallmeier-Tiessen, S., Khodiyar, V., Murphy, F., Nurnberger, A., Raymond, L., Whyte, A., 2017. Connecting Data Publication to the Research Workflow: A Preliminary Analysis, International Journal of Digital Curation, 12, https://doi.org/10.2218/ijdc.v12i1.533.
2. González-Beltrán, A., Li, P., Zhao, J., Avila-Garcia, M. S., Roos, M., Thompson, M., van der Horst, E., Kaliyaperumal, R., Luo, R., Lee, T.-L., Lam, T., Edmunds, S.C., Sansone, S.-A., Rocca-Serra, P, 2015. From Peer-

Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics, PLOS ONE 10, 7, pp. 1–20, https://doi.org/10.1371/journal.pone.0127612.

3. Mark D. Wilkinson et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3. https://doi.org/10.1038/sdata.2016.18
4. Pangaea, https://www.pangaea.org
5. ENA, https://www.ebi.ac.uk/ena
6. GenBank, https://www.ncbi.nlm.nih.gov/genbank/
7. GFBio Data Lifecycle, https://www.gfbio.org/training/materials/data-lifecycle
8. GFBio, https://www.gfbio.org

Session S1.6

SEMANTICS FOR BIODIVERSITY AND ECOSYSTEM RESEARCH

Semantics for biodiversity and ecosystem research

***Alsayed Algergawy*[1], *Clement Jonquet*[2], *Naouel Karam*[3], *Friederike Klan*[4], *Nikos Minadakis*[5], *Alessandro Oggioni*[6], *Caterina Bergami*[6], *and Ilaria Rosati*[6]**

[1] *FSU Jena, Germany*; [2]*University of Montpellier – LIRMM, France*; [3]*Fraunhofer FOKUS Berlin, Germany;* [4]*DLR Jena, Germany*; [5]*LifeWatch Greece, Greece*; [6]*LifeWatch Italy, Italy*

Corresponding author(s) e-mail: alsayed.algergawy@uni-jena.de

ABSTRACT:

Biodiversity research aims at comprehending the totality and variability of organisms, their morphology, genetics, life history, habitats and geographical ranges; including the network of interactions with the abiotic and biotic components. Ecosystem research puts its focus on how natural systems and their valuable resources can be protected and thus is tightly coupled to biodiversity. Both domains are outstanding not only with respect to their societal relevance, but also from a data science point of view. They deal with heterogeneous and distributed data resources generated from a large number of disciplines which need to be integrated to advance scientific knowledge in these areas. The presence of such a myriad of data resources makes integrative biodiversity and ecosystem research increasingly important, but at the same time very challenging. It is severely strangled by the way data and information are made available and handled today. Semantic Web techniques have shown their potential to enhance data interoperability, discovery and integration by providing common formats to achieve a formalized conceptual environment, but have not been widely applied to address open data management issues in the biodiversity domain as well as in ecosystem research.

This session aims at bringing together computer scientists, biologists and ecologists working on Semantic Web approaches in biodiversity and ecosystem research, including related areas such as agro-ecology. After the successful of a number of initiatives of the organizers, such as the “Thesauri & Semantics in the Ecological Domain”, “Ontology & Semantic Web for Web for Research” and “Semantics for Biodiversity” workshops, the goal of the session is to keep up the positive momentum and attempt to define a common strategy for advancing semantic web approaches in these domains. The goal is to present new ideas and early on experiences related to the design of high quality biodiversity and ecosystem information systems based on Semantic Web techniques and to foster the exchange on these topics between disciplines.

We welcome topics related to the development and application of semantic technologies to support research in the biodiversity and ecosystem domain and related areas. These include, but are not limited to the following areas:

- Applications of Semantic Web technologies for biodiversity
- Semantic data integration
- Development and design of domain specific ontologies
- Ontology-based applications
- Semantic annotation of biodiversity data
- Semantic approaches for the discovery of biodiversity data and research data services
- Semantic support for scientific workflows

· Data provenance and reproducibility

· Data lifecycle management

· Knowledge extraction and text mining

· Ontology learning

· Standards for biodiversity Data

· Linked Open biodiversity Data

· Ontology development for biodiversity

· Semantic representation of biodiversity and ecosystem data

· Interoperability of biodiversity and earth observation data

KEYWORDS: Semantic web, biodiversity data, ecosystem data, integrative research, semantic annotation, semantic data integration, semantic data interoperability, ontology based applications

Extending the Environment Ontology with Text-mined Habitat Mentions

***Riza Batista-Navarro*[1], *Marie-Angelique Laporte*[2], *Michael Regan*[3], *William Ulate*[4] and *Claus Weiland*[5]**

[1]*The University of Manchester, UK;* [2]*Bioversity International, France;* [3]*University of Colorado Boulder, USA;* [4]*Missouri Botanical Garden, USA;* [5]*Senckenberg Biodiversity and Climate Research Centre, Germany*

Corresponding author(s) e-mail: riza.batista@manchester.ac.uk

ABSTRACT:

Ontologies, i.e., formal specifications of concepts and relations relevant to a specialised domain of interest, are information resources which play a crucial role in the tasks of knowledge representation, management and discovery. Knowledge acquisition, the process of curating and updating them, is typically carried out manually, requiring human efforts that are tedious, time-consuming and expensive. This holds true especially in the case of ontologies which are continuously being expanded with new terms, in their aim to support a growing number of use cases. An example of such is the Environment Ontology (ENVO). Initially developed to support the annotation of metagenomic data, ENVO has more recently realigned its goals in support of the Sustainable Development Agenda for 2030 and thus is currently much broader in scope, covering the domains of biodiversity and ecology. As a result, there has been a dramatic increase with respect to ENVO's number of classes; hence the process of curating and updating the ontology can benefit from automated support. In this work, we aim to help in expanding ENVO in a more efficient manner by automatically discovering new habitat mentions. To this end, we developed a text mining-based approach underpinned by the following pipeline: (1) automatic extraction of habitat mentions from text using named entity recognition methods; (2) normalisation of every extracted mention, i.e., identification of the most relevant ENVO term based on the calculation of lexical similarity between them; (3) application of a filter to retain only habitat mentions that appear to not yet exist in ENVO; and (4) construction of clusters over the remaining mentions. The pipeline results in clusters consisting of potential synonyms and lexical variations of existing terms, as well as semantically related expressions, which can then be evaluated for integration into an existing ENVO class, or, on occasion, be indicative of a new class that could be added to the ontology.

Applying our approach to a corpus pertaining to the *Dipterocarpaceae* family of forest trees (based on documents from the Biological Heritage Library and grey literature), we generated more than 1,000 new habitat terms for potential incorporation into ENVO.

KEYWORDS: Ontology Curation, Text Mining, Environment Ontology, Concept Normalisation

A Semantic Big Biodiversity Data Integration Tool

Taysir H. Soliman[2]*, Alsayed Algergawy*[1]*, Birgitta König-Ries*[1]*, Majid Askar*[2]*, and Marwa Abdel Reheim*[2]

[1] *Friedrich-Schiller-University of Jena, Jena, Germany,*
[2] *Faculty of Computers and Information, Assiut University, Egypt*

Corresponding author: taysser.soliman@fci.au.edu.eg

ABSTRACT:

Our planet is facing huge effects of global climate changes that are threatening biodiversity data to be surviving. Biodiversity data exist in very complex characteristics, such as high volume, variety, veracity, velocity, and value, as Big data. The variety or heterogeneity of biodiversity data provides a very high challenging research problem since they exist in unstructured, semi-structured, quasi-structured, and generated in XML, EML, Excel sheets, videos, images, or ontologies. In addition, the availability of biodiversity data includes trait-measurements, species distribution, species' morphology, genetic sequences, phylogenetic trees, spatial data, and ecological niches; data are collected and uploaded in Bio Portals via citizen scientists, museums' collections, ecological surveys, and environmental studies. These data collections generate big data, which is an important current research. The first phase of Big data analytics life cycle discovers whether the data is enough to perform the analytics process, which takes more time than any other phase. In addition, Big biodiversity data management life cycle includes data integration as a main phase, affecting storage, indexing, and querying. In the data integration phase, we apply semantic data integration in order to combine data from different sources and consolidate them into valuable information that depends on semantic technologies. A number of research attempts have been achieved on semantic big data integration. For example, Ontology-Based Data Access (OBDA) has been proposed in relational schema and in NOSQL [1,2] databases since it provides a semantically conceptual schema over data repository. Another example is Semantic Extract Transform Load (ETL) framework [3], which integrates and publishes data from multiple sources as open linked data provides through semantic technologies. Moreover, Semantic MongoDB-based has been developed where researchers represented as an OWL ontology. However, the need for semantic big data integration tools becomes highly recommended because of the growth of biodiversity big data. In the current work, a semantic big data integration system is developed, which handles the following features: 1) Data heterogeneity, 2) NoSQL databases, 3) Ontology based Integration, and 4) User Interaction, where data integration components can be chosen. A proof-of-concept will be developed based on biodiversity data, having various data formats. In addition, related ontologies will be used from BioPortal.

Keywords: Semantic data integration, Big Data, Biodiversity, and NoSQL databases

REFERENCES:

1. Abbes, H. , Gargouri, G. 2018. MongoDB-Based modular ontology building for big data integration. Journal on Data Semantics. 7:1-27.
2. Kiran, VK., Vijayakumar, R. 2014. Ontology based data integration of NoSQL datastores. In:9th international conference on industrial and information systems (ICIIS).
3. Bansal, SK, Kagemann, S. 2015. Integrating big data: a semantic extract-transform-load framework. Computer 48(3):42–50

Acknowledgment: This work has been part of BioDialog Project funded by DAAD.

LakeBase Semantic Service

Jan Martin Keil

Heinz Nixdorf Chair for Distributed Information Systems,
Institute for Computer Science, Friedrich Schiller University Jena, Germany

Corresponding author(s) e-mail: jan-martin.keil@uni-jena.de

ABSTRACT:

We present the features and architecture of the LakeBase Semantic Service. The LakeBase Semantic Service is a web service for the management of semantic annotations and semantic search of ecological research data. It was developed in the scope of the LakeBase project to support the management of long-term research data of German lake ecosystems.

The basic workflow using the service consists of the following steps: (a) Suggestion of semantic annotations–represented by concept IRIs–for provided data and metadata, (b) optional adjustment of suggested annotations by the data provider, (c) storing the suggested and adjusted annotations, (d) auto completion of search queries typed by data users, (e) suggestion of semantic annotations for search queries, (f) optional selection of suggested annotations for search queries by the data users, (g) semantic search of annotated data with (optionally annotated) queries. To enable presentation and management of the annotations, the LakeBase Semantic Service further allows to output stored annotations, to copy annotation, to delete annotation, and to provide descriptions of annotation concepts.

There are several sources of concepts for the semantic annotation. We defined a default programming interface for semantic data sources that has to be implemented for each source. These implementations are provided for ontology files, SPARQL endpoints, and other services like the World Register of Marine Species (WoRMS) webservice and the GFBio Terminology Server. Due to this modular architecture, further semantic data sources can be added with only little programming effort.

The LakeBase Semantic Service will be used in the IGB Freshwater Research and Environmental Database (FRED). Further, the source code will be freely available to also allow the usage beyond the LakeBase project.

KEYWORDS: Semantic Annotation, Semantic Search, Ontology Based Applications

ACKNOWLEDGMENTS: This work was funded by DFG in the scope of the LakeBase project within the Scientific Library Services and Information Systems (LIS) program.

Towards an Interactive Approach for Ontology Recommendation and Reuse

Marwa Abdelreheim[1]*, Friederike Klan*[2]*, Taysir Soliman*[1]

[1]*Assuit University, Egypt*; [2]*(DLR) Jena, Germany*

Corresponding author(s) e-mail: marwa.abdelrehem, taysser.soliman}@fci.au.edu.eg, friederike.klan@dlr.de

ABSTRACT:

Ontologies are machine-comprehensible and reusable pieces of knowledge designed to explicitly define the semantics of an application domain, using a set of concepts, properties that relate concepts to each other or to literals, and a set of individuals. When deciding to develop an ontology for a new application domain, ontology engineers face the question whether to reuse existing ontologies or to build a new ontology from scratch. In conceptually diverse domains, such as biodiversity, building an ontology from scratch is an expensive and time-consuming process. In such a case, it is a better choice to reuse existing ontologies or parts of them. In general, ontology reuse is defined as the process where existing ontologies, along with possibly other non-ontological resources, are determined and used for building new integrated ontologies as part of a knowledge base. A case study on ontology reuse in different domains, that we conducted, revealed that ontology reuse is either done manually or semi-automatically with IT-support mainly focusing on the retrieval and recommendation of existing ontologies based on their conceptual coverage. This contrasts with the fact that manual ontology engineering and reuse, especially in complex domains, requires great efforts from both ontology engineers and domain experts. Moreover, the ontology reuse process is inherently incremental, as an ontology is developed step-by-step and evolves over time. This aspect is not considered by existing tools that typically make one-shot recommendations. In our talk, we present the concept of a tool which supports interactive ontology recommendation and reuse in order to assist ontology engineers and domain experts in their task to generate an ontological knowledgebase for a specific application domain. The tool will have the following features: a) it allows the user to specify a (potentially empty) seed ontology as a starting point for the new ontology, b) based on a set of candidate ontologies and textual input describing the specified domain, it identifies, extracts and recommends pieces of the candidate ontologies (properties, concepts, textual and formal specifications of concepts) that might be used to extend the seed ontology in an interactive and iterative process, the user selects recommended pieces, which are automatically integrated with the seed ontology. The system makes sure that the resulting ontology is consistent and complies with the domain semantics intended by the user. This will be achieved by the use of logical reasoning and the provisioning of explanations and proper visualizations.

KEYWORDS: Ontology engineering, Interactive Ontology Recommendation, Ontology Reuse

ACKNOWLEDGEMENT: This work is partly funded by the DAAD project “BioDialog”. We are particularly grateful to Prof. Birgitta König-Ries and Dr. Alsayed Algergawy

How to Search for Biological Data? A Comparison of User Interfaces in a Semantic Search

***Felicitas Löffler*[1], *Friederike Klan*[2], *Birgitta König-Ries*[1]**

[1]*Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Germany*;
[2]*German Aerospace Center (DLR), Germany*;

Corresponding author(s) e-mail: felicitas.loeffler@uni-jena.de

ABSTRACT:

Data discovery is a frequent task in a scholar's daily work. In biodiversity, data search is a particular challenge. Here, scholars have complex information needs such as the rich interplay of organisms and their environments that cannot be unambiguously expressed with a traditional keyword search, e.g., Does tree diversity reduce competition in a subtropical forest? Therefore, data repositories usually offer interfaces that enable users to browse datasets by a pre-determined set of categories or facets. Faceted search is a good compromise between cumbersome user interfaces for structured queries (e.g., using SPARQL) and natural language queries that are hard to interpret for machines. Thus, developers can specify relevant relationships between entities explicitly and users can filter search results by selecting proper categories. For the given query, appropriate categories could be *Organism* and *Habitat*.

However, there are two crucial design issues that have an impact on the effectiveness of category-based query interfaces: The choice of proper categories and the visual presentation of these categories in the query interface. In our work, we focus on the second aspect. We aim to develop two query interfaces: (a) a common one-box keyword search interface that analyzes the entered terms with respect to their categories automatically (b) a form-based query interface where users can enter their search keywords into a form with a query field per category. In both interfaces, the query keywords are matched against concepts in a knowledge base to make their semantics explicit. In case of a successful match the URI is used to obtain the labels of all sub-concepts to expand the query before sending it to the search engine. Retrieved results are displayed in a list. The aim of our system is not to answer the question completely but to support users in retrieving relevant datasets that give hints to answer a research question.

In our talk, we will introduce the two interfaces and invite the conference participants to give feedback. We are particularly interested in a discussion on the appropriateness of the suggested user interfaces. Do scholars prefer a form-based user interface or only a one-field search? What other functions might be helpful, e.g., providing more information about other relations and properties from the concept in the ontology? What kind of explanations might be helpful to understand why a certain result was returned?

KEYWORDS: user interfaces, semantic search, biological data, life sciences, biodiversity

Towards a harmonization of distributed trait datasets

***Florian D. Schneider*[1], *Malte Jochum*[2], *Gaëtane Le Provost*[3], *Andreas Ostrowski*[4],**
***Caterina Penone*[2] *and Nadja K. Simons*[5]**

[1] *Institute of Linguistics and Literary Studies, Technische Universität Darmstadt, Darmstadt, Germany;*
[2] *Institute of Plant Sciences, University of Bern, Bern, Switzerland;*
[3] *Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany;*
[4] *Department of Mathematics and Computer Science, Friedrich-Schiller-Universität Jena, Jena, Germany;*
[5] *Department for Ecology and Ecosystem Management, Center for Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany*

Corresponding author(s) e-mail: florian.dirk.schneider@gmail.com

ABSTRACT:

Trait-based research spans from evolutionary studies of individual-level properties to global patterns of biodiversity and ecosystem functioning. An increasing number of trait data is available for many different organism groups, being published as open access data on a variety of file hosting services. Thus, standardization between datasets is generally lacking due to heterogeneous data formats and types. The compilation of these published data into centralised databases remains a difficult and time-consuming task.

We reviewed existing trait databases and online services, as well as initiatives for trait data standardization. Together with data providers and users we identified a need for a minimal trait-data terminology that is flexible enough to include traits from all types of organisms but simple enough to be adopted by different research communities.

In order to facilitate reproducibility of analyses, the reuse of data and the combination of datasets from multiple sources, we propose a standardized vocabulary for trait data that is compatible with existing ontologies. We tested the vocabulary using trait datasets from several research groups working on different taxa and questions in a large project (the Biodiversity Exploratories, www.biodiversity-exploratories.de). By relying on unambiguous identifiers, the proposed minimal vocabulary for trait data captures the different degrees of resolution and measurement detail for multiple use cases of trait-based research. It further encourages the use of global Uniform Resource Identifiers (URI) for taxa and trait definitions, methods and units, thereby following the standards for a semantic web of scientific data.

In addition, we developed an R-based tool to convert any trait dataset into the proposed standard format. The R-package facilitates the upload of own data to hosting services but also simplifies the access to published trait data. It also offers direct access to trait datasets that have been published in the public domain or under creative commons licenses. All these products are available through the Github platform (https://github.com/EcologicalTraitData) with the aim of a continuous collaboration and improvement with the research community.

KEYWORDS: traits, standardization, ontology, semantic web, tools, distributed data, R package, Biodiversity Exploratories

Visualizing the research ecosystem of ecosystem research via Wikidata

***Daniel Mietchen*[1], *Finn Årup Nielsen, Egon Willighagen*[3]**

[1]*Data Science Institute, University of Virginia, Charlottesville, VA, USA*; [2]*Cognitive Systems, DTU Compute, Technical University of Denmark, Copenhagen, Denmark*; [3]*Department of Bioinformatics - BiGCaT, Maastricht University, Maastricht, The Netherlands*

Corresponding author(s) e-mail: dm7gn@virginia.edu

ABSTRACT:

Like research in general, biodiversity and ecosystem research takes place in a sociotechnical ecosystem that connects researchers, institutions, funders, databases, locations, publications, methodologies and related concepts with the objects of study and the world around them. Schemas for describing such concepts are growing in breadth and depth, number and popularity, as are mechanisms to persistently and uniquely identify the concepts, the schemas, their relationships or any of their components. In parallel, more and more data — and particularly metadata — are being made available under open licenses, which facilitates discoverability, reproducibility and reuse, as well as data integration.Wikidata is a community-curated open knowledge base in which concepts covered in any Wikipedia — and beyond — can be described in a structured fashion that can be mapped to RDF and queried using SPARQL as well as various other means. Its community of close to 20,000 monthly contributors oversees a corpus that currently comprises nearly 50 million 'items', i.e. entries about concepts. These items are annotated and linked via almost 5000 'properties' that describe relationships between items or between items and external entities or that express specific values. The items and properties have persistent unique identifiers, to which labels and descriptions can be attached in about 300 natural languages. For instance, Q61457 represents the item for 'acetaldehyde' and Q183339 'Antilope cervicapra', while P3063 stands for the property of 'average gestation period', and P3117 for 'DSSTOX substance identifier'. Besides taxa, chemical compounds, toxicology, geomorphological features or ecological interactions, Wikidata also contains information about researchers and many components of their research ecosystems, including a growing body of publications and databases, particularly in the life sciences. A range of open-source tools is available to interact with Wikidata — to enter information, curate and query it. One of them is Scholia [1], a frontend to Wikidata's SPARQL endpoint. Available via https://tools.wmflabs.org/scholia/ , it can be used to explore research publications and how they relate to authors, institutions, funders and other parts of the research ecosystem, as well as to taxa, metabolic networks, or geolocations.In this presentation — which will be given on the basis of https://github.com/Daniel-Mietchen/events/blob/master/ICEI2018-research-ecosystem.md — we will use Scholia as a starting point for exploring how information about biodiversity and ecosystem research is represented in Wikidata and how it can be explored, curated and reused.

KEYWORDS: Wikidata, SPARQL, research system, visualization, bibliometrics

REFERENCES:

1. Nielsen F.Å., Mietchen D., Willighagen E. (2017) Scholia, Scientometrics and Wikidata. In: Blomqvist E., Hose K., Paulheim H., Ławrynowicz A., Ciravegna F., Hartig O. (eds) The Semantic Web: ESWC 2017 Satellite Events. ESWC 2017. Lecture Notes in Computer Science, vol 10577. Springer, Cham . DOI: 10.1007/978-3-319-70407-4_36

Towards Interoperability for Observed Parameters: Position Statement of an Emerging Working Group

***Barbara Magagna*[1], *Markus Stocker*[2], *Michael Diepenbroek*[3]**

[1]*Environment Agency Austria, Austria;* [2]*TIB Leibniz Information Centre for Science and Technology, Germany*; [3]*PANGAEA Data Publisher for Earth & Environmental Science, Germany*

Corresponding author(s) e-mail: Barbara.magagna@umweltbundesamt.at

ABSTRACT:

For decades, many communities have worked on the definitions of parameters, specifically scientific observation and measurement parameters. A well known example are the climate and forecast standard names (CF) [1]. Controlled vocabularies (e.g. EnvThes [2], Anaeethes [3], BODC Parameter Usage Vocabulary [4], ...) are often used for describing parameters in different domains. PANGAEA [5] as a multidisciplinary data publisher for environmental sciences holds around 375 thousand citable data sets which have to be described with consistent semantics; this can be really challenging when dealing with complex parameters. Inconsistencies among existing parameter definitions as well as syntactic and semantic heterogeneity in their representation in systems prevent the integration of data about parameters from different providers. For individual providers, the growing number and complexity of observation and measurement parameters referred to in published data urgently demands viable approaches for their representation and organization. To address these problems and find common approaches, a group of interested scientists involved in different national and international initiatives and research infrastructures (PANGAEA, LTER-Europe [6], GFBio [7], BODC [8], ENVO [9], LifeWatch Italy [10], ICOS [11], AnaEE [12], AquaDiva [13], ...) decided to organize themselves as an RDA Working Group (WG). Having met several times via conference calls to present each other's related work, it became clear that the problem has been recognized and tackled in various ways, reflecting the specific needs of data and semantic infrastructures of varying maturity. In this talk, we will describe the process of defining a common strategy with a clear output that will be beneficial for all involved communities, and beyond. This entails a consistent terminology used within the group, thorough SWOT analysis of the different methodologies in use (within and outside the group) and a synopsis of the current state. The ultimate aim of this undertaking is to elaborate a common concept for the definition of parameters and develop best practices illustrated on a number of use cases. We will highlight the problem, present and discuss the findings of the current working group, and provide an outlook for the planned work, in particular also a possible work plan for the RDA WG. The talk is an opportunity for this working group to reach out to other potentially interested parties.

KEYWORDS: Network, Baltimore, Ecology, Long-term

REFERENCES:

1. Climate and Forecast Standard Name Table.
http://cfconventions.org/Data/cf-standard-names/49/build/cf-standard-name-table.html (accessed 10 April 2018).
2. EnvThes.
http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn (accessed 10 April 2018).

3. AgroPortal.
http://agroportal.lirmm.fr/ontologies/ANAEETHES (accessed 10 April 2018).
4. SeaDataNet.
http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P01 (accessed 10 April 2018).
5. PANGAEA.
https://www.pangaea.de/ (accessed 10 April 2018).
6. eLTER H2020 project.
http://www.lter-europe.net/lter-europe/projects/eLTER (accessed 10 April 2018).
7. GFBIO.
https://www.gfbio.org/ (accessed 10 April 2018).
8. BODC.
https://www.bodc.ac.uk/ (accessed 10 April 2018).
9. ENVO.
http://environmentontology.org/ (accessed 10 April 2018).
10. LifeWatch Italia.
https://www.lifewatch.eu/italy (accessed 10 April 2018).
11. ICOS.
https://www.icos-cp.eu/ (accessed 10 April 2018).
12. AnaEE.
https://www6.anaee.com/ (accessed 10 April 2018).
13. AquaDiva.
http://www.aquadiva.uni-jena.de/ (accessed 10 April 2018).

Extracting granular information on habitats and reproductive conditions of Dipterocarps through pattern-based literature analysis

***Roselyn Gabud*[1, 2] *and Riza Batista-Navarro*[2,3]**

[1]*University of the Philippines Diliman, Philippines*; [2]*University of the Philippines Los Baños, Philippines*;
[3]*The University of Manchester, UK*

Corresponding author(s) e-mail: rsgabud@up.edu.ph

ABSTRACT:

Lowland tropical rainforests in Southeast Asia primarily comprised of dipterocarp species are one of the most threatened ecosystems in the world. Belonging to the family *Dipterocarpaceae*, dipterocarps are economically and ecologically important due to their timber value as well as contribution to wildlife habitat. The challenge in the restoration and rehabilitation of these Dipterocarp forests lies in their complex reproduction patterns, i.e., supra-annual mass flowering events that may occur in irregular intervals of two to ten years, possibly synchronously across Asia. Understanding their regeneration to make plans for effective reforestation can be aided by providing access to a comprehensive database that contains long-term and wide-scale data on dipterocarps. The content of such a database can be enriched with literature-derived information on habitats and reproductive conditions of dipterocarps.

We aim to develop literature mining methods to automatically extract information relevant to the distribution and reproductive cycle of dipterocarps, in order to help predict the likelihood of their regeneration, and subsequently make informed decisions regarding species for reforestation. In previous work, we developed a machine learning-based named entity recognition (NER) model that automatically annotates entities relevant to species' distribution, e.g., taxon names, geographic locations, temporal expressions, habitats, authorities, and names of herbaria. Furthermore, the species' reproductive condition, e.g., whether it is sterile or in the state of producing fruit ("in fruit") or flower ("in flower"), was also automatically annotated to enable the derivation of phenological patterns. The model was trained on a manually annotated corpus of documents, e.g., scholarly articles and government agency reports.

In this work, we focus our efforts specifically on the extraction of relationships between habitats and their locations, and between reproductive conditions and temporal expressions. To this end, we have developed a syntactic pattern-based matching approach by building upon Grew (http://grew.fr/), a graph rewriting system for manipulating linguistic representations. For our purposes, patterns that made use of syntactic dependencies, part-of-speech tags and named entity types (derived from NER results) were designed. When fed into Grew, these patterns were able to analyse sentences in scholarly articles by associating habitats with their geographic locations, and by determining a species' reproductive condition at a specific point in time. The resulting relationships are then used to enrich information contained in a database of dipterocarp occurrences. Such a resource will provide more comprehensive ecological data that could form the basis of more informed reforestation decisions.

KEYWORDS: Text mining, Dipterocarps, Relation extraction, Syntactic patterns

Towards Semantic Data Management in LifeWatch Italy: the Phytoplankton Study Case

***Nicola Fiore*[1], *Caterina Bergami*[1], *Carla Carrubba*[1], *Alessandro Oggioni*[1], *Ilaria Rosati*[1], *Elena Stanca*[1], *Paolo Tagliolato*[1]**

[1]*LifeWatch Italy, Italy*

Corresponding author e-mail: nicola.fiore@lifewatch.eu

ABSTRACT:

LifeWatch Italy, the Italian node of LifeWatch-ERIC, has promoted and stimulated the debate on the use of semantic in the biodiversity data management. Actually, information from biodiversity and ecosystems is very heterogeneous and needs to be better managed in order to improve the actual scientific knowledge, as well as to address the urgent societal challenges concerning environmental issues.

Here we present the Phytoplankton Study Case, where the semantic approach was used to address data harmonisation, integration and discovery. An interdisciplinary team of LifeWatch Italy has developed a thesaurus on phytoplankton functional traits and linked its concepts to other existing conceptual schema related to the specific domain. In the meantime, the team has produced the LifeWatch Core Ontology, a customization of the OBOE core ontology, for the semantic description/capture of basic concepts and relationships in ecological studies. This framework ontology is based on 7 main concepts (classes) as Domain, Entity, Observation, Characteristic, Measurement, Protocol, Standard, providing a structured yet generic approach for semantic data annotation, and for developing domain-specific ecological ontologies as the Phytoplankton Trait Ontology (PhyTO).

To date, LifeWatch e-Infrastructure stores and manages data and metadata using an mix of Database Management Systems (the Relational MySQL and the NoSQL MONGO DB); for the purpose of the study case, we selected the VIRTUOSO Triple Store as semantic repository and we developed different modules to automate the management workflow.

A first software module has been developed to allow the data annotation with classes, subclasses and properties of the PhyTO (i.e. Semantic Annotation). The designed module allows to map metadata and data stored in the LifeWatch Data Portal with the OWL schema of the PhyTO and to produce .rdf output files. A second developed module uses as input the .rdf files and store the data in the VIRTUOSO Graph to make them available for the semantic search. Moreover, a user-friendly search interface (i.e. Java Portlet) has been implemented to retrieve annotated data with queries suggested by the data users.

This approach facilitates data discovery and integration, and can provide guidance for, and automate, data aggregation and summary.

KEYWORDS: Thesauri, Ontology, Semantic annotation, Semantic Search, Phytoplankton

Provenance-based Semantic Approach for the Reproducibility of Scientific Experiments

***Sheeba Samuel*, Birgitta König-Ries**

Heinz-Nixdorf Chair for Distributed Information Systems, Friedrich-Schiller University, Jena, Germany

Corresponding author(s) e-mail: sheeba.samuel@uni-jena.de

ABSTRACT:

Data provenance has become an integral part of the natural sciences where data flow through several complex steps of processing and analysis to generate intermediate and final results. To reproduce scientific experiments, scientists need to understand how the steps were performed in order to check the validity of the results. The scientific experiments consist of activities in the real world (e.g., wet lab or field work) and activities in cyberspace. Many scientists now write scripts as part of their field research for different tasks including data analysis, statistical modeling, numerical simulation, computation and visualization of results. Reproducibility of the computational and non-computational parts are important steps towards reproducibility of the experiments as a whole. In order to reproduce results or to detect which error occurred in the output, it is required to know which input data was responsible for the output, the steps involved in generating them, the devices and the materials used, the settings of the devices used, the dependencies, the agents involved and the execution environment etc.

The aim of our work is to semantically describe the provenance of the complete execution of a scientific experiment in a structured form using linked data without worrying about any underlying technologies. In our work, we propose an approach to ensure this reproducibility by collecting the provenance data of the experiment and using the REPRODUCE-ME ontology extended from the existing W3C vocabularies to describe the steps and sequence of steps performed in an experiment. The ontology is developed to describe a scientific experiment along with its steps, input and output variables and their relationship with each other. The semantic layer on top of the captured provenance provided with ontology-based data access allows the scientists to understand and visualize the complete path taken in a computational experiment along with its execution environment. We also provide a provenance-based semantic approach which captures the data from interactive notebooks in a multi-user environment provided by JupyterHub and semantically describe the data using the REPRODUCE-ME ontology.

KEYWORDS: Experiments, Ontology, Provenance, Scripts, Reproducibility

Extraction of terms highly associated with named rivers

Juan Rojas-García[1], Riza Batista-Navarro[2], Pamela Faber[1]

[1]*University of Granada, Spain*; [2]*The University of Manchester, UK*

Corresponding author(s) e-mail: juanrojas@ugr.es

ABSTRACT:

EcoLexicon [1] is an electronic, multilingual, terminological knowledge base on environmental sciences, whose flexible design permits the contextualization of data so that they are more relevant to specific subdomains and geographic areas [2] [3]. However, to facilitate the geographic contextualization of concepts such as those belonging to the semantic category of LANDFORM, it is necessary to know what terms are semantically related to each type of landform according to the research papers published by experts, and how those terms are related to each other (*ecologically sustainable development*, *river erosion*, *hurricane-generated seas*, etc).

This paper describes a semi-automatic method for extracting knowledge about terms related to rivers as a type of landform, from a specialized corpus of English Environmental Science texts. A GeoNames database dump [4] was first applied to automatically match the sequences of words in the corpus which are the proper names of rivers (*San Juan River*, *San Bernard River*, *Bird Foot River*, etc).

For all the named rivers recognized in the corpus, their respective geographic coordinates, i.e. longitude and latitude, were automatically retrieved from the GeoNames database dump and then automatically visualized on top of a static map. This type of visualization accounted for the representativeness of the corpus in reference to the location of rivers and the number of times that they were mentioned. Moreover, a hierarchical clustering technique was deployed in order to group the named rivers, based on their latitude and longitude. This allowed us to automatically annotate each river with the geographical area (California, Brazil, etc) it belongs to.

For each river, the contexts in which it appeared were automatically retrieved in such a way that all the contexts contained complete sentences. Subsequently, the subcorpus of contexts was lemmatized. The multi-word terms (i.e., a sequence of words that function as a single specialized term, such as *sediment transport process* or *beach profile change*), automatically collected from EcoLexicon and the Environment Ontology ENVO [5], were automatically matched in the corpus and joined with underscores.

Then, a document-term matrix of co-occurrences was obtained, and the terms in the columns transformed into binary variables (presence vs. absence). Finally, the clustering technique ROCK for categorical variables [6] was adopted to group the named rivers, based on the terms related to them, as reflected in the corpus data.

The preliminary results show that there is a slight association between the geographical areas of the named rivers and the processes mentioned by researchers affecting them. Once these experimental results were validated by Coastal Engineering experts, the knowledge extracted with this method facilitates the geographical contextualization of EcoLexicon with regard to rivers, in the sense that a specific named river can be linked to its more highly associated terms dealt with in the corpus data.

KEYWORDS: Terminological knowledge base, Geographical contextualization, Named river, Conceptual information extraction, Text mining

REFERENCES:

1. LexiCon Research Group, 2018. EcoLexicon: Terminological Knowledge Base on Environmental Sciences. http://ecolexicon.ugr.es (accessed 16 May 2018).

2. Faber, P., 2011. The Dynamics of Specialized Knowledge Representation: Simulational Reconstruction or the Perception-action Interface. Terminology 17 (1), 9-29.

3. León-Araúz, P., Reimerink, A., Faber, P., 2013**.** Multidimensional and Multimodal Information in EcoLexicon, in: Przepiórkowski, A., Piasecki, M., Jassem, K., Fuglewicz, P. (Eds.), Computational Linguistics. Springer, Studies in Computational Intelligence 458, Berlin, Heidelberg, pp. 143-161.

4. GeoNames, 2018. Geonames Data Dump. http://www.geonames.org/export/ (accessed 16 May 2018).

5. Buttigieg, P. L., Pafilis, E., Lewis, S. E., Schildhauer, M. P., Walls, R. L., Mungal, C. J., 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of Biomedical Semantics*, 7 (1), 57.

6. Guha, S., Rastogi, R., Shim, K., 2000. ROCK: A Robust Clustering Algorithm for Categorical Attributes. Information Science, 25 (5), 345-366.

Session S2.2

UNCERTAINTY AND BAYESIAN INFERENCE TECHNIQUES

Uncertainty Analysis and Bayesian Inference Techniques

George Arhonditsis[1]***, Alex Neumann***[1]***, Yuko Shimoda***[1]

[1]*University of Toronto, Canada*

Corresponding author(s) e-mail: georgea@utsc.utoronto.ca

ABSTRACT:

The credibility of the scientific methodology of environmental models and their adequacy to form the basis of public policy decisions have been frequently challenged. The current challenges make compelling the development of more realistic modelling platforms (i) to elucidate causal mechanisms, complex interrelationships, direct and indirect ecological paths; (ii) to examine the interactions among the various stressors (e.g., climate change, urbanization/land-use changes, alternative management practices, invasion of exotic organisms); and (iii) to assess their potential consequences on ecosystem functioning. The proposed session aims to provide insights into the current state of the field, and also highlight the major challenges and future directions of research. Special emphasis will be placed on studies that address topics, such as novel uncertainty analysis techniques, Bayesian inference methods (including Bayesian networks), development of new model formulations and proper representation of biotic functional types, emerging techniques of data assimilation and model optimization, effective integration of physics with biology, and strategies to improve the contribution of complex models to ecological theories. The proposed session encourages contributions from both mathematical and statistical ecosystem modellers.

KEYWORDS: Bayesian Inference, Uncertainty Analysis, Environmental Management, Policy Analysis

Predicting Ecological Responses to Climate Variability with a Dynamic Bayesian Network Model

***Neda Trifonova**[1, 3]*, ***Mandy Karnauskas**[2]*, ***Christopher Kelble**[3]*

[1]*University of Miami, USA*; [2]*Southeast Fisheries Science Centre, USA*; [3]*Atlantic Oceanographic and Meteorological Laboratory, USA*

Corresponding author(s) e-mail: neda.trifonova@noaa.gov

ABSTRACT:

The Gulf of Mexico is an ecologically and economically important marine ecosystem that is affected by a variety of natural and anthropogenic pressures. These complex and interacting pressures, together with the dynamic environment of the Gulf, present challenges for the effective management of its resources. The recent adoption of Bayesian networks to ecology allows for the discovery and quantification of complex interactions from data after making only a few assumptions about observations of the system. In this study, we apply Bayesian network models, with different levels of structural complexity and a varying number of hidden variables to account for uncertainty when modelling ecosystem dynamics. From these models, we predict focal ecosystem components within the Gulf of Mexico. The predictive ability of the models varied with their structure. The model that performed best was parameterized through data-driven learning techniques and accounted for multispecies associations and their interactions with human and natural pressures over time. Then, we altered sea surface temperature in the best performing model to explore the response of different variables to increased temperature. The magnitude and even direction of predicted responses varied by ecosystem components due to heterogeneity in driving factors and their spatial overlap. Our findings suggest that due to varying species sensitivity to drivers, changes in temperature will potentially lead to trade-offs in terms of population productivity. We were able to discover meaningful interactions between ecosystem components and their environment and show how sensitive these relationships are to climate perturbations, which increases our understanding of the potential future response of the system to increasing temperature. Our findings demonstrate that accounting for additional sources of variation, by incorporating multiple interactions and pressures in the model layout, has the potential for gaining deeper insights into the structure and dynamics of ecosystems.

KEYWORDS: Bayesian network, Gulf of Mexico, ecosystem dynamics, hidden variable, climate variability

Uncertainty assessment of scenarios on climate and land use changes for the Millbrook catchment - reservoir system simulated by the model ensemble SWAT-SALMO

***Hanh Hong Nguyen*[1]*, Friedrich Recknagel*[1]*, Wayne Meyer*[1]

[1] *Department of Ecology and Environmental Sciences, University of Adelaide, SA 5005*

Corresponding author(s) e-mail: hanh.nguyen@adelaide.edu.au

ABSTRACT:

In this study, we analyse the uncertainty of eutrophication effects of ongoing environmental and climate changes on the Millbrook reservoir simulated by the model ensemble SWAT-SALMO. The semi-arid Millbrook catchment-reservoir system provides drinking water to the north-eastern region of Adelaide, South Australia. The Soil and Water Assessment Tool (SWAT) simulated flow as well as nitrate and phosphate loadings originating from the catchment before entering the reservoir. The lake model SALMO received the simulated nitrate and phosphate loadings as input and determined daily phosphate, nitrate, and chlorophyll-a concentrations in the reservoir. This integrated modelling framework was key for simulating complex scenarios on impacts of future climate and land use changes on the whole catchment-reservoir system.

The uncertainty of simulation results has been taken into account by complex statistical algorithms, including the Sequential Uncertainty Fitting (SUFI2) of the SWAT calibration wizard, and multi-objective parameter optimisation of SALMO by means of the Hybrid Evolutionary Algorithm (HEA). In view of the large number of data processing steps required for the integrated simulations, the uncertainty assessment focused on the five best simulations results from the SWAT to be utilised for the parameter optimisation of SALMO.

The uncertainty of the model ensemble has been quantified as envelope of the fifty best iterations of nitrate, phosphate, and chlorophyll-a concentrations based on daily time steps for a typical “dry” and a typical “wet” year. The synergized envelop was further used to compare with the results of prediction of impacts of climate and land use changes on the Millbrook catchment - reservoir system. Overall, the estimation of uncertainty bound from the catchment-reservoir model ensemble may improve the credibility of the model predictions to be further considered in decision-making.

KEYWORDS: SWAT, SALMO, catchment-reservoir system, uncertainty, climate and land use change

What is a Prior and How to Find One?

