# Smart4RES

## *Identification of new and enhanced forecasting products*

**D1.2 Identification of new forecasting products**

**WP1, T1.2**

Version V2.0

Authors:

Pierre PINSON, DTU

Amandine PIERROT, DTU

Liyang HAN, DTU

Ricardo BESSA, INESC

José Gouveia, INESC

Carlos Moreira, INESC

Simon CAMAL, ARMINES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337

*Disclaimer*

The present document reflects only the author's view. The European Innovation and Networks Executive Agency (INEA) is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337

## Technical references

| | |
|---|---|
| Project Acronym | Smart4RES |
| Project Title | Next Generation Modelling and Forecasting of Variable Renewable Generation for Large-scale Integration in Energy Systems and Markets |
| Project Coordinator | ARMINES – MINES ParisTech |
| Project Duration | November 2019 – April 2023 |

| | |
|---|---|
| Deliverable | D1.2: Identification of new forecasting products |
| Dissemination level [1] | PU |
| Nature [2] | R |
| Work Package | WP 1 – Design of Forecasting Framework and Data Management Platform |
| Task | T 1.2 – Identification and standardization of new forecasting products |
| Lead beneficiary | DTU |
| Contributing beneficiary(ies) | ALL |
| Reviewers | Gregor GIEBEL (DTU)<br>Matthias LANGE (EMSYS) |
| Due date of deliverable | 30 09 2020 |

[1] PU = Public
PP = Restricted to other program participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

[2] R = Report, P = Prototype, D = Demonstrator, O = Other

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337

| Document history | | |
|---|---|---|
| V | Date | Description |
| 0.1 | 21/09/20 | First version containing contributions from all partners |
| 0.2 | 26/11/20 | Second version after analysis by Task leader (DTU) |
| 1.0 | 31/12/20 | Review by Gregor Giebel (DTU) and Matthias Lange (EMSYS) |
| 2.0 | 15/02/21 | Final public version released by Coordinator |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337

**Abstract:**

Forecasting of renewable energy generation is a mature R&D area, also with a lot of operational experience and commercial offering. As the share of renewable energy generation in the energy mix increases, in parallel to a liberalization of electricity markets, new needs for forecasting have appeared. Throughout the last few decades, emphasis was placed on deterministic and probabilistic forecasting, extremes, ramps, etc. As of today, we argue that new developments and proposals for forecast products should happen within a problem-oriented approach: there, forecast products are at the interface between forecasters and forecast users, and driven by the decision problem at hand. In this report, we review the current evolution in market, operations, etc. that call for new forecast products. We also look at new views coming from the R&D side, with the aim to introduce new forecast products and show their interest for forecast users. Eventually, we argue that, beyond forecast products only, there are also a number of new business models of relevance being developed, e.g. related to collaborative analytics and data markets.

# Contents

# 1 Introduction

The state of the art in renewable energy forecasting has evolved tremendously over the last few decades, with a very strong push for new developments driven by wind power prediction mainly at the beginning (e.g. probabilistic forecasting, ramp forecasting, etc.), though now also with a strong contribution from solar energy forecasting. The reader is referred to [1–3] to give a representative coverage of recent state of the art overviews in renewable energy forecasting. A lot of those developments have come from the R&D side through various national and international (e.g., at European level) projects, with the investigation of new methods for statistical and machine learning for instance, use of meteorological forecasts including ensemble forecasts, accommodating of new types of data as input, etc. In parallel, commercial and operational forecast providers have also done their fair share in developing solutions and services in renewable energy forecasting, based on the needs and requirements expressed by their forecast users.

At the core of those developments is the concept of *forecast products*, which are at the interface between the forecasters and the forecast users, or in other words, between the forecasting process and the use of forecasts in decision-making. Therefore, when thinking of future needs and potential development in renewable energy forecasting, we believe that a strong emphasis ought to be placed on potential new forecast products. The importance of forecast products has long been realized by the meteorological forecasting community – for instance, the European Centre for Medium-range Weather Forecasts (ECMWF) regularly looking into new forecast products, while updating their user guide to ECMWF forecast products[1]. It should be noted that while the term of "forecast product" is popular within the meteorological forecasting community and for various meteorological applications e.g. energy, different terminology may be used by other forecasting communities. We will stick here to that terminology.

Thinking of forecast products, their origination may relate to new needs expressed by forecast users, new needs as imposed by changes in markets and in regulation (e.g. change in the