Song S. Qian[1],

[1]*The University of Toledo, USA*

Corresponding author(s) e-mail: song.qian@utoledo.edu

ABSTRACT:

Application of Bayesian statistics requires eliciting prior distributions, an important first step that is often ignored. The difficulty in prior elicitation is largely due to the vague definition of the prior. Furthermore, formal methods for deriving priors are mostly focused on deriving priors with least amount of information (e.g., the reference prior). In practice, we often resort to a class of "non-informative" or "vague" priors when using relatively simple models. These priors are usually informative in some way and can lead to unintended consequences. In this presentation, I discuss the meaning of a prior distribution from an empirical Bayes perspective, which is the "centre of gravity" of similar (exchangeable) units. Based on this definition, I present a Bayesian network based method to derive prior distributions for relatively complex models. The method borrows the Bayesian network model approach of using a directed acyclic graph to summarize our knowledge on the subject of interest and extends the Bayesian network to accommodate continuous variables. Continuous variables can enter the network through empirical models based on exploratory data analysis through existing models. The continuous variable Bayesian network modelling approach is illustrated using three examples – a model for evaluating the risk of Cryptosporidium contamination in US drinking water systems, model-based nutrient criteria for small rivers and streams in Ohio, and assessing water availability to meet the use of both societal and ecological needs in the southeastern US.

KEYWORDS: Bayesian network, empirical Bayes, exchangeability, hierarchical model

Uncertainty Analysis by Bayesian Inference

George Arhonditsis[1]

[1]*University of Toronto, Canada*

Corresponding author(s) e-mail: georgea@utsc.utoronto.ca

ABSTRACT:

The scientific methodology of mathematical models and their credibility to form the basis of public policy decisions have been frequently challenged. The development of novel methods for rigorously assessing the uncertainty underlying model predictions is one of the priorities of the modelling community [1]. Striving for novel uncertainty analysis tools, I present the Bayesian calibration of process-based models as a methodological advancement that warrants consideration in ecosystem analysis and biogeochemical research [2]. This modelling framework combines the advantageous features of both process-based and statistical approaches; that is, mechanistic understanding that remains within the bounds of data-based parameter estimation. The incorporation of mechanism improves the confidence in predictions made for a variety of conditions, whereas the statistical methods provide an empirical basis for parameter value selection and allow for realistic estimates of predictive uncertainty [3]. Other advantages of the Bayesian approach include the ability to sequentially update beliefs as new knowledge is available, the rigorous assessment of the expected consequences of different management actions, the optimization of the sampling design of monitoring programs, and the consistency with the scientific process of progressive learning and the policy practice of adaptive management. I illustrate some of the anticipated benefits from the Bayesian calibration framework, well suited for stakeholders and policy makers when making environmental management decisions, using the Hamilton Harbour and the Bay of Quinte– two eutrophic systems in Ontario, Canada – as case studies [4].

KEYWORDS: Bayesian Inference, Mechanistic Modelling, Uncertainty Analysis, Risk Assessment, Adaptive Management Implementation

REFERENCES:

1. Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modelling. Mar. Ecol. Prog. Ser. 271, 13-26.

2. Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. Ecol Model. 208, 215-229.

3. Arhonditsis G.B., Kim, D-K., Kelly, N., Neumann, A., Javed, A., 2017. Uncertainty Analysis by Bayesian Inference, in Recknagel, F., Michener, W., (Eds), Ecological Informatics. 3rd Edition Springer., Cham, pp. 215-249.

4. Recknagel F., Arhonditsis, G.B., Kim, D-K ., Nguyen H.H., 2017. Strategic Forecasting in Ecology by Inferential and Process-based Models. in Recknagel, F., Michener, W., (Eds), Ecological Informatics. 3rd Edition Springer., Cham, pp. 341-372.

Beyond Allopatric Speciation: Testing for Genetic Homogeneity in *Duttaphrynus melanostictus* in Relation to Human-induced Dispersal

***Siti N. Othman*[1], *Yi-Huey Chen*[2], *Desiree Andersen*[1], *Ming-Feng Chuang*[3], *Yikweon Jang*[1], *Amaël Borzée*[1]**

[1]Department of Life Sciences and Division of EcoScience, Ewha Womans University, Seoul, Republic of Korea; [2] Department of Life Science, Chinese Culture University, Taipei, Taiwan; [3] Department of Life Science, Tunghai University, 1727 Taiwan Boulevard, Taichung, Taiwan

Corresponding author(s) e-mail: amaelborzee@gmail.com

ABSTRACT:

Human-induced dispersal of species is accelerating along with the increase in human movements. This unnatural dispersal contributes to range expansions, such as for the Asian black-spined toad, Duttaphrynus melanostictus. The species became invasive in numerous ecosystems, such as Madagascar and the Komodo islands area, where it has a strong ecological impact. Here, we investigated the origin of D. melanostictus in Taiwan using statistical parsimony networks together with joint Bayesian inference of phylogeny and population clustering approaches. We tested four major dispersal hypotheses to explain the relation between the Taiwanese and other D. melanostictus populations: 1) the species originates from South East Asia (SEA) and it is invasive in Taiwan as a result of human-induced dispersal, 2) the species originates from South East Asia, and dispersed over land bridges, 3) the species comes from the Chinese mainland through human-induced dispersal, 4) the species originates from the Chinese mainland, and dispersed over land bridges during glacial maxima. Our unrooted haplotype network based on a fragment of contiguous tRNA Gly-ND3 mitochondrial DNA from 22 individuals, together with homologous sequences extracted from GenBank for China and SEA, confirmed the non-clustering of haplotypes from SEA and Taiwan. Likewise, Bayesian phylogenetic inferences further clarified the absence of genetic segregation between Taiwanese and mainland Chinese populations. The haplotypes from Taiwan were segregated within a monophyletic clade, shared with mainland Chinese populations. The origin of the Taiwanese D. melanostictus is consequently linked to the Chinese clade, and rejects our first and second hypothesis. The clustering with haplotypes from geographically close localities in mainland China, as well as with haplotypes present in geographically distant but active trading areas, here Hong-Kong, suggests both past dispersion over land-bridges, and recent human-induced dispersal. Interestingly, we also found four haplotypes from the SE Asian clade clustering with the mainland Chinese clade, and two mainland Chinese haplotypes clustered within the SE Asian clade. This is another potential indicator of additional human-induced dispersal events. Past dispersal over land bridges reflects the highly vagrant behaviour of the species, followed by inadvertent translocations, favoured by the resilience of the species in dry environments. Our study emphasises the risk of future hybridisation between Taiwanese, SE Asian and mainland Chinese clades as a result of human activities.

KEYWORDS: Human-induced dispersal, Bayesian inference, Phylogenetic analysis, *Duttaphrynus melanostictus*, Invasive species

Integrating Hierarchical Bayes with Limnological Modelling

***Yuko Shimoda* and George B. Arhonditsis**

University of Toronto Scarborough, Canada

Corresponding author(s) e-mail: yshimoda@utsc.utoronto.ca

ABSTRACT:

The tendency to invoke complexity as a means for improving the learning capacity of our models is primarily prompted by the need to make complex environmental policy decisions. Problems related to eutrophication and lake management are often supported by complex ecosystem models with limited ability to properly constrain their inputs by available empirical knowledge and data. Thus, the resulting poor identifiability undermines their credibility as management tools. In this regard, our study argues in favour of simple models that remain within the bounds of empirical parameter estimation. We also examine the potential of Bayesian hierarchical inference to offer a convenient means for addressing a variety of limnological modelling problems in which partial, but not complete, commonality can be assumed among the modelled units. The Bayesian hierarchical proposition is a compromise between site-specific (where limited local data is a problem) and globally-common (where heterogeneous systems in wide geographical areas are assumed to be identical) models. Under the hierarchical structure, the models are dissected into levels (hierarchies) that explicitly account for the role of significant sources of variability, e.g., geographical location, trophic status, morphometry, mixing regime, or even watershed land use patterns. Our study presents two case studies that use a Bayesian hierarchical strategy; a spatially-explicit simple TP mass-balance model forced with idealized sinusoidal loading, and a multivariate regression model that aims to elucidate the importance of different abiotic factors (light attenuation, water temperature, phosphorus, and ammonium) on cyanobacteria biovolume in the Bay of Quinte, Lake Ontario, Canada.

KEYWORDS: Hierarchical Bayes, Eutrophication, Cyanobacteria, Lake management

Predicting the spatial and temporal dynamics of hypoxia in Hamilton Harbour, Ontario, Canada: A Bayesian modelling framework

***Dong-Kyun Kim*[1], *George B. Arhonditsis*[1]**

[1]*Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, M1C 1A4, Canada*

Corresponding author(s) e-mail: dkkim1004@gmail.com

ABSTRACT:

Our study focuses on the development of a hypoxia model founded upon a Bayesian framework. The objectives of our study are to depict the spatiotemporal dynamics of dissolved oxygen concentrations as well as to rigorously assess the degree of hypoxia under different nutrient loading scenarios in the Hamilton Harbour Area of Concern, Ontario, Canada. Our model postulates two basic processes as causal factors of hypoxia: (i) the stratification driven by the vertical water temperature gradient, and (ii) the sediment oxygen demand (SOD) as predicted by the areal carbon loading. Based on these fundamental mechanisms, we constructed three statistical formulations to predict daily dissolved oxygen (DO) concentrations for five stations. Using Bayesian Kriging, the DO predictions were projected spatially over the entire system during the summer period. Our results showed that hypoxia (i.e., minimum daily DO <2 mg L-1) manifests itself in mid-June and dissipates by late September. In particular, we demonstrate that the hypoxic areal extent can account for up to 93% of the entire harbour (or 43% of the transitional zone) in early August under the present nutrient loading conditions. In contrast, under a 30% carbon loading reduction, we found that an approximately 10% increase of DO concentrations is achievable across the entire harbor and that the hypoxic areal extent may decrease to 21-47%. Finally, we discuss the implications of hypoxia for fish population integrity, and shed light on the likelihood of alleviating the current stress through the on-going remedial efforts in Hamilton Harbour.

KEYWORDS: Bayesian Kriging, dissolved oxygen, Hamilton Harbour, hypoxic areal extent, loading reduction

Overview of Bayesian inference techniques for conceptual semi-empirical watershed models (SPARROW, GREEN)

***Alex Neumann*[1], *Dong-Kyun Kim*[1], *Feifei Dong*[1], *George Arhonditsis*[1]**

[1]*University of Toronto, Canada*

Corresponding author(s) e-mail: george.arhonditsis@utoronto.ca

ABSTRACT:

The parsimonious model structures of semi-empirical conceptual watershed models (e.g., SPARROW, GREEN) offer considerable advantages over physically-based watershed models to incorporate stream water monitoring data and effectively accommodate rigorous error analysis. Nonetheless, even these model structures demonstrate intrinsic equifinality problems due to multicollinearity of model parameters. Bayesian inference techniques offer a robust and formal statistical calibration methodology to address model equifinality issues with watershed inverse analysis. In our presentation, we first summarise known case-studies of Bayesian inference implementation for semi-empirical watershed models (USA, Canada, China). We then provide an overview of relevant Bayesian statistical formulations to explicitly consider watershed spatial heterogeneity, serial correlation, and inter- and intra-annual dynamics. Finally, we outline the strengths and weaknesses of inverse watershed models within a Bayesian inference context for recursive calibration and data assimilation, including hotspot identification, loading source apportionment, representation of legacy nutrients, and quantification of all major sources of uncertainty.

KEYWORDS: SPARROW, GREEN, watershed, source apportionment

Bayesian belief networks for integrated ecological modelling to assess communities and ecosystem services

***Marie Anne Eurie Forio*[1]*, Wout Van Echelpoel*[1]*, Niels De Troyer*[1]*, Arne Deknock*[1]*, Luis Dominguez*[2] *and Peter Goethals*[1]**

[1]*Ghent University, Belgium;* [2]*ESPOL, Ecuador*

Corresponding author(s) email: peter.goethals@ugent.be

ABSTRACT:

Community traits are interesting to investigate how rivers function differently as a result of environmental disturbance. However, only few studies made a quantitative investigation between trait combinations in function of environmental conditions. In this study, macroinvertebrate traits were linked to environmental river conditions in the Guayas basin via the use of BBN models. This study is an important step towards more quantitative insights in river functioning resulting from alterations in land use, installation of hydropower dams and the increase of sewage from households, and can serve as a basis for an improved environmental management in this key basin of Ecuador.

Modeling a Species Identification Process as a Bayesian Inference Problem

***Oliver Bley*[1], *Patrick Mäder*[1]**

[1]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: oliver.bley@tu-ilmenau.de

ABSTRACT:

Plant species identification typically involves a key that guides the user through a series of questions to the unknown species. Per question the reader makes a decision and is being redirected in the tree-like key structure to the next question until eventually a species is uniquely be identified. Every step in this process essentially works as filter on the whole species set. Therefore, plant identification keys are mainly targeted towards experts, where even one wrong answer throughout the identification process will not lead to the correct species.

In this talk we will discuss how this key-based approach can be reformulated as a retrieval problem using Bayesian methods. We are proposing an interactive process, that presents a dynamic sequence of questions to the user, which depend on the given answers but without discarding species that do not correspond to already given answers. The approach additionally allows multiple and no choice per question, if the user is uncertain or unable to answer.

The proposed online process is guided by an expert knowledge graph and a catalogue of possible questions that the algorithm can arbitrarily choose from. We discus which measures work best in utilizing prior-knowledge to find the species the user is looking for. Another focus of the talk will be strategies for determining the next question. We will also have an in-depth look at the sources of errors that occur in a plant identification process and will discuss the dimensions of uncertainty the algorithm has to cope with.

KEYWORDS: Bayesian Network, Species Identification, Identification Key, Question Strategy Optimization

Evaluating explanations of land-cover change in the Iberian agricultural revolution using approximate Bayesian computation

***Andrew Lane*[1], *James Millington*[2], *Simon Miles*[3]**

[1, 2, 3] *King's College London, UK*

Corresponding author(s) e-mail: andrew.lane@kcl.ac.uk

ABSTRACT:

The ecology of the Mediterranean region today is the product of thousands of years of human intervention, combined with the effects of climatic change on its flora, and natural disturbance arising from its active fire regime. Understanding the complex ways in which anthropogenic land-cover change, climate, and fire interact with each other could help guide public policy and land management decision-making, especially important to ensure sustainability in an uncertain future of climate change. A fruitful step in this endeavour will be to explain changes to landscape-scale vegetation ecology arising from the time when humans first began to compete with fire as a source of ecological disturbance 10,000 years ago – the agricultural revolution.

Paleoecological data provide a means to study ecosystem change in the distant past, and over timescales longer than a human lifetime. In particular, analysis of how the abundance of pollen produced by different plant functional types has changed over time can tell us much about the evolution of the composition of those functional types in the land-cover around a study site. Pollen time series can be viewed as fingerprints of the interacting processes of anthropogenic land-cover change, fire and ecological succession, against the backdrop of a changing climate. However, no amount of scrutiny of the data alone can elucidate the causal role each of the interacting factors had in creating them.

We have developed a model which integrates theoretical understanding of ecological and anthropogenic processes via an agent-based modelling framework. By simulating both natural and anthropogenic processes and allowing them to interact with each other in a spatially explicit virtual landscape, we can generate dynamic pictures of landscapes evolving in time which are consistent with socio-ecological theory. However, as scientists we must be vigilant in our reporting of uncertainty about our models. Historical sciences provide a unique challenge in this regard, since theories cannot be tested by direct experimentation.

In this talk, I will discuss the use of approximate Bayesian computation (ABC) as a means to unite acausal paleoecological data with a socio-ecological simulation model which encodes causal hypotheses about Neolithic land-cover change, but whose likeness to reality is uncertain. ABC will be considered as a tool to support inference about which models best explain available empirical data, as well as indicating which aspects of those models are most uncertain.

KEYWORDS: Mediterranean, Fire, approximate Bayesian computation, Paleoecology, agent-based modelling

Session S2.3

QUANTIFYING THE FUNCTIONS IN TERRESTRIAL ECOSYSTEMS: FROM CONCEPTS TO DATA DRIVEN METHODS

Quantifying the Functions in Terrestrial Ecosystems: From Concepts to Data Driven Methods

***Talie Musavi*[1]*, Xuanlong Ma*[1]*, Miguel Mahecha*[1,2]*, Mirco Migliavacca*[1]**

[1] *Max-Planck Institute for Biogeochemistry, Germany*; [2] *German Centre for Integrative Biodiversity Research (iDiv), Germany*

Corresponding authors e-mail: tmusavi@bgc-jena.mpg.de, xma@bgc-jena.mpg.de

ABSTRACT:

One argument for nature conservation efforts is preserving e.g. ecosystem services such as carbon sequestration potential, water and nutrient retention, among many others. To quantify such processes, we not only need a clear conceptual understanding of ecosystem functioning but also novel ways to quantify them and understand controlling factors at various spatial and temporal scales.

In this session we invite contributions on conceptual advances to define and identify “Ecosystem Functioning”. We also aim to discuss latest developments in observing ecosystem functions from in-situ data or with proximal or remote sensing from the site level to the global scale. Innovative advances on nonlinear statistical methods, model-data integration, or inversion studies that help us to constrain ecosystem functioning or to retrieve functional properties of the ecosystems (e.g. radiation use efficiency, nitrogen use efficiency, water use efficiency and so forth) – even if in early stages of development are also encouraged to participate as we strive for a broad session. The goal of our session is to provide a stage for those interested in understanding the ecosystem functioning combining diverse data streams or analytic frameworks.

KEYWORDS: Ecosystem function, Ecosystem processes, flux, remote sensing, model, data, site level, global scale

Sensitivity of potential gross primary productivity to climate drivers

Daniel E. Pabon-Moreno*[1], *Talie Musavi*[2, 3], *Mirco Migliavacca*[3], *Miguel D. Mahecha

[1]*Max Planck Institute for Biogeochemistry, Germany*

Corresponding author(s) e-mail: dpabonw@gmail.com

ABSTRACT:

“Ecosystem Functional Properties” (EFPs) are thought to be static quantities that can be used to characterize ecosystem processes and predict their response to environmental drivers. They reflect the interaction between the biotic and abiotic components of the ecosystems. One example is the “Maximal rate of Gross Primary Productivity” during the growing season (GPPmax). This metric is very useful to characterize an ecosystems CO2 uptake capacity via photosynthesis. While instantaneous GPP measurements at ecosystem scale covary with meteorological conditions, the GPPmax may only be affected by long-term climatic conditions to which the ecosystem is adapted. Here we present a global study for (1) understanding the regional patterns of GPPmax in relation to different climatic drivers (2) understanding the sensitivity of timing of GPPmax. The results could be used to improve our perception of the ecosystem changes under future global climate change as well as identifying areas where photosynthetic capacity dynamics is controlled by temperature or water availability.

KEYWORDS: Ecosystem Functional Properties, Potential Gross Primary Production, Climate drivers, Climate Change

What controls global fire activity? Evaluating emergent responses in satellite observations and ecosystem models using machine learning

Matthias Forkel[1], Sandy P. Harrison[2], Niels Andela[3], Gitta Lasslop[4], Margreet van Marle[5], Wouter Dorigo[1], Matthew Forrest[4], Stijn Hantson[6], Angelika Heil[7], Fang Li[8], Stephane Mangeon[9], Joe Melton[10], Chao Yue[11], Almut Arneth[12]

[1]TU Wien, Austria; [2]University of Reading, UK; [3]NASA Goddard Space Flight Center, USA; [4]Senckenberg Biodiversity and Climate Research Institute, Germany; [5]Deltares, The Netherlands; [6]University of California Irvine, USA; [7]Max Planck Institute for Chemistry, Germany; [8]International Center for Climate and Environmental Sciences, China; [9]Imperial College London, UK; [10]Environment Canada Victoria, Canada; [11]Laboratoire des Sciences du Climat et de l'Environnement, France; [12]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Germany

Corresponding author e-mail: matthias.forkel@geo.tuwien.ac.at

ABSTRACT:

Fire is a major disturbance agent in terrestrial ecosystems. The occurrence and spread of wildfires is controlled by the interplay of human activities, weather conditions, and the conditions of vegetation and litter fuels. Most state-of-the-art global ecosystem models represent such controls to simulate fire effects on vegetation dynamics and global carbon cycling. However, global fire models poorly reproduce the observed dynamics and variability of fire burned area. Here we aim to identify and evaluate functional responses of global burned area to environmental and human controls.

We use several global satellite, climate, and socioeconomic datasets, and simulations from the Fire Model Inter-comparison Project (FireMIP) [1] to predict the observed or modelled burned area with the random forest machine-learning algorithm. We then derive from the trained random forests individual conditional expectation curves [2], which represent emergent functional responses of burned area to controlling factors. These functional responses allow us to compare data- and model-derived sensitivities.

FireMIP models mostly represent the emergent responses to climate variables but show diverse responses to human population, land cover, and vegetation. The models especially underestimate the emergent strong increase of burned area with increasing precedent plant productivity in many semi-arid ecosystems. The results suggest that FireMIP models misrepresent the links between plant productivity, biomass allocation, litter turnover, and fuel production. Additionally, the good performance of data-driven modelling approaches [3] suggests to develop hybrid global fire models to better represent and predict the role of fire dynamics for ecosystem functioning and vegetation-climate interactions.

KEYWORDS: burned area, remote sensing, vegetation-fire models, random forest, individual conditional expectations

REFERENCES:

1. Rabin, S.S., Melton, J.R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J.O., Li, F., Mangeon, S., Ward, D.S., Yue, C., Arora, V.K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G.A., Sheehan, T., Voulgarakis, A., Kelley, D.I., Prentice, I.C., Sitch, S., Harrison, S., Arneth, A., 2017. The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions. Geosci Model Dev 10, 1175–1197. https://doi.org/10.5194/gmd-10-1175-2017

2. Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2013. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. ArXiv13096392 Stat.

3. Forkel, M., Dorigo, W., Lasslop, G., Teubner, I., Chuvieco, E., Thonicke, K., 2017. A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). Geosci Model Dev 10, 4443–4476. https://doi.org/10.5194/gmd-10-4443-2017

Insight into hydrochemistry: a multi-catchment comparison using Horizontal Visibility Graphs

***Eva Klotzbücher*[1]*, Holger Lange*[2]*, Sebastian Sippel*[2, 3]*, Michael Hauhs*[1]*, Britta Aufgebauer*[1]**

[1]*University of Bayreuth, Germany*; [2]*Norwegian Institute of Bioeconomy Research, Norway*; [3]*ETH Zürich, Switzerland*

Corresponding author(s) e-mail: holger.lange@nibio.no

ABSTRACT:

Long time series of environmental variables are reflecting the dynamics and functioning of ecosystems. Here, we investigate data from a long-term monitoring site in Germany, the Bramke valley in the Harz mountains, where time series of ion concentrations in stream water are obtained since the 1970ies at four measurement locations from three small adjacent forested catchments. Since for (only) one of the catchments daily runoff rates are also available, we invent a method to generate time series of nutrient output from the catchments. Both concentrations and outputs show a number of remarkable long-term changes, including ones not obviously related to changes in atmospheric deposition, management or properties of the forest stands.

For the analysis of the Bramke data, we investigate Horizontal Visibility Graphs (HVGs), a recently developed method to construct networks based on time series. Values (the nodes of the network) of the time series are linked to each other if there is no value higher between them. The network properties, such as the degree and distance distributions, reflect the nonlinear dynamics of the time series. For certain classes of stochastic processes and for periodic time series, analytic results can be obtained for some network properties. HVGs have the potential to discern between deterministic-chaotic and correlated-stochastic time series.

We classify the Bramke series according to their stochastic nature, with a focus on inter-catchment comparison on one hand, on different nutrients for one catchment on the other, and conclude on possible reasons for the observed changes and their ecological interpretation.

KEYWORDS: Horizontal Visibility Graphs (HVGs), network analysis, time series analysis, nonlinear dynamics, catchment similarity, catchment comparison, degree distribution, long-range correlations

Nutrient induced changes in Sun-Induced Fluorescence emission in a Mediterranean grassland

***David Martini*[1], *Javier Pacheco-Labrador*[1], *Oscar Perez-Priego*[1], *Christiaan van der Tol*[2], *Tarek El Madany*[1], *Tommaso Julitta*[3], *Micol Rossini*[4], *Anatoly Gitelson*[5], *Markus Reichstein*[1] and *Mirco Migliavacca*[1]**

[1]*Max Planck Institute for Biogeochemistry, Jena, Germany*; [2]*Twente University, Enschede, Netherlands*;
[3]*JB hyperspectral devices, Düsseldorf, Germany*; [4]*University of Milano Bicocca, Milan, Italy;*
[5]*University of Nebraska - Lincoln, Lincoln, USA*

Corresponding author e-mail: dmartini@bgc-jena.mpg.de

ABSTRACT:

Sun induced fluorescence (SIF), the radiation flux emitted by plant chlorophylls molecules in the 650-800 nm spectral window, is considered an indicator of photosynthetic performance. Recently it has been shown that SIF can track changes in light use efficiency (LUE), and therefore it is a good predictor of gross primary production (GPP) at various scales, from leaves and ecosystem to regional and global scale.

Although SIF has been successfully used to predict GPP in various ecosystems, the mechanistic link between GPP and SIF remains not fully understood, and especially the effect of function and structure on SIF at the canopy scale remains an active area of research.

SIF is emitted by the whole canopy, but only a fraction of the total emission is observed with remote sensing techniques. The escape probability of SIF (Fesc) controls the amount of SIF scattered by the canopy and is integral to separate the effect of canopy structure and function on the fluorescence signal.

In this contribution we make use of data collected at the research site Majadas del Tietar, a Mediterranean grassland manipulated with Nitrogen and Phosphorus. Using the SCOPE model (Soil Canopy Observation Photochemistry and Energy fluxes) we obtain Fesc and we analyse how Top of canopy SIF and emitted SIF vary in response to the fertilization.

With a combination of processes-based modelling and data driven analysis, such as relative importance analysis and structural equation modelling, we unravel the processes and causal relationship that are at the base of the GPP - SIF relationship.

We show that the nutrient fertilization had an effect on plant composition, and therefore canopy structure, but also plant functioning. Nitrogen induced changes in biodiversity mainly affect leaf angle distribution of the canopy and therefore scattering properties such as Fesc. The nitrogen fertilization is also responsible for a change in plant functioning, with altered SIF emission.

The simultaneous change of both canopy and structure causes the fertilization effect to be visible mainly at the emission level, but not at top of canopy, as the variation in canopy structure masks the change observed at leaf level.

This contribution advances the knowledge of the highly complex dynamics involved in the GPP-SIF relationship. In depth understanding of the mechanistic processes is required to fully take advantage of the increasingly prevalent SIF data streams.

KEYWORDS: Sun-Induced Fluorescence, Fesc, Nutrient Fertilization

Analyzing the time variant causality in ecological time series: a time-frequency approach

***Maha Shadaydeh*[1]*, Yanira Guanche Garcia*[1,2]*, Miguel Mahecha*[2,3]*, Markus Reichstein*[2,3]*, and Joachim Denzler*[1,2]**

[1]*FSU Jena, Germany;* [2]*Michael Stifel Center Jena for Data-driven and Simulation Science, Jena, Germany;* [3]*Max Planck Institute for Biogeochemistry, Jena, Germany.*

Corresponding author e-mail: maha.shadaydeh@uni-jena.de

ABSTRACT:

Attribution in ecosystems aims to identify the cause-effect relationships between the variables involved. The availability of high temporal resolution data along with the powerful computing platforms further enhance the capacity of data-driven methods in capturing the complex relationships between the variables of the underlying system. Time series of ecological variables most often contain different periodical components that can significantly mask the underling causality structure in time domain. This motivates the use of time-frequency processing techniques such as wavelet analysis or short time Fourier transform. In this study we present a time-frequency approach for causality analysis where the coupling between the variables is assumed to follow a locally time-invariant multivariate autoregressive (MVAR) model. We propose a sliding time window approach to examine the change of interactions, i.e. direction and strength of causality, between the different variables over seasons. The cause-effect relationships are extracted using the frequency domain representation of the MVAR Granger causality (MVAR-GC) [1,2] based on the generalized partial directed coherence (gPDC) [3]. We have first applied the proposed method to synthetic data to evaluate its sensitivity to different issues such as the selection of the model order, the sampling frequency, the absence of cause as well as the presence of non-linear coupling. The method is then applied to half-hourly meteorological observations and land flux eddy covariance data to investigate the causal-effect relationships between global radiation (Rg), air temperature (Tair), and the CO_2 land fluxes: gross primary productivity (GPP), net ecosystem exchange (NEE) and ecosystem respiration (Reco). The results show that time-frequency analysis based on MVAR-GC has promising potential in identifying the time variant causality structure within these variables along with the main time delay between different cause-effect pairs. Further research work is currently going for the investigation of the selection criteria of the model order, the sampling frequency, and the size of the time window at different time scales of causality analysis. This study is carried out within the framework of the project BACI which in part aims at developing an attribution scheme for changes in ecosystem functioning and studying the impacts of these changes on biodiversity patterns.

KEYWORDS: Time-frequency analysis, MVAR Granger causality, Partial directed coherence, Ecological time series.

REFERENCES:

1. *Granger, C. W. J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica. 37 (3): pp. 424–438.*

2. *Geweke J., 1982. Measurement of linear dependence and feedback between multiple time series. J Am Stat Assoc; 77:304–313.*

3. Baccala L. A., K. Sameshima and D. Y. Takahashi, 2007. Generalized Partial Directed Coherence. *15th International Conference on Digital Signal Processing*, Cardiff, 2007, pp. 163-166.

Early Indicators of High Impact of an Invasive Ecosystem Engineer on Ecosystem Functioning From Leaf to Landscape Scale

***André Große-Stoltenberg*[1], *Christine Hellmann*[2], *Jan Thiele*[3], *Christiane Werner*[2], *Jens Oldeland*[4]**

[1]*WWU Münster, Germany*; [2]*ALU Freiburg, Germany*; [3]*Thünen Institute of Biodiversity Braunschweig, Germany;* [4]*University of Hamburg, Germany*

Corresponding author(s) e-mail: ags@uni-muenster.de

ABSTRACT:

Invasive ecosystem engineers, such as the N-fixing tree *Acacia longifolia*, are a major threat to ecosystem functioning across the globe. The local impact of *A. longifolia* on ecosystem structure and functioning in Mediterranean dunes has been well characterized by in-situ measurements, e.g. on water and N cycling. However, novel approaches are required for early detection of its impact at larger spatial scales. Therefore, our objective was to assess the impact of the invader on ecosystem functioning from the leaf to the landscape level applying sensor-based methods. To achieve this aim, we focused on three research questions: Can contrasts in leaf traits (e.g. leaf N content) between the invader and native species be retrieved from hyperspectral data? Can the invader's spatial impact on N cycling be mapped at stand level using functional tracers and remote sensing? Finally, how can *A. longifolia*'s alterations of ecosystem structure and functioning be tracked at landscape scale?

First, leaf traits differed between *A. longifolia* and the native species, especially regarding leaf N content [1]. This trait dissimilarity can be an early warning sign for invaders with a significant impact on N cycling. It can be derived from hyperspectral data at both leaf and canopy scale. Therefore, there is potential for mapping.

Second, we traced the invader's impact on N cycling at the stand scale [2]. For this purpose, we combined spatial data on the distribution of a functional tracer of N-fixation, $\delta^{15}N$, with geospatial data on environmental heterogeneity derived from airborne LiDAR. The values of foliar $\delta^{15}N$ of the non-fixing, native shrub *Corema album* are naturally quite low in this ecosystem. However, foliar $\delta^{15}N$ of *C. album* clearly increased for shrubs growing with a margin of 5 – 8 m around *A. longifolia* stands. This indicated an uptake of N previously fixed by the invader. Adding LiDAR metrics to the spatial prediction model enabled mapping of foliar $\delta^{15}N$ of *C. album*.

Third, *A. longifolia* was detected at landscape level by integrating airborne hyperspectral imagery with LiDAR data [3]. Gross Primary Production (GPP) increased significantly after invasion even at early invasion stages when *A. longifolia* cover was below 10%, which indicated a regime shift from a dune to a forest-type ecosystem. Thus, early warning signs of high impact caused by invasive ecosystem engineers on ecosystem functioning can be retrieved from remote sensing data across spatial scales. This offers promising possibilities for monitoring high impact invasive plant species in threatened ecosystems.

KEYWORDS: Hyperspectral, LiDAR, Invasive Species, Acacia, Nitrogen, Stable Isotopes, GPP

REFERENCES:

1. Große-Stoltenberg, A., Hellmann C., Thiele, J., Oldeland, J., Werner C., 2018. Invasive acacias differ from native dune species in the hyperspectral/biochemical trait space. J. Veg. Sci. 29, 325-335.
2. Hellmann, C., Große-Stoltenberg, A., Thiele, J., Oldeland, J., Werner C., 2017. Heterogeneous environments shape invader impacts: integrating environmental, structural and functional effects by isoscapes and remote sensing. Sci. Rep. 7, 4118.
3. Große-Stoltenberg, A., Hellmann, C., Thiele, J., Werner C., Oldeland, J., 2018. Early detection of GPP-related regime shifts after plant invasion by integrating imaging spectroscopy with airborne LiDAR. Remote Sens. Environ. 209, 780-792.

Representativeness-Based Sampling Network Design for the Arctic

Forrest M. Hoffman[1,2], Jitendra Kumar[1,2], William W. Hargrove[3], Martijn Pallandt[4], Mathias Goeckedei[4]

[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*; [2]*University of Tennessee, Knoxville, TN, USA*;
[3]*U.S. Department of Agriculture Forest Service, Asheville, NC, USA*;
[4]*Max Planck Institute for Biogeochemistry, Jena, Germany*

Corresponding author(s) e-mail: forrest@climatemodeling.org, jkumar@climatemodeling.org, hnw@geobabble.org, mpall@bgc-jena.mpg.de, mgoeck@bgc-jena.mpg.de

ABSTRACT:

Resource and logistical constraints limit the frequency and extent of environmental observations, particularly in the Arctic, necessitating the development of a systematic sampling strategy to maximize coverage and objectively represent environmental variability at desired scales. Required is a quantitative methodology for stratifying sampling domains, informing site selection, and determining the representativeness of measurement sites and networks. Multivariate spatiotemporal clustering was applied to down-scaled general circulation model results and data for the State of Alaska at 2 km × 2 km resolution to define multiple sets of bioclimatic ecoregions across two decadal time periods. Maps of ecoregions for the present (2000–2009) and future (2090–2099) were produced, showing how combinations of 37 bioclimatic and permafrost characteristics are distributed and how they may shift in the future. Representative sampling locations are identified on present and future ecoregion maps. A representativeness metric was developed, and representativeness maps for eight candidate sampling locations were produced. This metric was used to characterize the environmental similarity of each site. This analysis provides model-inspired insights into optimal sampling strategies, offers a framework for up-scaling measurements, and provides a down-scaling approach for integration of models and measurements. These techniques can be applied at different spatial and temporal scales to meet the needs of individual measurement campaigns. More recently, we have extended this approach to investigate pan-Arctic and tropical forest representativeness, employing remote sensing and other data products, to quantify coverage of spatial heterogeneity from international monitoring and sampling efforts. New results describing global forest site constituency and Arctic sampling regimes will be presented.

KEYWORDS: spatiotemporal, network analysis, representativeness, ecoregions

Session S2.4

DEEP LEARNING FOR ENVIRONMENTAL SCIENCE & ECOLOGY

Deep Learning for Environmental Science & Ecology

***Christian Requena Mesa*[1, 2], *Basil Kraft*[2], Marco Körner[3], Joachim Denzler[1], Markus Reichstein[2]**

[1]*FSU Jena, Germany*; [2]*Max Planck Institute for Biogeochemistry, Germany*; [3]*TU Munich, Germany*

Corresponding author(s) e-mail: crequ@bgc-jena.mpg.de; bkraft@bgc-jena.mpg.de

ABSTRACT:

Deep learning is an extremely active research area in machine learning and pattern recognition communities. It has gained huge success in areas such as speech recognition, computer vision and natural language processing. Applications of deep learning in geosciences and ecology like extracting knowledge from big-data, short-term forecasting or anomaly detection are promising, in particular since it can deal very well with space-time structures.

In this session, we invite contributions on the use of deep learning in ecology and environmental science. A series of oral presentations builds the basis to discuss the latest developments in deep learning for insights into and prediction of ecological systems. We welcome contributions covering all aspects of ecology and environmental science, including biodiversity, climate impact, ecosystem and organismal ecology, biogeography and others.

KEYWORDS: Machine learning, Deep learning, Big-data, Environmental science, Ecology

Extracting Trait Data from Digitized Herbarium Specimens Using Deep Convolutional Networks

Sohaib Younis[1,2]***, Marco Schmidt***[1,3]***, Claus Weiland***[1]***, Stefan Dressler***[4]***, Susanne Tautenhahn***[5]***, Jens Kattge***[5]***, Bernhard Seeger***[2]***, Thomas Hickler***[1,6]***, Robert Hoehndorf***[7]

[1]*SBiK-F, Germany*; [2]*University of Marburg, Germany*; [3]*Palmengarten, Germany*; [4]*SRI-NHM, Germany;* [5]*MPI-BGC, Germany*; [6]*Goethe University, Germany*; [7]*KAUST, Saudi Arabia*

Corresponding author e-mail: Muhammad-Sohaib.Younis@senckenberg.de

ABSTRACT:

Herbarium collections have been the foundation of taxonomical research for centuries and become increasingly important for related fields such as plant ecology or biogeography. Herbaria worldwide are estimated to include c. 400 million specimens, by inclusion of type specimens cover with few exceptions all known plant taxa (c. 350 000 species) and have a temporal dimension that is reached by only few other botanical data sources.

Presently, c. 13.5 million digitized herbarium specimens are available online via institutional websites or aggregating websites like GBIF. We used these specimen images in combination with morphological trait data obtained from TRY and the FLOPO knowledge base in order to train deep convolutional networks to recognize these traits as well as phenological states from specimen images. To improve trait recognition, we expanded our approach to include high resolution scans to enable fine grain feature extraction. Furthermore we analyze differences in recognizability of traits depending on trait group (e.g. leaf traits) or higher taxa. Newly mobilized trait data will be used to improve our trait databases. Our approach is described in detail and performance in the recognition of different traits is analyzed and discussed.

KEYWORDS: Trait Recognition, Deep Convolutional Neural Network, Plant Phenotyping, Digitized Natural History Collections, Image Processing

Habitat-Net: Habitat Interpretation Using Deep Neural Nets

Anand Vashishtha[1], **Jesse F. Abrams**[2], **Azlan Mohamed**[2], **Andreas Wilting**[2], **Anirban Mukhopadhyay**[1]

[1]*Technische Universität Darmstadt, Germany*; [2]*Leibniz-Institut für Zoo- und Wildtierforschung (IZW) im Forschungsverbund Berlin e.V., Germany*

Corresponding author(s) e-mail: anirban.mukhopadhyay@gris.tu-darmstadt.de; anandvashistha@gmail.com

ABSTRACT:

Biological diversity is decreasing at a rate of 100-1000 times pre-human rates [1] [2], and tropical rainforests are among the most vulnerable ecosystems. To avoid species extinction, we need to understand factors influencing the occurrence of species. Fast, reliable computer-assisted tools can help to describe the habitat and thus to understand species habitat associations. This understanding is of utmost importance for more targeted species conservation efforts. Due to logistical challenges and time-consuming manual processing of field data, months up to years are often needed to progress from data collection to data interpretation. Deep learning can be used to significantly shorten the time while keeping a similar level of accuracy. Here, we propose Habitat-Net: a novel Convolutional Neural Network (CNN) based method to segment habitat images of rainforests. Habitat-Net takes color images as input and after multiple layers of convolution and deconvolution produces a binary segmentation of an image. The primary contribution of Habitat-Net is the translation of medical imaging knowledge (inspired by U-Net [3]) to ecological problems. The entire Habitat-Net pipeline works automatically without any user interaction. Our only assumption is the availability of annotated images, from which Habitat-Net learns the most distinguishing features automatically. In our experiments, we use two habitat datasets: (1) canopy and (2) understory vegetation. We train the model with 800 canopy images and 700 understory images separately. Our testing dataset has 150 canopy and 170 understory images. We use the Dice coefficient and Jaccard Index to quantify the overlap between ground-truthed segmentation images and those obtained by Habitat-Net model. This results in a mean Dice Score (mean Jaccard Index) for the segmentation of canopy and understory images of 0.89 (0.81) and 0.79 (0.69), respectively. Compared to manual segmentation, Habitat-Net prediction is approximately 3K – 150K times faster. For a typical canopy dataset of 335 images, Habitat-Net reduces total processing time to 5 seconds (15 milliseconds/ image) from 4 hours (45 seconds/ image). In this study, we show that it is possible to speed up the data pipeline using deep learning in the ecological domain. In the future, we plan to create a freely available mobile app based on Habitat-Net technology to characterize the habitat directly and automated in the field. In combination with ecological models our tools will help to understand the ecology of some poorly known, but often highly threatened, species and thus contribute to more timely conservation interventions.

KEYWORDS: Canopy Interpretation, Understory Interpretation, Deep Learning, Image Segmentation, Convolutional Neural Network, U-Net

REFERENCES:

1. Sachs et al. "Biodiversity conservation and the millennium development goals." Science 325.5947 (2009): 1502-1503.
2. Chapin Iii, F. Stuart, et al. "Consequences of changing biodiversity." Nature 405.6783 (2000): 234.

3. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

Structured observations for automated plant identification

Michael Rzanny[1], Jana Wäldchen[1], Alice Deggelmann[1], David Boho[2], Marco Seeland[2], Patrick Mäder[2]

[1]*MPI for Biogeochemistry, Germany*; [2]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: mrzanny@bgc-jena.mpg.de

ABSTRACT:

Deep learning algorithms for automated plant identification are usually trained on the basis of diverse and unstructured image collections. We develop an image-capturing scheme to collect plant images comprising multiple views of one and the same plant individual. These so called structured observations, capture five predefined perspectives (habitus, flower top view, flower side view, leaf top side, leaf back side). We collect a dataset, covering about 70 species with an emphasis on visually very similar, conspecific species, including a number of grass species. We compare the prediction accuracy of these structured observations to the prediction accuracy of unstructured images collections, typically used for the training of image recognition algorithms.

We identify the most important perspectives and analyse the increase in accuracy by combining different perspectives.

KEYWORDS: automated plant identification, image acquisition, deep learning, image recognition
structured observations

Evaluating State-of-the-art Object Detection Methods for Plant Organ Detection

***Minqian Chen*[1], *Marco Seeland*[1], *Patrick Mäder*[1]**

[1]*TU Ilmenau, Germany*

Corresponding author(s) e-mail: marco.seeland@tu-ilmenau.de

ABSTRACT:

Region-based ConvNets (R-CNNs) allow for object detection in images, i.e., recognition of an object and its simultaneous localization within the image. In contrast to previous detection approaches using CNNs or ConvNet features, Fast R-CNN uses the whole image at a glance, forward it through a single network and use a softmax classifier for recognition in parallel to linear bounding-box regression for localization in one and the same network. Hence, the detection process can be trained end-to-end while network parameters can still be updated during the training. This allows for better tuning towards the specific task as well as faster training. Instead of using selective search as preceding step for creating costly region proposals, Faster R-CNN (Fast R-CNN's successor) uses the image features of the first forward pass of the CNN for computing region proposals, which further speeds the training up and allows for almost real-time detection.

We comparatively evaluate Faster R-CNN for detecting and recognizing plant organs in images. We utilize a dataset of annotated images of flowers and leaves. Faster R-CNNs are compared to alternative approaches in terms of standard metrics for object detection, i.e., precision, recall and mean average precision. Using the same framework, the performance of a combined organ and plant taxon detection is evaluated.

KEYWORDS: Bounding box, Faster R-CNN, Detection, Classification

Deep Learning Approach for Mapping Arctic Vegetation using Multi-Sensor Remote Sensing Fusion

***Jitendra Kumar*[1], *Zachary L. Langford*[2], *Forrest M. Hoffman*[1]**

[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*; [2]*University of Tennessee, Knoxville, TN, USA*

Corresponding author(s) e-mail: jkumar@climatemodeling.org, zlangfor@vols.utk.edu, forrest@climatemodeling.org

ABSTRACT:

Land cover datasets are essential for modeling Arctic ecosystem structure and function and for understanding land–atmosphere interactions at high spatial resolutions. However, most Arctic land cover products are generated at a coarse resolution, and finding quality satellite remote sensing datasets to produce such maps is difficult due to cloud cover, polar darkness, and poor availability of high-resolution imagery. A multi-sensor remote sensing-based deep learning approach was developed for generating high-resolution (5 m) vegetation maps for the western Alaskan Arctic on the Seward Peninsula, Alaska. The datasets from hyperspectral, multispectral, synthetic aperture radar (SAR) platforms, and terrain datasets were fused together using unsupervised and supervised classification techniques over a 343 km^2 region to generate high-resolution (5 m) vegetation type maps. A unsupervised technique was developed to classify high-dimensional remote sensing datasets into cohesive clusters and a quantitative technique to add supervision to the unlabeled clusters was employed, producing a fully labeled vegetation map. Deep neural networks (DNNs) were developed using multi-sensor remote sensing datasets to map vegetation distributions using the original labels and the labels produced by the unsupervised method for training [1]. Fourteen different combinations of remote sensing imagery were analyzed to explore the optimization of multi-sensor remote sensing fusion. To validate the resulting DNN-based vegetation maps, field vegetation observations were conducted at 30 plots during the summer of 2016 and developed vegetation maps were evaluated against them for accuracy. Our analysis showed that the DNN models based on hyperspectral EO-1 Hyperion, integrated with the other remote sensing data, provided the most accurate mapping of vegetation types, increasing the average validation score from 0.56 to 0.70 based on field observation-based vegetation.

KEYWORDS: deep learning, arctic vegetation, remote sensing, multi-sensor fusion

REFERENCES:

1. Langford, Z. L., Kumar, J., and Hoffman, F. M., "Convolutional Neural Network Approach for Mapping Arctic Vegetation Using Multi-Sensor Remote Sensing Fusion," 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, 2017, pp. 322-331. doi: 10.1109/ICDMW.2017.48

Identification and Segmentation of Whale Call Events Using Convolutional Neural Networks

Carlos de Obaldía[1], Udo Zölzer[1]

[1] *Helmut Schmidt University, Hamburg, Germany*

Corresponding author(s) e-mail: deobaldia@hsu-hh.de

ABSTRACT:

The detection of acoustic events in underwater environments is of particular importance for the conservation and behavioural research of marine mammals. In this work, a deep learning method for the identification of whale calls is proposed which uses convolutional neural networks for the classification and segmentation of the events. Input sound waveforms are transformed to a time-frequency representative space which is then processed for the detection and segmentation of events like sweeps, whistles and clicks. For this purpose, a regression and an autoencoder network are evaluated for performance and accuracy using signals recorded under different conditions and with different active species. The results show good outcomes for source separation tasks and high accuracy for acoustic event classification. Furthermore, the networks perform robust segmentation of the whale calls also yielding features like center frequency, time related bandwidth and call duration which can be used by standard classifiers for species identification.

KEYWORDS: whale classification, click segmentation, source separation, deep learning.

Session S3.2

CITIZEN SCIENCE MEETS INFORMATICS: DATA SCIENCE CHALLENGES IN ECOLOGICAL RESEARCH WITH PUBLIC PARTICIPATION

Citizen Science Meets Informatics: Data Science Challenges in Ecological Research with Public Participation

Friederike Klan[1], Jana Wäldchen[2]

[1]*Institute of Data Science, German Aerospace Center (DLR),Germany*; [2]*MPI for Biogeochemistry, Germany*

Corresponding author(s) e-mail: friederike.klan@dlr.de, jwald@bgc-jena.mpg.de

ABSTRACT:

During the last decade, Citizen Science, i.e. the involvement of laymen in scientific research, has gained great attention, both from the public and within the scientific community. Particularly the life sciences benefit from this development as citizen scientists contribute environmental observations of high resolution, analyze large amounts of ecological data or raise entirely new research questions. In doing so, they help to tackle pressing societal challenges such as loss of biodiversity and climate change.

In addition to the issue of how to engage and empower volunteers, data science aspects are major challenges in Citizen Science projects. This includes the following questions

1. How to make sure that data collected by citizen scientists are useful and relevant for addressing scientific questions?
2. How can data collected by citizen scientists be found, accessed, interpreted and used by other (citizen) scientists?
3. How to acknowledge citizen contributions in the scientific world?
4. How to facilitate the integration and analysis of data collected by the public with other data sources?
5. How to assess and improve the quality and reliability of Citizen Science data?
6. How to increase the credibility of data collected by volunteers?
7. How to enable citizen scientists to gain insights from data?

While some of these topics are specific to Citizen Science, they often share challenging aspects of data-intensive science in general (e.g. How to make data findable, accessible, interoperable and re-usable?). However, despite first cross-disciplinary and Citizen Science specific initiatives such as FORCE11 and the FAIR Data Principles [1,3], the Cost Action “Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe” [4] or first ideas on a EU Citizen Science Gateway for Biodiversity Data [2] where basic data science challenges are jointly discussed and best practices are collected, Citizen Science practitioners often address these core questions in an ad-hoc manner and individually in the context of specific projects.

We believe that the Citizen Science community would greatly benefit from an intensified scholarly exchange on data science topics related to Citizen Science across projects and disciplines. Thus, the objective of this special session is to bring together practitioners in Citizen Science projects as well as “traditional” scientists to discuss basic data science challenges arising in Citizen Science projects, to share best practices and lessons learned as well as to identify next steps towards a regular exchange on these topics and to initiate joint efforts on systematically addressing these challenges.

KEYWORDS: Citizen Science, Data Science, Ecology, FAIR principles, data quality, public participation

REFERENCES

1. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., C ’t Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018, DOI doi:10.1038/sdata.2016.18.

2. Runnel, V., Wetzel, F.T., Groom, Q.J., Koch, W., Pe’er, I., Valland, N., Panteri, E., Kõljalg, U., 2016. Summary report and strategy recommendations for EU citizen science gateway for biodiversity data. Research Ideas and Outcomes 2. e11563, https://doi.org/10.3897/rio.2.e11563.

3. FORCE11, 2011. FAIR Data Principles. https://www.force11.org/about (accessed 28 May 2018).

4. Citizen Science COST Action CA15212. https://www.cs-eu.net/ (accessed 28 May 2018).

Support of Forest Inventory Data Collection by Citizen Scientists

***Christian Thiel*[1], *Friederike Klan*[1], *Carsten Pathe*[2], *Christiane Schmullius*[2], *Jussi Baade*[2]**

[1]*German Aerospace Center (DLR), Germany*; [2]*Friedrich-Schiller-University Jena, Germany*

Corresponding author(s) e-mail: Christian.Thiel@dlr.de

ABSTRACT:

Precise forest inventory data are requested by a wide range of users such as scientists, politicians, administrators, forest owners, or the forest industry. One forest inventory parameter of great importance is the forest stem volume (or growing stock volume, GSV). On the one hand, GSV is related to the monetary value of a forest. On the other hand, the amount of bound carbon can be estimated based on GSV. For the determination of the GSV the stem diameter (usually diameter at breast height, DBH), the tree height, the number of trees per unit area, and a species and forest stand specific form factor are required. In forestry, sample based approaches are used to gather these parameters. For minimizing effort and expense, the number and dimensions of these samples are small compared to the total forest area. Also, the repeat time between two inventories is rather large (in the order of ten years). Accordingly, relative GSV errors of approximately 20% have to be accepted.

There exists a great interest to minimize both, effort and inventory errors. Precise inventory data are of particular interest in the research domain. For instance, satellite based methods aiming at GSV estimation suffer from inaccurate reference measurements, as the inventory errors propagate to the final satellite based estimates. Airborne light detection and ranging data (LiDAR) can be utilized to detect single trees and to measure the corresponding tree heights with sufficient accuracy for forestry applications. In some Scandinavian countries forest inventories are supported by LiDAR campaigns by default. Moreover, most European countries execute regular and country-wide LiDAR acquisitions, thus LiDAR based tree height measurements could be achieved. For instance, the LiDAR campaign repetition rate in Germany is five years. However, the stem diameter cannot be measured using airborne LiDAR data. Although some technical ground- and low altitude airborne solutions have been proposed, currently the most efficient approach is manual DBH measurement.

The simplicity of DBH measurements makes this task an excellent citizen science exercise. To assess the achievable DBH measurement precision, an experiment involving students of a secondary school was carried out in late 2017. The test site “Roda Forest” is located 20 km in the Southeast of Jena. The selected stand is dominated by pine with an age of 60 years. The reference data for the experiment was generated by means of a terrestrial laser scanner (TLS). Based on the TLS data the precise location and the GSV of approximately 200 trees were delineated. The students were equipped with a smartphone application to localize the single trees. During the campaign the circumference of approximately 100 trees was determined using simple measuring tape. These measurements were converted to DBH after the field campaign. The measured DBH varied between 7 cm and 38 cm. In overall, TLS-based and student campaign based measurements were in great agreement ($R^2 = 0.98$). Nevertheless, the identification of the correct trees by the students during the campaign was challenging, which was related to general orientation difficulties and a weak GPS signal underneath the forest canopy. This resulted in a remarkable offset between GPS-based and real coordinates. Forthcoming campaigns have to deal with this issue. One option we will explore in the future is the absolute calibration of the GPS signal using checkpoints with precise coordinates.

KEYWORDS: Citizen Science, Forest parameters, Data Quality

Beyond Data and Quality – Unleashing the Value of Citizen Contributions

***Friederike Klan*[1]**

[1]*German Aerospace Center (DLR), Germany*

Corresponding author(s) e-mail: friederike.klan@dlr.de

ABSTRACT:

Citizen Science projects generate outcomes that have the potential to be highly valuable both for science and society. Those contributions are not restricted to data, but can be other kinds of results, e.g. best practices, new insights or research questions.

Letting Citizen Science have a true impact requires us to unleash the value of citizen contributions. This touches a wide range of points and goes beyond mere data quality considerations. Challenges include, but are not restricted to

- How cannot just data, but other outcomes of Citizen Science projects be captured and made accessible to others?
- How to make sure that citizen contributions are findable, accessible and interpretable by others and how to enable interoperability with other data and knowledge sources (i.e. adhere to the FAIR-Principles)?
- How to assess and improve the quality and reliability of Citizen Science outcomes and increase their credibility?
- How to make sure that data collected by citizen scientists are useful and relevant for addressing scientific questions?
- How to enable citizen scientists to gain and apply insights from project outcomes?

This talk is meant as an impulse for further discussion and scholarly exchange on these topics to foster a systematic approach to address these challenges.

KEYWORDS: Citizen Science, Data Science, FAIR data, data quality

Citizen science and science-policy interface: Towards sustainable forest managements

***Ryo Kohsaka*[1], *Shuichiro Kajima*[1], *Yuta Uchiyama*[1]**

[1]*Tohoku University, Graduate School of Environmental Studies, Japan*

Corresponding author e-mail: kohsaka@hotmail.com

ABSTRACT:

The forest policy research with quantitative approaches is limited especially in the research fields of policy transfer and science-policy interface at the local government level. As a context of the citizen science and the science-policy interface, the attitudes of the Japanese local governments vary from municipality to municipality. For example, certain local governments proactively introduce participation of citizen in forest policy making and its implementation, while such attitudes or policies are absent for others. Where comes such differences amongst municipalities? This study conducts empirical analysis at local policy level, which has been largely overseen. In concrete terms, trends of adoption of local ordinances of forest planning in Japanese prefectures and participatory monitoring activities in forest lands are reviewed and analyzed. This study examines the relationships between political factors, social economic factors, and policy diffusion. We have looked at the various factors including local demographics, size of administrative areas, government structure, percentage of forest lands and net forestry production in local government to examine the differences. The results reveal the significant impacts of behavior of neighboring local governments. In other words, if a neighboring body acts, others will follow. The perceptions of the governors who are not from conservative political party seems to be a factor to motivate local governments to introduce the local ordinances of forest planning which encourage citizen to participate forest managements and forest policy making processes. Furthermore, participatory monitoring is a useful tool in citizen science and the number of biodiversity monitoring activities is increasing in Japan as well as other Asian countries such as Korea. The local ordinances of forest planning can influence the monitoring activities. The focus of recent monitoring activities based on citizen science in Japan is changing from “scientists use citizens as data collectors” to “citizens as scientists” under the recent development of data science. Alternatively, sharing, collecting and analyzing data in effective manner with participations remains as future challenges for data science.

KEYWORDS: Citizen science, forest management, policy transfer, participatory monitoring, Japan

Crowdsourcing participatory millet variety selection in Hoima, Uganda for climate change adaptation through Triadic Comparisons of Technologies

***Tobias Recha*[1], *Gloria Otieno*[2,], *Carlo Fadda*[3]**

[1]*Bioversity International Kenya,* [2]*Bioversity International Uganda, Bioversity International Ethiopia*

Corresponding author(s) e-mail: tobiasrecha@gmail.com; g.otieno@cgiar.org

ABSTRACT:

Climate change is a key issue affecting agriculture in Uganda. A base line survey done in Hoima indicates that there have been changes in climate, such as shifting seasons, erratic rainfall; an increase in temperature; and, pests and diseases. This has led to loss of agricultural productivity and loss of genetic diversity. Among the strategies identified for coping with the effects of climate change is the access of genetic diversity necessary for communities to adapt to climate change and development of new varieties of crops through plant breeding.

Using geographical information systems (GIS) and climate profiles consisting similarities in temperature and precipitation; 43 millet landraces were identified from national gene bank collections in Uganda and Tanzania. These were exchanged Standard Material Transfer Agreements (SMTAs) and multiplied at National Agricultural Research Organization and were then distributed to 200 farmers for participatory trials. Triadic Comparison of Technologies (TRICOT) Methodology, farmers were randomly allocated three blind varieties coded A, B and C and trained on the data collection methods. They were tasked to evaluate six aspects which included faster maturity, pest resistance, disease resistance, drought resistance, yield and overall performance. ClimMob Manager – a software for crowd sourcing climate smart agriculture, was then used to collect and analyze the data. Varieties performance ranking was done and analyzed using Bradley-Terry Models.

Results indicated farmers' preferences for faster maturing, high yielding and pest and disease resistance. The analysis also provided three top performing varieties. These are then ranked against local varieties. From the results, TZA 1695 and UNGB 4146 matured fastest. In the category of pest resistance, TZA 1700 and UNGB 4146 were not affected by pests. However, these varieties are still not the best in terms of overall performance. The overall performance of the 43 tested varieties showed that UNGB 43, UNGB 4400 and TZA 186 were the top three best varieties recommended for Hoima. Top ranking varieties for each trait are then isolated to produce elite lines for participatory plant breeding.

KEYWORDS: Crowdsourcing, ClimMob, Millet Varieties, Adaptation

Wikimedia projects as citizen science platforms

***Daniel Mietchen*[1]**

[1]*Data Science Institute, University of Virginia, Charlottesville, VA, USA*

Corresponding author(s) e-mail: dm7gn@virginia.edu

ABSTRACT:

When searching the web for information on a given topic — say, citizen science — Wikipedia will often come up high in the search results. What is less known is that Wikipedia forms the nucleus of an entire ecosystem of Wikimedia projects that are roughly organized by information channel. Besides the encyclopedia, this includes an archive (Wikisource) and a dictionary (Wiktionary), along with sites dedicated to quotes (Wikiquote), taxa (Wikispecies), media files (Wikimedia Commons), books (Wikibooks), coursework (Wikiversity), news (Wikinews), structured data (Wikidata), software (MediaWiki) or technical infrastructure (Wikimedia Labs).

While citizen science-related topics form only a small sliver of the content found on these open knowledge platforms, the practice of contributing is similar to that of many citizen science projects: contributors are volunteers who participate in areas of their own interest, in a language of their choice and on their own timeline. The similarities go further, with community-driven collaborations forming around issues that are also in the focus of citizen science projects: the identification and disambiguation of taxa, people or locations, the collection and categorization of photographs, audiovisual materials and historic records, the transcription and translation of such records, the integration and visualization of data, or identifying and locating suitable resources that can help verify information related to any of the above.

This contribution is intended as a guided tour around such citizen science activities taking place in the framework of Wikimedia projects, both within and across domains of knowledge. Focusing on examples from biodiversity and ecosystems research, it will highlight some data-related issues (e.g. data and metadata quality, data integration) and how they are handled in Wikimedia contexts. The presentation will be given on the basis of https://github.com/Daniel-Mietchen/events/blob/master/ICEI2018-citizen-science.md , which will provide further information ahead of the event. Participants are encouraged to bring their laptops, as they will be invited to actively engage with Wikimedia-based information in their domain, as well as to entertain and share ideas on how they could integrate it with their own activities.

KEYWORDS: Wikipedia, Wikimedia Commons, Wikidata, collaboration, open knowledge

Mobile app and platform development in citizen science

Ulrike Sturm

Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Germany

Corresponding author(s) e-mail: ulrike.sturm@mfn.berlin

ABSTRACT:

In the past two decades citizen science is increasingly perceived and acknowledged. However data quality concerns, occur when data are provided by many, mostly anonymous people. Data produced by volunteers can have high quality, depending on task difficulty and volunteer experience. Therefore it is especially important to use and develop suitable methods and data collection tools that support participants. Meanwhile digital technologies are more and more used for data collection and management. Hence support for planning, design and data management of mobile apps and platforms in the citizen science community is needed. We developed 43 recommendations during two workshops, where citizen science practitioners with experience in mobile application and web-platform development and implementation came together [1].

I will present these recommendations and demonstrate their implementation using the example of the mobile app “Naturblick”. The smartphone app “Naturblick” was developed with a user centred approach and is being continuously improved based on users’ feedback of both the content and technology. Users are able to identify animals and plants and share observations. For species identification we developed auditory and visual pattern recognition tools and multi-access identification keys for animals and plants. As part of the citizen science project “Forschungsfall Nachtigall” (nightingale research case) participants record nightingale songs with “Naturblick”. The recordings are analysed in regard to song dialects and habitat preferences.

KEYWORDS: citizen science, mobile app and platform development, participant centred design

REFERENCES:

1. Sturm, U., Schade, S., Ceccaroni, L., Gold, M., Kyba, C., Claramunt, B., Haklay, M., Kasperowski, D., Albert. A., Piera, J., Brier, J., Kullenberg, C., Luna, S., 2018. Defining principles for mobile apps and platforms development in citizen science. Research Ideas and Outcomes 4: e23394. https://doi.org/10.3897/rio.4.e23394.

PlanktonID – Combining in situ Imaging, Deep Learning and Citizen Science for Global Plankton Research

***Rainer Kiko*[1], *Svenja Christiansen*[2], *Simon-Martin Schröder*[3], *Reinhard Koch*[3], *Lars Stemmann*[4]**

[1]*GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*; [2]*University of Oslo, Norway;* [3]*Christian-Albrechts University Kiel, Germany;* [4]*Sorbonne Universités, UMPC Univ Paris 06 and Observatoire Océanologique Villefranche sur Mer, France*

Corresponding author(s) e-mail: rkiko@geomar.de

ABSTRACT:

Recent publications revealed the global importance of single-celled zooplankton, belonging to the super group Rhizaria and highlighted the need of in-situ imaging to study these fragile organisms. The advance of in situ plankton imaging techniques leads to increasing amounts of image data sets that require identification to different taxonomic levels. Automatic classification by computer algorithms provides the means for fast data availability, however the accuracy of those algorithms still requires manual identification by humans. We combined state of the art automatic image classification by convolutional neural networks (deep learning) with a citizen science project to classify a large dataset of ~ 3 million images from an Underwater Vision Profiler 5 (UVP5). On our website https://planktonid.geomar.de, citizen scientists can confirm or reject the automatic assignment of UVP5 images to different plankton categories in a memory-like game. Inbuilt quality controls and multiple validations per image enable scientific analysis of the citizen science data. In total more than 500 users have validated more than 300.000 images until now. We will present further data on citizen scientist engagement, data quality assessment and the distribution analysis of large protists (Rhizaria) in the Mauretanian, Benguela and Humboldt Current upwelling systems.

KEYWORDS: Network, Plankton, Citizen Science, Deep Learning, Rhizaria, Global distribution

sMon - Trend analysis of German biodiversity data

Aletta Bonn[1,2,3]***, David Eichenberg***[4,1,3]***, Helge Bruelheide***[5,3]***, Florian Jansen***[6]***, Marten Winter***[4,3]

[1]*Helmholtz-Centre of Environmental Research-UFZ, Germany*; [2]*Friedrich Schiller University Jena, Germany*; [3]*German Centre for integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany;* [4]*University of Leipzig, Germany*; [5]*Martin Luther University Halle, Germany*; [6]*University of Rostock, Germany*

Corresponding author(s) e-mail: aletta.bonn@idiv.de

ABSTRACT:

Most biodiversity data are collected by volunteers organised in natural history societies or citizen science projects, often closely aligned with (sub)national agencies and local authorities. Data may be heterogeneous in space, time and quality. Here, we present first results of trend analyses of joint work with German natural history societies and state and national conservation agencies through the sMon synthesis project within iDiv. We combine and harmonize exemplary datasets of different taxa and habitats to evaluate the potentials and limits for analysing changes in the state of biodiversity in Germany. We show trend analyses of occupancy frequency data for 60 dragonfly, 42 grasshopper species and amphibia across 3 federal states 1980-2015, using Bayesian hierarchical trend analyses that build on occupancy detection models. Based on these insights and evaluation of citizen science programmes globally, we derive principles for good practice citizen science project design, data collection and archiving and explore methodologies that can deal with fragmented data of different spatio-temporal resolution and quality. This includes harnessing the potentials offered by modern technology. Combined with experiences of joint working of volunteer experts, agencies and academic scientists, this informs perspectives for future biodiversity monitoring programmes in Germany.

Flora Capture - An Adaptive Digital Herbarium Using Mobile Devices

David Boho[1]***, Michael Rzanny***[2]***, Jana Wäldchen***[2]***, Patrick Mäder***[1]

[1]*Technische Universität Ilmenau, Germany*; [2]*Max Planck Institut for Biogeochemistry, Germany*

Corresponding author e-mail: patrick.maeder@tu-ilmenau.de

ABSTRACT:

With the advent of automated species identification and automated trait analysis, researchers need large amounts of training data to improve methods and to gain more insights into the plasticity of traits. Herbaria all over the world have invested large amounts of money and time in collecting samples of plants. However, recent studies have shown that today's methods are not sufficiently able to generalize information acquired on traditional herbarium samples and to transfer them to living samples. This finding stimulated our research into a digital herbarium that fulfils the requirements of traditional collections, but also preserves an individual plant in its living state. Utilizing the various sensors of modern mobile devices, we developed an multi-platform app that seamlessly allows acquiring a digital herbarium sheet. The user initially decides about the life form of a plant to collect and is then being directed through an adaptive serious of image perspectives to be acquired. Completed observations are uploaded to the projects server and become part of a botanists work queue, who is carefully checking and correcting observation details and eventually approving or rejecting it. The observation's status is synced back to the user's device along with detailed information on potential problems discovered by the expert in order motivate the user and to improve observation quality. Approved observations are transferred to conservations bodies. We will demonstrate the end-to-end scenario of capturing an observation, reviewing its correctness, and propagating feedback back to the user. Through Flora Capture, users acquired more than 5,200 observations of more than 500 species already.

KEYWORDS: citizen science, mobile apps, herbarium, images

Session S3.3

OPERATIONALIZING ESSENTIAL BIODIVERSITY VARIABLES: ASPECTS OF DATA INTEGRATION, PRODUCTION AND DISSEMINATION

Operationalizing Essential Biodiversity Variables: aspects of data integration, production and dissemination

***Néstor Fernández*[1], *W. Daniel Kissling*[2], *Robert Guralnick*[3], *Alex Hardisty*[4]**

[1]*iDiv, Germany*; [2]*University of Amsterdam, The Netherlands*; [3]*Florida Museum of Natural History, USA;* [4]*Cardiff University, UK*

Corresponding author(s) e-mail: Nestor.Fernandez@idiv.de; HardistyAR@cardiff.ac.uk

ABSTRACT:

The Group on Earth Observations – Biodiversity Observation Network (GEO BON) has the impetus to develop conceptual and technical approaches to the production of Essential Biodiversity Variables (EBVs). However, the ecological scientific community as well as those who are responsible for acquiring, curating, publishing, processing and using heterogeneous biodiversity and ecological data must invest into supporting a systematic production and use of EBVs. Such information products should be applicable to any geographic area, covering time-period(s) of interest for detecting biodiversity change at policy-relevant time scales, with data that is held in any or multiple repositories, and produced by appropriately skilled persons anywhere in the world. Within constraints of specific data types, EBV information products should be harmonised and comparable at various scales from local to global and across time, such that they can be used to monitor and measure biodiversity change.

By showcasing what has been done so far, and guided by principles of global coordination of biodiversity monitoring, this session intends to foster scientific and technical exchange and build communities of practice to support production, delivery, use and sustainability of Essential Biodiversity Variables (EBVs) data products. This session specifically aims to:

- Showcase the role and need for informatics to develop and support the production and delivery of Essential Biodiversity Variables (EBVs) information products at scales from local to global;

- Showcase approaches for utilising and sustaining EBV information products and their adherence to FAIR (Findable, Accessible, Interoperable, Re-usable) principles.

- Provide recommendations for the production and dissemination of biodiversity observations under the EBV framework.

KEYWORDS: Biodiversity monitoring, data standards, GEO BON, interoperability, workflows

Ten outcomes to improve informatics interoperability in cyber/e-Infrastructures for biodiversity and ecological sciences through the Essential Biodiversity Variables (EBV) use case

***Alex Hardisty*[1]**

[1]*Cardiff University, UK*

Corresponding author(s) e-mail: hardistyar@cardiff.ac.uk

ABSTRACT:

Data products for Essential Biodiversity Variables (EBV) must be (re-)producible and comparable for any geographic area, small or large, fine-grained or coarse; at temporal scale determined by need and/or the frequency of available observations; at a point in time in the past, present day or in the future; as appropriate, for any species, assemblage, ecosystem, biome, etc.; using data for that area/topic that may be held by any and across multiple research/data infrastructures; using harmonized, widely accepted protocols (workflows); capable of being executed in any infrastructure; by any (appropriate) person anywhere.

To date, the GLOBIS-B project (www.globis-b.eu) established there are technical needs for: i) common dimensional structure, packaging and metadata descriptions of EBV data products; ii) consistent quality checking and assertion across data from different sources that contribute to EBVs; iii) EBV workflows with common representation that is independent of underlying computational infrastructure; and iv) use of standard mechanisms for recording provenance of EBV data products.

However, too little is presently known about how the technical production of EBV data products will work in practice. Experimental implementation work is necessary, both to show what is technically feasible and useful, and to reveal what is really needed. We must, for example agree details of both the compact data/file structure for EBV data products, and programmatic interfaces to those data products. Experimental work must lead eventually to formal standardisation. Scientists, infrastructure providers, informaticians and GEO BON Working Groups must jointly address the specific problems of moving from limited, experimental, proof-of-concept type studies (such as the Atlas of Living Australia / Global Biodiversity Information Facility (GBIF) invasive species case study) to first trials producing and using real data products with real users. Beyond first trials, they must jointly move to more robust solutions that scale out and up, as well as providing the basis for the long-term support to GEO BON across a wide range of EBVs classes.

Satisfying the EBV use case acts for generally improving informatics interoperability among diverse cyber / e-Infrastructures supporting biodiversity science and ecology. It is desirable to guide participating providers without restricting their autonomy to achieve what is needed in ways appropriate to their own business. We show ten specific outcomes we want to see achieved, with the mission being the ability to deploy and execute standard workflows for preparing, publishing and preserving fit-for-use EBV data products that are comparable with one another. Achieving such outcomes significantly improves the ability of infrastructure providers to support the EBV production process.

KEYWORDS: Essential Biodiversity Variables, EBV, informatics, infrastructure, interoperability

Towards an EBV Analyzer based on VAT

Christian Beilschmidt[1]***, Johannes Drönner***[1]***, Néstor Fernández***[2]***,***
Christian Langer[2]***, Michael Mattig***[1]***, Bernhard Seeger***[1]

[1]*University of Marburg, Germany*; [2]*German Centre for Integrative Biodiversity Research (iDiv)*
Halle-Jena-Leipzig, Germany

Corresponding author(s) e-mail: beilschmidt@mathematik.uni-marburg.de

ABSTRACT:

The Essential Biodiversity Variables (EBVs) proposed by GEO BON offer great benefits for decision makers and scientists if they are made readily available. Currently, the usage of EBVs is, however, still a challenge. The required data sets are very large, heterogeneous and temporal. This makes the usage in common tools like GIS very cumbersome. We propose the VAT system, a web-based processing engine for spatio-temporal data, as an ideal tool for EBV visualization and analysis. VAT is being developed as part of GFBio, but was designed to work independently from the start. Here, users can access a rich set of environmental layers and combine them with their own, private data. So-called exploratory workflows allow users to process data in an interactive fashion while guaranteeing reproducibility of the result. In cooperation with GEO BON we are working towards using VAT as the future EBV Browser & Analyzer.

In our first use case we considered the global forest change EBV that measures the yearly global loss and gain of tree cover. In particular we wanted to investigate the change over time for a given region of interest. This information is highly relevant for decision makers as well as scientists. In our application users can choose between forest loss and forest gain and then specify their region of interest. This is either done by selecting a country from a list, or by drawing a custom polygon on the map. The relevant part of the EBV data set is added to the map as a new layer. In addition, a plot visualizes the aggregated changes over time. On the technical side this requires intersecting the raster data from the EBV with the user-defined polygon. The size of the raster makes it necessary to work with overviews in order to achieve a low-latency computation. For plotting the temporal development, we make use of our interface to R which allows us exploiting its powerful plotting functionality.

In our ongoing work we will implement a public EBV catalogue that allows users to search and submit EBVs. Quality indicators will assist users to decide on the suitability of the data for their given use case. The data from the EBV catalogue will be available in the VAT system for scientists. In addition, VAT also offers simplified report views that hide the complexity of the actual computations and are thus suitable for decision makers and even the general public.

KEYWORDS: Essential Biodiversity Variables, GIS, Interactive Data Exploration, Spatio-temporal processing

Minimum Information Standards for Essential Biodiversity Variables

***Néstor Fernández*[1], *W. Daniel Kissling*[2], *Robert Guralnick*[3]**

[1]*iDiv, Germany*; [2]*University of Amsterdam, The Netherlands*; [3]*Florida Museum of Natural History, USA;*
[4]*Cardiff University, UK*

Corresponding author(s) e-mail: Nestor.Fernandez@idiv.de; HardistyAR@cardiff.ac.uk

ABSTRACT:

Minimum Information Standards (MIS) are sets of specifications for describing datasets that aim to standardize data reporting and to maximize its discoverability and interoperability. While MIS have greatly contributed to enhance the distribution and reuse of datasets in the biological and biomedical sciences, their adoption in ecology and biodiversity sciences is still incipient. Here we present a community effort to generate minimum standards for Essential Biodiversity Variables (EBVs). The operationalization of EBVs require integrating heterogeneous datasets of disparate origin and, often, to combine information from different geographic areas and time periods. Furthermore, developing policy-relevant indicators from EBVs requires an additional level of integration between datasets that inform on different facets of biodiversity, e.g. at levels from species to ecosystems. MIS for Essential Biodiversity Variables is founded in the description of the EBV-data cube as the unifying framework to deliver interoperable biodiversity observations. They summarize aspects of the spatial and temporal domains of the datasets, as well as uncertainty and bias reporting. MIS also incorporate the GEOSS proposed principles for data management. Finally, a metadata publishing toolkit will be developed in order to ensure that EBVs are discoverable and used under the auspices of GEO BON.

KEYWORDS: Biodiversity monitoring, data standards, GEO BON, data management

Towards structured biodiversity data

Henrique Pereira[1]

[1]*German Centre for Integrative Biodiversity Research (iDiv), Germany*

Corresponding author(s) e-mail: hpereira@idiv.de

ABSTRACT:

Over the last 15 years there was a huge progress in mobilising biodiversity data through the Global Biodiversity Information Facility (GBIF). However, the standard adopted for this work, the Darwin Core, was designed with museum collection data in mind and is only appropriate for capturing species occurrence data. The Group on Earth Observations Biodiversity Observation Network (GEO BON) aims at improving biodiversity monitoring and making the data available. Most biodiversity monitoring protocols include some sampling procedure and repeated observations. Similarly, much biodiversity data collected by ecologists falls under the form of species per site matrices, where the absences and presences are recorded for the whole community at each site. The original Darwin Core was not appropriate for this kind of structured biodiversity data. Here we present recent developments in extending the Darwin Core to handle this data, the Darwin Sampling Event, and other related efforts. We also present the initiative being developed by GEO BON, iDiv and GBIF, to allow scientists around the world to publish directly on a iDiv repository, the APIS repository, structured biodiversity data.

KEYWORDS: Biodiversity, Darwin Core, GEO BON

Developing and implementing a general standard for inventory data

Robert Guralnick[1], **Ramona Walls**[2] and **Walter Jetz**[3]

[1]*University of Florida Museum of Natural History, USA;* [2]*CyVerse, USA;* [3]*Yale University, USA*

Corresponding author(s) e-mail: robgur@gmail.com

ABSTRACT:

Species inventories, i.e., the recording of multiple species for a specific place and time, are routinely performed and offer particular value for characterizing biodiversity and its change. However, a general reporting standard allowing these inventories to be re-used, compared to one another, and further integrated with other sources of biodiversity data are lacking, impeding their broadest utility. Humboldt Core is a conceptual standards framework for capturing core information about processes underpinning inventory work. Our framework is based on a community input process, followed by extensive refinement and testing using published inventories. In this talk, we first provide a typology of inventories and inventory processes, distinguishing between single, elementary inventories, extended and summary inventories, representing increasing levels of sampling event aggregation. Next, we cover the key content needed for reporting inventory process and how that reporting can be used to document species absence and use in occupancy modeling. This content includes terms to capture geospatial, temporal, taxonomic, and environmental conditions, along with methodological descriptors related to the assessment of sampling effort and inventory completeness. We demonstrate examples of Humboldt Core reporting and its use in existing infrastructure. We also discuss ways to further integrate Humboldt Core broadly into global biodiversity data sharing networks and how this can help catalyze best re-use of data for global monitoring. Humboldt Core will benefit from further enhancements based on community testing and input but represents a step toward significantly expanding biodiversity dataset discovery, interoperability, and modelling utility for a type of data essential to the assessment of biodiversity variation in space and time.

KEYWORDS: Humboldt Core, Species inventories, Standards framework

Session S3.5

Plant Traits and Biogeochemical Cycles

Plant traits and biogeochemical cycles

***Jens Kattge*[1,2], *Susanne Tautenhahn*[1], *Michael Bahn*[3], *Han Wang*[4] *and Oskar Franklin*[5]**

[1]*Max Planck Institute for Biogeochemistry, Germany;* [2]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany;* [3]*Institute of Ecology, University of Innsbruck, Austria;* [4]*College of Forestry, Northwest A & F University, Yangling, China;* [5]*International Institute for Applied Systems Analysis (IIASA), Austria*

Corresponding author(s) e-mail: jkattge@bgc-jena.mpg.de

ABSTRACT:

Plant traits extend the range of earth observations to the level of individual organisms, providing a link to ecosystem function and modeling in the context of rapid global changes. However, overcoming the differences in temporal and spatial scales between plant trait data and biogeochemical cycles remains a challenge.

This session will address the role of plant species, biodiversity and adaptation / acclimation / optimality and their connection to the biogeochemical cycles of water, carbon, nitrogen and phosphorus.

We welcome conceptual, observational, experimental and modeling approaches, and studies from the local to the global scale, including e.g. remote sensing observations and novel concepts and tools for the acquisition, management, analysis and synthesis of trait data.

KEYWORDS: Plant Traits, Biogeochemical Cycles, Functional Biogeography, Ecosystem Modeling, Plant Adaptation / Acclimation / Optimality

Biophysical modelling of risks and feedbacks from forest fires: the role of plant traits

Philip Zylstra[1]

[1]*University of Wollongong, Australia*

Corresponding author e-mail: pzylstra@uow.edu.au

ABSTRACT:

Climatic and demographic changes across the globe are altering the wildfire environment, with significant implications for carbon, water and nutrient cycles. This emphasises the need to more completely understand fire and its drivers, however, recent work has identified a number of fundamental controls and dynamics that play little or no role in traditional models. In some cases, these even run counter to model expectations.

Changes in live plant moisture for example can limit or facilitate landscape-scale fire [1], and widespread forested areas long believed to increase in flammability due to fuel accumulation have been found to be least flammable in their long-unburnt states [2]. At a finer scale, explanations for the flammability of plant parts have been found from traits such as specific leaf area [3], dimensions [4], curliness and chemistry [5]. These findings have prompted the growing recognition that simple biomass accumulation is a poor explanation of ecosystem flammability, and that an adequate model must properly account for the various roles of plant traits [6]. Even in the simplest arrangements however, trait effects are not additive [7], and scaling from leaf to plant flammability is highly complex [8].

The Forest Flammability Model [9] provides a biophysical, mechanistic approach to modelling this complex system. The flammability of leaves is modelled in its separate components of ignitability, combustibility and sustainability, then scaled upward by calculating the ignition of new leaves from the flames produced by leaves that are already burning. By explicitly finding the effect of individual plant traits, changes in these traits such as those caused by altered atmospheric CO_2 levels or phenological variation can be integrated into predictions of future fire. All sub-models of trait effects or physical processes can be updated and replaced by new research as it arises.

Validation across diverse eucalypt communities and wildfire conditions found that the model explained 80% of the variability in flame heights when all plant traits were considered, but only 11% when surface fuel loads were the sole representation of forest flammability. Positive fire-flammability feedbacks in one forest community were also explained using trends in post-fire plant growth and species' succession [10].

Due to the modelling process, flames are calculated dynamically, and the areas of burning vegetation are explicitly located within the forest profile along with the resulting flames, on a second-by-second basis. The large body of heating information that this produces is currently being constructed into a risk model for wildlife, including habitat effects such as fire severity and soil heating.

Historically, fire ecology has focused on the effects of fire on flora and fauna. This body of work extends that field to quantify the full feedback, by providing a platform to integrate work on plant traits and ecosystem dynamics and thereby find the ways in which flora and fauna in turn affect the fire regimes of a forest.

KEYWORDS: Fire, flammability, positive feedbacks, fire ecology

REFERENCES:

1. Nolan R.H., Boer M.M., Resco de Dios V., Caccamo G., Bradstock R.A. 2016. Large scale, dynamic transformations in fuel moisture drive wildfire activity across south-eastern Australia. Geophys Res Lett. 43, 4229–4238.

2. Zylstra P. 2018. Flammability dynamics in the Australian Alps. Austral Ecol. doi:10.1111/aec.12594

3. Grootemaat S., Wright I.J., van Bodegom P.M., Cornelissen J.H.C., Cornwell W.K. 2015. Burn or rot: leaf traits explain why flammability and decomposability are decoupled across species. Funct Ecol. 29, 1486-1497.

4. Cornwell W.K., Elvira A., van Kempen L., van Logtestijn R.S.P., Aptroot A., Cornelissen J.H.C. 2015. Flammability across the gymnosperm phylogeny: the importance of litter particle size. New Phytol. 206, 672–681.

5. Cornelissen J.H.C., Grootemaat S., Verheijen L.M., Cornwell W.K., van Bodegom P.M., van der Wal R., et al. 2017. Are litter decomposition and fire linked through plant species traits? New Phytol. 216, 653–669.

6. Archibald S., Lehmann C.E.R., Belcher C.M., Bond W.J., Bradstock R.A., Daniau A., et al. 2018. Biological and geophysical feedbacks with fire in the Earth system. Environ Res Lett. 13, 033003.

7. de Magalhães R.M.Q., Schwilk D.W. 2012. Leaf traits and litter flammability: evidence for non-additive mixture effects in a temperate forest. J Ecol. 100, 1153–1163.

8. Pérez-Harguindeguy N., Díaz S., Garnier E., Lavorel S., Poorter H., Jaureguiberry P., et al. 2013. New handbook for standardised measurement of plant functional traits worldwide. Aust J Bot.61, 167–234.

9. Zylstra P., Bradstock R.A., Bedward M., Penman T.D., Doherty M.D., Weber R.O., Gill, A.M., Cary, G.J. 2016. Biophysical mechanistic modelling quantifies the effects of plant traits on fire severity: species, not surface fuel loads determine flame dimensions in eucalypt forests. PLoS One.11: e0160715.

10. Zylstra P. 2013. The historical influence of fire on the flammability of subalpine Snowgum forest and woodland. Vic Nat.130, 232–239.

Variability in Community Productivity: Mediation by Vegetation Traits

H. Wayne Polley[1]***, Brian J. Wilsey***[2]

[1]*USDA–Agricultural Research Service, USA*; [2]*Iowa State University, USA*

Corresponding author(s) e-mail: wayne.polley@ars.usda.gov

ABSTRACT:

Plant productivity varies through time in response to environmental fluctuations. Learning to reduce temporal variability in primary productivity is a frequent goal of management with clear relevance to sustaining ecosystem services for an expanding human population. Reducing variability in productivity requires an improved understanding of how plant community traits interact with environmental fluctuations to influence plant growth dynamics. We evaluated links between two community traits, species diversity and species abundance-weighted values of specific leaf area (SLA; leaf area per unit of leaf dry weight), and temporal variability in grassland productivity at patch (local) and aggregate (multi-patch) spatial scales. Aggregate communities were created by combining patches of spatially-distinct communities of perennial plant species from grassland biodiversity experiments in Texas, USA. Inter-annual variability in aboveground net primary productivity (ANPP) of aggregate communities was analyzed as a function of two multiplicative components, mean temporal variability in the ANPP of patches and temporal synchrony in ANPP dynamics among patches. We used regression analyses to determine whether temporal variability in aggregate ANPP and its components were correlated with either species diversity or community-weighted SLA over 5 years.

We found that temporal variability in ANPP of aggregate communities, calculated as the square of the temporal CV $[(\delta/\mu)^2]$ of ANPP, was strongly correlated with temporal variability in patch ANPP. Increasing mean SLA reduced ANPP variability of aggregate communities by increasing mean productivity (μ). Increased temporal changes in patch-scale SLA further reduced temporal variability in aggregate ANPP by reducing effects of precipitation fluctuations on productivity. Conversely, increasing species diversity over the narrow range measured increased temporal variability in aggregate ANPP. High diversity was associated with reduced dominance of temporally-stable C_4 grasses. Our results implicate means and patch-scale temporal dynamics in community SLA as potential indicators of variability in grassland primary productivity through time.

KEYWORDS: Diversity, Functional trait, Grassland, Spatial scale, Specific leaf area

Inconstant exponents of scaling leaf nitrogen to phosphorus

***Di Tian*[1]*, Zhengbing Yan*[2]*, Jingyun Fang*[2]**

[1] *College of Life Sciences, Capital Normal University, China;* [2] *Institute of Ecology, College of Urban and Environmental Sciences, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, China*

Corresponding author(s) e-mail: jyfang@urban.pku.edu.cn

ABSTRACT:

Nitrogen (N) and phosphorus (P), especially the N in Rubisco that drives photosynthesis and the P in rRNA that drives the generation and maintenance of proteins, are essential nutrients for plants. As an important plant leaf trait and allometric "rule" in ecology, the scaling relationship between leaf N and P concentrations is crucial to modelling N and P cycles in terrestrial ecosystems. Previous studies have generalized an invariably "constant" law that N scales roughly as the 2/3 or 3/4 power of P (i.e., $N \propto P^{\alpha=2/3 \text{ or } 3/4}$). However, whether the numerical value of the scaling exponent is constant remains unclear and is one of key issues in plant ecology. To address how the numerical value of the scaling exponent changes with functional groups and environmental conditions, we compiled a global data set and found that the exponent varied significantly across different functional groups, latitudinal zones, ecoregions (continents), and sites. The exponents of herbaceous and woody plants were 0.659 and 0.705, respectively. Among woody plants, the exponents of coniferous, deciduous and evergreen broad-leaved species were 0.610, 0.712 and 0.731, respectively. The exponents also showed significant latitudinal patterns, decreasing from tropical to temperate to boreal zones. Further, across the ecoregions of North America, Europe, Asia, Oceania, Africa, and South America, the exponents were 0.603, 0.672, 0.712, 0.786, 0.835, and 1.071, respectively. At sites with a sample size >10, the values fluctuated from 0.366 to 1.928, with an average of 0.841. Such large numerical variations of the N vs. P scaling exponents likely reflect species composition, P-related growth rates, relative nutrient availability of soils and a number of other factors. Our results therefore indicated that there is no canonical numerical value for the leaf N vs. P scaling exponent. The traditional analysis of pooled data at global scale for this scaling relationship hides biologically and ecologically significant variation. This finding has a critical bearing on the parameterization of N and P biogeochemical models.

KEYWORDS: Nitrogen, phosphorus, scaling exponent, functional group, latitudinal pattern, leaf

How do nitrogen and phosphorus supply affect elemental stoichiometry in plant leaves? Using experimental manipulations of *Arabidopsis thaliana*

Zhengbing Yan[1]***, Wenxuan Han***[2]***, Di Tian***[3]***, Jingyun Fang***[1]

[1]*Peking University, China*; [2]*China Agricultural University, China,* [3]*Capital Normal University, China*

Corresponding author(s) e-mail: jyfang@urban.pku.edu.cn

ABSTRACT:

Ambient nutrient changes influence the coupling of nitrogen (N) and phosphorus (P) in terrestrial ecosystems, but whether it could alter the scaling relationship of plant leaf N to P concentrations remains unclear. Moreover, knowledge about how multi-elemental stoichiometry responds to varying N and P availabilities remains limited. Here we conducted experimental manipulations using Arabidopsis thaliana, with five N and P addition levels and nine repeated experiments. We found that the scaling exponent of leaf N to P concentrations decreased with increasing N levels, but increased with increasing P levels. This suggests that high nutrient availability decreases the variability of its own concentration, but promotes the fluctuation in another tightly associated nutrient concentration in leaves among plant individuals. We call this as Nutrient Availability–Individual Variability Hypothesis. In addition, N and P supply exerted differential influences on the concentrations of multi-elements in leaves. Compared with the green-leaves, the senesced-leaves had higher variability of C, N, P, K and Mg concentrations but lower variability of Ca concentration under varying nutrient availabilities. This suggests that stage-dependent pattern of leaf stoichiometric homeostasis relies on the type of elements, and the elemental feature should be considered when choosing a more favorable tissue in plants for diagnosing the nutrient availability in ambient environments. These findings provide a novel mechanism for understanding the dynamic processes of population structure and functioning under global nutrient changes, which should be incorporated into modeling stoichiometric growth in terrestrial ecosystems. Furthermore, our study can advance the holistic understanding about plant eco-physiological response and adaption under global nutrient changes from the stoichiometric perspective of multiple elements beyond N and P.

KEYWORDS: Arabidopsis thaliana, multiple elements, N and P supply, Nutrient Availability-Individual Variability Hypothesis, plant stoichiometry, scaling relationship

The TRY Database

***Gerhard Boenisch*[1], *Jens Kattge*[1,2], *Christian Wirth*[1,2,3]**

[1] *Max Planck Institute for Biogeochemistry, Hans Knöll Str. 10, 07745 Jena, Germany*
[2] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*
[3] *University of Leipzig, Department Systematic Botany and Functional Biodiversity, Johannisallee 21, 04103 Leipzig, Germany*

Corresponding author(s) e-mail: boenisch@bgc-jena.mpg.de

ABSTRACT:

The TRY initiative (www.try-db.org) is a network of vegetation scientists providing curated plant trait data for the scientific community. The TRY Database currently contains about 7 million trait records for nearly 3000 different traits. The flexible database structure can hold any number of traits and a generic program can import any kind of data without requiring a template. About 10 million trait records for about 100 requests are released on a monthly basis. This is organized via the TRY Data Portal, which facilitates data contribution, exploration and customized requests. The Dataset Custodian Centre allows managing the status of contributed datasets and monitoring the use of these data from requests to scientific publications. The Request PI Centre allows managing and monitoring requests. Both centres facilitate direct contact of data contributors and users. In addition to the TRY Database we have established a file archive, which facilitates publication and DOIs for else unpublished plant trait datasets. The TRY Data Portal has evolved toward a long-term scientific data infrastructure, which combines the advantages of easy access to curated plant trait data almost ready for analyses, with direct contact of data providers and users, the opportunity for data providers to publish individual datasets and track the use of their data. This presentation will introduce details of the TRY database system.

KEYWORDS: Database, traits

Assessing among-lineage variability in phylogenetic imputation of functional trait datasets

***Rafael Molina-Venegas*[1], *Juan Carlos Moreno-Saiz*[2], *Isabel Castro Parga*[2], *T. Jonathan Davies*[3], *Pedro R. Peres-Neto*[4], *Miguel Á. Rodríguez*[1]**

[1]*Universidad de Alcalá, Spain*; [2]*Universidad Autónoma de Madrid, Spain*; [3]*McGill University, Canada;* [4]*Concordia University, Canada.*

Corresponding author e-mail: rmolina@us.es

ABSTRACT:

The distribution of plant functional diversity across spatial scales is of fundamental interest to ecologists. However, even one of the largest and most comprehensive trait databases compiled to date (i.e. the TRY Plant Trait Database [1]) is highly incomplete for many species, and trait-based studies often have to deal with missing data. Recently, phylogenetic imputation has emerged as a promising tool for filling such gaps [2,3], and thus understanding the limitations of phylogenetic modelling in predicting missing data is critical if we are to use this information in subsequent analyses. Previous studies have focused on the relationship between phylogenetic signal and clade-level prediction accuracy [4], yet variability in prediction accuracy among individual tips of phylogenies remains largely unexplored. Here, we used simulations of trait evolution along the branches of phylogenetic trees to show how the accuracy of phylogenetic imputations is influenced by the combined effects of (i) the amount of phylogenetic signal in the traits and (ii) the branch length of the tips to be imputed. Specifically, we conducted cross-validation trials to estimate the variability in prediction accuracy among individual tips on the phylogenies (hereafter “tip-level accuracy”). We found that under a Brownian motion model of evolution (BM, Pagel’s $\lambda = 1$), tip-level accuracy rapidly decreased with increasing tip branch-lengths, and only tips of approximately 10% or less of the total height of the trees showed consistently accurate predictions (i.e. cross-validation R-squared > 0.75). When phylogenetic signal was weak, the effect of tip branch-length was reduced, becoming negligible for traits simulated with $\lambda < 0.7$, where accuracy was in any case low. Our study shows that variability in prediction accuracy among individual phylogenetic tips should be considered when evaluating the reliability of phylogenetically imputed trait values. To address this challenge, we describe a Monte Carlo-based method that allows one to estimate the expected tip-level accuracy of phylogenetic predictions for continuous traits. Our approach identifies gaps in functional trait datasets for which phylogenetic imputation performs poorly, and will help ecologists to design more efficient trait collection campaigns by focusing resources on lineages whose trait values are more uncertain.

KEYWORDS: Phylogenetic imputation, Missing data, Traits, Prediction accuracy, Phylogenetic tips

REFERENCES:

1. Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bonisch, G., et al. 2011. TRY – a global database of plant traits. Glob. Change Biol. 17, 2905–2935.
2. Swenson, N.G. 2014. Phylogenetic imputation of plant functional trait databases. Ecography. 37, 105–110.

3. Swenson, N.G., Weiser, M.D., Mao, L., Araújo, M.B., Diniz-Filho, J.A.F., Kollmann, J., et al. 2017. Phylogeny and the prediction of tree functional diversity across novel continental settings. Glob. Ecol Biogeogr. 26, 553–562.
4. Guénard, G., Boisclair, D., Legendre, P. 2015. Phylogenetics to help predict active metabolism. Ecosphere. 6, 1–11.

SCOPE model inversion for Sentinel-3 data retrieval

Egor Prikaziuk[1], **Christiaan van der Tol**[1]

[1]*Department of Water Resources, Faculty ITC, University of Twente, Netherlands*

Corresponding author(s) e-mail: e.prikaziuk@utwente.nl

ABSTRACT:

The SCOPE is a coupled radiative transfer and energy balance model used for simulation of vegetation optical properties and temperature at leaf and canopy level over a spectral range from 0.4 to 50 µm. Inversion of the model allows retrieving a number of plant traits: pigments (Cab, Car, Cant), dry matter content (Cdm), water content (Cw), leaf area index (LAI) and others. Subsequent forward simulation can calculate photosynthesis, evapotranspiration (ET) and a fraction of absorbed photosynthetically active radiation (fAPAR) that can be used further for integrated water use efficiency (WUE) and light use efficiency (LUE) calculations, respectively. The higher the accuracy in retrieved parameters is achieved the higher precision in calculated ecosystem functional properties will be.

This work aimed to develop a model-based retrieval algorithm from multispectral satellite data. The initial retrieval algorithm used numerical optimization of residuals squared sum and operated over the spectral range from 0.4 to 2.4 µm. First, the algorithm was extended to the thermal domain (up to 50 µm) and validated against open-source spectral measurement datasets (SPECCHIO). As the SCOPE model operates at both leaf and canopy levels, we had to use different cost functions and constraints for each level. Having validated the hyperspectral retrieval algorithm, we tried to make a convolution to the multispectral case of Sentinel-3 satellite sensors: ocean and land colour instrument (OLCI) and sea and land surface temperature radiometer (SLTR). Finally, parameter retrieved with the algorithm from Sentinel-3 images were used for a forward simulation of the SCOPE model and calculation of integrated WUE and LUE at few selected FLUXNET towers. The results of the simulation were validated against data from FLUXNET eddy-covariance towers.

KEYWORDS: Remote sensing, Ecosystem, Retrieval, Sentinel-3, SCOPE

Plant functional groups for carbon and nitrogen cycle modelling and diversity estimation in boreal forest ecosystems

***Larisa Khanina*[1], *Maxim Bobrovsky*[2], *Vadim Smirnov*[1, 3], *Natalia Lukina*[3], *Michael Romanov*[1]**

[1]*Institute of Mathematical Problems of Biology RAS – Branch of the M.V. Keldysh IAM RAS, Russia*; [2]*Institute of Physico-Chemical and Biological Problems of Soil Science RAS, Russia*; [3]*Center for Forest Ecology and Productivity RAS, Russia*

Corresponding author(s) e-mail: khanina.larisa@gmail.com

ABSTRACT:

We have developed a system of plant functional groups (PFGs) for modelling dynamics and biodiversity estimation of forest ground vegetation. Species Ellenberg ecological values and phytocoenotic characteristics together with morphological traits were used to classify plant species into the functional groups [1]. Ground vegetation patches dominated by species of different PFGs were distinguished in the forest floor. We have characterized biomass of these patches performing statistical analysis of biomass values taken from published data and results of field investigations combined into the database "Biomass" [2]. The database contains data on aboveground and belowground biomass of species and groups of species occurred in the ground layer of boreal and hemiboreal forests of European Russia. There are more than 7000 records on biomass and productivity for 196 vascular species, 32 bryophytes and 4 lichen species and for tens groups of species. Biomass of vascular plants and bryophytes quite well differed between the patches dominated by species of different PFGs. According to statistical analysis results, the patches explain much more of biomass variance than the forest types; differences in the influence of the factors are more pronounced for belowground than for aboveground biomass. The values of the belowground biomass for all groups were higher than the values of the aboveground biomass, and differed between the analyzed groups to a greater extent than the values of the aboveground biomass. Database "Elements" on element concentrations of plants was also developed on the basis of results of field investigations in the North and the Middle Taiga. The database contains data on more than 2500 samples of 80 plant species. The PFGs also differed well in the content of the chemical elements in species that comprise the group. The groups differed most in the following elements: N, Mg, Ca, P, S, Mn, Al, Zn, Fe and slightly less C. We have proposed a conceptual model of dynamics of ground forest vegetation; a model compatible with the individual-based EFIMOD model of forest growth, carbon and nitrogen cycles in forest ecosystems [3]. Spatial unit of the ground vegetation model is the same as spatial unit of the EFIMOD (0.5×0.5 м); it is a patch of the forest floor dominated by species from one or two functional groups. Development and analysis of the database “Elements” was supported by the Russian Science Foundation, project 16-17-10284. The results on modelling of ground vegetation were received under umbrella of POLYFORES project (ERA-NET Sumforest).

KEYWORDS: forest-soil system, ground vegetation, life forms, plant biomass, element concentration

REFERENCES:

1. Khanina, L.G., Bobrovsky, M.V., Smirnov, V.E., Grozovskaya, I.S., Romanov, M.S., Lukina, N.V., Isayeva, L.G., 2015. Ground vegetation modeling through functional species groups and patches in the forest floor. Mathematical Biology and Bioinformatics. 10, 15–33.

2. Khanina, L.G., Grozovskaya, I.S., Smirnov, V.E., Romanov, M.S., Bobrovsky, M.V., 2013. Analysis of database on the biomass of forest ground vegetation for modelling plant dynamics in forest ecosystem models. Khvoinye borealnoi zony (Coniferous in the Boreal Region). 31, 22–29. (in Russ.)

3. Komarov, A.S., Chertov, O.G., Zudin, S.L., Nadporozhskaya, M.A., Mikhailov, A.V., Bykhovets, S.S., Zudina, E.V., Zoubkova, E.V., 2003. EFIMOD 2 – a model of growth and cycling of elements in boreal forest ecosystems. Ecol. Model. 170, 373–392.

A general framework for global mapping of plant traits with operational satellites and climatological data

***Alvaro Moreno-Martinez*[1,2], *Gustau Camps-Valls*[2], *Nuno Carvalhais*[3], *Jens Kattge*[3], *Markus Reichstein*[3], *Emma Izquierdo*[4], *Daniel Heestermans Svendse*[2], *and Steven W. Running*[1]**

[1] *Numerical Terradynamic Simulation Group (NTGS), University of Montana, USA.*
[2] *Image Processing Laboratory (IPL), Universitat de València, Spain.*
[3] *Max-Planck-Institute for Biogeochemistry (MPI-BGC), Jena, Germany*
[4] *Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands.*

Corresponding author(s) e-mail: alvaro.moreno@ntsg.umt.edu

ABSTRACT:

Plant traits characterize species and have been used intensively for ecosystem modeling. In most of these models functional properties of vegetation are over simplified to a reduced number (~10) of plant functional types (PFTs) with similar physical, phenological and physiological traits. Such groupings ignore important variability within individual PFTs, and constitute a significant source of errors in model estimates. Solutions to this issue include the use of plant trait databases and potential covariates (eg. meteorological data) to move forward from the static PFT parameter paradigm and create continuous maps of plant traits. We propose and validate a general framework to estimate global maps (500 m) of plant traits using modern machine learning techniques along with remote sensing (MODIS/Landsat) and climate data. Due to the massive amount of data needed and the high computational demands, most of the work has been implemented exploiting the Google Earth Engine (GEE) cloud platform. The developed processing chain uses a gap filling to maximize the global representativeness of the global plant trait database (TRY). We aggregate plant trait species measurements to Plant Functional Types (PFT) to calculate the spatial abundance of each PFT at 500 m spatial resolution. These abundances are computed by means of a global high resolution (30 m) PFT map developed using Landsat data. Based on these PFT abundance estimates, canopy representative trait values are calculated at a 500 m spatial resolution with nearby plant trait observations. Finally, the spatialization of these canopy representative trait values is carried out using random forests in combination with satellite and climatological data. The validation and assessment of our methods indicate that, on average, the models explain 65% of the spatial variance with low RMSE (<15%) and virtually no bias. The produced maps offer opportunities to study the patterns of variation of plant traits at global coverage with a very high spatial resolution from a modeler's perspective.

KEYWORDS: Plant Traits, Remote Sensing, MODIS, Machine Learning, Cloud computing.

Potentials and challenges of remote sensing functional trait diversity by integrating high-resolution spaceborne multispectral and radar measurements

Xuanlong Ma[1,2]***, Miguel Mahecha***[1, 2]***, Mirco Migliavacca***[1,2]***, Talie Musavi***[1]***, Fons van der Plas***[3]***, Ronny Richter***[2, 3]***, Sophia Ratcliffe***[4]***, Raquel Benavides***[5]***, Daniel Pabon***[1]***, Timo Domisch***[6]***, Leena Finér***[6]***, Jens Kattge***[1]***, Markus Reichstein***[1,2]***, Christian Wirth***[2,3]

[1] Max Planck Institute for Biogeochemistry, Hans Knoll Strasse 10, 07745, Jena, Germany; [2] German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103, Leipzig, Germany; [3] University of Leipzig, Johannisallee 21-23, 04103, Leipzig, Germany; [4] National Biodiversity Network, 14-18 St Mary's Gate, Lace Market, NG1 1PF, Nottingham, UK; [5] National Museum of Natural Science, CSIC, Serrano 115 dup., 28006, Madrid; [6] Natural Resources Institute Finland, Yliopistokatu 6, FI-80100, Joensuu, Finland

Corresponding author e-mail: xma@bgc-jena.mpg.de

ABSTRACT:

Assessing functional trait diversity from space can improve our understanding about the large-scale relationship between functional diversity and environmental factors and provide a complementary approach to field experiments for assessing the biodiversity-ecosystem functioning links across space and time. Here we aim to develop the statistical method for upscaling functional diversity to a regional scale using ESA's (European Space Agency) high-resolution Sentinel-1 and Sentinel-2 measurements (hereafter S1 and S2). The S2 spaceborne spectroradiometer provides global coverage of up to 10 m spatial resolution over 13 spectral bands, suitable for resolving fine details of variations in vegetation dynamics, and the C-band SAR (synthetic aperture radar) instrument onboard S1 satellites provides information about landscape topography and canopy structure that is complementary to the S2 optical instrument. We plan to compute community weighted means and diversity of functional traits based on a combination of *in-situ* measured and compiled trait and species abundance data over 209 forest plots across Europe. For these plots, we will extract S1 radar backscatter and S2 spectral reflectance measurements. We will explore the link between functional traits/diversity and S1/S2 signals with the aim of developing a robust statistical algorithm for upscaling functional diversity to a regional scale. The expected technical challenges include 1) to what extent we could resolve variations in canopy structure and physiological traits using spaceborne measurements with limited spatial and spectral resolutions; 2) how to statistically identify the correlations among different predictors and their non-linear relationships with response variables. The expected outcomes include functional diversity derived from field plot data across a wide range of climate conditions representing major European forest types and improved understanding of the link between spaceborne optical and radar signals with functional traits/diversity over different forest structures.

KEYWORDS: plant traits, functional diversity, remote sensing, functional biogeography

How are spectrally relevant plant traits distributed across plant functional gradients?

***Teja Kattenborn*[1], *Sebastian Schmidtlein*[1]**

[1]*Institite of Geography and Geoecology (IFGG), Karlsruhe Indtitute of Technology (KIT), Germany*

Corresponding author(s) e-mail: teja.kattenborn@kit.edu

ABSTRACT:

Various plant traits that affect the spectral properties of plant canopies can be retrieved using optical earth observation data and thus enable to map plant functional properties (Kattenborn 2017). From a remote sensing perspective, the mechanistic response of these optically relevant plant traits is already quite well understood and formulated in process-based models, i.e. canopy radiative transfer models (RTMs). The latter model the reflectance of plant properties using the sun and observer (sensor) orientation and defined plant traits. However, the relationship of these traits towards plant functioning was not systematically assessed. Thus, the present study examines how spectrally relevant traits (those implemented in PROSAIL) are related to two established plant functional schemes, i.e. the leaf economic spectrum (LES) and CSR plant strategies. The trait expressions were measured in-situ in 42 cultivated herbaceous plants. As expected these plant traits indeed relate to the assessed gradients of plant functioning (LES and CSR). As expected traits related to leaf properties (e.g. pigments and dry matter content) show clear correspondence to the LES. Traits related to the canopy structure show no or very little correspondence to the LES but clearly relate to CSR plant strategies which reflect plant functioning at the level of plant individuals or communities. Multiple trait expressions such as LAI, canopy foliage mass (LMA * LAI), faPAR and fAPAR integrated over a growing season feature comparable or even higher correlations to the CSR space than traits that were originally used to allocate CSR scores (e.g. LMA or LDMC). Our results therefore highlight that spectrally relevant plant traits are a valuable alternative or addition to traits traditionally used in trait-based ecology. These traits might not only enrich the suite of potential indicators to characterize plant functional gradients using EO data; they also allow to establish physical and therefore explicit relationships which advance our theoretical understanding as well as the operationalization of such knowledge into mapping and monitoring approaches.

KEYWORDS: Plant traits, Remote Sensing, Radiative Transfer Model, PROSAIL, Hyperspectral

REFERENCES:

1. Kattenborn, T., Fassnacht, F. E., Pierce, S., Lopatin, J., Grime, J. P., Schmidtlein, S. 2017. Linking plant strategies and plant traits derived by radiative transfer modelling. Journal of Vegetation Science, 28(4).

Remote sensing for the observation of senescence in Conference pear trees

Laura Paladini[1,2]***, Stephanie Delalieux***[1]***, Laurent Tits***[1]***, Ben Somers***[2]

[1]*VITO - Flemish Institute for Technological Research, Belgium*; [2]*K.U. Leuven, Division of Forest, Nature and Landscape, Belgium*

Corresponding author e-mail: laura.paladini@vito.be

ABSTRACT:

Leaf senescence in trees is the phenological stage during which nutrient resorption happens. In this process, part of the nutrients is transferred to the perennial organs of the plant, contributing to tree vitality and, in pome fruit trees, to flowering intensity the following year. Another share of the nutrients remains inside leaf litter and enters the agroecosystem's nutrient cycles. The timing and duration of leaf senescence influences the ratio between the two parts of nutrients and thus influences nutrient cycles in the agroecosystem.

Among innovative techniques to investigate these processes, satellite remote sensing has proved a valid tool in natural ecosystems. The same cannot be said about fruit orchards, because of the image quality of the satellites active before Sentinel-2, often deemed insufficient for agricultural studies. The features of Sentinel-2, instead, offer new possibilities for monitoring phenology in agricultural environments.

This research aims to study senescence in Conference pear trees, in three regions of Flanders (Belgium). One cloud-free Sentinel-2 image, acquired in the middle of the senescence period, was analysed, by means of different spectral indices. Ground data was collected through a network of 34 webcams with an RGB camera. A visual analysis was performed, to determine the beginning of senescence (the moment in which the first yellow/red leaves appear in the canopy) and the end of senescence (the moment in which the entire canopy turns yellow/red).

Webcam data showed that leaf (dis)colouration started between September and October, during a one-month time span. Full discolouration of the canopy, occurring at the end of November, was instead more synchronous. Moreover, some trees only turned yellow, while others showed red leaves, probably a stress indicator.

Sentinel-2 data revealed that spectral indices correlate well with the date of the beginning of senescence, thus suggesting that it would possible to map it.

These results already offer evidence that monitoring variability in the dynamics of senescence is possible from satellite remote sensing. Current focus is on the link between canopy colour, as it appears in the webcam imagery, and satellite data.

KEYWORDS: phenology, senescence, pear, Sentinel-2, remote sensing.

Towards the Automatic Extraction of Plant Traits from Textual Descriptions

Onatkut Dagtekin[1], William Ulate[2] and Riza Batista-Navarro[1]

[1]The University of Manchester, UK; [2]Missouri Botanical Garden, USA

Corresponding author(s) e-mail: riza.batista@manchester.ac.uk

ABSTRACT:

Many ecological restoration programmes are informed by evidence coming from empirical research. Specifically, such programmes analyse species traits in order to differentiate species that are suitable for restoration from unsuitable ones. Indeed, understanding plant traits (and their relationships with each other) informs research into vegetation modelling and environmental change prediction, which in turn help in answering many ecological questions. In 2006, the Center for Tropical Forest Science (CTFS) formulated recommendations in support of their research programme, the foremost of which is the creation of trait databases by building upon published information catalogued by existing herbaria.

In this work, we aim to enrich World Flora Online (WFO), a web-based inventory of known plant species, by integrating trait information contained in data sets coming from botanical institutions all over the world. This poses a few challenges, as trait information tends to be buried within verbose textual descriptions and do not conform with conventions of writing. Specifically, they typically do not come in the form of full sentences and look like long-winded enumerations of various types of plant attributes or characteristics. Such descriptions are difficult to search and understand unless decomposed into meaningful units.

In order to decompose textual descriptions of plant species into spans pertaining to specific types of attributes, we have developed a machine learning-based approach to automatic text segmentation. Casting the problem as a sequence labelling task, we have investigated a number of probabilistic classifiers including conditional random fields (CRFs), hidden Markov models (HMMs) and naïve Bayes (NB). To train our models, we utilised data contributed by the South African National Biodiversity Institute (SANBI) which contain traits labelled as one of the following trait categories: morphology, habitat and distribution. To help the models discriminate between these categories, we designed features capturing word characteristics (e.g., n-grams at the character and word level), context (i.e., surrounding words within a predefined window), as well as domain knowledge (i.e., words that match terms in plant-related ontologies). In this way, we can automatically elucidate exactly which parts of the original descriptions pertain to plant traits such as morphology, habitat or distribution. By applying the resulting models on textual descriptions coming from several botanical institutes, we can facilitate the automatic population of WFO with plant traits for a number of species.

KEYWORDS: Text Segmentation, Plant Traits, Machine Learning, World Flora Online

Phenological diversity is linked to the diversity of functional traits in alpine grasslands

Gianluca Filippa[1], Edoardo Cremonese[1], Marta Galvagno[1], Mirco Migliavacca[2], Christine Römermann[3], Solveig Franziska Bucher[4], Talie Musavi[2]

[1]*Environmental Protection Agency Aosta Valley, Italy - Climate change unit*; [2]*Max Planck Institute for Biogeochemistry Department Biogeochemical Integration, Jena, Germany*; [3]*Institut für Ökologie und Evolution Friedrich-Schiller Universität, Jena, Germany;* [4]*Institut fur Spezielle Botanik Friedrich-Schiller Universität Jena, Germany*

Corresponding author(s) e-mail: gian.filippa@gmail.com

ABSTRACT:

Climate change is exerting a profound influence on the timing of seasonal development of vegetation, i.e. phenology, worldwide - and particularly in mountain grasslands. A key feature controlling the resilience of alpine plant communities to climate change is phenological plasticity: the ability to express different phenologies within a given plant assemblage can help the community to better cope with climate shifts and extremes. Recent work has shown a tight relationship between phenology and certain plant functional traits, especially those associated to competition and growth rate.

Phenocameras proved to be an effective mean to monitor community-level phenology by retrieving average phenological signals across a portion of the field of view of digital cameras. Here we used pixel-level information to quantify spatially-explicit phenology from multi-year imagery acquired over 5 alpine grasslands in the western Alps. Concurrently, based on site-specific species inventories we retrieved information about functional traits from the global plant trait database TRY.

The objectives of this work are:

1) to assess the link between phenological diversity and plant functional trait diversity in alpine grasslands;

2) to test the consistency in space and time of the functional diversity-phenological diversity relationship.

A total of 19 year-sites of phenocam images were processed to obtain maps of phenological metrics. We focused on either spring (start of season, moment of greenness peak, spring recovery rate), autumn (end of season, start of senescence, autumn senescence rate) or full season (length of season, greenness integral) metrics. To describe the spatial distribution of phenological metrics we used indexes such as the Moran Index and the Entropy index. The FD R package was used to compute multidimensional functional diversity.

We will illustrate the traits that best correlate with phenological diversity across the 5 different grasslands included in this study and discuss inter-year and inter-site variability in the relationships in the light of climate variability.

KEYWORDS: Phenocameras, Leaf N content, Climate change, TRY.

Spatiotemporal trends of vegetation carbon residence time across forest biomes: linking plant traits, soil properties, and climate change

***Kailiang Yu*[1,2], *William Anderegg*[1], *Tom Crowther*[2]**

[1]*University of Utah, USA*; [2]*ETH Zürich, Switzerland*

Corresponding author(s) e-mail: kailiang.yu@usys.ethz.ch

ABSTRACT:

Forests cover ~33% of the land's surface area and they represent a prominent control on global carbon cycling. The capacity of forests to sequester atmospheric CO_2 in a changing climate depends not only on the response of carbon uptake but also on the simultaneous change in carbon residence time. Unlike carbon uptake, spatiotemporal trends of carbon residence time remain poorly studied and we lack a basic understanding of the dominant drivers governing carbon residence times across landscapes [1,2]. This constrains our capability to predict future changes of the forest carbon sink and its feedbacks to climate. While global climate change drivers, plants traits, and soil properties are expected to considerably influence carbon residence time, studies are scarce and their impacts remain poorly quantified. In this study, we used long-term forest plot data, remote sensing, and Earth system models to quantify spatiotemporal trends in vegetation carbon residence time and its drivers across forest biomes. We found the pervasive vegetation carbon residence time across forest biomes, despite of the large spatial variation in temporal changes of vegetation carbon residence time across landscapes. We also examined how this large spatial variation in temporal changes of vegetation carbon residence time could be linked to plant traits, soil properties and climate change.

KEYWORDS: Carbon residence time, forest biomes, plant traits, soil, climate change

REFERENCES:

[1] Carvalhais N, et al. (2014). Nature. 514, 213–217.

[2] Friend AD, et al. (2014). PNAS. 111, 3280–3285.

Session S3.6

EFFICIENT DATA AND WORKFLOW MANAGEMENT FOR REACHING SUSTAINABLE DEVELOPMENT GOALS (SDGS) TARGETS ASSOCIATED TO BIODIVERSITY AND ECOSYSTEMS

Efficient data and workflow management for reaching Sustainable Development Goals (SDGs) targets associated to biodiversity and ecosystems

***Jan Bumberger*[1], *Jitendra Gaikwad*[2], *Aidin Niamir*[3]**

[1]*Helmholtz Centre for Environmental Research – UFZ, Germany*; [2]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany*; [3]*Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Germany*

Corresponding author(s) e-mail: jan.bumberger@ufz.de

ABSTRACT:

The Sustainable Development Goals (SDGs) are a collection of 17 global goals, including zero hunger, good health and well-being, climate action, clean water and sanitation, affordable and clean energy that has been agreed by international communities to be hopefully achieved by 2030. Monitoring progress towards these goals require a reliable data and information which are accessible and reproducible over time and space. Earth Observation (EO) products can potentially address such a need for trusted sources of data to monitor the trends of environmental conditions (i.e. essential variables), and also inform models to predict progress (i.e. indicators) towards policy targets over variety of scenarios. However, efficient management of big earth observation datasets and reproducible modeling workflows remains a challenge.

In this session we want to bring together experts representing broad range of experience in applications related to SDGs with a special focus on biodiversity and ecosystem services. We invite contributions presenting challenges, solutions, cases studies, and best practices dealing with big data and modeling workflows management. We intend to discuss on the list (not excluding) of following topics:

- Efficient data management approaches along the chain of information from field data to derived indicators taking into account the uncertainties.

- Standardized and operationalized data quality assurance and fusion approaches for biotic, abiotic and other EO data.

- From data to variable, and from variable to indicator workflows for biodiversity and ecosystem services.

- Best practices to improve interoperability.

- Uncertainty/data reliability.

KEYWORDS: Sustainable Development Goals (SDGs), Earth Observation (EO), Data management, Data quality assurance, Data Fusion, Modeling workflows management, Indicator workflows, Biodiversity, Ecosystem services, Uncertainty/data reliability

Data lifecycle is not a cycle, but a plane!

***Javad Chamanara*[1], *Birgitta König-Ries*[1]**

[1]*Friedrich Schiller University Jena, Germany*

Corresponding author(s) e-mail: javad.chamanara@uni-jena.de

ABSTRACT:

Most of the data-intensive scientific domains, e.g., life-, natural-, and geo-sciences have come up with data life cycles. These cycles feature, in various ways, a set of core data-centric steps, e.g., planning, collecting, describing, integrating, analyzing, and publishing. Although they differ in the steps they identify and the execution order, they collectively suffer from a collection of short-comings.

They mainly promote a waterfall-like model of sequentially executing the lifecycles' steps. For example, the lifecycle used by DataOne suggests that “analyze” happens after "integrate". However, in practice, a scientist may need to analyze data without performing the integration. In general, scientists may not need to accomplish all the steps. Also, in many cases, they simply jump from, e.g., "collect" to "analyze" in order to evaluate the feasibility and fitness of the data and then return to "describe" and "preserve" steps. This causes the cycle to gradually turn into a mesh. Indeed, this problem has been recognized and dealt with by the GFBio and USGS data lifecycles. The former has added a set of direct links between non-neighboring steps to allow shortcuts, while the later has factored out cross-cutting steps, e.g., "describe" and "manage quality" and argued that these tasks must be performed continually across all stages of the lifecycle. Although aforementioned lifecycles have realized these issues, they do not offer customization guidelines based on, e.g., project requirements, resources availability, priority, or effort estimations.

In this work, we propose a two-dimensional Cartesian-like plane, in that the x- and y-axes represent phases and disciplines, respectively. A phase is a stage of the project with a predefined focus that that leads the work towards achieving a set of targeted objectives in a specific timespan. We identify four phases; conception, implementation, publishing, and preservation. Phases can be repeated in a run, and do not need to have equal timespan. However, each phase should satisfy its exit criteria to be able to proceed to the next phase. A discipline, on the vertical axis, is a set of correlated activities that, when performed, makes a measurable progress in the data-centric project. We have incorporated these disciplines: plan, acquire, assure, describe, preserve, discover, integrate, analyze, maintain, and execute.

An execution plan is developed by placing required activities in their respective disciplines' lanes on the plane. Each task (activity instance) is visualized as a rectangle that its width and height respectively indicate the duration and effort estimation needed to complete it. The phases, as well as the characteristics of the project (requirements, size, team, time, and budget), may influence these dimensions.

It is possible for a discipline or an activity to be utilized several times in different phases. For example, a planning activity gains more weight in conception and fades out over the course of the project, while analysis activities start in mid-conception, get full focus on implementation, and may still need some attention during publishing phases. Also, multiple activities of different disciplines can run in parallel. However, each task's objective should remain aligned according to the phase’s focus and exit criteria. For instance, an analysis task in the conception phase may utilize multiple methodologies to perform experimentation on a small sample of a designated dataset, while the same task in the implementation phase conducts a full-fledged analysis using the chosen methodology on the whole dataset.

KEYWORDS: Data Lifecycle, Data Management, Research Data Management, Scientific Data Management

Data science for environmental health

Angela Lausch[1/2], Peter Dietrich[1], Toralf Kirsten[3], Sonja Knapp[1], Josef Settele[1], Steffen Zacharias[1], Jan Bumberger[1]

[1]*Helmholtz Centre for Environmental Research – UFZ, Germany;* [2]*Humboldt University, Germany;*
[3]*University of Applied Sciences, Mittweida*

Corresponding author(s) e-mail: angela.lausch@ufz.de

ABSTRACT:

Ecosystems fulfill a whole host of ecosystem functions that are essential for life on our planet. However, an unprecedented level of anthropogenic influences is reducing the resilience and stability of our ecosystems as well as their ecosystem functions. The relationships between drivers, stress and ecosystem functions in ecosystems are complex, multi- faceted and often non-linear and yet environmental managers, decision makers and politicians need to be able to make rapid decisions that are data-driven and based on short- and long-term monitoring information, complex modeling and analysis approaches. A huge number of long-standing and standardized ecosystem health approaches like the essential variables already exist and are increasingly integrating remote-sensing based monitoring approaches [1-2]. Unfortunately, these approaches in monitoring, data storage, analysis, prognosis and assessment still do not satisfy the future requirements of information and digital knowledge processing of the 21st century. This presentation therefore discusses the requirements for using Data Science as a bridge between complex and multidimensional Big Data for environmental health.

It became apparent that no existing monitoring approach, technique, model or platform is sufficient on its own to monitor, model, forecast or assess vegetation health and its resilience. In order to advance the development of a multi-source ecosystem health monitoring network, we argue that in order to gain a better understanding of ecosystem health in our complex world it would be conducive to implement the concepts of Data Science with the components: (i) digitalization, (ii) standardization with metadata management adhering to the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles, (iii) Semantic Web, (iv) proof, trust and uncertainties, (v) complex tools for Data Science analysis and (vi) easy tools for scientists, data managers and stakeholders for decision-making support [3-4].

KEYWORDS: Ecosystem health; Data Science; Semantic Web; Spectral Traits/Spectral Trait Variation approach; Remote Sensing; Ecosystem Integrity

REFERENCES:

1.Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E.; Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., Cord, A.F., 2016. Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. Ecol. Indic. 70, 317–339. doi:10.1016/j.ecolind.2016.06.022.

2.Lausch, A., Erasmi, S., Douglas, J., King, Magdon, P., Heurich, M., 2016. Understanding forest health with remote sensing - Part I - A review of spectral traits, processes and remote sensing characteristics. Remote Sens. 8, 1029; doi:10.3390/rs8121029.

3.Lausch, A.; Bastian O.; Klotz, S.; Leitão, P. J.; Jung, A.; Rocchini, D.; Schaepman, M.E.; Skidmore, A.K.; Tischendorf, L.; Knapp, S. 2018. Understanding and assessing vegetation health by in-situ species and remote sensing approaches. Methods Ecol. Evol. 00, 1–11. doi:10.1111/2041-210X.13025.

4.Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., Jung, A., Klenke, R., Knapp, S., Mollenhauer, H., Paasche, H., Paulheim, H., Pause, P., Schweitzer, C., Schmulius, C., Settele, J., Skidmore, A.K.,, Wegmann, M., Zacharias, S., Kirsten, T.; Schaepman, M.E., 2018. Understanding forest health with remote sensing -Part III - Requirements for a scalable multi-source forest health monitoring network based on Data Science approaches. (Remote Sens., in review).

Green Infrastructures and Essential Variables Workflows towards SDG 15

***Erica Honeck*[1], *Anthony Lehmann*[1], *Diana-Denisa Rodila*[1], *Gregory Giuliani*[1]**

[1]*University of Geneva, Institute for Environmental Sciences, EnviroSPACE, GRID-Geneva, Geneva, CH*

Corresponding author(s) e-mail: erica.honeck@unige.ch

ABSTRACT:

The Sustainable Development Goals (SDGs) established to be achieved by 2030 are an ensemble of 17 goals to address global environmental and social economic concerns [1]. SDG 15 concerns the protection of terrestrial ecosystems to halt biodiversity loss. Target 15.9 states that by 2020, ecosystem and biodiversity values should be integrated into national and local planning, and is related to Aichi Biodiversity Target 2 of the Strategic Plan for Biodiversity 2011-2020, which also involves integrating biodiversity values into national accounting and reporting systems [2].

The importance of maintaining ecosystem integrity is becoming widely recognized, not only to halt biodiversity loss, but also to preserve Nature's benefits to human well-being, and has been included in many other targets such as the EU 2020 Biodiversity Strategy's target 2, which requires the restoration of at least 15% of degraded ecosystems as well as the establishment of green infrastructures to enhance ecosystem services (ES) [3].

The Green Infrastructures (GI) framework is used as a policy tool and promotes the multi-functional use of landscapes to improve biodiversity conservation and benefits to society. It is formulated as a "strategically planned network of natural and semi-natural areas" [4] and is based on three main pillars: key habitats for target species, connectivity and ES [5].

As part of ERA-PLANET's GEOEssential project (Essential Variables workflows for resource efficiency and environmental management), our study aims at demonstrating how the GI framework can be implemented at any geographical area or time-period through reproducible modeling workflows from field data to Essential Variables (EV) data products and policy relevant indicators to monitor and inform advances towards environmental targets.

A proof of concept workflow was already set in place for computing the indicator 15.1.2: *Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem*, while other workflows will follow. The execution platform is the GEOEssential Virtual Laboratory, a cloud-based virtual platform which enables access to, and execution of workflows for the ecosystem science community of practice and even more.

KEYWORDS*:* Green infrastructures, Sustainable Development Goals (SDGs) indicator workflows, Essential Biodiversity Variables (EBVs), ecosystem services, GEOEssential

REFERENCES:

1. UNSD, 2016. Sustainable Development Goals Report.
https://unstats.un.org/sdgs/report/2016/ (accessed 18 May 2018).

2. CBD Secretariat, 2010. The Strategic Plan for Biodiversity 2011-2020, and the Aichi Biodiversity Targets. Secretariat of the Convention on Biological Diversity, Nagoya.

3. European Commission, 2011. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Our life insurance, our natural capital: an EU biodiversity strategy to 2020, Brussels.

4. European Commission, 2013. Green infrastructure (GI) - Enhancing Europe's Natural Capital, Brussels.

5. Liquete, C., Kleeschulte, S., Dige, G., Maes, J., Grizzetti, B., Olah, B., & Zulian, G., 2015. Mapping green infrastructure based on ecosystem services and ecological networks: A Pan-European case study. Environmental Science & Policy, 54, 268–280.

moveStore: an extensible cloud-based framework for applications dealing with movement data

***Babak Naimi*[1,2], *Kamran Safi*[2]**

[1]*ETH Zurich, Switzerland*; [2]*Max Planck Institue for Ornithology, Germany*

Corresponding author(s) e-mail: naimi.b@gmail.com

ABSTRACT:

In recent decades, huge efforts have been put to make multiple-sources of spatial and temporal data available through data repositories, followed by explosive progress in technological developments of data providers, crowdsourcing, drones, sensor networks, etc. that can be expected to continue for the next decades. Examples are global initiatives such as the movebank data repository containing animal movement data at large spatial scales over long time periods, as well as the ICARUS satellite-based animal tracking technology (Wikelski et al., 2007). The question is that whether and to what extent analytical methods are developed/adopted to deal with such progress in data? This research aims to introduce moveStore, a platform that is designed on a cloud to support and boost the progresses in methods/workflows/applications developments, followed the progresses in data developments. Analogous to a data repository, moveStore provides a specific repository for methodological developments that have been implemented as a suit of functions and classes that may also be framed into an application. moveStore is an open-source and extensible framework (Naimi and Araújo, 2016) that intends to facilitate sharing and distributing of software applications dealing with movement data. Users can share the analytical functions, procedures, workflows either packaged as software applications (Apps) or developed as simple functions. The contributions, followed by appropriate metadata, can then be accessible by other users. The applications in moveStore are usually developed using R shiny web framework, and the platform is implemented on a cloud-based computational engine with a user interface on the web through which users can access to all the Apps registered on the platform and search for a specific App (same as in an App store).

KEYWORDS: movement, movebank, cloud computing, web application, shiny

REFERENCES:

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39, 368-375.

Wikelski, M., Kays, R.W., Kasdin, N.J., Thorup, K., Smith, J.A., Swenson, G.W., 2007. Going wild: what a global small-animal tracking system could do for experimental biologists. J Exp Biol 210, 181-186.

Community-curated Linked Open Data about the Sustainable Development Goals, their targets and indicators

***Daniel Mietchen*[1]**

[1]*Data Science Institute, University of Virginia, Charlottesville, VA, USA*

Corresponding author(s) e-mail: dm7gn@virginia.edu

ABSTRACT:

The targets and indicators associated with the Sustainable Development Goals form a complex network of relationships between policy and practice, aspirations and achievements across domains and around the globe. This poses challenges for monitoring, evaluating and communicating progress towards these goals. Wikidata is a Linked Open Data platform that is developed and curated collaboratively by the Wikimedia community as a sister project to Wikipedia. Extending across disciplinary boundaries, it contains information about many of the concepts related to the Sustainable Development Goals, and the community has begun to map the network of relationships between these concepts.

Focusing on examples related to biodiversity and ecosystems, this presentation will explore how Wikidata, its semantic core Wikibase and its global multilingual community can be leveraged to represent and contextualize the Sustainable Development Goals as well as the associated targets and indicators in a way that allows people to make use of that information in the language of their choice, for a given country or on a global level. It will be given on the basis of https://github.com/Daniel-Mietchen/events/blob/master/ICEI2018-SDGs.md, which will be updated on a regular basis until the event. Special emphasis will be given to practical examples for how Wikidata and Wikibase can help integrate information about efforts addressing different components of the SDGs, or in different contexts.

KEYWORDS: Wikidata, Wikibase, multilinguality, Linked Open Data, SPARQL

Methods of calculating indicators of the sustainable development goals using biophysical modelling and classification of the land cover in Nexus Approach

Leonid Shumilo[1,2]***, Nataliia Kussu***[1]***, Mykola Lavreniuk***[1,2]

[1]*Space Research Institute NAS Ukraine and SSA Ukraine, UKR*; [2] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", UKR*;

Corresponding author(s) e-mail: shumilo.leonid@gmail.com

ABSTRACT:

During the GeoEssential project of Era-Planet there was a need for the development of intelligent systems and data processing chains to calculate essential for indicators of sustainable development goals. In this paper, we pro-pose the use of the Nexus approach based on the methodology of classification of satellite images of high resolution using neural network methods of machine learning, analysis of NDVI trend, biophysical model of vegetation growth named Wofost and Stochastic Climate Simulation Model. On the basis of the results of this Nexus approach, it is possible to obtain 15 essential variables related to the state of food, water and energy, while calculating the indicators of sustainable development goals: 15.1.1. Forest area as a proportion of total land area, 15.3.1. Proportion of land that is degraded over total land area,2.4.1. Proportion of agricultural area under productive and sustainable agriculture. This nexus approach is not only effective in monitoring the state of the environment and important natural resources, but also to provide their forecast through the use of the statistical climate model, which is a very useful feature of the proposed approach to the task of adapting humanity to climate change and developing effective control systems.

KEYWORDS: Classification, ERA-Planet, GeoEssential, essential variables, Nexus Approach, machine learning

Automated and efficient workflow for large airborne remote sensing vegetation mapping and research of Natura 2000 habitats

***Adam Kania*[1], *Dominik Kopeć*[2], *Jan Niedzielko*[1], *Łukasz Sławik*[3, 1]**

[1]*MGGP Aero, Poland*; [2]*University of Lódź, Poland;* [3]*University of Warsaw, Poland*

Corresponding author e-mail: kania@definity.pl

ABSTRACT:

The goal of HabitARS (Habitats Airborne Remote Sensing) project is the implementation of remote sensing methods for Natura 2000 habitat monitoring in terms of environmental protection and sustainable agriculture. The methodology of identification of non-forest Natura 2000 habitats and their threats (desiccation, succession, invasive/expansive species) will be developed.

Very extensive field campaign and remote sensing scanning make it probably one of the largest remote sensings project in Europe. Ground truths were collected on multiple sites (spring/summer/autumn) over two vegetation seasons (2016-2017). 200 field campaigns collected 31.500 reference samples.

Aerial scanning was performed using multi-sensor platform, integrating Riegl full-wave LiDAR, 50Mpix RGB camera and HySpex hyperspectral scanner (SWIR-384, VNIR-1800). In 400 h flights hours 2300 km2 of data was acquired, achieving 7 ALS points/m2 and ground pixels sizes of 10 cm (RGB) and 1 m (hyperspectral).

We automated classification workflow to free teams from repetitive work. A system based on Vegetation Classification Studio software was deployed in datacenter, integrating all datasets and facilitating remote access to research teams from 6 universities.
Simple text language defines input data, goals and parameters of scientific experiments. Batches of 10-100s of related experiments can be defined. The basic process automates e.g. preparation of reference data, splitting into training/validation sets, rasterization, model learning & validation, quality assessment & reporting and produces a set of final vegetation maps in multiple formats. Often in minutes.

Many procedures/algorithms performing multiple cycles of classification, prediction and accuracy assessment – like feature selection, optimizations in search for specific target, or search over multiple parameters – are automated. Modern approaches e.g. fuzzy prediction and multiple fuzzy visualizations, dimensionality reduction and feature engineering algorithms are supported..

That way of work revolutionized the daily routine of research, allowing research teams to shift focus from performing experiments to concentrate on analysis and understanding of results, and designing new approaches. It brought more confidence to our results – we now base our conclusions on thousands of experiments, not just a few as before.

For many vegetation maps, results are delivered in hour(s). Larger experiment batches, spanning multiple flights and study sites, are often ready “next morning”. 9 months after introducing VCS system, all teams created over 300.000 classifications (compared to just 600 during 1st year of the project) and 17.500 vegetation maps.

While being an obvious success, such a big shift in organization of research work lead to its own problems. The 100x-1000x increase in number of experiments and results was quite disruptive for whole project organization. Besides many benefits and obvious improvements, it uncovered unexpected bottlenecks, creating need for further automation of related activities like data management, error detection, organization of results and better visualization and analysis of results.

Research has been carried out under the Biostrateg II Programme of the Polish National Centre for Research and Development, project DZP/BIOSTRATEG-II/390/2015: The innovative approach supporting monitoring of non-forest Natura 2000 habitats, using remote sensing methods.

KEYWORDS: Vegetation mapping, Airborne remote sensing, Classification, Automation, Workflow

ProSIt: An Ontology-based Reference Process to Provide Interoperability and Foster Database Integration for Biodiversity Data

***Fabiana Santana*[1], *Silvia Scheunemann*[2], *Anarosa A.F. Brandão*[2], *Antonio Saraiva*[2]**

[1]*University of Canberra, Australia*; [2]*University of Sao Paulo, Brazil*

Corresponding author(s) e-mail: Fabiana.Santana@gmail.com

ABSTRACT:

Citizen science and scientific and applied research worldwide are continually producing large volumes of biodiversity-related data. This data represents an important source of knowledge, which can be used for several purposes, such as the assessment of climate change impacts on the environment and the definition of public policies and scenarios for sustainable use of the biodiversity. Unfortunately, the data is being collected and stored in different forms, formats and standards. Many researchers organise their data to attend immediate research purposes without taking the time to organise them properly, even if the research community demands the publicity of such data, to enable reproduction, continuity and better evaluation of research contributions. Free and open access data facilities have been created to store and publish discoverable biodiversity-related data, such as GBIF (Global Biodiversity Information Facility) and ALA (Atlas of Living Australia). These facilities usually store the data based on standards such as the Darwin Core, which is used by both GBIF and ALA. If the data is not compliant with the standards adopted by the facility, a specific solution for database integration must be developed. Some of the original data might become useless in the process if a technical solution to store and provide discoverability of non-standardised data is not available, potentially hindering access to information. The data should be standardised from the very early stages of a project, but solutions are also required to standardise the data that have already been collected. A semantic approach can achieve this goal by applying ontologies to improve the understanding of the available data and metadata. Ontologies have been advocated as a powerful technique to provide interoperability among datasets and information systems. This paper presents ProSIt, an ontology-based reference process (workflow) to guide the creation of a semantic approach to provide biodiversity data interoperability based on the semantic integration of biodiversity standards. A functional ontology was built as a case study to evaluate the reference process by providing interoperability between the ABCD and Darwin Core standards, which are the TDWG (Biodiversity Information Standards) currently recommended standards and among the most adopted worldwide. The ontology and reference process were evaluated and proved effective, representing a promising solution for biodiversity-related data interoperability. The reference process can be applied to other existing standards and ad hoc databases that can be semantically interpreted, eliminating the structural barrier that prevents simplified access and availability of information between heterogeneous and distributed databases.

KEYWORDS: Data and Workflow Management, Biodiversity data standards, Semantic and Interoperability, Ontology, Database integration.

Session S3.8

USING BEXIS 2 TO FACILITATE SCIENCE FROM DATA COLLECTION TO DATA PUBLICATION

Using BEXIS 2 to facilitate science from data collection to data publication

Roman Gerlach[1]***, Javad Chamanara***[1]***, Sven Thiel***[1]***, David Schöne***[2]***, Nafiseh Navabpour***[1]

[1]*Friedrich Schiller University Jena, Germany*; [2]*Max-Planck-Institute for Biogeochemistry Jena, Germany*

Corresponding author(s) e-mail: roman.gerlach@uni-jena.de

ABSTRACT:

This session is dedicated to researchers using the BEXIS 2 research data management platform. We invite short presentations from the user community that showcase the usage of BEXIS 2. Showcases can be related to any data management aspect (e.g. metadata creation, data structure definition, data publication) supported by BEXIS 2. In addition, we encourage users to contribute potential features and improvements that they would like to see in BEXIS 2.

Contributed presentations should, first, describe a specific data management problem, and then demonstrate a solution within BEXIS 2. Besides such showcase we encourage any other experience report. Although presentations are expected primarily from the users community (i.e. researchers), we also expect a number of developers and data managers of BEXIS 2 instances to be present in the session, which may lead to some fruitful discussion. After each talk there will be time for discussion.

This session is intended to provide a forum for the BEXIS community, which in previous years met at the annual BEXIS User and Developer Conference. This year the conference has been suspended in favour of the ICEI conference. However, the session is open to anyone else interested in BEXIS 2.

KEYWORDS: research data management, BEXIS 2 platform, data sharing, data publication, open source software

Using BEXIS 2 as efficient research data management system for the ATTO research project

***Marcus Guderle*[1], *Jošt V. Lavrič*[1], *David Schöne*[1], and *Susan Trumbore*[1]**

[1] *Max Planck Institute for Biogeochemistry*

Corresponding author(s) e-mail: mguderle@bgc-jena.mpg.de

ABSTRACT:

The Amazon Tall Tower Observatory (ATTO) is a joint German-Brazilian project launched in 2009 and funded by MCTI (Brazil), BMBF (Germany) and the Max-Planck Society. ATTO is with its 325 m-tall tower, the associated research infrastructure and nearby scientific plots a unique multidisciplinary scientific research platform in a region of global significance. Located at the centre of the world’s largest continuous tropical forest, ATTO allows the observation of geo-/bio-/atmosphere interactions and their impact on climate, atmospheric chemistry, aerosols and clouds, and greenhouse gases in near-pristine conditions.

From the data management point of view, the main challenge is to provide a functional platform for consortium-internal and -external transparency, traceability, and data exchange between a large number of institutions and research groups that are involved in the project in order to foster collaboration and maximize the output of the project.

We use BEXIS 2, a flexible, interoperable and modular data management software, to provide efficient data management and exchange within such a large research consortium; and to allow long-term data archiving, documentation, secondary analysis, data reuse and accessibility for scientific community.

We will present a showcase how the BEXIS 2 instance was modified in order to manage highly diverse data of a large research consortium ranging from i.e. micrometeorological data and greenhouse gas measurements as well as remote sensing data towards soil and water samples. The main focus will be on the flexibility of the metadata definition and the implementation of an automatic DOI registration, which allows BEXIS 2 to act as a data repository for the ATTO project.

KEYWORDS: multidisciplinary, research platforms, Amazon

Showcase: Biodiversity Exploratories Information System – Report of our data migration to the new BEXIS2 instance

***Andreas Ostrowski*[1], *Eleonora Petzold*[1], *Sirko Schindler*[1]**

[1]*FSU Jena, Germany*

Corresponding author(s) e-mail: andreas.ostrowksi@uni-jena.de

ABSTRACT:

The development of the Biodiversity Exploratories Information System (here BExIS1) started more than 10 years ago and since then has been functioning as the data management platform and information system for the SPP 1374 "Biodiversitäts-Exploratorien" (BE). Since then, more than 1,000 datasets with a total of more than 20,000 variables and several million data rows were uploaded to the system. A dataset consists of metadata, a data structure, and the research data itself. Besides the storage of tabular data, it is also possible to upload unstructured data (files) to the system. Stored data is subject to common operations like updating, editing, and deletion. These actions led to various changes inside the datasets, which are reflected as versions or archived data. Of course, a lot of additional information to run such a system is stored. This mostly concerns authorization and authentication, but also plenty of personal information is included.

Migrating data from one system to another is always a major and sometimes thrilling task. Data needs to be transferred without alterations. BEXIS2 uses a different schema to store and handle its data. Furthermore, some functionality is implemented differently from BExIS1. Examples include the concept of variables or the user management. These changes need to be considered during data migration and the transferred data needs to be adjusted to adhere to changed model.

We implemented a system to transfer data from a BExIS1 instance to a BEXIS2 instance. BExIS1 data is accessed directly or by using BExIS1 functionalities. The data is then stored in BEXIS2 by making use of BEXIS2 API calls and by using auxiliary information. The main auxiliary source is an instance specific mapping of BExIS1 variables to the different BEXIS2 variable concept.

We show our implementation of the system to migrate a production level Bexis1 instance into a newly set up Bexis2 instance. We show steps necessary and obstacles encountered together with our implemented solutions.

We intend to make our implementation available to other BExIS1 instance users to facilitate and ease the migration to the new BEXIS2 system.

KEYWORDS: BEXIS, Biodiversity Exploratories, data management system, data migration, project repository

Enhance BEXIS2: from pure data management to an information system

***Andreas Ostrowski*[1], *Eleonora Petzold*[1]**

[1]*FSU Jena, Germany*

Corresponding author(s) e-mail: andreas.ostrowksi@uni-jena.de

ABSTRACT:

More than 10 years ago, the first version of the Biodiversity Exploratories Information System (here BExIS1) went online to act as the data management platform and information system for the SPP 1374 “Biodiversitäts-Exploratorien” (BE). With its major function as data management system it facilitates data sharing among the projects and data re-use across all phases. In addition, BExIS1 offers access to further resources and has many extra functionalities to facilitate the collaboration between people and to support the project as a whole. It provides a lot of information related to field work like interactive maps of all research plots, and a booking system for field station resources and to announce plot visits. Further functionalities include a management system for publications, a document uploading mechanism, an event registration tool, and a question-answer system as an interface to stakeholders, in particular landowners. With these functionalities, BExIS is more than a conventional data management system. It acts as an information or even better project assisting system.

Since release 2.11.0 the BEXIS2 system is fully modularized. It allows to un-/install and de-/activate modules in a running instance. This was a major advance because it allows to implement features also from 3rd parties.

We re-developed a couple of BExIS1 functionalities as separate modules. It involves fieldwork management related features like resource booking management, and the plot map management. Further features are related to event management, information/document sharing, land use intensity calculation, and dataset overview computation. The development followed a generic, modular and extensible approach. Therefore these modules are adaptable to project specific needs and could be of high interest for other BEXIS2 instance users.

KEYWORDS: Biodiversity Exploratories, BEXIS, data management system, information system, project repository, fieldwork, resource booking management, plot management, event management

Research data management with BEXIS 2 – An overview and introduction to the special session

***Roman Gerlach*[1], *David Schöne*[2], *Javad Chamanara*[1], *Nafiseh Navabpour*[1], *Sven Thiel*[1], *Martin Hohmuth*[1]**

[1]*Friedrich Schiller University Jena, Germany*; [2]*Max-Planck-Institute for Biogeochemistry Jena, Germany*

Corresponding author(s) e-mail: roman.gerlach@uni-jena.de

ABSTRACT:

In this talk, we will introduce the data management platform BEXIS 2 and provide an overview of its features and capabilities. We will demonstrate how BEXIS 2 can support researchers in managing their data throughout the different aspects of the data life-cycle. Since BEXIS 2 has been designed for large collaborative projects with a central data management, we will also address features relevant to decision makers and system administrators.

KEYWORDS: BEXIS 2, **r**esearch data management, data life-cycle, software development

Data Visualization: a new module for BEXIS 2

***Nafiseh Navabpour*, Roman Gerlach**

Friedrich Schiller University Jena, Germany

Corresponding author(s) e-mail: nafiseh.navabpour@uni-jena.de

ABSTRACT:

The visual representation of data, or in other words a visualization is a very common way to represent data in an easily comprehensible form. Especially in situations where a user would like to get a quick overview of a single dataset (e.g. a distribution of observations) or a system administrator tries to monitor the number of uploaded or downloaded datasets in a system, visualizations come in very handy.

In this talk, we introduce the new visualization module developed for the data management platform BEXIS 2. We will discuss the concept of the module and demonstrate the first implementation of the user interface. This first implementation helps administrators to trace the system's behaviour and provides visualizations such as number of datasets created over time. In the next step, the module will be further developed to provide also visualizations more relevant to end users (i.e. researchers) such as providing information on the completness of metadata in a dataset.

KEYWORDS: BEXIS 2, Visualization, Data management platform, Module

POSTERS

Towards truly automatic bird audio detection: an international challenge

Dan Stowell

Machine Listening Lab, Centre for Digital Music, Queen Mary University of London, UK

Corresponding author(s) e-mail: dan.stowell@qmul.ac.uk

ABSTRACT:

Automatic detectors of animal sounds should be useful in acoustic monitoring of ecosystems. In some specific cases, hand-made custom techniques work well, while machine learning has also brought improved flexibility even in adverse noise conditions. However, for global-scale monitoring of many species, there is a problem: advanced methods often need expert tuning in order to work well - and even then, there may be many false positives and false negatives. "Automatic" detectors don't create a truly automatic workflow. This is why we introduced the Bird Audio Detection challenge - a data challenge in which we invite signal-processing and machine-learning researchers worldwide to innovate by creating detectors that are truly hands-free, high quality and adaptable to new situations [1]. In this talk we will describe the outcomes of the first challenge [2], and the hot-off-the-press progress of the second 2018 challenge. We will distil this experience into concrete recommendations for practitioners wanting to deploy ecoacoustic remote monitoring.

KEYWORDS: ecoacoustics, bird, detection, machine learning, evaluation

REFERENCES:

1. Stowell, D., Wood, M., Stylianou, Y., Glotin, H., 2016. Bird detection in audio: a survey and a challenge. In *Proceedings of MLSP 2016*.
2. Stowell, D., Wood, M., Stylianou, Y., Glotin, H., submitted. Automatic acoustic detection of birds with deep learning: the first Bird Audio Detection challenge.

Assessing Ecosystem Change using Soundscape Analysis

Diana Carolina Duque-Montoya[1]***, Claudia Isaza***[1]***, Juan M. Daza***[1]

[1]*Universidad de Antioquia, Colombia*

Corresponding author(s) e-mail: diana.duquem@udea.edu.co

ABSTRACT:

Soundscape is already a recognized term in different acoustics sciences. In ecology, it is defined as the collection of biophony (sounds produced by living organisms), geophony (sounds produced by geological phenomena) and technophony (sounds produced by human technology). This term, and the associated science, soundscape ecology, were first introduced in landscape ecology 7 years ago, however there is still not a clear consensus in how to quantify soundscape and establish its relation to ecosystem changes. Acoustic indices such as acoustic entropy (H) [1], or acoustic complexity index (ACI) [2] have been mainly applied to biodiversity assessment, but their connection to soundscape complexity is not clear yet [3]. The ratio of biophony to technophony (ρ) [3] and the Normalized Difference Soundscape Index (NDSI) [4] are the only indices that use technophony besides biophony to measure the soundscape; but geophony is not included and the criteria to quantify biophony and technophony are not accurate.

To find a general measure of soundscape, we have tested different types of features, psychoacoustical and visual (spectrogram based), on 125,259 recordings collected for one year, every 20 minutes in six sites from the northern Andes of Colombia. Using a perturbation gradient, we deployed the recorders in three contrasting habitats (forest, stubble and open areas) and have trained machine learning techniques for automatically classifying the sites and finding structural soundscape patterns in tropical landscapes.

On a first approach, the use of contour plots of normalized power spectral density (nPSD) followed by the extraction of SIFT descriptors, a bag of words model and a support vector machine led to the classification of four sites with 97% accuracy [5]. Currently, the goal is to identify which are the best acoustical features to characterize a soundscape (meaning to reduce the computational cost of graphical approaches) and detect ecosystem changes through space and time. To achieve this goal, acoustic biodiversity indices as the mentioned above, and other acoustical measures have been tested on our dataset. Particularly, psychoacoustical soundscape studies have highlighted the importance of considering three components of sound perception: time, frequency and intensity. Using acoustical features, sequential learning algorithms such as hidden markov models are trained to identify spatial and temporal patterns in the recordings. Consequently, we expect to understand the connection between ecosystem health and soundscape, and thus be able to foresee and prevent structural ecological changes by the analysis of sound.

KEYWORDS: Soundscape, Ecosystem Change, Acoustic Indices, Machine Learning

REFERENCES:

[1] Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid Acoustic Survey for Biodiversity Appraisal. PLoS One 3, e4065. https://doi.org/10.1371/journal.pone.0004065

[2] Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). Ecol. Indic. 11, 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005

[3] Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust. united with Acust. 100, 772–781.

[4] Kasten, E.P., Gage, S.H., Fox, J., Joo, W., 2012. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. Ecol. Inform. 12, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001

[5] Gaitan, D., Isaza, C., Gomez, W., Daza, J., 2016. Categorization of Ecosystems Based on Soundscape Analysis: A Perspective from Image Classification, in: 2016 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, pp. 762–766. https://doi.org/10.1109/CSCI.2016.0148

MAAD, a rational unsupervised method to estimate diversity in ecoacoustic recordings

Juan Sebastian Ulloa[1,2]*, *Thierry Aubin[2]*, *Sylvain Haupert[1]*, *Chloé Huetz[2]*, *Diego Llusia[3]*, *Charles Bouveyron[4]*, *Jerome Sueur[1]

[1]*Muséum national d'Histoire naturelle, FRANCE*; [2]*Université Paris-Saclay, FRANCE*; [3]*Universidad Autónoma de Madrid, SPAIN;* [4]*Université Côte d'Azur, FRANCE*

Corresponding author(s) e-mail: juan.ulloa@mnhn.fr; sueur@mnhn.fr

ABSTRACT:

Ecoacoustic research mainly relies on signal and data analysis. Beyond manual inspection, the current solutions to decipher the content of population, community or soundscape recordings either refer to supervised classification methods that need labelled data (e.g. random forest) or to global diversity indices that totally avoid species identification (e.g. acoustic complexity index). Between these far-off strategies, unsupervised learning methods aim at returning automatically and without a priori clusters that correspond to relevant sound structures (e.g. species-specific signals). Here, we report a new unsupervised learning method named MAAD (Multiresolution Analysis of Acoustic Diversity) that combines signal analysis and multivariate clustering. We designed MAAD to decompose an acoustic recording into few elementary components based on their time-frequency attributes. First, we used the short-time Fourier transform to detect regions of interest (ROIs). Then, we characterized these ROIs by estimating the median frequency and by running a 2D wavelet analysis at multiple scales and angles. Finally, we grouped the ROIs using a model-based subspace clustering technique so that ROIs were automatically annotated. To test the performance of the method, we applied MAAD to two distinct tropical environments in French Guiana, a lowland high rainforest and a rock savannah, and we compared manual and automatic annotations using the adjusted Rand index. For both environments, the similarity between the manual and automated partitions was high and consistent indicating that MAAD can generate useful clusters, which may represent a general analytic solution for ecoacoustic research. The weight of the features estimated by the clustering process revealed important information about the structure of the acoustic communities; in particular the median frequency had the strongest effect on modelling the clusters, and on classification performance. The implementation of a free and open Python module opens the possibility to use MAAD for other ecoacoustic research programs.

KEYWORDS: ecoacoustic monitoring, multiresolution analysis, unsupervised learning, wavelets, subspace clustering

Underwater acoustic habitats: towards a toolkit to assess acoustic habitat quality

***Irene T. Roca*[1], *Ilse Van Opzeeland*[1,2]**

[1]Helmholtz Institute for Functional Marine Biodiversity (HIFMB), Carl von Ossietzky University, Oldenburg, Oldenburg, Germany; [2]Ocean Acoustics Lab, Alfred-Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Alten Hafen 26, Bremerhaven, Germany

Corresponding author(s) e-mail: ilse.van.opzeeland@awi.de

ABSTRACT:

Underwater marine environments are filled with sounds. Many aquatic organisms produce and rely on acoustic cues to communicate and as primary source of information about their environment. However, to date, there has been comparatively little research aimed at understanding the variations in acoustic features of marine habitats and their relation with the ecosystem biodiversity structure and dynamics [1]. The newly founded Helmholtz Institute for Functional Marine Biodiversity (HIFMB) in Oldenburg, Germany, has initiated a project to investigate how environmental acoustic features relate to habitat use and variations in behaviour of multiple marine mammal species. The project's resources include basin-wide passive acoustic monitoring (PAM) data sets from the Southern Ocean spanning 11 years at 20 sites collected by AWI's Ocean Acoustic Lab. Because these data were obtained in one of the most pristine ocean locations world-wide, this project offers a unique opportunity to analyze habitat acoustics in virtual absence of anthropogenic sound effects. By drawing from terrestrial research, we aim to adapt and further develop existing metrics on e.g., acoustic complexity [2] to analyze passive acoustic data from the Southern Ocean to develop maps comprehensively describing the temporal, spectral and spatial characteristics of the underwater acoustic environment, to assess variations in acoustic assemblages' composition [3,4], and ultimately, to develop a comprehensive "toolkit" to assess acoustic habitat quality.

KEYWORDS: Passive acoustic monitoring, Habitat quality, Marine mammals, Ambient noise, Southern Ocean

REFERENCES:

1.Risch, D., Parks, S.E., 2017. Biodiversity Assessment and Environmental Monitoring in Freshwater and Marine Biomes using Ecoacoustics, in: Farina, A., Gage, S.H. (Eds.), Ecoacoustics: The Ecological Role of Sounds. John Wiley & Sons Ltd., New Jersey, pp. 145–168.

2.Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust united Ac, 100, 772–781.

3.Van Opzeeland, I., Boebel, O., 2018. Marine soundscape planning: Seeking acoustic niches for anthropogenic sound. Journal of Ecoacoustics. 2. 5GSNT.

4.Roca, I.T., Proulx, R., 2016. Acoustic assessment of species richness and assembly rules in ensiferan communities from temperate ecosystems. Ecology, *97*, 116–123.

Focus on geophony: what weather sounds can tell

Roberta Righini[1] ***and Gianni Pavan***[1]

[1]*University of Pavia, Italy*

Corresponding author e-mail: roberta.righini01@universitadipavia.it

ABSTRACT:

Nowadays, one of the most important biodiversity conservation challenges is to manage and mitigate the threats of climate change and the high rate of species loss caused by it.

In this context, Ecoacoustics is an increasingly emerging interdisciplinary science that investigates natural and anthropogenic sounds and their relationship with the environment. It can become a powerful tool for management and conservation efforts. By the use of passive acoustic monitoring of a habitat, it is possible to get a picture of its soundscape, that is composed of the sounds produced by the animals (biophony), the noise produced by atmospheric and physical events (geophony: wind, rain, water, etc..) and by human activity (anthropophony: cars, airplanes, etc..; other activities to be referred as technophony, if electro-mechanical noise overbears).

But, if on the one hand, there are more and more studies concerning the analysis of the biophony component that are translated into the detection new indices for the determination of different bio-diversity and bio-richness levels in different habitat, and, more recently, a great attention is paid to the human-activity sounds, on the other hand, the third main soundscape component, geophony, is a rich but still unexplored source of information in the environmental acoustics recordings.

The aim of the work is to focus the analysis on natural background noise of recording collected in an Italian National Park in order to investigate the acoustic characteristics of a natural environment in different weather conditions (good, wind or rainy conditions).

Moreover, since passive acoustic monitoring generates large data sets of audio recordings that have to be stored and processed, the possibility of discriminating the recordings according also to different weather conditions, will contribute to make processing time manageable and to reduce analysis effort.

KEYWORDS: Ecoacoustics, Soundscape ecology, Geophony, Rain, National park

Using robustness as a tool for the assessment of ecosystem health – a case study from Hooghly Matla estuary, India

Nabyendu Rakshit ***and Santanu Ray***

Systems Ecology & Ecological Modelling Laboratory, Department of Zoology, Visva-Bharati

Corresponding author(s) e-mail: Nabyendu.rakshit@visva-bharati.ac.in

ABSTRACT:

Modern aquatic resource management is based on ecosystem approach as it includes fish as well as different organisms and different natural processes. In this regard, static mass balanced modelling approach, well-grounded with realistic data, can give answers to many salient questions for analysing ecosystem structure and function holistically through different ecological indicators. The Hooghly Matla estuarine ecosystem with adjacent mangroves is one of the largest detritus based ecosystems of the world and has great importance in costal landscape of India. It provides diverse habitat for wide variety of aquatic resources having ecological and economical significance. Recent years have seen that this ecosystem has been degrading gradually owing to different anthropogenic factors mainly overfishing. To understand the situation, two mass-balanced network models of this ecosystem, from two different time periods (less exploited phase 1985-1990 and highly exploited phase 1998-2003) have been constructed using Ecopath with Ecosim (EwE) software. Thereafter, different network indices such as total system throughput (TST), redundancy (R), Finn's Cycling Index (FCI) and ascendency over development capacity ratio (A/C) are calculated. The network indices indicate that the system is approaching towards degradation. But the indices are not able to assess the overall ecosystem organization, function and magnitude of stress imposed on system. It is well known that a 'healthy' ecosystem have an efficient diversity of components and flows while maintaining an insurance i.e. redundancy as protection against stress; thus a measure of ecosystem robustness can successfully be used to measure the system health. In this present work, system level assessment of robustness for these two different periods is conducted. Specifically, we investigate the response of the system to overfishing and other anthropogenic stress. These are considered to answer the question: how does the system robustness change as a result of exploitation?

KEYWORDS: Overfishing, Ecopath with Ecosim, Network analysis, Ascendency, Finn Cycling index

DeltaMP: A flexible, reproducible and resource efficient pipeline for the analysis of high-throughput amplicon sequencing data of eukaryotes

Christina Weißbecker*[1,2], ***Anna Heintz-Buschart***[1,3], ***Tesfaye Wubet***[1,3], ***François Buscot***[1,2,3], ***Guillaume Lentendu***[1,4]

[1]Helmholtz-Centre for Environmental Research, Germany; [2]University of Leipzig, Germany; [3] German Centre for Integrative Biodiversity Research (iDiv) Jena – Halle – Leipzig, German; [4]University of Kaiserslautern, Germany

Corresponding author(s) e-mail: christina.weissbecker@ufz.de

ABSTRACT:

Due to ever decreasing costs in next and third generation sequencing, multiplexed high-throughput amplicon sequencing became accessible for a broad clientele of researchers and companies. However, there are no common standards for data processing. Several data analysis platforms and a plethora of single tools are available. We established DeltaMP a new pipeline connecting the tools that perform best on amplicon data from eukaryotic microorganisms generated by 454 or Illumina technologies in a wrapper program. The program development is undertaken in the framework of the German Federation for Biological Data and complies with the associated guidelines of research data management. DeltaMP is programmed for linux high performance computer clusters. It efficiently parallelizes the workload on several computer nodes in accordance with the memory requirements, minimizing analysis time and optimizing the use of computational resources. Input data can be provided as local files or as sequence data publicly available from the ENA short read archive. DeltaMP allows for a smooth adjustment of the analysis workflow for specific datasets: checkpointing of results allows for the reanalysis of data with modified parameters reusing as much of the already conducted steps as possible. Depending on the specific steps modified, only a fraction of computational resources are needed to run the workflow. DeltaMP offers a user-friendly operation: desired analysis tools and steps are chosen via a configuration file, the pipeline is started by a single command, it generates logfiles of all steps and provides final results together with workflow documentation and intermediate statistics as well as standard output in the BIOM format. Besides, standard/example configuration files are provided to allow analyses following the workflow of reference publications for multiple taxonomic groups and targets. We present a next generation amplicon dataset analyzed with deltaMP. We assessed soil fungi in experimental subtropical Chinese forests applying 454 pyrotag sequencing of the internal transcribed spacer (ITS2) region of the rRNA operon. About 1.2 million raw sequences from 400 individual samples were generated and processed. During the publication process, several analyses options like trimming parameters, clustering algorithm and taxonomic database version had to be compared. The workflow of DeltaMP will be continuously updated and further functionalities (eg. single nucleotide variant based read binning) will be included.

KEYWORDS: high-throughput amplicon sequence analysis, eukaryotes, resource efficient analysis pipeline

Extracting information on bat activities from long-term ultrasonic recordings through sound separation

***Chia-Yun Lee*[1]*, Tzu-Hao Lin*[2]*, Mao-Ning Tuanmu*[1]**

[1] *Academia Sinica, Taiwan*; [2] *Japan Agency for Marine-Earth Science and Technology, Japan*

Corresponding author(s) e-mail: cylee918@gate.sinica.edu.tw

ABSTRACT:

Soundscape monitoring has become a powerful method for exploring the interactions among organisms, environmental dynamics, and human activities. This monitoring approach is especially suitable for bats because most bats rely on echolocations and they are sensitive to changes in the surrounding environment. Many acoustic indices have been developed for measuring characteristics of soundscapes in the intensity, frequency, or time domain. However, most acoustic indices calculated from recordings collected in the field may not well reflect bat activities when geophony, anthrophony or non-target biophony appears. To overcome these challenges, we used a periodicity-coded non-negative matrix factorization (PC-NMF) for mining long-term eco-acoustic recordings and separating signals from different sound sources. Calculated from the signals solely from bats, the acoustic index values would more reliably reflect bat activities. We applied the PC-NMF method on nine-month ultrasonic recordings collected at three sites at different elevations to separate signals from three sources, including bat echolocation calls, sounds produced by other organisms and those from other sources. We then analyzed the separated bat signals to investigate spatio-temporal patterns of bat activities. Results showed that bat activities increased with decreasing elevation. The temporal pattern of bat activities at the middle elevation site differed from the patterns at high and low elevations. Variations in frequency compositions among the sites were also observed in this study. In short, PC-NMF can effectively capture characteristics of long-term eco-acoustic recordings and extract the information of interest by sound source separation. It could be an effective and efficient tool for analyzing a massive amount of eco-acoustic monitoring data to understand spatial and temporal patterns of organism activities.

KEYWORDS: ultrasound, echolocation calls, Chiroptera, sound separation, machine learning

Reverse Wavelet Interference Algorithm for Detection of Avian Species and Characterization of Biodiversity

***Sajeev C Rajan*, Athira K, R Jaishanker**

C V Raman Laboratory of Ecological Informatics, Indian Institute of Information Technology and Management-Kerala, Thiruvananthapuram, Kerala, India

Corresponding author(s) e-mail: jrnair@iiitmk.ac.in

ABSTRACT:

The authors introduce reverse wavelet interference algorithm to identify avian (bird) species from sonic data. It employs the principle of phase cancellation, in which two identical, but inverted waveforms, summed together mutually cancel to result in silence. Reverse wavelet for each avian species, generated from sonic library is panned across the spectrogram. Minimal residual amplitude or silence indicates presence of particular species. Sequential analysis of sonic records of all species in library will reveal the number of bird species in the study region. Representative sound samples were recorded with sound recorder (Zoom® H1 sonic recorder) during morning and evening from two agricultural landscapes at Perumbavoor, and Kollam, Kerala, India. The reverse wavelet interference algorithm identified 80% of bird species (out of 10 bird species) at both the locations. Avian diversity at these locations were expressed using sonic derived indices Acoustic Complexity Index, (ACI), Acoustic Diversity Index, (ADI), and Normalized Difference Soundscape Index (NDSI). It offers a promising means for rapid characterization of landscapes to aid research, conservation and policy framing.

KEYWORDS: Sound, Reverse Wavelet Interference Algorithm, Bird, Diversity

Automatic Bird Sound Detection: Logistic Regression Based Acoustic Occupancy Model

***Yi-Chin Tseng*[1], *Bianca N.I. Eskelson*[1], *Kathy Martinl*[2]**

[1]*Department of Forest Resources Management, University of British Columbia, Vancouver, Canada*
[1]*Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada*

Corresponding author(s) e-mail: sunnyyctseng@gmail.com

ABSTRACT:

Avian bioacoustics research was revolutionized by the introduction of the digital audio recorder, which not only allows remote-monitoring but also makes large-scale studies possible. However, manual inspection of recordings is not feasible given increasingly larger datasets. The need for an automatic bird sound detection algorithm as a pre-filtering step prior to other acoustic analyses has arisen. In this research, we developed logistic models to predict the probability of bird presence in a recording using frequency percentiles. Our smartphone-derived recordings covered a wide range of environments (e.g., grassland, forest, urban areas) with the presence of noise due to weather, traffic, insects, and human speech. A wide range of bird species was included. Based on leave-one-out validation and using an optimal cut-off of 0.35 (i.e., probability ≥ 0.35 indicates the presence of birds), our model resulted in 70% accuracy with a 14% false negative rate. Our model took less than one hour to process 15,690 ten-second recordings (i.e., 44 hours in duration). This acoustic occupancy model opens up new possibilities toward automatic, accurate, and efficient analysis of large remote-monitoring audio datasets.

KEYWORDS: Bird sound, Detection, Binary, Logistic regression, Occupancy model

A software detector for monitoring endangered common spadefoot toad populations

Guillaume Dutilleux[1], **Charlotte Curé**[2]

[1]Acoustics research centre, NTNU/IES, Norway; [2]UMRAE, Cerema Ifsttar, France

Corresponding author(s) e-mail: guillaume.dutilleux@ntnu.no

ABSTRACT:

In the context of global amphibian decline population monitoring is relevant. For the rapidly declining European common spadefoot toad (Pelobates fuscus) that vocalizes underwater, unattended acoustic monitoring using programmable recorders is a possible approach. This kind of monitoring is likely to generate a large amount of data. Therefore special attention must be paid to limit human post-processing time.

In this communication we present a dedicated software detector that targets the advertisement call of P. fuscus. Since the primary goal of our detector is to deliver presence information the false positive rate is a key performance measure here. But the detector provides also call counts to inform about the strength of the signal.

The detector is based on both frequency- and time-domain features. From data collected over several months in north-eastern France, a ground truth was built in order to test the performance of the detector. The detector returned false positive rates below 1.5 % and true positive rates above 50 %. Moreover the full-scale evaluation demonstrated that the associated computation time for a 3-month campaign was less than 1 day per site. These figures are compatible with presence monitoring on several sites in parallel.

The signal strength feature of the detector allowed in addition to investigate vocal activity patterns. Our results point out that they more complex than previously recognized. Over the whole ostensible breeding season the actual time window can vary broadly and may be split in distinct episodes. The circadian vocal activity occasionally proceeded uninterrupted for 24 hours but usually a several-hour-long lull occurred prior to sunset.

KEYWORDS: acoustic monitoring, amphibian, automated species detection, pelobates fuscus, underwater sounds

Query Processing in Ontology Based Data Access

Majid Askar[2], Alsayed Algergawy[1], Taysir Soliman[2], Birgitta König-Ries[1], Adel Sewisy[2]

[1] *FSU Jena, Germany*; [2]*Assiut University, Egypt*

majid.askar@aun.edu.eg; alsayed.algergawy@uni-jena.de; Taysser.soliman@fci.au.edu.eg; birgitta.koenig-ries@uni-jena.de; sewisy@aun.aun.edu.eg

ABSTRACT:

A large number of ontologies are developed and used to represent and annotate biodiversity data. There are different schemes proposed to capture ontologies to enhance the management of biodiversity data. Ontology Based Data Access (OBDA) is a common scheme that allows the access such heterogeneous data via ontologies and provides an elegant way to cope with inconsistencies and incomplete data (i.e. instance morphology, genetics, and any other related information regarding organisms in biodiversity domain). The idea behind OBDA is inspired from the efficiency of relational database management systems in answering conjunctive queries. It can be used as a part of the data discovery component of the data management lifecycle. In general, OBDA has three main layers: a conceptual layer, a data layer, and a mapping layer that mediate between the first two layers. Query processing is a bottleneck in OBDA-based systems, in terms of supporting the end user to input her query in her own language, transforms this user query to an acceptable format to be executed. To this end, we introduce a tool based on the proposed pipeline, the tool is developed, and it is under testing. This tool can be used to translate from natural language query to SPARQL query, where the tool consists of four components: User Query Handler, Ontology Reader, User Interaction, SPARQL Query Builder. The user query passed through tokenizer, stop word remover, and stemmer that are included in the User Query Handler unit. The resulting items are used by the Ontology Reader unit which compares these items with those in the ontology. Then, the resulting matches are presented to the user to get his feedback using the User Interaction unit. Finally, the SPARQL query is built based on deliverables of the User Interaction unit, using the SPARQL Query Builder unit. A part of this research was supported by DAAD funding through the BioDialog project.

KEYWORDS: Semantic web, Ontology Based Data Access (OBDA), Query processing

Satellite-based monitoring of invasive woody species in central Chile

Anne Clasen[1], **Michael Förster**[1], **Tobias Schmidt**[1], **Fabian E. Fassnacht**[2], **Birgit Kleinschmit**[3]

[1]*TU Berlin, Germany*; [2]*KIT, Germany*

Corresponding author(s) e-mail: anne.clasen@tu-berlin.de

ABSTRACT:

Chile's rich flora is characterized by a large number of endemic species, which are increasingly endangered by a variety of invasive species. This development is further intensified by high disturbance rates due to land-use changes, forest plantation industries and wild fires. In this context, the distribution patterns of three invasive woody plants - *Acacia dealbata*, *Pinus radiata* and *Ulex europaeus* – are subject of the presented study. Long-term and recent satellite data (Landsat & Sentinel-2) is used to map the current and retrospective distribution patterns of the three target species using spectral unmixing and the one class classifier Maxent.

The three target species show different spatial and temporal spreading strategies. In the study the current spreads of the target species are assessed using multiple Sentinel-2 images of the whole phenological cycle of the year 2016. Similar to the Landsat-mission, the Sentinel-mission offers a range of wavebands adapted to vegetation studies, but at a finer spatial resolution of 10 m. A process chain for linking reference data from multi and hyperspectral unmanned aerial systems (UAS) data with the satellite data of the Sentinel mission was developed. Classification accuracies of 0.74 to 0.98 were achieved. The suitability of the individual image acquisition timeframes for the detection of the target species was examined. For *Ulex* classification the summer and spring scenes had the highest importance. While an average RSME of 20.4 was achieved with a monotemporal model, the accuracy improved significantly with the addition of further images. However, with an increasing number of scenes saturation occurred.

For an improved understanding of the invasion dynamics over the last few decades, remote sensing data from the Landsat archive was accessed and allowed a retrospective view on the distribution patterns back to the late 1980s. We combined three vegetation indices and the two time-series-analysis algorithm BFAST and LandTrendr to reconstruction of the disturbance history. The results show that Landsat-based time series analysis is suitable for detection, mapping and categorization of disturbances like wood fires and clear-cuts.

With our results, we could contribute to a better understanding of the invasion dynamic of three different widespread woody invasive species in central Chile. With that knowledge, we can make a contribution to a better understanding of invasion ecology and adapt monitoring strategies to stop further expansion. Based on the current and retrospective spreading, the potential future distribution can be estimated for more efficient planning of control strategies.

KEYWORDS: Invasive Species, Remote Sensing, UAS, Maxent, Random Forest

Biodiversity implications of international trade: comparison of species threatened from alternative electricity production with natural gas or concentrated solar power

***Irene Rodríguez-Serrano*[1], *Natalia Caldés*[1], *Yolanda Lechón*[1]**

[1]*Energy System Analysis Unit, CIEMAT. Avenida Complutense 40, Madrid (Spain)*

Corresponding author(s) e-mail: irene.rodriguez@ciemat.es

ABSTRACT:

International trade frequently implies that parts of the value chain of products and services consumed in developed nations take place in developing countries. The latter have abundant natural resources but also weaker environmental and social protection laws, which often lead to social and labor rights infringements as well as uncontrolled air and water pollution. However, the implications that such international trade have on biodiversity have been less analyzed for various reasons. The lack of detailed and comprehensive data, the complexity in assessing the interlinked causes of species threats, and being considered low priorities in policy-makers´ agendas could be some of the reasons that partially explain the limited consideration of such type of impacts. The implications of any activity in ecosystems and biodiversity should be equally analyzed as also non-human species deserve to exist in a good quality environment and because there is a tight relation between humans, other species and ecosystems and the affection in one of them implies consequences for the others, in a direct or indirect way.

To shed some light to this issue, this work analyzes, using a Multiregional Input Output (MRIO) framework, the biodiversity implications in terms of affection to threatened species along the supply chain of two electricity generation systems in Mexico: Concentrated solar Power (CSP) vs. Natural Gas Combined Cycle (NGCC). Besides the fact that the installation and operation of the hypothetical electricity plants are located in Mexico, the assessment also accounts for the required international trade of products and services, both directly (such as imports of plant components) and indirectly (e.g. main inputs needed for manufacturing the plant components). Based on the Framework for Integrated Sustainability Assessment (FISA) [1], the global economic stimulation from the supply chains demands is estimated through a MRIO analysis. Next, the associated effects on biodiversity are estimated, as every human production activity has, in a higher or lesser extent, an impact in the ecosystems and its involved species. This relation has been conducted using the threatened species database EORA [2], which relates the IUCN red list of species threats [3] to the responsible economic sectors. By adapting EORA to the MRIO database used in this work [4], results show the amount of threatened species that would be affected by the global trade involved along the two electricity production supply chains; as well as the main effects (direct/indirect, national/international or investment/operational project phases) and economic sectors responsible of such threats.

Main results show that activities throughout the supply chain of the electricity produced with STE would be affecting to less than one threatened species, while in the case of the NGCC supply chain the figure increases to more than five species. This difference is caused by the larger stimulation of the “Electricity, gas and water supply” sector in Mexico, mainly from the direct effects of the natural gas activities needed for the operation in the electricity production. This affection to species could be related to long term consequences of climate change, which according to the IUCN threat classification scheme, include habitat shifting and alteration, droughts, temperature extremes, storms and flooding [3]. These types of analysis could help identify and raise awareness about the harmful effects on biodiversity of the embedded international trade and start including such implications in decision makers´ agendas in order to find solutions to minimize the riskiest activities and damages to threatened species. Additionally, the wider dissemination of the results could help consumers make more informed decisions about the biodiversity implications of their products and services demands and hopefully a progressive change in their consumption patterns.

KEYWORDS: Biodiversity impacts, international trade, natural gas electricity, concentrated solar power, Multiregional Input Output MRIO

REFERENCES:

1. Rodríguez-Serrano, I., Caldés, N., De la Rúa, C., Lechón, Y., Garrido, A., 2016. Using the Framework for Integrated Sustainability Assessment (FISA) to expand the Multiregional Input-Output analysis to account for the three pillars of sustainability, Environ. Dev. Sustain. doi:10.1007/s10668-016-9839-y.

2. EORA, 2016. Biodiversity footprints. http://worldmrio.com/ (accessed 28 March, 2016).

3. The IUCN Red List of Threatened Spcies, 2018. Threats Classification Scheme. http://www.iucnredlist.org/technical-documents/classification-schemes/threats-classification-scheme/ (accessed 22 May, 2018).

4. WIOD, 2016. World Input Output Database. http://www.wiod.org/new_site/home.htm (accessed 13 March, 2016).

Ecological observations and conservation strategies of a threatened - endemic plant (*Wendlandia angustifolia* Wight ex Hook.f.) in the forests of Western Ghats, India

***Chellam Muthumperumal*[1] , *Paramasivam Balasubramanian*[1]**

[1]*Salim Ali Centre for Ornithology and Natural History, Coimbatore, Tamil Nadu, India*

Corresponding author(s) e-mail: perumal.ecology@gmail.com

ABSTRACT:

The present paper discusses about a rediscovered threatened endemic plant species found in Southern Western Ghats, India. *Wendlandia angustifolia* Wight ex Hook.f. is a narrow endemic, large shrub belongs to the family Rubiaceae. This species was first collected in Courtallum, Western Ghats, India by Wight and described by Hook .f. (1880) [1]. Rangachari collected this plant from Kannikatti in 1917. After a long gap of 81 years later, Viswanathan et al. (2000) rediscovered it Inchikuli in Kalakkad-Mundanthurai Tiger Reserve, Western Ghats [2]. This endemic plant species is placed under Extinct category of the IUCN Red List (Version 2017-3, www.iucnredlist.org) due to lack of data deficiency of the species population [3]. Hence the present study quantified the distribution of W. *angustifolia* in the adjoining areas of the previously recorded localities. Ninety three quadrats (each measuring 0.1 ha area) were laid to gather data on the endemic species and associated taxa. A total of 2676 plant individuals belonging to 131 species were enumerated. This includes 94 trees, 26 lianas, and 13 shrubs. A healthy population of 1091 individuals of targeted endemic species were recorded in a 9.3 ha area in seven different localities of the Kalakkad-Mundanthurai Tiger Reserve, Western Ghats. Periodic field visits documented the phenology such as leafing, flowering and fruiting. Pollinator visits were quantified and it was recorded that flies contributed 35% of visits, followed by bees (25%), wasps (24%) and butterflies (16%). Dusty seeds are facilitated for water and wind dispersal. Natural disturbance such as heavy water flow in the rainy season and extreme heat in summer season seems to inhibit the species population. The seeds of this endemic species were subjected to germination tests. Stem cuttings and air layering technique were tried to develop saplings. Barcode of this species was developed and submitted to gene bank of the National Center for Biotechnology Information (NCBI; accession number MF581792). The present study made a holistic approach of habitat analysis along with species biology to facilitate and improve the population of a targeted endemic plant species in the Western Ghats.

KEYWORDS: Endemic plants, *In-situ* conservation, Phenology, Seed dispersal, Western Ghats, India.

REFERENCES:

1. Hook, f., 1880. Flora of British India.3: 40.

2. Viswanathan, M. B., Premkumar, E. H., Ramesh, N., 2000. Rediscovery of *Wendlandia angustifolia* Wight ex Hook.f. (Rubiaceae) from Tamil Nadu – a species presumed extinct. J. Bombay Nat. Hist. Soc. 97, 311- 313.

3. IUCN Red List of Threatened Species 2018. http://www.iucnredlist.org/

Automating the integration and accessibility of diverse geospatial data cubes

***Heather Savoy*[1], *Debra P.C. Peters*[1.2]**

[1]*New Mexico State University, Jornada Experimental Range Unit, and Jornada Basin Long Term Ecological Research Program, USA*; [2]*USDA-ARS, Jornada Experimental Range Unit, and Jornada Basin Long Term Ecological Research Program*

Corresponding author(s) e-mail: hsavoy.jrn.lter@gmail.com

ABSTRACT:

Continental-scale geospatial data, such as daily precipitation or soil properties across North America, have the capability to serve the data needs of a wide variety of ecological projects, but their cost in terms of time and memory can hinder efficient integration. To reduce the time of downloading big data sets from online repositories, cropping the data to desired extents, harmonizing the resolutions and coordinate reference systems, and performing statistical analyses for multiple data sets, we aim to create a portal that automates these steps. This Data Access and Spatiotemporal Harmonization (DASH) portal is built in Shiny and R with packages for GIS-like functionality and fast memory access for large data sets which allows the portal to be more computationally efficient than using GIS software directly for big data. DASH balances this computational efficiency with the inherent ease of use that graphical user interfaces have such that users can use the tool without knowing R.

This poster will detail the design and functionality of DASH and describe examples of how ecological research projects motivated or benefited from its implementation. These projects use data sets representing climate, soil properties, hydrology, agricultural assets, and land use to address questions pertaining to changes in vegetation, and the role of climate, surface water, topography, and other landscape features in the spread of livestock disease. These individual projects have overlapping data needs, which can result in redundant data acquisition, standardization, harmonization, and storage if approached individually. With the creation of DASH, current and future projects will benefit from an organized repository of these data sets with an interface to provide a suite of functions to support their specific data needs.

KEYWORDS: DASH Portal, R, Shiny, Big Data

Improved environmental monitoring of air and water quality by combination of high throughput image cytometry and deep learning

Susanne Dunker[1,2], **David Boho**[3], **Jana Wäldchen**[4], **Patrick Mäder**[3]

[1]Helmholtz-Centre for Environmental Research, Germany; [2]German Centre for Integrative Biodiversity Research, Germany; [3]Technical University Ilmenau, Germany; [4]Max-Planck-Institute for Biogeochemistry, Germany

Corresponding author(s) e-mail: susanne.dunker@ufz.de

ABSTRACT:

Good air and water quality as ecosystem services are crucial aspects of our daily life and directly related to our well-being. An impairment of these ecosystem services has dramatic consequences for health and quality of life, e.g. in form of respiratory diseases or restricted drinking water supply. Regular monitoring of air and water quality is therefore extremely important and traditionally mainly done via microscopy. Microscopic investigations are of high quality but come with some restrictions like taxonomic expert requirement, difficulties in providing standardized taxonomic knowledge and taking long time, limiting the samples which can be processed in total [1]. A new series of flow cytometers is capable of automatically taking images of particles at a maximum rate of 5000 particles/ sec. Images of referenced species (phytoplankton or pollen) can ideally be used for training a deep learning network. First training and validation of a dataset of phytoplankton species similar in size and shape revealed high accuracy (97 %) of correct identification [2]. The innovative combination of high throughput image cytometry and deep learning is therefore suggested as a promising tool to monitor air and water quality in future.

KEYWORDS: Image flow cytometry, Deep learning, CNN, images, image-based identification, machine learning, high throughput, pollen, phytoplankton

REFERENCES:

1. First, M. R., Drake, L. A., 2012. Performance of the human “counting machine”: evaluation of manual microscopy for enumerating plankton. *Journal of plankton research, 34*(12), 1028-1041.

2. Dunker, S., Wäldchen, J., Boho, D., Mäder, P., submitted. Combining image-based high-throughput cytometry and machine learning for efficient species and life-cycle stage identification of phytoplankton.

Identification of the Green Infrastructure in Switzerland and its implementation at different scales

Arthur Sanguet[1,2]*, Erica Honeck*[2]*, Nicolas Wyler*[1]*, Pascal Martin*[1]*, Martin Schlaepfer*[2]*, Anthony Lehman*[2]

[1]*Botanical Conservatory of Geneva, CH*; [2]*University of Geneva, CH*

Corresponding author(s) e-mail: Arthur.sanguet@unige.ch

ABSTRACT:

In February 2018, the canton of Geneva – Switzerland – has established a roadmap to implement the national and state laws on biodiversity conservation. The country strives to have a functional green infrastructure network of connected habitats to protect biodiversity and ecosystem services. Green Infrastructure (GI) are defined as "a strategically planned network of natural and semi-natural areas with other environmental features designed and managed to deliver a wide range of ecosystem services" [1].

It is composed of core areas, concentrating biodiversity and ecosystem services, and links between them allowing daily movements of animals and habitats connection. Basically, it's a network of protected and conserved natural areas that aimed to be integrated into landscape management to reduce biodiversity loss [2].

While GI is getting more and more attention these last few years, there is no consensus on the method needed to find the best areas to integrate into GI landscape management. Indeed, some articles concentrate their database and analysis on ecosystem services only, others on biodiversity and a few mix these two concepts together (without using the same kind of data, services or group species). Thus, the results of such studies are difficultly comparable and some might not even be as relevant as expected.

To identify areas that might be part of GI, we propose to study 3 main pillars: biodiversity patterns, ecosystem services and finally landscape connectivity & structure. Then, we can assess a spatial prioritization analysis as each element of the landscape is expected to be multifunctionnal and might be interesting for one or several pillar(s). We aim at creating a universal framework to identify GI at different scales and using various sets of variables.

The poster presents the concepts taking the example of the Green Infrastructure identification currently at work in the Geneva Canton (CH).

KEYWORDS: green infrastructure, biodiversity, ecosystem services, connectivity, fragmentation, conservation

REFERENCES:

1. European Environment Agency (2014). Spatial analysis of green infrastructure in Europe. EEA Technical report No 2/2014, ISSN1725-2237, pp. 53..

2. Liquete, C., Kleeschulte, S., Dige, G., Maes, J., Grizzetti, B., Olah, B., and Zulian, G. (2015). Mapping green infrastructure based on ecosystem services and ecological networks: A Pan-European case study. Environ. Sci. Policy 54, 268–280.

Analyzing the effect of armed conflict, agriculture and fire on the movement and migratory behaviour of white eared kob and Roan antelope in the Boma-Gambella landscape of Ethiopia and South Sudan

***Kassahun Abera*[1], *Afework Bekele*[2]**

[1]*GIZ Biodiversity and Forestry Programme, Addis Ababa, Ethiopia*; [2]*Addis Ababa University, Addis Ababa, Ethiopia*

Corresponding author: e-mail: afeworksimegn@gmail.com: kassahun.abera@giz.de; kasahungis@gmail.com

ABSTRACT:

Wildlife migrations are indicators for functionality and connectivity of land or seascapes. A species movement promotes nutrient cycling, seed dispersal, prey predator interaction, as well as species fitness and survival. The disruption of wildlife migration phenomena are alarms to an underlying and complicated socio-political and environmental crisis yet to come. In this thesis I evaluated the disturbance on White eared kob (*Kobus kob leucotis*) and Roan antelope (*Hippotragus equinus)* generated by the armed conflict, livestock encroachment and agricultural investments in the Boma-Gambella landscape. Four years of wildlife telemetry data was collected to analyse and detect the movement and migratory behaviour of White eared kob and Roan antelope against armed conflict, agriculture, fire, livestock and settlements in the landscape. The telemetry data was complemented by the data from five systematic reconnaissance aerial flights. Kernel density estimates, overlay analysis, statistical derivatives and GIS illustration methods were applied to produce results. Migratory kobs have shown behavioural response to the ongoing armed conflict and livestock encroachments by shrinking their former home ranges and making sudden displacements away from conflict hotspot areas. Human footprints and their effect on wildlife conservation in the landscape are evident. The home ranges and movement patterns of resident roan antelopes overlap with the major agricultural concessions and their cropping seasons. The key wildlife core areas, the main migration routes and home ranges exist outside of formal protected areas. The migration route analysis resulted with the longest terrestrial mammal migration and the second largest wildlife migration in the continent.

KEYWORDS: White eared kob, Roan antelope, Migration, Home range, Armed Conflict

Towards FAIR Data in the Biodiversity Exploratories Project

Andreas Ostrowski[1], ***Eleonora Petzold***[1], ***Cornelia Fürstenau***[1]

[1]*Friedrich Schiller University Jena, Germany*

Corresponding author(s) e-mail: andreas.ostrowski@uni-jena.de

ABSTRACT:

More than 10 years ago, the first version of the Biodiversity Exploratories Information System (here BExIS 1) went online to act as the data management platform and information system for the SPP 1374 “Biodiversitäts-Exploratorien” (BE). With its major function as data management system it facilitates data sharing among the BE projects and data re-use across all phases. The BE are data intensive. Availability of data across space and time is a major value of the project design. Following the project data policy, the data has to be made publically available after a specific time. First datasets were made public available on April 25th, 2016 via BExIS 1. Datasets are released at least once a year. To ensure that the data potential can be leveraged a proper way following the FAIR principles is needed.

With our poster, we want to bring our way to make public data available in arrangement with the FAIR data principles. We show for each FAIR element the current situation with our public data, reason this classification, indicate obstacles, and show possible approaches to overcome missing links. We demonstrate the effort undertaken and the open tasks that needs to be done.

Furthermore we want to create awareness of the large number of datasets available from the BE.

KEYWORDS: research data management, FAIR principles, public data, BEXIS, Biodiversity Exploratories

Towards Automatic Identification of Elephants in the Wild

***Matthias Körschens*, Björn Barz, Joachim Denzler**

Computer Vision Group, Friedrich Schiller University Jena, Germany

Corresponding author(s) e-mail: matthias.koerschens@uni-jena.de

ABSTRACT:

Identifying animals from a large group of possible individuals is very important for biodiversity monitoring and especially for collecting data on a small number of individuals that have become particularly interesting for the research, for example by showing an uncommon behaviour. The identification task can be very time-consuming, as in most cases the features of the individual have to be compared with the registered features of all individuals observed before. This task becomes even harder, if the animals look very similar and have only a small number of obvious distinctive features, like elephants do. Some elephant individuals can be identified by the size of their tusk or holes in their ear, but for many others the distinctive features are much more subtle.

In the case of elephant monitoring there is usually only a small group of researchers on site and the animals often stay at the same place only for a short period of time. Therefore, the researchers have to focus their attention on a small set of particular animals most important for research. Thus, these individuals have to be identified quickly out of all the animals on site. For this reason, a system supporting the researchers in identifying elephants would aid and speed up the research progress immensely.

We present such a system for identifying elephants in the face of a large number of individuals with only few training images per individual. For that purpose, we combine object part localization, image features extracted from pre-trained convolutional neural networks, and support vector machine classification to provide field researchers with proposals of possible individuals given new images of an elephant.

The performance of our system is demonstrated on a dataset comprising a total of 2078 images of 276 individual elephants, where we achieve 56% top-1 test accuracy and 80% top-10 accuracy.

To deal with occlusion, varying viewpoints, and different poses present in the dataset, we furthermore enable the analysts to provide the system with multiple images of the same elephant to be identified and aggregate confidence values generated by the classifier.

With that, our system achieves a top-1 accuracy of 74% and a top-10 accuracy of 88% on the held-out test dataset.

KEYWORDS: individual identification, wildlife monitoring, elephants, computer vision, machine learning

Long-term continuous ecoacoustic monitoring of a pristine coral reef: when data-intensive ecology responds directly to ecosystem management objectives

Simon Elise[1, 2]***, Marion Couëdel***[1]***, Joanna Kolasinski***[3]***, Sophie Marinesque***[3]***, Henrich Bruggemann***[1]

[1]*UMR ENTROPIE, La Réunion, France*; [2]*NORTEKMED, France;* [3]*Terres Australes et Antarctiques Françaises (TAAF), La Réunion, France*

Corresponding author(s) e-mail: simon_elise@hotmail.com

ABSTRACT:

The few percent of the world's coral reefs that remain in near pristine conditions represent absolute conservation priorities. Their relative inaccessibility is key to their healthy state but also poses challenges for their management. In the South-West Indian Ocean, the pristine reefs around Europa island have benefited from occasional surveys by in situ traditional methods. While accurate, these scarce surveys fail to provide crucial information on the temporal dynamics of these ecosystems and are particularly ill-suited for evaluating ecosystem responses to intense disturbances (e.g. coral bleaching, hurricanes) that are projected to increase in frequency. Effective management of Europa island coral reefs requires regular feedback from the field in order to: i) survey the natural dynamics of their biological communities; ii) detect disturbances, and understand their origins and mechanisms; iii) define technical instructions of authorized and regulated human activities (eco-tourism, research) to avoid any environmental disturbances; and iv) implement ecosystem conservation and restoration actions to counteract biodiversity and ecosystem functions losses. Ecoacoustics has recently emerged as a promising tool for monitoring marine habitats and associated biodiversity. The CORCOPA project, funded by the BEST 2.0 European Union programme, proposes the long-term application of this method to provide Europa island's managers (the French Southern and Antarctic Lands, i.e. Terres Australes et Antarctiques Françaises - TAAF) with information to implement their conservation strategy. Soundscapes were characterised at various sites on the fringing reefs around Europa island in April 2018, while an autonomous station for continuous passive acoustic monitoring was installed. Here we present the first results of this project. The variability of commonly used ecoacoustic indices (Sound Pressure Level and Acoustic Complexity Index calculated on several frequency bands) from seven sites around the island are described and temporal variations of these indices on the permanent monitoring site are presented. The critical importance of learning methods and classification tools in disentangling useful information from underwater soundscapes in support of management will be discussed.

KEYWORDS: coral reefs, long-term monitoring, ecoacoustic indices, event detection, ecosystem management

BIOfid, a Platform to Enhance Accessibility of Biodiversity Data

Claus Weiland[1], Christine Driller[1], Markus Koch[1], Marco Schmidt[1], Giuseppe Abrami[2], Sajawel Ahmed[2], Alexander Mehler[2], Adrian Pachzelt[3], Gerwin Kasperek[3], Angela Hausinger[3], Thomas Hörnschemeyer[4]

[1]*Senckenberg Biodiversity and Climate Research Centre, Germany;* [2]*Text-technology Lab, Faculty of Computer Science and Mathematics, Goethe-University Frankfurt, Germany;* [3]*University Library Johann Christian Senckenberg, Goethe-University Frankfurt, Germany;* [4]*Senckenberg Research Institute and Natural History Museum Frankfurt, Germany*

Corresponding author(s) e-mail: claus.weiland@senckenberg.de; christine.driller@senckenberg.de

ABSTRACT:

With the ongoing loss of global biodiversity, long-term recordings of species distribution patterns are increasingly becoming important to investigate the causes and consequences for their change. Therefore, the digitization of scientific literature, both modern and historical, has been attracting growing attention in recent years. To meet this growing demand the *Specialised Information Service for Biodiversity Research* (BIOfid) was launched in 2017 with the aim of increasing the availability and accessibility of biodiversity information. Closely tied to the research community the interdisciplinary BIOfid team is digitizing data sources of biodiversity related research and provides a modern and professional infrastructure for hosting and sharing them. As a pilot project, German publications on the distribution and ecology of vascular plants, birds, moths and butterflies covering the past 250 years are prioritized. Large parts of the text corpus defined in accordance with the needs of the relevant German research community have already been transferred to a machine-readable format and will be publicly accessible soon. Software tools for text mining, semantic annotation and analysis with respect to the current trends in machine learning are developed to maximize bioscientific data output through user-specific queries that can be created via the BIOfid web portal (https://www.biofid.de/). To boost knowledge discovery, specific ontologies focusing on morphological traits and taxonomy are being prepared and will continuously be extended to keep up with an ever-expanding volume of literature sources. Here we present the key elements of the BIOfid pipeline with emphasis on a practical approach to develop domain-specific ontologies, to deal with the dynamic nature of taxonomies, and to promote interoperability and standardization of terms and definitions. The taxonomic ontologies are designed to fetch any taxon-related literature entry. In this respect, scientific names, synonyms, vernaculars, taxonomic ranks and other classifications peculiar for the respective organism group are compiled. The sources to be used are common open access platforms like *Catalogue of Life* and the *Global Biodiversity Information Facility* (GBIF), complemented by databases more narrowly focusing on certain taxa, such as the *International Ornithological Congress* (IOC) *World Bird List*. We further give insight into our work on the OBO Foundry *Flora Phenotype Ontology* (FLOPO) and the *Lepidoptera Anatomy Ontology* (LepAO). The latter is build on the already existing *Hymenoptera Anatomy Ontology* (HAO) and developed within the framework of an international collaboration that aims at designing a unified *Insect Anatomy Ontology*. BIOfid is co-funded by the Deutsche Forschungsgemeinschaft (DFG) under project number 326061700.

KEYWORDS: Bio-Ontologies, Text Mining, Machine Learning, Specialised Information Service, Biodiversity Knowledge Base

Drivers of diurnal and seasonal soundscape patterns in temperate grasslands

***Sandra Mueller*[1], *Michael Scherer-Lorenzen*[1]**

[1] *University of Freiburg, Freiburg, Germany*

Corresponding author(s) e-mail: sandra.mueller@biologie.uni-freiburg.de

ABSTRACT:

Soundscape ecology aims to understand how sounds from various sources - biological, geophysical and anthropogenic - can be used to understand coupled natural-human dynamics across different spatial and temporal scales. It has a high potential for large scale ecosystem monitoring especially if combined with other remote sensing tools. Vocalizing animals contribute to the biophonic component with different taxonomic groups (e.g. birds, orthoptera) having characteristic seasonal and diurnal patterns. The acoustic niche hypothesis postulates that a higher organismic diversity should result in a higher acoustic richness and complexity. Land-use intensity affects biodiversity in various ways and therefore could have an indirect effect on acoustic diversity. Abiotic factors such as temperature and rain as well as noise from nearby roads and machinery have been shown to influence the calling behaviour of vocalizing animals and therefor are hypothesized to have additional indirect effects on acoustic diversity.

To test this conceptual framework we set up autonomous recording systems on a total of 300 plots in grasslands and forests established as long term research plots along a land-use and biodiversity gradient by the German Biodiversity Exploratories. Using autonomous recording devices, scheduled to record one minute every ten minutes, we collected data for a whole annual cycle. We applied structural equation modelling to analyse the underlying drivers of these patterns and will present seasonal and diurnal patterns of the soundscape.

Acoustic diversity was highest in May and June, with peaks during dawn and dusk. Bird diversity was an important driver for these patterns, but also rain, distance to the nearest road and topology explained an important part of the variation in acoustic diversity among plots. It could be verified that Land-use intensity acted as an indirect driver of acoustic diversity as it influenced the diversity of birds and orthoptera. Orthoptera, if present, contributed to an additional acoustic diversity peak in July.

KEYWORDS: soundscape ecology, acoustic diversity, acoustic diversity indices, land-use intensity, German Biodiversity Exploratories

Traditional vs. modern Citizen Sciences in Phenology – an example of DWD, ZAMG and naturgucker.de

***Saskia Pietzsch*[1], *Thomas Hübner*[2], *Gaby Schulemann-Maier*[3]**

[1]*Deutscher Wetterdienst (DWD), Germany*; [2]*Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Austria*; [3] naturgucker.de*, Germany*

Corresponding authors e-mail: saskia.pietzsch@dwd.de; thomas.huebner@zamg.ac.at; g.schulemann-maier@naturgucker.de

ABSTRACT:

DWD practices traditional Citizen Sciences for a long time. One example therefor is its phenological observation network. Currently it consists of round about 1200 stationary voluntary observers that report up to 168 plant stages during the year. Data from this network exist since 1951. The observations are made on certain plants in a radius of 5 km and the same height around each observation station. To achieve a valid coverage, Germany is divided into natural regions depending on their specific climatic and geographic characteristics. In each of these natural regions ideally 3 phenological observation stations are located. The volunteers are trained, constantly in contact with DWD and do their observations according to an observation guideline. They get an expense allowance when they do their reports properly and in time. That helps to control the system and the data quality. Disadvantages are that the structures are quite rigid, most of the observers are very old and it is not easy to find successors.

So DWD is interested in checking out modern methods of Citizens Sciences to complement the traditional network. In Phenology there are a lot of activities where Citizen Scientists make their reports online or via apps any time and from all over the world. ZAMG and naturgucker.de already work with these modern methods. The advantages are that the observers are often young, highly motivated, present in social media and easy to reach, so the data access works very quickly. Problems are that they often don´t stay for a long time, the data coverage is irregular, the data are not from defined stations or areas and it is harder to train and support the volunteers. So the quality of data is not ensured.

In an ongoing project since 2018 the users of naturgucker.de as modern Citizen Scientists report a small selection of DWD phenological phases. So DWD can compare these data to the data from its traditional Citizen Scientists in terms of difference, quality and applicability. ZAMG already works with both systems of citizen sciences but in smaller dimensions and can contribute its experiences.

Without knowing the outcome of the before mentioned project yet, the expected conclusion is that both variants of Citizen Sciences can complement each other. In both cases the sense of responsibility of the volunteers to nature is strengthened and they have an individual effect on science. The systems should be optimized by raising the volunteers awareness for the special subject and its importance. The quality control of data especially of the modern Citizen Scientists needs further improvement. And a way needs to be found of using both systems parallel efficiently.

KEYWORDS: Phenology, traditional Citizen Science, modern Citizen Science, observation network, data quality

Galaxy-E, towards a complete open source workflow oriented e-infrastructure in Ecology

***Yvan Le Bras*[1]*, Valentin Chambon*[1]*, Alan Amossé*[1]*, Elisa Michon*[1]*, Clara Urfer*[1]*, Julien Salanikone*[2]*, Elise Brax*[2]*, Marianne Linares*[2]*, Marie Delannoy*[2]*, Grégoire Loïs*[2]*, Romain Julliard*[2]**

[1]*French National Museum of Natural History, Station de Biologie marine, Concarneau - CESCO UMR 7204 Paris, France* ; [2]*French National Museum of Natural History, Station de Biologie marine, Concarneau - CESCO UMR 7204 Paris, France*

Corresponding author(s) e-mail: yvan.le-bras@mnhn.fr

ABSTRACT:

Analysis steps in Ecology require a substantial level of integration of data as well as complex preparation and processing steps. Historically, Ecology labs have produce sets of computer ‘scripts’ written in a variety of programming languages and usable mainly on an interactive unreproducible way. Thus, a challenge consists in developing and preserving data access and processing steps as an open-source and freely available collaborative service. The Galaxy-E project proposes to face it building the first bricks of a workflow oriented e-infrastructure to support the automation of routine tasks and improve accessibility, reproducibility and transparency of ecological research. In addition to this global orientation, and considering the wonderfull Essential Biodiversity Variables (EBVs) concept, Galaxy-E seems to represent an intersting initiative for EBVs workflows creation, share and use to produce biodiversity indicators from raw data.

KEYWORDS: Galaxy, Workflow, Data access, Data analysis, EBV, Accessibility, Reproducibility, Transparency, Species Distribution Modeling, R shiny, Docker, Openstack, SLURM

A Data Mining Approach for Species Co-occurrence Prediction

***Dina Sharafeldeen*, Alsayed Algergawy, Birgitta König-Ries**

Heinz Nixdorf Chair for Distributed Information Systems, Jena, Germany

Corresponding author(s) e-mail: dina.sharafeldeen@uni-jena.de

ABSTRACT:

Considering the rapid loss in biodiversity we evidence right now, understanding the mechanisms behind ecological community assembly is crucial. Ecosystems contain a huge amount of complex interactions between different species (like complex food webs). Therefore, it is quite challenging to study and prove all these interactions empirically and theoretically to enable community assembly modeling. In recent years, enormous development in complex data analysis has been accomplished using data mining techniques in various domains. Data mining pursues the objective of extracting knowledge and analyzing complex data to find existing associations, to extract structures, patterns, and regularities in large and complex databases. We propose to leverage data mining techniques to support biodiversity scientists in community assembly modeling and (positive and negative) species co-occurrence prediction. Our proposed approach consists of three main phases: a pre-processing phase, a processing phase, and a post-processing phase. In the pre-processing phase, the input dataset is transformed to the appropriate format for mining. In the processing phase, association rule mining (Apriori algorithm) is applied on species occurrences datasets to extract associations between species. For example, a pairwise association rule is (sp1 then sp2) is explained as a positive association between sp1 and sp2. These associations rules are statistically evaluated by statistical measures. A prediction model based on the evaluated association rules is built to predict species co-occurrences. As a post-processing, these rules are visualized as a graph to help scientists better understands community assembly. Our next step will be to further evaluate our approach using a wider variety of existing datasets. We then aim to extend it to a broader set of questions.

KEYWORDS: Data mining, Association rules, Prediction, Community assembly, Species co-occurrence

Using multi-temporal RapidEye remote sensing data to map semi-natural grassland communities

Christoph Raab[1]***, Hans Georg Stroh***[1]***, Bettina Tonn***[1]***, Marcus Meißner***[2]***, Nina Rohwer***[2] ***, Niko Balkenhol***[1]***, Johannes Isselstein***[1, 2]

[1]*University of Goettingen, Germany*; [2]*Institut für Wildbiologie Göttingen und Dresden e.V. Germany*

Corresponding author(s) e-mail: craab@uni-goettingen.de

ABSTRACT:

Semi-natural grasslands are habitats with high biodiversity. Nature conservation and management initiatives, such such as the European Habitats Directive (Council Directive 92/43/EEC 1992) impose legal obligations to develop a monitoring framework for grasslands. However, monitoring through field surveys is time- and labour-intensive. In addition, field mapping is difficult to reproduce and prone to subjective interpretation. Remote sensing offers unique possibilities to map and monitor vegetation at large scales. We present a remote-sensing-based monitoring framework for semi-natural grasslands on a Natura 2000 site, using the Random Forest algorithm. The Grafenwoehr military training area is located in the south-east of Germany (Bavaria). A multi-temporal Rapid Eye time series (2014-2017) of 17 images was acquired covering different phenological phases. Field mapping for the two focus areas (Sommerhau: about 140 ha, 11 grassland community classes and Hoehenberg: about 71 ha, 10 grassland community classes), was carried out between 2014 and 2017. Based on the derived grassland community reference map, a stratified random sampling was implemented and repeated 100 times. Each training data candidate set was subsequently screened for potential outliers using the Random Forest proximity measure. The performance of each reduced training data candidate set was estimated using a 5-fold cross-validation approach. The final training data set was selected based on the ratio of Overall Accuracy and the range of the class specific F-score performance measure. Automated training data selection was successfully implemented and revealed good Overall Accuracies, ranging from 77.5 to 86.5%. Image acquisition dates from onset of vegetation (prespring, first spring) and senescence (late summer, first autumn) were identified as important phenological seasons. The estimated probability maps were able to describe transition zones between different grassland communities and can be seen as a better representation of real world conditions, compared to discrete maps. Incorporating future remote sensing data to the presented mapping strategy can identify hot spot areas of change. This will support the monitoring and reporting obligations as required under Ar.-17 of the EU Habitats Directive.

KEYWORDS: European Habitats Directive, monitoring, Random Forest, proximity, training data, variable importance, probability maps

Refined Methodology for Accurately Detecting Objects from Digitized Herbarium Specimens

***Abdelaziz Triki*[1], *Bassem Bouaziz*[1], *Jitendra Gaikwad*[2]**

[1] *MIRACL Laboratory Sfax, Tunisia*; [2] *Friedrich Schiller University Jena, Germany*

Corresponding author(s) e-mail: abdelaziz.triki@yahoo.fr; bassem.bouaziz@isims.usf.tn

ABSTRACT:

There are global initiatives such as TRY trait database that are making efforts to minimize the paucity of functional trait data. The herbarium specimens provide valuable information on functional traits including length, width, and size of leaves and petiole length. Traditionally, scientists extract such information manually, which is time consuming and prone to errors.

To overcome these limitations, worldwide scientists are applying computer vision techniques to automatically extract trait data from digitized specimen images. However, to extract the trait values pre-requisite is to accurately detect objects present on the specimen images. In the Managing Multimedia Data for Science (MAMUDS) project, we have refined a deep learning technique, which is efficiently able to detect objects from the specimen images provided by the herbarium Haussknecht, as compared to other state-of-the-art techniques.

The herbarium Haussknecht in Germany provide access to more than 30000 scanned type specimen images to researchers and public. However, the varied placement and diversity of objects such as plant specimen, scale bar, color pallet, specimen label, envelope, barcode and stamp on the specimen sheet make the task of automatic detection challenging. To address this challenge, we have developed a refined methodology RefYOLO, which is based on the pioneer object detection system called You Only Look Once (YOLO). The refinement was done by editing the activation function of VGG16 model using the Parametric ReLu (PReLU). In our study, we observed that YOLO and other state-of-the-art techniques such as Region-based Convolutional Network (RCNN) and its variants are unable to robustly detect objects from digitized specimen images. RefYOLO efficiently localizes object classes located in specimen images and scales by encoding contextual feature information about classes such as shape, contours and their appearance. After training, the performance of RefYOLO was enhanced using Average-Max pooling method (AM-PM). Overall, RefYOLO provides higher detection accuracy and need less processing time.

KEYWORDS: Functional trait data, Herbarium specimens, deep learning, You Only Look Once (YOLO), Object detection

Assessing biodiversity from space: An example from the diverse tropical mountain rainforests

***Christine I. B. Wallis*[1], *Nina Farwig*[2], *Roland Brandl*[3], *Jörg Bendix*[1]**

[1]LCRS, *University of Marburg, Germany*; [2]Conservation ecology, *University of Marburg, Germany*; [3] Ecology, *University of Marburg*

Corresponding author(s) e-mail: christine.wallis@staff.uni-marburg.de

ABSTRACT:

Remote sensing provides opportunities to monitor measures of biodiversity in space and time. In particular in diverse and highly remote regions remote sensing should be a valuable tool. Recently, most studies investigate the power of airborne active sensors, hyperspectral sensors or at least multispectral sensors at very high resolution which are limited in their access and availability. In comparison, operational spaceborne data have an appropriate life span and repetition rate which enables long-term monitoring. Therefore, we here investigated the performance of multispectral Landsat 8 OLI data to model biodiversity within a tropical rainforest. We considered different biological levels and measures of biodiversity all sampled along an elevation gradient in southern Ecuador. We used partial least squares regressions to model the richness of trees, ants, moths and birds at the species level and forest productivity and four canopy traits (measured as community weighted means) at the community level of trees. We found varying predictability among taxonomic groups as well as among productivity and trait measures. The models indicated differences in their predictor importance unveiling that certain taxa could be modeled using topographical metrics only, while others benefit from the inclusion of spectral and textural metrics. All response variables were mapped across our study area. We extracted the upper 5 % from each map and classified them as hotspot areas. We built a map of combined hotspot criteria and analyzed this map with respect to the existing protected areas and conservation prioritization areas in the tropical mountain rainforests. Our study highlights the importance of remote sensing data to model biodiversity and how these models can be translated into knowledge and prioritization decisions.

KEYWORDS: Remote sensing, Species diversity, Ecosystem functions, Functional leaf traits, Forest productivity

Interactive Plant Identification in Flora Key

Oliver Bley[1], *Jana Wäldchen[2]*, *Patrick Mäder[1]*

[1]*TU Ilmenau, Germany*; [2]*MPI BGC, Germany*

Corresponding author(s) e-mail: oliver.bley@tu-ilmenau.de

ABSTRACT:

Flora Key is a software system with a mobile app interface that brings key-based plant identification towards a general audience.

Plant identification keys help to identify species in the wild based on observing characteristic plant traits. Most often these keys are large books that pack their knowledge into a tome of questions, which by themselves form a tree-like information structure that the reader has to follow carefully. The key user has to be trained to correctly assess the plant character that is asked for in the key with regard to the body of plants included. Therefore these kind of keys are specialized tools for botanical experts.

Consequential, one of the main challenges in designing an interactive plant identification key for experts and laymen alike is reacting more dynamically upon the users' responses. The uncertainty of the identification process increases due to the nature of the data collection through observations by potential inexperienced users. Each step is not anymore about filtering through the plant set quickly but about asking the right question to isolate the likely species. Therefore the actual sequence of questions might be different between users and sessions even though they want to identify the same plant.

Finding the right balance between asking about general plant characteristics and inquiring about the already collected data can also be seen as a multi-step decision process that has to maintain an optimal trade-off between exploration and exploitation within each step. Following this abstraction we are looking for the best strategy in Flora Key which question to ask next based on the provided answers. To accomplish this goal we are using a planning strategy that is backed by a model.

To this end we have gathered a corpus of several hundreds of species with their respective characters to support our decision making process. The focus of this expert database is the plant wildlife in Thuringia. In the Flora Key algorithm the model is used as prior knowledge for determining the most likely species as well as for the reasoning which question to ask next. Our baseline strategy will evaluate the question with the optimal expected information gain. This is with regard to a species candidate list that will be narrowed down over time until we can make a final suggestion.

Our poster will describe the problem domain and the main components of our solution, as well as additional goals we wanted to achieve with Flora Key. We visualized the results of our approach, and how it performs according to the problem we have described. We are also giving an outlook on future research.

KEYWORDS: Species Identification, Identification Key, Model-based Decision Planning, Question Strategy

PylotWhale a python package for automatically annotating bioacoustic recordings

***Florencia Noriega[1,2]*, Heike Vester [3], Marc Timme[1,2]**

[1]*Chair for Network Dynamics, Center for Advancing Electronics Dresden (cfaed), TU Dresden, Germany;* [2]*Institute for Theoretical Physics, TU Dresden, Germany;* [3]*Ocean Sounds, Marine Mammal Research & Conservation Charity*; *Germany*

Corresponding author(s) e-mail: floreencia@gmail.com

ABSTRACT:

Bioacoustic data contain information essential for understanding animals as well as the interactions within their ecosystems. Yet to date, there is no standard approach for extracting this information. Annotating recordings is one way of transforming acoustics data into information. Here we present a novel python package PylotWhale for automatically annotating bioacoustic recordings with a machine learning classifier. The framework relays on text annotations for audio — a standard and flexible format, handleable by many audio processing platforms. The classifier is trained with annotated recordings and its output is text annotations. The package supports two classification modes. In the first mode, classification instances are time frames of equal length from a recording. This classification scheme is capable of detecting or segmenting animal sounds. We illustrate this segmenting mode classification by detecting whale calls and bat chirps. In the second mode, the classification instances are recording segments of varying durations. This classification scheme is capable of classifying different types of vocal units. We illustrate this mode by classifying calls of pilot whales into predefined call types. The package is written in Python, an open source programming language that has gained great popularity among the scientific community. The modularity of the package enables us to easily tune the feature extraction procedure and type of classifier to suit the classification task at hand. Automating the annotation of bioacoustic recordings enables researchers to process larger volumes of data in an objective manner, thereby paving the way for understanding complex eco-acoustic interactions, with impact on mitigation and supporting conservation efforts.

KEYWORDS: automatic annotation, classification, whale call, bat calls, python package, segmentation

Effects of Environmental Variables and Zooplankton on Picocyanobacteria Dominance in Two Tropical Mesotrophic Reservoirs

Maria Isabel Rocha[1], Friedrich Recknagel[2], Ricardo Minoti[3], Vera Huszar[4], Betina Kozlowsky-Suzuki[1], Hongqing Cao[2], Fernando Starling[5], Ewerton Oliveira[1], Christina Castelo Branco[1]

[1]*Universidade Federal do Estado do Rio de Janeiro, Brazil;* [2]*University of Adelaide, Australia;* [3]*Universidade de Brasília, Brazil;* [4]*Universidade Federal do Rio de Janeiro, Brazil;* [5]*Companhia de Saneamento Ambiental do Distrito Federal, Brazil.*

Corresponding author(s) e-mail: cbranco.unirio@gmail.com

ABSTRACT:

Picocyanobacteria are ubiquitous in lakes of different trophic status but the conditions favouring their occurrence in these environments are not yet established. Whereas cyanobacteria-forming blooms are well known from an ecological and taxonomic point of view, the main components of the so-called picocyanobacteria are often underestimated in water samples, because of either their small size or analytical difficulty in evaluating them. Meanwhile, the dominance of pelagic picocyanobacteria has been recorded in tropical and sub-tropical lakes and reservoirs in Africa, Australia and South America. They can be present in several kinds of aquatic ecosystems and parallel information on their toxicity and geosmin synthesis has increased. Although evolutionary computation has been applied to predict the occurrence of massive cyanobacteria proliferations, in the present study this tool was used to explore the factors responsible for maintaining picocyanobacteria dominance.

This study focused on the diagnosis of the relationships among picocyanobacteria, zooplankton and environmental variables using evolutionary computation. Aiming to increase the understanding of factors promoting dominance of picocyanobacteria in tropical mesotrophic reservoirs, we choose two reservoirs used for water supply and in different regions of Brazil. One is Ribeirão das Lajes Reservoir, located in the Southeast part of the country and constructed for energy production. Now its waters are used now for domestic supply for about 1 million people, in beer industry, sport fishing and fish cage culture. The other is Descoberto Reservoir, in the Central-west part of Brazil and created for domestic supply. Currently it supplies 62% of the drinking water for the city of Brasília (around 1.9 million people). Data comprise 241 monthly samplings near the dam area of both systems taken from 2005 to 2014.

The modelling approach was conducted using the hybrid evolutionary algorithm HEA. The integrated data analysis performed by our study was very successful to elucidate dynamics of picocyanobacteria density variation influenced both by abiotic and biotic factors by the modelling approach. The relative water column stability - RWCS and electrical conductivity were highlighted as the most important environmental drivers for picocyanobacteria peaks. The Hybrid Evolutionary Analysis models for the two reservoirs indicated that rotifers, small-seized cladocerans and copepods can act directly or indirectly in the control of the picocyanobacteria in tropical mesotrophic reservoirs, depending on the conditions of RWCS and electrical conductivity. However, this control is also modulated by threshold of pH, water transparency and water temperature.

KEYWORDS: Cyanobacteria, Hybrid Evolutionary Analysis, Hydroelectric Reservoir, Water Quality

CONNECTing remote sensing and *in situ* measurements in river-lake chain research

Christine Kiel[1], Stella A Berger[1], Sabine Wollrab[1], Jens C Nejstgaard[1], Hans-Peter Grossart[1], Gabriel Singer[1], Franz Hölker[1], Andreas Jechow[1], Jürgen Fischer[2], Thomas Ruhtz[2], Peter Gege[3], Torsten Sachs[4], Matthias Labrenz[5], Gunnar Lischeid[6], Rüdiger Röttgers[7], Thomas Schneider[8]

[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin/Stechlin, Germany; [2]Free University of Berlin (FUB), Germany; [3]German Aerospace Center (DLR), Oberpfaffenhofen, Germany; [4]German Research Centre for Geosciences (GFZ), Potsdam, Germany; [5]Leibniz Institute for Baltic Sea Research (IOW), Warnemünde, Germany; [6]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany; [7]Helmholtz Centre Geesthacht (HZG), Germany; [8]Technical University Munich (TUM), Germany

Corresponding author e-mail: kiel@igb-berlin.de

ABSTRACT:

Lake systems are facing multiple stressors like anthropogenic eutrophication and an increasing frequency of extreme weather events due to climate change. Eutrophication increases the likelihood of harmful algal blooms, which might lead to elevated greenhouse gas emissions from lakes. Heavy rain events and thus nutrient run-off from the catchment area can specifically influence nutrient levels and bloom development of downstream lakes along river-connected lake chains. However, phytoplankton dynamics and algal bloom development along river-connected lake chains are currently poorly understood. This is of special concern, since the water quality of river-connected German lowland lakes has been reported as in poor or bad ecological status according to the European Water Frame Directive.

The interdisciplinary project CONNECT therefore, aims at getting comprehensive insights on lake-chain ecology, i.e. lake coherence along river-connected lake systems, by using high frequency *in-situ* measurements in combination with remote sensing data, enabling lake monitoring on high spatial and temporal scales. We hypothesize that lake connectivity is of major importance for spreading of eutrophication events and algal blooms revealing coherent behavior of lake ecosystems.

We will test our hypotheses of lake-to-lake connectivity in a field campaign on river-lake chains in the NE German lowlands and a large-scale enclosure experiment in the highly instrumentally equipped LakeLab where we will manipulate lake connectivity. This combination of high frequency *in-situ* measurements, high throughput approaches by FlowCam, HPLC algal pigment analyses and remote sensing will improve environmental monitoring and management of river-lake chain ecosystems.

KEYWORDS: river-lake chains, phytoplankton dynamics, remote sensing, high-frequency *in situ* measurements

Spatial localization of vocalizations of Spotted Towhee (Pipilo maculatus) in playback experiments using robot audition techniques

***Shinji Sumitani*[1], *Reiji Suzuki* [1], *Shiho Matsubayashi*[2], *Takaya Arita*[1], *Kazuhiro Nakadai*[3, 4], *Hiroshi G. Okuno*[5]**

[1]*Nagoya University, Japan*; [2]*Osaka University, Japan*; [3]*Tokyo Institute of Technology, Japan*; [4]*Honda Research Institute Japan Co., Ltd., Japan*; [5]*Waseda University, Japan*

Corresponding author(s) e-mail: sumitani@alife.cs.is.nagoya-u.ac.jp

ABSTRACT:

It is important to clarify fine-scaled relationships between characteristics of songs and their behavioral contexts in order to understand ecological functions and semantics of bird songs. For this purpose, we are developing a portable system for bird song localization called HARKBird (Suzuki et al., 2017), which automatically extract sound sources and their direction of arrival (DOA). HARKBird consists of a laptop PC with an open source software for robot audition HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) (Nakadai et al., 2017)) combined with a low-cost microphone array. We conducted a real-time 2D localization and offline classification of vocalizations of Spotted Towhee (Pipilo maculatus) in California during playback experiments using a dimension reduction algorithm t-SNE and a clustering algorithm DBSCAN. We could estimate the position of a loudspeaker replaying conspecific vocalizations and classify some vocalization types of the target individuals semi-automatically with high accuracy (Sumitani et al., in press). We discuss differences in the spatial distributions of these vocalization types, and also discuss effects of vocalization types of replayed vocalizations on the behaviour of the target individuals. Acknowledgements: We thank Charles E. Taylor and Martin L. Cody (UCLA) for field experiments. This work was supported in part by JSPS/MEXT KAKENHI: JP16K00294, JP17H06841, JP18K11467, and JP17H06383 in #4903.

KEYWORDS: Sound source localization, Birdsongs, Spotted Towhee, HARK, t-SNE

REFERENCES:

1. Suzuki, R., Matsubayashi, S., Nakadai, K., Okuno, H.G., 2017. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. Journal of Robotics and Mechatronics. 27(1), 213-223.

2. Nakadai, K., Okuno, H.G., Mizumoto, T., 2017. Development, deployment and applications of robot audition open source software HARK. Journal of Robotics and Mechatronics, 27(1), 16-25.

3. Sumitani, S., Suzuki, R., Matsubayashi, S., Arita, T., Nakadai, K., Okuno, H.G., in press. Extracting the relationship between the spatial distribution and types of bird vocalizations using robot audition system HARK. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018).

An overview of the Saint-Pierre-and-Miquelon marine soundscape

***Paul Nguyen Hong Duc*[1], *Dorian Cazau*[2], *Olivier Adam*[1], *Odile Gérard*[3], *Paul White*[4]**

[1]*Institut Jean Le Rond d'Alembert, Sorbonne University, CNRS UMR 7190, France*; [2]*IMT Atlantique, Lab-STICC, CNRS UMR 6285, France*; [3]*DGA-TN, France*; [4]*Institute of Sound and Vibration Research, University of Southampton, UK*

Corresponding author(s) e-mail: p.nguyenhongduc@gmail.com

ABSTRACT:

Underwater soundscapes are valuable source information as they are the result of complex interactions between biophony, geophony and anthrophony at various spatio-temporal scales. Describing the soundscape and then understanding relationships between these different types of sounds may help to protect marine ecosystems. Underwater soundscapes can bring complementary information about life activities from underwater or aerial visual observations. Sometimes because of the access to the ocean (due to bad weather conditions, lack of observers, or too much expensive outcomes to organise field missions), passive acoustic monitoring becomes an attractive alternative. We deployed a semi-permanent acoustic observatory off the Saint-Pierre-and-Miquelon Archipelago (French islands close to Newfoundland, Canada) for three reasons: firstly, this geographic site is the hot-spot for all different cetacean species, especially the migratory species which come to eat from June to October, including humpback whales, fin whales and blue whales. Secondly, the weather conditions (cold, fog, rain) are not appropriate for boat trips and to collect recurrent visual observations, even at the sea surface. Thirdly, human activities are present, including fisheries and marine traffic to the US and Canada countries, and will potentially dramatically increase in the next decade (project to build a hub to optimically organize the marine trafic in the St Lawrence channel). Our acoustic observatory is used to provide the presence of vocal cetaceans off Saint-Pierre-and-Miquelon islands and also to inform about the anthropogenic sounds. This poster presents the first analysis of the Saint-Pierre-and-Miquelon (SPM) archipelago soundscape. Acoustic recordings were sampled at different time periods to highlight seasonal variations over two years. A standard soundscape analysis workflow was used to compute acoustic metrics such as Power Spectral Density (PSD), third-octave levels and sound pressure level (SPL). Results show that the SPM marine soundscape is made up of three main sound sources. Several cetacean species were encountered in the audio recordings such as sperm whales (which were never seen by visual observations, except the stranding of a male individual in 2014), humpback and blue whales. Seasonality was observed in their sound activity. Moreover, an increase of marine traffic was noticed over the two years. Finally, weather conditions, especially the high intensity of rain, came acoustically masking the study of the SPM soundscape. This study can help to regulate human activities, for example to set up new shipping roads to avoid collision with cetaceans during their migration.

KEYWORDS: underwater, soundscape, passive acoustic monitoring

RSonde Tool: A project for developing a reservoir monitoring tool coupling remote sensing and high frequency in situ data

Carmen Cillero[1], Jordi Delgado[2], Ramón A. Díaz Varela[3], José Antonio Domínguez[4], Boris Hinojo[1], José Luis Cereijo[2], David García[2], Federico Cheda[1], Marco Rubinos[1]

[1]R&D Department. 3edata S.L., Spain.[2]Civil Engineering School. A Coruña University. Spain.[3]Botany Department. Santiago de Compostela Unversity. Spain. [4]Crop Research Institute Praha. Czech Republic.

Corresponding author(s) e-mail: carmen.cillero@3edata.es

ABSTRACT:

In the European Union, the Water Framework Directive (2000/60/EC) requires periodic monitoring of reservoirs and water bodies. Currently, most of the reservoir monitoring programmes are based on specific measures taken in the field, which make it difficult to capture the spatial and temporal variability of phenomena such as cyanobacterial blooms, with irregular spatio-temporal patterns.

The overall objective of the project is to study the viability of combining three types of data: Satellite imagery (Landsat 8 and Sentinel 2), UAV multispectral imagery and high frequency in situ water quality data (HF data) through the entire water column, for detailed monitoring of the water quality status in reservoirs. The project aims at developing a methodology for spatially-explicit modelling of relevant parameters (namely transparency, Chla, phycocyanin, CDOM and temperature) as a tool for supporting reservoir management and decision making, focused on problems with high social and economic impact (eutrophication, harmful algal blooms).

In this work we present the first results obtained in the spatial modelling of Chlorophyll-a with OLI sensor imagery (on board Landsat 8) and with the images obtained with the first testing UAV flights, that were taken with a commercial multispectral sensor with 5 bands of acquisition in the visible and near-infrared (Rededge. Micasense) on board an octocopter (Atyges. System FV8). We have tested the performance of 6 published algorithms and 1 normalized index with Landsat 8 imagery (years 2016 and 2017) in two reservoirs in the same catchment in Spain (Galicia. NW Iberian Peninsula), an area with a high frequency of cloud cover. All of them showed significative correlation with surface in situ Chl-a but not for both reservoirs. The first testing UAV flight was done in one reservoir during september 2017, when a state of alert for drought and cyanobacterial bloom was declared and no satellite images were available for a period of 24 consecutive days due to cloud cover. The previously validated relationships were tested with the UAV imagery data and chlorophyll in situ data obtained both with an YSI6026 Chlorophyll probe attached to a YSI6600 V2 sonde at different depths, and Chlorophyll samples taken at 0,6*Secchi Depth (SD). The best performances were obtained with two published indexes: SABI (4 bands) and 2BDA (2 bands); and the in situ chlorophyll data obtained with the probe at depth of 0,6*SD.

KEYWORDS: water quality, HF data, remote sensing, UAV, monitoring tool.

Analysis of almost a hundred long distance migration paths of Italian and Swiss barn swallows (Hirundo rustica) reconstructed on the basis of light level geolocators

Mattia Pancerasa[1]*, *Roberto Ambrosini[2]*, David W. Winkler[3], Eldar Rakhimberdiev[4], Diego Rubolini[2], Felix Liechti[5], Nicola Saino[2], *Renato Casagrandi[1]

[1]*Politecnico di Milano, IT*; [2]*Università degli Studi di Milano, IT*; [3]*Cornell University, USA;*[4]*NIOZ Royal Netherlands Institute for Sea Research, NL;* [5]*Swiss Ornithological Institute, CH*

Corresponding authors e-mail: mattia.pancerasa@polimi.it; roberto.ambrosini@unimi.it; renato.casagrandi@polimi.it

ABSTRACT:

Barn swallow (Hirundo rustica) is one of the most iconic migratory bird, worldwide known for its fascinating seasonal journeys. Each year, every individual encompasses thousands of kilometres looking for favourable environmental conditions where it could overwinter in Africa, returning the typically at the same nest in Europe the subsequent spring for breeding. Despite the numerous attempts to track barn swallow's migration throughout the last century (e.g. by ringing data, genetic analyses and stable isotopes), the reconstruction of the first complete paths of very few individuals (six individuals in total) has been possible just in recent years, thanks to the use of light level geolocators. These devices, used for measuring daylight throughout the whole migration of each specimen, are now a robust technology applied for tracking migratory animals, in particular those that are too small to carry satellite tags (e.g. passerine birds). In the last decade, different methods have been proposed to reconstruct migration paths from geolocators data, based on astronomical equations binding variation of solar irradiance, the hour and the date during twilight events. Here we analysed almost a hundred of light-level geolocators data of Italian and Switzerland barn swallows, tagged in the years 2010, 2011 and 2012. We used `TWGeos`, `FLightR` and `trajr` R packages respectively to elaborate geolocators data, to perform path estimation and to compute trajectories indices. The analysis of migration trajectories revealed that spring migration display more complex paths than autumn migration, not only in terms of inter-individual geographical variability, but also in terms of path tortuosity. We noticed that more individuals made a westward detour during their spring migration, some of them performing a journey even double compared to the great circle distance between wintering and breeding positions. A first set of statistics on trajectories is presented and studied in relation to timing of departure and arrival. Statistical analysis reveal that the year of migration has a significant effect on trajectories indices, while there is less influence from the population or sex of the individual. This could be a clue of the importance of *en-route* weather conditions on barn swallow migration. Our reconstruction may serve as a basis for more in depth and possibly mechanistic investigation on the movement ecology of European barn swallows.

KEYWORDS: Barn swallow, Bird migration, Light-level geolocator, Migration path, Trajectories analysis

Multimodal Plant Species Identification

Hans Christian Wittich[1], David Boho[1], Patrick Mäder[1]

[1]*TU Ilmenau, Germany*

Corresponding author e-mail: hans-christian.wittich@tu-ilmenau.de

ABSTRACT:

Wild flowering plants represent a complex biological system, their accurate identification requiring expert knowledge only available to an ever-decreasing number of botanists. Machine learning can be applied to plant identification tasks, typically using a photo taken of the plant as the most discriminative information source. While it is possible to use an arbitrary photo for image recognition, a single data sample alone does not generally provide a complete representation of the complex system necessary for high classification accuracy. A photo taken from a single perspective may therefore not depict all features relevant for successful identification; even for taxa reliably distinguishable by visible features alone it is often necessary to analyze specific plant organs.

We propose an interactive guided plant observation process integrating multimodal data sources, each characterizing complementary aspects of a plant observation and combine machine learning-based image classification with contextual analysis of spatio-temporal metadata. In a two-fold strategy we analyze images of plant organs taken by users in the field, estimate their predictive uncertainty and select those modalities providing the highest information gain, informing additional iteration steps if necessary. At the same time, we analyze the spatio-temporal context under which observations are made. By modeling environmental similarity and geographical distance to known observations we predict the plant taxa most likely to be observed at the current geographical location and day of year.

While it is important to achieve high accuracy, consolidating information from multiple modes also allows estimating predictive uncertainty, thereby providing an important measure about the confidence of the prediction.

KEYWORDS: Plant Species Identification, Deep Learning, Spatio-temporal Context, Multimodal, Plant Species Distribution

Content-based Image Retrieval for Biodiversity: A Comparative Study

Nora Youssef[1], **Alsayed Algergawy**[2], **Ibrahim F. Moawad**[1], **EL-Sayed M. EL-Horbaty**[1], **Rania Elgohary**[1]

[1]*Faculty of Computer and Information Sciences, Ain Shams University, Egypt*; [2]*Institute of Computer Science, Friedrich Schiller University of Jena, Germany*

Corresponding author(s) e-mail: nora.youssef@cis.asu.edu.eg

ABSTRACT:

Biodiversity resources are rapidly growing with a huge increase of visual contents. Thus, searching these images accurately and efficiently for all types of datasets becomes one of the most challenging tasks. Content-based image retrieval (CBIR) is the technique that retrieves images based on their visual contents. So that, selecting appropriate features that describe an image sufficiently is a clue for a successful retrieval system. To this end, a comparative study to investigate the effect of using a single and a combined set of features in the context of a CBIR is presented. To achieve this goal, several features including, edge histogram (EHD), color layout (CLD), and fuzzy color texture histogram (FCTH) as well as different combinations of these features such as, all edges (local, global and semi-global edges), all edges with CLD. Moreover, all edges with FCTH have been exploited. To demonstrate the effectiveness of the proposed method, a set of experiments utilizing different ImageNet biodiversity synsets (Animal plant flora, Plant life, Owl, Elephant, and etc.), have been carried out. The results in terms of precision, recall, F-measure, and mean average precision, show a higher retrieval accuracy while using a set of combined features compared to exploiting only single features for the same retrieval task.

KEYWORDS: Biodiversity CBIR, Combined Features, EHD, CLD, FCTH.

Nocturnal bird migration on the Baltic Sea coast

***Hanna Pamuła*[1], Maciej Kłaczyński[1]**

[1] *AGH University of Science and Technology, Kraków, Poland*

Corresponding author(s) e-mail: hanna.pamula@agh.edu.pl

ABSTRACT:

Passerines migration is studied for over 50 years on the Polish Baltic Sea coast. Bird-ringing programmes and direct observations are the most popular bird research methods. However, these techniques are usually not useful to track the nocturnal migration, so modern technologies such as tracking geolocators, radar monitoring, thermal imaging and acoustic recordings are considered.

This study presents the first results of the research on audio bird monitoring. Two months of autumn migration was recorded and part of the data was manually annotated. As night flight calls are soft and impulsive sounds, their detection is a very challenging task. We tried to parameterise the calls and check whether simple classification methods such as logistic regression, support vector machines or k-neighbours models can detect the birds sounds with acceptable accuracy. The results indicate that such standard classifiers are too simple and therefore more advanced techniques such as machine or deep learning should be applied, such as e.g. convolutional neural networks. To conclude, the acoustics methods combined with advanced classification techniques are showing great potential to supplement the research on nocturnal bird migration.

KEYWORDS: Acoustic Monitoring, Night Flight Calls, Bird Calls Detection, Bioacoustics

Plant Diseases prediction using Deep Learning

Jihen amara[1]***, Bassem Bouaziz***[1]***, Alsayed Algergawy***[2]

[1]*Miracl Laboratory, University of Sfax*; [2]*FSU Jena, Germany*

Corresponding author(s) e-mail: bassem.bouaziz@isims.usf.tn

ABSTRACT:

Crop diseases are major sources of famine and food insecurity on our planet. In fact, it is estimated that plant pathogens may account for annual crop yield losses of up to 16% globally. Furthermore, the current solutions to fight different diseases need the massive use of crop protection products, which are dangerous for the environment and the user. While Microscope and DNA sequencing-based methods are effective to identify and discover different kinds of diseases, many of the farmers over the world do not have access to these diagnostics tools but the vast majority of them possesses a cell phone. Hence, a phone-based tool that helps in diagnosing crop diseases based on capturing and analyzing automatically a picture of a plant leaf is a promising solution. To deal with the mentioned challenges, we introduce a deep learning-based approach to classify and identify banana leaves diseases.

Convolutional Neural Networks (CNNs) are known for their robustness toward low variation in inputs, they require low pre-processing for their execution. They are also able to extract appropriate features while simultaneously performing discrimination. More specifically, in the current implementation we updated LeNet architecture. Our model is composed of three main parts which are convolution, pooling and fully connected layers. The convolution and pooling layers act as self-taught feature extractors from the input images while the fully connected layer acts as a classifier.

To validate the performance of the proposed approach, we conducted a set of experiments using a real dataset of banana diseases obtained from the PlantVillage project (https://www.plantvillage.org/en/). The plant village project contains thousands of images of healthy and diseased crop plants that are open and available on the web. The images in our dataset are annotated as belonging to three different categories which are healthy (1643 images), blacksigatoka(725images) and blackspeckle(1332images). These images are captured with different sizes, orientation, poses, backgrounds and illumination.

In our implementation we used the deeplearning4j as an open source deep learning library which supports the use of GPUs to make the execution of the deep learning algorithms faster. The obtained results confirm the effectiveness of our model in term of accuracy, recall, precision and F1-mesure evaluated respectively by 0.9861, 0.9867, 0.9861 and 0.9864.

The proposed model can serve as a decision support tool to help farmers to identify the disease in the banana plant.

KEYWORDS: plant deseases, deep learning, classification, CNN

Specimen-GT tool: Ground Truth Annotation tool for herbarium Specimen images

Bassem Bouaziz[1], ***Rochdi ben Ali***[1], ***Abdelaziz Triki***[1], ***Jitendra Gaikwad***[2]

[1]*Miracl Laboratory, University of Sfax*; [2]*FSU Jena, Germany*

Corresponding author(s) e-mail: bassem.bouaziz@isims.usf.tn

ABSTRACT:

Digitized herbarium specimens are very important for scientists belonging to medical, biological, chemical, botanical and in general bio scientist’s communities. Extracting traits and identifying species of plants are highly demanded even for species inventory and traits extraction.

Information on specimen are considered as metadata and can be then stored in a standard format (ie ABCD). As for the specimen images content, it needs tools and techniques to make it available for scientist especially the morphological traits (https://www.try-db.org). Measuring leaves for example need at first use of segmentation techniques before the measurement process.

Recently, with the success demonstrated by machine learning techniques to solve segmentation challenges there is a growing need for ground truth data not only for the purpose of evaluation but also for labelling learning datasets. The labelled datasets are then stored in standard and structured format linking data and metadata.

Specimen-GT tool is a system for editing and reviewing specimen image data and metadata. It provides the following functionalities:

1. Generate data structure to store and manage 'ground truth' data.
2. Generate and respect naming convention between specimen images data and annotation files
3. Provide a tool box to extract measurement from specimen leaves.
4. Interact with Bexis platform
5. Organize annotation data and metadata into files according to a dynamic structure:
6. Generate file SHAPE-data
7. Generate Content annotated file DATA.
8. Generate ABCD-data to manage store and share meta-data
9. Generate resized specimen image data
10. Visualize annotated data from specimen
11. Provide a toolbox to customize (add-update-delete) attributes and associated types (shape, number, text, etc).

Actually, the tool is used to create a full annotated dataset with (leaves measurements, missed part leaf, perfect leaf, herb, overlapped leaves). The annotated images count over 4000.

KEYWORDS: digitized specimen, dataset, annotation, tool, ABCD

Proceedings of the 10th International Conference on Ecological Informatics

Developing semantic interoperability in ecology and ecosystem studies: the AnaEE infrastructure framework

Christian Pichot[1]***, Cécile Callou***[2]***, André Chanzy***[3]***, Philippe Clastre***[1]***, Mohamed El-Hamadry***[4]***, Benjamin Jaillet***[1]***, François Lafolie***[3]***, Jean-François Legalliard***[5]***, Chloé Martin***[2]***, Dalien Maurice***[6]***, Nicolas Moitrier***[3]***, Ghislaine Monet***[7]***, Hélène Raynal***[4]***, Antoine Schellenberger***[8]***, Rachid Yahiaoui***[8]

[1]*INRA URFM, Avignon,France;* [2]*UMS BBEES, Paris, France;* [3]*INRA UMR EMMAH, Avignon, France;* [4]*INRA MIAT, Auzeville, France;* [5]*UMS CEREEP-Ecotron, St Pierre les Nemours, France;* [6]*INRA UMR EEF, Champenoux, France;* [7]*INRA UMR CARRTEL, Thonon-les-bains, France;* [8]*INRA INFOSOL, Orléans, France*

Corresponding author(s) e-mail: christian.pichot@inra.fr

ABSTRACT:

The AnaEE (Analysis and Experimentation on Ecosystems, ESFRI) Research Infrastructure offers experimental facilities for studying ecosystems and biodiversity [1].

A distributed Information System (IS) is developed, based on semantic interoperability of its components and the use of common vocabularies (AnaeeThes thesaurus and an OBOE-based ontology). Discovery and access portals are fed by information (rdf triples) produced by the semantic annotation of AnaEE distributed resources: relational databases and modeling platforms.

A first pipeline was developped for the automation of the annotation process and the production of the semantic data, annotation that may represent a huge conceptual and practical work. A second pipeline is devoted to the exploitation of these semantic data through the generation i) of standardised ISO and GeoDCAT metadata records and ii) of data files (NetCDF format) from selected perimeters (experimental sites, years, experimental factors, measured variables…). These tools are part of the ENVRIplus service portfolio and will be usable in different contexts of ontologies and databases.

KEYWORDS: Ecology, Interoperability, Linked data, Ontology, Distributed architecture

REFERENCES:

1. Clobert , J., Chanzy, A., Le Galliard, J.-F., Chabbi, A., Greiveldinger, L., Caquet, T., Loreau, M., Mougin, C., Pichot, C., Roy, J., Saint-André, L., 2018. How to integrate experimental research approaches in ecological and environmental studies: AnaEE France as an example. Frontiers in Ecology and Evolution, 6 (43). , DOI : 10.3389/fevo.2018.00043

Improving the Quality of Biodiversity Data Through Semantic Web Standards

Marcos Zárate[1, 2], **Mirtha Lewis**[1], **Pablo Fillottrani**[3, 4], **Claudio Delrieux**[5]

[1]Centre for the Study of Marine Systems, Patagonian National Research Centre, Argentina; [2]Laboratory for Computer Science, National University of the Patagonia San Juan Bosco, Argentina; [3]Computer Science and Engineering Department, Universidad Nacional del Sur, Bahia Blanca, Argentina; [4]Comisión de Investigaciones Científicas, Provincia de Buenos Aires, Argentina; [5]Electric and Computer Engineering Department, Universidad Nacional del Sur, Bahia Blanca, Argentina

Corresponding author e-mail: zarate@cenpat-conicet.gob.ar

ABSTRACT:

The lack of accurate spatial geographic information in the occurrence data of species generates problems in many conservation activities, such as systematic planning for the protection of endangered species. In this abstract we describe the experiences made to improve the location quality of biodiversity data extracted from an Integrated Publishing Toolkit (IPT) http://ipt.cenpat-conicet.gob.ar:8081/ belonging to the Patagonian National Research Centre in Darwin Core Archive format (DwC-A) [1]. Our approach is based on previous work [2] where we published a set of biodiversity data using Resource Description Framework (RDF) [3], a standard model for data interchange on the Web. The main scientific questions to answer currently are: (1) *how can we integrate biodiversity data from different sources using its geographic location?* (2) *how can we check if locations of a DwC file are consistent?* (3) *How can we correct the locations that are wrong?*

While there are tools to check the quality of biodiversity data, there is a gap to ensure the quality of the georeferenced data in a dataset. To answer these questions, we added more semantics to geographic locations using ontologies such as Linked GeoData http://linkedgeodata.org/About, GeoNames http://www.geonames.org/ and GeoSPARQL http://www.opengeospatial.org/standards/geosparql which allows us check on if a certain position (latitude and longitude) is located in the spatial coverage of the region described. For example, with a simple SPARQL query [4] we know if a position that claims to belong to an area/region is correct or not. Another important improvement is that the values of certain fields of DwC that were previously literal, are now replaced by references to URI, such as the field *dwc:country* containing the literal *Argentina*, was replaced by the URI http://sws.geonames.org/3865483/.

In addition, the use of the GeoSPARQL standard allows complex semantic queries, an example of this is: *find all non-native or invasive species that have occurrences within a certain region defined by the user*.

Although the quality of the data depends on many factors and previous controls, we believe that taking advantage of the Semantic Web [5] and in particular of GeoSPARQL can help to address these problems. However, widespread adoption and implementation remain a challenge. As future work, we intend to extend our current implementation with more advanced requests, in partnership with biodiversity researchers to develop, improve and test this tool of quality control for species location. We also aim to build a benchmark to assess the accuracy and recall of our queries.

KEYWORDS: Data Quality, Darwin Core, Semantic Web, GeoSPARQL

REFERENCES:

1. Remsen D, K.D.M.R.T. Braak (2011). Darwin Core Archive How-To Guide. https://github.com/gbif/ipt/wiki/DwCAHowToGuide (accessed 26 July 2018).

2. Zárate, M., Braun, G., & Fillottrani, P. (2017). Adding Biodiversity Datasets from Argentinian Patagonia to the Web of Data, 2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference 2017. Vienna, Austria: URL: http://ceur-ws.org/Vol-1933/paper-6.pdf

3. Lassila, O., & Swick, R. R. (1999). Resource description framework (RDF) model and syntax specification.

4. Harris, S., Seaborne, A., & Prud'hommeaux, E. (2013). SPARQL 1.1 query language. W3C recommendation, 21(10).

5. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 34-43.

ADOnIS – An ontology-based information system providing seamless integration of structured and unstructured data

Alsayed Algergawy[1], _Friederike Klan_[1,2], _Erik Fäßler_[3], _Bernd Kampe_[3], _Hamdi Hamed_[1], _Birgitta könig-Ries_[1], _Udo Hahn_[3]

[1]*Fusion group, FSU Jena, Germany*; [2]*DLR Jena, Germany*; [3]*Julie Lab, FSU Jena, Germany*

Corresponding author(s) e-mail: firstname.lastname@uni-jena.de

ABSTRACT:

The CRC AquaDiva [1] is a large collaborative project spanning a variety of domains, such as biology, geology, chemistry and computer science with the common goal to better understand the Earth's critical zone in particular how environmental conditions and surface properties shape the structure, properties, and functions of the subsurface. This necessitates the collection and integration of large volumes of heterogeneous observational data. Besides this structured data, knowledge is also encoded in an unstructured form in scientific publications. Ideally, scientists should be able to seamlessly access both types of information.

To this end, we are developing the AquaDiva Ontology-based Information System, ADOnIS. This system gives scientists various ways to upload their datasets into a common repository based on the BExIS framework [2]. To enhance the integration process and to resolve conflicts among heterogeneous datasets, we build a conceptual, ontology-based layer on top of the common repository. Finally, the system grants different mechanisms to search and look for a specific piece of information and/or knowledge, including keyword search, semantic search [2], and location-based search. In all cases, search results will contain structured data as well as publications obtained from PubMed and the open access subset of Pubmed Central by using SeMedico [3].

The normal search provides the possibility for a scientist to enter a keyword (or a set of keywords) looking for the existence of this keyword either in the set of available metadata or primary data applying an exact match technique. This kind of search completely ignores the semantics of keywords as well as their relationships. Therefore, ADOnIS also provides semantic search exploiting features introduced by the conceptual layer. Finally, interactive search offers a view covering all the geo-related datasets displayed on a map.

KEYWORDS: Ontology-based systems, structured data, unstructured data

REFERENCES:

1. CRC AquaDiva: http://www.aquadiva.uni-jena.de/
2. BEXIS 2: http://bexis2.uni-jena.de/
3. Klan, F., Faessler, E., Algergawy, A., König-Ries,B., Hahn, U.: Integrated Semantic Search on Structured and Unstructured Data in the ADOnIS System. S4BioDiv@ISWC 2017 (http://ceur-ws.org/Vol-1933/paper-9.pdf)
4. Faessler, E., Hahn, U.: Semedico: A Comprehensive Semantic Search Engine for the Life Sciences. ACL (System Demonstrations) 2017: 91-96 (http://aclweb.org/anthology/P17-4016)

DiSSCo – The Distributed System of Scientific Collections

Niels Raes*[1]*, Wouter Addink*[1]*, Dimitris Koureas*[1] *and the DiSSCo consortium

[1]*Naturalis Biodiversity Center, Leiden, the Netherlands, P.O. Box 9517*

Corresponding author(s) e-mail: niels.raes@naturalis.nl

ABSTRACT:

DiSSCo (Distributed System of Scientific Collections; http://dissco.eu/) is a **Research Infrastructure (RI)** aiming to **mobilise**, **unify** and **deliver** bio- and geo-diversity information at the *scale*, *form* and *precision* required by scientific communities. DiSSCo will provide virtual (digital) access to the approximately 1.5 billion biological and geological specimens in collections across Europe and bring natural science collections to the information age, investing in a *linked open data* approach. DiSSCo represents the largest ever formal agreement between natural science museums (115 organisations across 21 European countries). With the recent acceptance of DiSSCo on the European Strategy Forum on Research Infrastructures (ESFRI) roadmap, the political and financial support across 14 European governments, and a robust governance model DiSSCo will deliver, by 2025, a series of innovative end-user discovery, access, interpretation and analysis services for natural science collections data.

KEYWORDS: Research Infrastructure, DiSSCo, collections, digitisation, biodiversity

COMPUTER DEMONSTRATIONS

BirdNET: Real-time Bird Sound Identification using Convolutional Neural Networks

***Stefan Kahl*[1], *Thomas Wilhelm-Stein*[1], *Holger Klinck*[2], *Danny Kowerko*[1], *Maximilian Eibl*[1]**

[1]*Chemnitz University of Technology, Germany*; [2]*Cornell Lab of Ornithology, USA*

Corresponding author(s) e-mail: Stefan.kahl@informatik.tu-chemnitz.de

ABSTRACT:

Reliable identification of bird species in recorded audio files and live streams would be a transformative tool for researchers, conservation biologists, and birders. Deep learning techniques have greatly affected the field of audio event detection in recent years. However, bringing those technologies to application is still challenging. Our demo provides an overview of our research in this area and covers two use-cases: Analysis of pre-recorded audio files (mainly directional and mono-species) and the real-time analysis of audio streams (mainly omnidirectional microphones with overlapping species). We developed web-interfaces for the presentation of both scenarios, making results easily accessible. Additionally, all prototypes are publicly available online at **birdnet.cornell.edu**

KEYWORDS: Bioacoustics, Bird Sounds, Soundscapes, Convolutional Neural Networks

Unfolding existing Data Publication Practice in Research Data Workflows in the Biological and Environmental Sciences

***Felicitas Löffler*[1], *Tina Astor*[2], *Claudia Müller-Birn*[3]**

[1]*Heinz-Nixdorf-Professur für verteilte Informationssysteme, Friedrich-Schiller-Universität Jena, Germany*; [2]*Georg-August-Universität Göttingen, Germany*; [3]*Freie Universität Berlin, Germany*

Corresponding author(s) e-mail: felicitas.loeffler@uni-jena.de

ABSTRACT:

In recent years, data publication workflows get more and more attention [1,2]. In order to obtain FAIR data [3], reviewers, data curators and other stakeholders have realized that not only the submitted data matter but also the underlying process to create that data within existing research practice. A better understanding of existing data publication practices in research workflows will help service providers such as data repositories (Pangaea [4], ENA [5], GenBank [6]) to support their users with more appropriate services and tools when submitting data, and otherwise, will sustain the role of data repositories in research practice. Such improved coordination will minimize the workload of researchers and data curators and will facilitate the review process of all stakeholders with respect to reproducibility.

One obstacle towards comprehensible and properly described research workflows is the fact that data publication workflows in the life sciences are hard to define. Scholars have their very individual disciplinary background, research skills and experiences. In some domains such as biodiversity, scholars work from several weeks to years to collect and analyze often heterogeneous data from various sources, such as collections, environmental or molecular data repositories. Thus, reconstructing their work process after the project is finalized is very difficult if not impossible. However, our goal is to reveal the state of the art on how scholars manage their data in their research practices. We are in the process of setting up a survey whose general structure is organized according to the GFBio Data Lifecycle [7]. The results will allow us to reveal typical data practices workflows that can be used to evaluate the suitability of existing data repository portals, such as GFBio [8].

In the poster and demo session, we would like to invite the audience on a discussion about research workflows in the life sciences. What are main and frequent tasks in what research field? Do our findings reflect the audience' own experiences? What prevents researchers from sharing their data and workflows? What concepts, policies or tools are necessary to leverage data sharing? Conference participants are also welcome to take part in the survey.

KEYWORDS: data publication workflows, data practices, biological and environmental data, green life sciences, biodiversity

REFERENCES:

1. Dallmeier-Tiessen, S., Khodiyar, V., Murphy, F., Nurnberger, A., Raymond, L., Whyte, A., 2017. Connecting Data Publication to the Research Workflow: A Preliminary Analysis, International Journal of Digital Curation, 12, https://doi.org/10.2218/ijdc.v12i1.533.
2. González-Beltrán, A., Li, P., Zhao, J., Avila-Garcia, M. S., Roos, M., Thompson, M., van der Horst, E., Kaliyaperumal, R., Luo, R., Lee, T.-L., Lam, T., Edmunds, S.C., Sansone, S.-A., Rocca-Serra, P, 2015. From Peer-

Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics, PLOS ONE 10, 7, pp. 1–20, https://doi.org/10.1371/journal.pone.0127612.

3. Mark D. Wilkinson et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3. https://doi.org/10.1038/sdata.2016.18
4. Pangaea, https://www.pangaea.org
5. ENA, https://www.ebi.ac.uk/ena
6. GenBank, https://www.ncbi.nlm.nih.gov/genbank/
7. GFBio Data Lifecycle, https://www.gfbio.org/training/materials/data-lifecycle
8. GFBio, https://www.gfbio.org

Galaxy-E, open source workflow oriented platform for Ecological data access and analysis

***Valentin Chambon*[1]*, Alan Amossé*[1]*, Elisa Michon*[1]*, Clara Urfer*[1]*, Yvan Le Bras*[1]**

[1]*French National Museum of Natural History, Station de Biologie marine, Concarneau - CESCO UMR 7204 Paris, France*

Corresponding author(s) e-mail:

ABSTRACT:

The French national project "65 Million Observers" through the Galaxy-E project allows macro-ecologists to share and analyse data obtained from biodiversity observatories. The Galaxy-E project proposes to build the first bricks of a workflow oriented e-infrastructure to support the automation of routine tasks and thus improve accessibility, reproducibility and transparency of ecological research. As Ecological data analysis tasks intensively use R products, it appears as a first challenge to better integrate R system on Galaxy. From developing Galaxy tools to manage Rdata or GIS objects to the integration of complexe R Shiny apps for Species Distribution Modeling, visualization or geographical exploration of the data, we propose to demonstrate the power of Galaxy to analyse environmental data! In particular, for the demo session, we propose attendees to 1/post-process data from animal detection on acoustic recordings, 2/ analyze SDM data using Wallace, 3/ treat butterfly phenology data and compute abundance index, 4/ use common community ecology functions through vegan GUI Shiny app and/or “simply” 5/visualize and describe data.

KEYWORDS: Galaxy, Workflow, Data access, Data analysis, EBV, Accessibility, Reproducibility, Transparency, Species Distribution Modeling, R shiny, Docker, Openstack, SLURM

Demonstration of the R package nbaR: Automated access to the Netherlands Biodiversity Data Services

***Hannes Hettling[1]**, Maarten Schermer[1], Rutger Vos[1], Daphne Duin[1]*

[1]*Naturalis Biodiversity Center, Leiden, P.O. Box 9517, 2300 RA, The Netherlands*

Corresponding author(s) e-mail: hannes.hettling@naturalis.nl; daphne.duin@naturalis.nl

ABSTRACT:

Here, we demonstrate some of the functionality of the R package *nbaR* in an evolutionary biology analysis workflow. The Netherlands Biodiversity data API (NBA, see also http://docs.biodiversitydata.nl) is a collection of public services for retrieving biodiversity related data, such as museum specimens, species occurrences, and taxonomic checklists. *nbaR* (https://github.com/naturalis/nbaR) is an official client for the NBA implemented in the R programming language, which has gained wide acceptance in the research community. This demo directly links to our oral presentation *"The Netherlands Biodiversity Data Services and the R package nbaR: Automated workflows for biodiversity data analysis"* in the *data integration* session. Prior experience of the attendants with R is beneficial, but not required.

The demo will consist of the following parts:

1. A short introduction about the NBA explaining the principles and characteristics of the data that can be obtained.
2. A demonstration of how to install the package and access documentation.
3. A demonstration of basic usage of the package: Introduction to the main data types and functions and how to build and execute simple and complex queries.
4. An example of a scientific workflow that uses the package nbaR for data retrieval. To this end, we will use our data to obtain divergence time estimates for a (molecular) phylogeny. Given a phylogenetic tree for a higher taxon, we will design queries to automatically retrieve data for specimens in that taxon that have assigned ages determined by geochronology and stratigraphy. Subsequently, we will show how to use the obtained data as an input for the popular phylogenetic analysis package *ape*, which we will use for time-calibration of the phylogenetic tree.
5. A discussion with the audience about the potential to use the package *nbaR* for their own research.

KEYWORDS: R package, API, automated analysis workflows, biodiversity data, phylogenetic inference

Showcase: Towards an EBV Data Portal

Christian Langer[1], Néstor Fernández[1], Christian Beilschmidt[2], Johannes Drönner[2], Michael Mattig[2], Bernhard Seeger[2]

[1] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany*; [2]*University of Marburg, Germany*

Corresponding author(s) e-mail: christian.langer@idiv.de

ABSTRACT:

Essential Biodiversity Variables (EBVs), defined as the minimum set of measurements required to study, report, and manage biodiversity change, are designed to provide decision makers with spatially and temporally consistent information on biodiversity states, connecting their needs with multiple-purpose monitoring initiatives. EBVs provide the first level of abstraction between low-level primary observations and high-level biodiversity indicators. Producing and delivering EBVs remains challenging due to the large data requirements, heterogeneous nature of the data, and data access via an API interface. We designed a GEO BON EBV Metadata catalogue with the aim of assist users in the production and delivery of EBV data products. This catalogue consists on a web-based system for spatiotemporal metadata. Here, users will be able to access EBVs under different production levels and catalogued into 6 different EBV classes. The system provides display and metadata download capacity, as well connection to EBV-related OGC Web Services. So-called exploratory workflows allow users to search data in an interactive Graphic User Interface (GUI).

In this showcase we will use Ecosystem Structure EBV data. The datasets of the “Global Forest Watch” is used as an input for the EBV “Ecosystem extent and fragmentation” from the yearly global loss and gain of tree cover. This information is highly relevant for decision makers as well as scientists. The EBV catalogue retrieves metadata including the associated OGC Web Services via the CSW Interface. Users can search for keywords referring to specific EBVs, display/download the metadata and interact with the associated Web Services. In addition, the system can connect to an EBV Spatial Analyzer under development that will allow to plot summary of changes in ecosystem structure over time and by reporting units specified by the user.

All datasets from the EBV catalogue will be available using a visualization and analysis system for scientists developed by University of Marburg (VAT). In addition, VAT also offers simplified report views that simplify the complexity of the actual computations in order to improve the usability and user experience for a wide range of potential users.

KEYWORDS: Essential Biodiversity Variables, GIS, Data Harvesting, OGC Web Services, Metadata Catalogue

TraitEx: tool for measuring morphological functional traits from digitized herbarium specimens

***Jitendra Gaikwad*[1], *Abdelaziz Triki*[2], *Bassem Bouaziz*[2], *Hamdi Hamed*[1], *Jörn Hentschel*[1]**

[1]*Friedrich Schiller University Jena, Germany*; [2]*Higher Institute of Computer science and multimedia, Sfax, Tunisia*

Corresponding author(s) e-mail: Jitendra.gaikwad@uni-jena.de

ABSTRACT:

Functional trait data is of significant importance to understand how species respond to biotic and abiotic factors. However, despite several global efforts, functional trait data is largely fragmented and available in heterogeneous formats. Initiatives such as TRY (https://www.try-db.org), which is a global network of vegetation scientists, are making efforts by applying novel methods in machine learning to fill gaps in the observed trait matrix and predict plant traits at continental to global scale. To complement existing efforts and increase plant trait data, within the Managing Multimedia Data for Science (MAMUDS) project, we have developed a software tool TraitEx to measure functional traits such as length, width, area and size of leaves from the digitized type specimens stored in herbarium Haussknecht (http://www.spezbot.uni-jena.de/herbarium/). To preserve specimens and make associated information easily accessible to researchers and public, herbarium Haussknecht from Germany, have digitized more than 30000 type specimens. TraitEx is a semi-automatic and open source tool, developed using computer programming language Java. The main features of TraitEx are (1) ability to handle high resolution specimen images, (2) supports extraction of trait measurements from multiple and diverse leaf shapes, (3) efficiently extract measurements from leaves of specimens that are mounted using white tapes, (3) ImageJ functions integrated for pre-processing and editing of images, (4) functionality to export measured values in CSV format along with original image, thumbnail and coordinates of the measured leaves, (5) user-friendly viewer to visualize measured leaves and values and (6) facilitates easy and quick measurement of leaf traits with high accuracy as compared to manual measurements process. As part of the demo we will (1) introduce TraitEx and use digitized type specimens from herbarium Haussknecht to show the processes involved in extraction of morphological trait extraction, (2) show how to download and install TraitEx if user want to try on their laptops and (3) document user experience as feedback.

KEYWORDS: TraitEx, Digitized specimens, Herbarium Hasuuknecht, Functional traits, MAMUDS, Software

BioOntoVis : Ontologies Visualization Tool for biodiversity

***Nassira Achich*[1], *Bassem Bouaziz*[1] , *AlsayedAlgergawy*[2] *and Birgitta König-Ries*[2]**

1Miracl laboratory, University of Sfax, Tunisia; [2]FSU Jena, Germany

Corresponding author(s) e-mail: achichnassira@gmail.com; bassem.bouaziz@isims.usf.tn; alsayed.algergawy@uni-jena.de; birgitta.koenig-ries@uni-jena.de

ABSTRACT:

Ontologies introduce a formal scheme for defining and representing terms and their relationships in different domains, such as ecology and biology, which are inherently cross disciplinary. The use of ontologies in such domains enhances the mapping, interpretation and integration of relevant databased on its intended meaning. Therefore, there are a large number of ontologies have been developed for such kind of interdisciplinary domains. For example, BioPortal (http://bioportal.bioontology.org/) contains 716 ontologies with 8,387,594 classes (with an average 11,714 classes per ontology). It is important to represent such ontologies in a visualized scheme to support the mapping and integration process by involving domain experts. However, ontology visualization is not a simple task and most of existing ontology visualization tools lack either coping with visualizing big ontologies or providing multiple views depending on ontology features and based on the user requirements. To this end, in this work, we introduce a new interactive Web based ontology visualization tool, called BioOntoVis, exploiting a combination of visualization techniques to reflect mentioned challenges. In particular, the proposed tool follows three main steps, including: (i) reading and parsing an input ontology, (ii) a set of processing aspects to make the ontology suitable for visualization, and (iii) finally, selection of a suitable visual representation to visualize the ontology. The BioOntoVis tool supports node-link and tree, zoomable, and 3D information landscape. Furthermore, the tool also provides the possibility to edit the visualized ontology and add/remove entities while checking the consistency. There automatic consistency checking is triggered when such operations are handled.

KEYWORDS: Ontology, visualization, biodiversity, editing, interpretation, integration

Research data management in action – demonstrating the full range of features of the BEXIS 2 platform

Roman Gerlach[1], *Andreas Ostrowski*[1], *Jitendra Gaikwad*[1, 2], *Alsayed Algergawy*[1]

[1]*Friedrich Schiller University Jena, Germany*; [2]*German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany*

Corresponding author(s) e-mail: roman.gerlach@uni-jena.de

ABSTRACT:

BEXIS 2 is a data management platform facilitating research data management in biodiversity, ecology and neighbouring fields. Following the data-life-cycle concept, the system provides features for describing, structuring, searching, publishing and accessing datasets; it includes features for dataset versioning, quality assurance, notification, and identity and permission management. BEXIS 2 is developed by a community of data managers and scientific users and is available as open source software under a GNU LGPL License.

In this demonstration, the core development team will join with the contributing community to present the latest release of the core system as well as a number of advanced modules developed by the community. For example, attendees may try the semantic search or the automatic data annotation tool developed within the AquaDiva project. In addition, visitors will be able to experience different operational instances of BEXIS 2 and see how they have been customized for specific communities.

Highlights of this demonstration will be:

- Resource booking management module (developed within the Biodiversity Exploratories project): This is a calendar based tool to manage any kind of resource. It is highly configurable to cover multiple scenarios. It helps to communicate the bookings and activities to all users (or colleagues and other stakeholders). For the demo the module is configured to support fieldwork, especially the management and sharing of field equipment. Researchers are able to book field station resources and announce field visit activities.
- Graphical research area management module (developed within the Biodiversity Exploratories project): This tool will assist researchers in organizing and documenting experiments on plots and subplots. The tool enables users to virtually create research plots, define their shape, plan and localize experiments within these areas, create maps for orientation in the field, and keep track of the history of all subplots to avoid conflicts.
- Multimedia Module (developed within the MAMUDS project): This is an extension to the BEXIS 2 system to display images, videos and audio files. The highlights of the module are (1) the media files in a dataset are shown as a preview in the list of uploaded files, (2) audio-video files can be played directly from the preview, (3) it provides a slideshow mode for images and full screen mode for videos and (4) ability to import and display annotated digitized herbarium specimen images that are exported as bundle from the trait extraction tool TraitEx.

KEYWORDS: research data management, software demonstration, data life-cycle, BEXIS 2

Program Committee

Aidin Niamir	Senckenberg Biodiversity and Climate Research Centre (SBiK-F)
Alberto Lanzoni	Dipartimento di Scienze Agrarie – Entomologia, Università di Bologna
Alessandro Oggioni	CNR - Institute for Electromagnetic Sensing of the Environment (IREA)
Alex Hardisty	Cardiff University
Alex Neumann	University of Toronto
Alexander Brenning	Friedrich Schiller University Jena
Alsayed Algergawy	Friedrich-Schiller-University Jena
Anton Güntsch	Botanischer Garten und Botanisches Museum Berlin
Antonino Staiano	Università di Napoli Parthenope
Basil Kraft	Max Planck Institute for Biogeochemistry
Bassem Bouaziz	ISIMS
Bernhard Seeger	University of Marburg
Bert Bredeweg	University of Amsterdam
Birgitta König-Ries	Friedrich Schiller University of Jena
Carole Goble	The University of Manchester
Christian Requena-Mesa	Friedrich Schiller University Jena
Christoph Steinbeck	Friedrich-Schiller-University (FSU) Jena
Clemen Thomas	
Clement Jonquet	University of Montpellier - LIRMM
Colin Smith	Environmental Data Initiative
Concetto Spampinato	DIEEI - University of Catania
Corinna Gries	University of Wisconsin-Madison
Dag Endresen	GBIF Norway, UiO Natural History Museum, University of Oslo
Dan Stowell	Queen Mary University of London
David Schöne	Max-Planck-Institute for Biogeochemistry Jena
Diepenbroek Michael	
Dreyfus Michel	
Duccio Rocchini	University of Trento
Elizabeth Arnaud	Bioversity International
Francois Pinet	Cemagref
Friederike Klan	German Aerospace Center (DLR)
Friederike Klan	Friedrich Schiller University of Jena
Friedrich Recknagel	The University of Adelaide
George Arhonditsis	University of Toronto
Guido Kraemer	Max Planck Institute for Biogeochemistry
Gustau Camps-Valls	Universitat Politècnica de València
Han Wang	Tsinghua University
Ibrahim Moawad	Ain shams univeristy

Ilaria Rosati	LifeWatch Italy
Ilia Ostrovsky	Israel Oceanographic and Limnological Research
Jan Bumberger	Helmholtz Centre for Environmental Research - UFZ
Jana Wäldchen	Max Planck Institute for Biogeochemistry
Janet Anstee	CSIRO Ocean & Atmosphere
Javad Chamanara	Friedrich Schiller University of Jena
Jeffrey Tuhtan	Tallinn University of Technology
Jens Kattge	Max Planck Institute for Biogeochemistry, Jena
Jérôme Sueur	Muséum national d'Histoire naturelle
Jitendra Gaikwad	Friedrich-Schiller-Universität Jena
Joachim Denzler	Friedrich Schiller University Jena
Klaus Joehnk	CSIRO
Kristin Vanderbilt	Environmental Data Initiative
Kühn Ingolf	Helmholtz Centre for Environmental Research - UFZ
Kyriakos Kritikos	Institute of Computer Science, FORTH
Ma Keping	
Marco Körner	Technical University of Munich
Margaret O'Brien	Environmental Data Initiative
Marie Roch	San Diego State University
Marie-Angélique Laporte	Bioversity International
Mark Schildhauer	NCEAS
Markus Reichstein	Max Planck Institute for Biogeochemistry
Michael Bahn	University of Innsbruck
Michael Diepenbroek	PANGAEA, MARUM, University Bremen
Miguel Mahecha	Max Planck Institute for Biogeochemistry
Mirco Migliavacca	Max-Planck Institute for Biogeochemistry
Nafiseh Navabpour	Friedrich Schiller University of Jena
Naouel Karam	Fraunhofer
Natalia Estrada Carmona	Bioversity International
Néstor Fernández	German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
Nicole van Dam	German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
Nikos Minadakis	FORTH
O'Connell Mark	
Oskar Franklin	iiasa
Pereira Henrique	iDiv
Peter Goethals	Ghent University
Peter Whigham	University of Otago

Pier Luigi Buttigieg	Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung
Qinghua Cai	Institute of Hydrobiology, Chinese Academy of Sciences
Qiuwen Chen	CEER Nanjing Hydraulic Research Institute
Ramona Walls	CyVerse
Riza Batista-Navarro	The University of Manchester
Robert Guralnick	University of Florida
Rocchini Duccio	University of Trento
Roeland Kindt	World Agroforestry Center
Roman Gerlach	Friedrich-Schiller-Universität Jena
Sandro Bimonte	irstea
Santanu Ray	Visva-Bharati University
Sarah Jones	Bioversity International
Saso Dzeroski	Jozef Stefan Institute
Scardi Michele	Department of Biology, University of Rome -Tor Vergata, Rome, Italy
Shinji Fukuda	Tokyo University of Agriculture and Technology
Simon Brewer	University of Utah
Song Lirong	
Stafford Richard	Bournemouth University
Steffen Neumann	Leibniz Institute of Plant Biochemistry (IPB) Halle
Susanne Tautenhahn	Max Planck Institute for Biogeochemistry, Jena
Sven Thiel	Friedrich Schiller University of Jena
Tae-Soo Chon	Ecology and Future Research Association
Talie Musavi	Max-Planck Institute for Biogeochemistry
Tao Tang	Institute of Hydrobiology, Chinese Academy of Sciences
Taysir Soliman	Assiut University, Egypt
Thomas Evert	Bioversity International
Tiffany Knight	German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
Tobias Fremout	Leuven University/Bioversity International
W. Daniel Kissling	University of Amsterdam
Xuanlong Ma	Max-Planck Institute for Biogeochemistry
Young-Seuk Park	Kyung Hee University
Yuko Shimoda	University of Toronto

Participants List

First Name	Last Name	Affiliation
Adam	Kania	MGGP Aero Sp. z o.o.
Ajung	Kim	Ewha Womans University
Alberto	Basset	Università del Salento
Alberto	Lanzoni	Dipartimento di Scienze e Tecnologie Agro-Alimentari
Aleksandra	Wolanin	Helmholtz Centre Potsdam German Centre for Geosciences (GFZ)
Aletta	Bonn	Helmholtz-Centre for Environmental Research - UFZ
Alex	Hardisty	Cardiff University
Alsayed	Algergawy	Friedrich Schiller University Jena
Alvaro	Moreno	Image Processing Laboratory (IPL) - UV
André	Große-Stoltenberg	University of Münster
Andreas	Ostrowski	Friedrich Schiller University Jena
Andrew	Lane	King's College London
Angela	Lausch	Helmholtz Centre for Environmental Research - UFZ
Angélica Paola	Diaz-Pulido	Universidad de Antioquia
Anirban	Mukhopadhyay	Technische Universität Darmstadt
Anke	Kügler	Hawai'i Institute of Marine Biology, University of Hawai'i at Manoa
Anne	Clasen	Technische Universität Berlin
Anton	Güntsch	Freie Universität Berlin
Antonino	Staiano	Università di Napoli Parthenope
Anubhav	Gupta	University of Zurich
Anusuriya	Devaraju	MARUM - University of Bremen
Arefeh	Bahrami	Friedrich Schiller University Jena
Arthur	Sanguet	University of Geneva
Astrid	Schmidt-Kloiber	University of Natural Resources and Life Sciences - BOKU Vienna
Ayoub	Haggui	University of Sfax
Ayoub	Haggui	University of Sfax
Babak	Naimi	Max Planck Institute for Ornithology
Barbara	Magagna	Umweltbundesamt GmbH
Basil	Kraft	Max Planck Institute for Biogeochemistry
Bassem	Bouaziz	MIRACL-ISIMS-USF
Bettina	Hesse	Friedrich Schiller University Jena
Birgitta	König-Ries	Friedrich Schiller University Jena
Blanka	Weber	journalist - MDR Kultur radio
Brigitte	Nixdorf	BTU Cottbus - Senftenberg - LS Gewässerschutz
Brigitte	Nixdorf	BTU Cottbus - Senftenberg - LS Gewässerschutz
Camilo	Sanchez Giraldo	Universidad de Antioquia
Carlos	de Obaldia	Helmut Schmidt University Hamburg
Carmen	Cillero castro	3edata. Ingeniería Ambiental S.L.
Carol	Bedoya	University of Canterbury
Carsten	Neumann	Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences

Caterina	Bergami	National Research Council of Italy
Chia-yun	Lee	Biodiversity Research Center - Academia Sinica
Chris	Reudenbach	Universität Marburg
Christian	Langer	Martin-Luther-Universität Halle-Wittenberg
Christian	Pichot	Ecologie des Forêts Méditerranéennes (UR 629 -) INRA PACA
Christian	Requena Mesa	Friedrich Schiller University Jena
Christian	Thiel	DLR - Institute for Data Science
Christina	Castelo Branco	Fundação Bio Rio
Christina	Weißbecker	Helmholtz Zentrum für Umweltforschung - UFZ
Christine	Kiel	Leibniz-Institut für Gewässerökologie und Binnenfischerei (IGB)
Christine Isabeau Bernarde	Wallis	LCRS – Universität Marburg
Christoph	Raab	University of Goettingen
Claudia Victoria	Isaza Narvaez	Universidad de Antioquia
Claudio	Barberato	
Claus	Weiland	Senckenberg Biodiversity and Climate Research Centre
Corinna	Gries	University of Wisconsin
Cornelia	Fürstenau	Friedrich Schiller University Jena
Dagmar	Triebel	Staatliche Naturwissenschaftliche Sammlungen Bayerns
Dainis	Jakovels	Institute for Environmental Solutions
Dan	Stowell	Queen Mary University of London
Daniel	Doktor	Helmholtz Centre for Environmental Research - UFZ
Daniel	Gabriel	Tokyo Institute of Technology
Daniel	Mietchen	University of Virginia
Daniel Ernesto	Pabon Moreno	Max Planck Institute for Biogeochemistry
David	Boho	Technische Universität Ilmenau
David	Eichenberg	iDiv - sDiv Group
David	Martini	Max Planck Institute for Biogeochemistry
Debra	Peters	USDA-ARS and Jornada LTER Program
Desiree	Andersen	Ewha Womans University
Di	Tian	Capital Normal University
Diana	Bowler	Friedrich Schiller University Jena
Diana Carolina	Duque Montoya	Universidad de Antioquia
Dimitri	Belousow	Technische Universität Ilmenau
Dimitri	Brosens	EVINBO
Dina	Sharafeldeen	Friedrich Schiller University Jena
Dong-Kyun	Kim	University of Toronto
Duccio	Rocchini	University of Trento
Edoardo	Cremonese	Environmental Protection Agency of Aosta Valley (ARPA VdA)
Egor	Prikaziuk	University of Twente
Elena	Catucci	University of Rome Tor Vergata
Eleonora	Petzold	Friedrich Schiller University Jena
Elizabeth	Znidersic	Charles Sturt University
Emma	Izquierdo-Verdiguier	University of Natural Resources and Life Sciences - Vienna (BOKU)

Estefany	Cano Rojas	Universidad de Antioquia
Fabiana	Santana	University of Canberra
Felicitas	Löffler	Friedrich Schiller University Jena
Florian	Jansen	University of Rostock
Florian	Schneider	
Forrest	Hoffman	Oak Ridge National Laboratory
Francesco	Mattei	University of Rome Tor Vergata
Francesco Maria	Sabatini	Martin-Luther Universität Halle-Wittenberg
Frank	Boons	University of Manchester
Frank	Dziock	University of Applied Sciences HTW Dresden
Frank	Löffler	Friedrich Schiller University Jena
Franziska	Zander	Friedrich Schiller University Jena
Friederike	Klan	Deutsches Zentrum für Luft- und Raumfahrt
Friedrich	Recknagel	The University of Adelaide
George	Arhonditsis	University of Toronto
Gerhard	Boenisch	Max-Planck Institute for Biogeochemistry
Gianluca	Filippa	Environmental Protection Agency of Aosta Valley
Guido	Kraemer	Max-Planck-Institut fuer Biogeochemie
Guillaume	Dutilleux	NTNU - Norway
Gustau	Camps-Valls	Universitat de València
Hamdi	Hamed	Friedrich Schiller University Jena
Hanh	Nguyen	The University of Adelaide
Hanna	Meyer	Philipps University Marburg
Hanna	Pamuula	AGH University of Science and Technology in Krakow
Hannes	Feilhauer	FAU Erlangen-Nürnberg
Hannes	Hettling	Naturalis Biodiversity Center
Hans Christian	Wittich	TU Ilmenau
Heather	Savoy	USDA Agricultural Research Service
Henrique Miguel	Pereira	Martin-Luther-Universität Halle-Wittenberg
Hiroshi G.	Okuno	Waseda University
Holger	Lange	Norwegian Institute of Bioeconomy Research
Hyo Gyeom	Kim	Pusan National University
Ibrahim	Moawad	Ain Shams Univeristy
Ilaria	Rosati	National Research Council of Italy
Irene	Rodríguez Serrano	CIEMAT
Irene	T. Roca	Universität Oldenburg
Jan	Bumberger	Helmholtz Centre for Environmental Research - UFZ
Jan Martin	Keil	Friedrich Schiller University Jena
Jana	Wäldchen	Max Planck Institute for Biogeochemistry
Janet	Anstee	CSIRO Oceans and Atmosphere
Janno	Harjes	Universität Bayreuth
Jean-Francois	Julien	CESCO Museum National d'Histoire Naturelle
Jeffrey	Tuhtan	Tallinn University of Technology
Jens	Kattge	Max Planck Institute for Biogeochemistry

Jens	Koblitz	Wildlife Acoustics
Jens	Nejstgaard	Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB)
Jens	Nieschulze	Georg-August University Goettingen
Jérôme	Sueur	MNHN-CNRS ISYEB UMR 7205
Jitendra	Gaikwad	Friedrich Schiller University Jena
Joachim	Denzler	Friedrich Schiller University Jena
Joachim	Gröger	Thünen-Institute of Sea Fisheries
Johannes	Drönner	Philipps-Universität Marburg
Jonas	Jäger	Hochschule Fulda
Josephine	Ulrich	Friedrich Schiller University Jena
Juan	Rojas Garcia	University of Granada
Juan Manuel	Daza Rojas	Universidad de Antioquia
Julien	Radoux	Université catholique de Louvain
Kailiang	Yu	ETH Zurich
Kamal kanti	Banik	Save the water sources
Kasahun	Abera	Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ)
Kazuhiro	Nakadai	Honda Research Institute Japan Co. Ltd.
Klaus	Joehnk	CSIRO Land and Water
Larisa	Khanina	IMPB RAS - Branch of the Keldysh IAM RAS
Laura	Paladini	VITO
Leila	Feddoul	Friedrich Schiller University Jena
Lina M.	Estupinan-Suarez	Max Planck Institute for Biogeochemistry
Maarja	Kruusmaa	Tallinn University of Technology
Maha	Shadaydeh	Friedrich Schiller University Jena
Majid	Askar	Faculty of Computers
Majid	Askar	Assiut University
Marco	Körner	Technical University of Munich
Marco	Schmidt	Senckenberg Biodiversität und Klima Forschungszentrum (SBiK-F)
Marco	Seeland	Technische Universität Ilmenau
Marcus	Guderle	Max Planck Institute for Biogeochemistry
Maria	Londono Noriega Romero	Max Planck Institute for Biogeochemistry
Maria Florencia	Vargas	Technische Universität Dresden
Marie	Roch	San Diego State University
Marie Anne Eurie	Forio	Ghent University
Markus	Stocker	German National Library of Science and Technology (TIB)
Markus	Weiss	Staatliche Naturwissenschaftliche Sammlungen Bayerns
Martin	Hofmann	Technische Universität Ilmenau
Martin	Hohmuth	FSU Jena
Marvin	Ludwig	Philipps-Universität Marburg
Marwa	Abdelreheim	Assuit University
Masahiro	Ryo	Freie Universitaet Berlin
Masatoshi	Denda	Public Works Research Institute
Matthew	McKown	Conservation Metrics Inc.

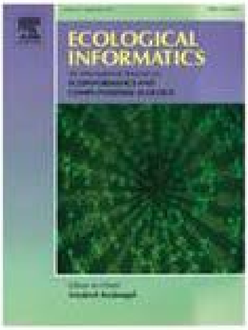
Matthias	Forkel	TU Wien – Department of Geodesy and Geoinformation
Matthias	Körschens	Friedrich Schiller University Jena
Mattia	Pancerasa	Politecnico di Milano
Maximilian	Konzack	Martin Luther University Halle-Wittenberg
Michael	Hauhs	University of Bayreuth
Michael	Rzanny	Max-Planck-Institute for Biogeochemistry
Michael	Towsey	Queensland University of Technology
Michela	Sammartino	Italian National Research Council (CNR)
Michele	Scardi	University of Rome 'Tor Vergata'
Michele	Torresani	University of Bolzano
Michiel	Stock	Ghent University
Miguel	Mahecha	Max Planck Institute for Biogeochemistry
Minqian	Chen	Technische Universität Ilmenau
Mirco	Migliavacca	Max Planck Institute for Biogeochemistry
Muhammad Sohaib	Younis	Senckenberg BiK-F
Muthmainnah	Muthmainnah	Stkip Muhammadiyah Enrekang
Nabyendu	Rakshit	Systems Ecology and Ecological Modelling Laboratory
Nafiseh	Navabpour	Friedrich Schiller University Jena
Naresh	Rijal	Support Umbrella Nepal
Nassira	Achich	Higher Institute of Computer Science and Multimedia of Sfax
Natalia	Petrovskaya	University of Birmingham
Neda	Trifonova	NOAA/University of Miami
Neha	Gupta	Friedrich Schiller University Jena
Néstor	Fernández	Martin-Luther-Universität Halle-Wittenberg
Nicola	Fiore	LIFEWATCH ERIC
Niels	Raes	Naturalis Biodiversity Center
Nora	Abdelmageedr	Ain Shams University
Nora	Linscheid	Max Planck Institute for Biogeochemistry
Oliver	Bley	TU Ilmenau
Olli	Malve	Finnish Environment Institute
Onatkut	Dagtekin	University of Manchester
Patrick	Mäder	Technische Universität Ilmenau
Patrick	Schratz	Department of Geography
Paul	Nguyen Hong Duc	Sorbonne University
Peter	Desmet	EVINBO
Peter	Goethals	Ghent University
Philip	Zylstra	University of Wollongong
Philipp	Kahn	Friedrich Schiller University Jena
Quentin	Groom	Meise Botanic Garden
Rainer	Kiko	GEOMAR Helmholtz Centre for Ocean Research Kiel
Rainer	Krug	University of Zürich
Rana	Mahmoud	Nile University
Reed	Anderson	UFZ
Reiji	Suzuki	Nagoya University

Ritika	Srinet	Indian Institute of Remote Sensing - Dehradun
Riza	Batista-Navarro	University of Manchester
Roberta	Righini	University of Pavia Italy
Roland	Cochard	Université de Lausanne
Roman	Gerlach	Friedrich Schiller University Jena
Roselyn	Gabud	University of the Philippines Diliman
Ryan	Nel	University of South Africa
Ryo	Kohsaka	Tohoku University
Sabine	Wollrab	Leibniz-Institute of Freshwater Ecology and Inland Fisheries
Sajeev	C Rajan	Kerala Thiruvananthapuram
Samira	Babalou	Friedrich Schiller University Jena
Sandra	Mueller	Albert-Ludwigs-University Freiburg
Sandra	Skowronek	German Federal Agency for Nature Conservation
Saskia	Pietzsch	Deutscher Wetterdienst
Sheeba	Samuel	Friedrich Schiller University Jena
Shinji	Fukuda	Tokyo University of Agriculture and Technology
Shinji	Sumitani	Graduate School of Informatics
Simon	Elise	Université de La Réunion
Siti Norhidayah Binti	Othman	Ewha Womans University
Song	Qian	The University of Toledo
Stefan	Halle	Friedrich Schiller University Jena
Stefan	Kahl	Chemnitz University of Technology
Stella	Berger	Leibniz-Institute of Freshwater Ecology and Inland Fisheries
Stephanie	Bohlmann	University of Florida
Stijn	Bruneel	Ghent University
Stijn	Van Hoey	EVINBO
Sungwon	Hong	Pusan National University
Susanne	Dunker	Helmholtz-Centre for Environmental Research GmbH – UFZ
Susanne	Tautenhahn	Max Planck Institut für Biogeochemie Jena
Sven	Thiel	Friedrich Schiller University Jena
Sylvain	Haupert	MNHN - CNRS - ISYEB UMR7205
Tae-Soo	Chon	Ecology and Future Research Association
Talie	Musavi	Max Planck Institute for Biogeochemistry
Tanja	Milotic	EVINBO
Teja	Kattenborn	Karlsruher Institute of Technology (KIT)
Thilo	Wellmann	Helmholtz-Zentrum für Umweltforschung GmbH - UFZ
Thomas	Moeckel	University Kassel
Thomas	Nauss	Philipps-Universität Marburg
Thomas	Wutzler	Max Planck institute for Biogeochemistry
Tim	Walles	Vorschungsverbund Berlin E.V. IGB
Trupti	Padiya	Friedrich Schiller University Jena
Tzu-Hao	Lin	Japan Agency for Marine-Earth Science and Technology
Ulrike	Selig	Philipps-Universität Marburg
Ulrike	Sturm	Museum für Naturkunde Berlin

Vanessa	Bremerich	Forschungsverbund Berlin e.V. (IGB)
Veronica	Morfi	Queen Mary University of London
W. Daniel	Kissling	University of Amsterdam
Wayne	Polley	USDA - Agricultural Research Service
Wesley	Tack	Botanic Garden Meise
William	Michener	University of New Mexico
William Esteban	Gomez	Universidad de Antioquia
Witold	Arndt	WWU Münster
Xuanlong	Ma	Max-Planck Institute for Biogeochemistry
Yuko	Shimoda	University of Toronto Scarborough
Yves	Bas	CESCO-MNHN
Zhengbing	Yan	Peking University

Supporters

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA


ECOLOGICAL INFORMATICS

iDiv
German Centre for Integrative Biodiversity Research (iDiv)
Halle-Jena-Leipzig

MichaelStifelCenterJena
for Data-Driven and Simulation Science

AquaDiva

DFG Deutsche Forschungsgemeinschaft

We are also grateful for the support by:

Max Planck Institute for Biogeochemistry

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

biodiversity exploratories

THE JENA EXPERIMENT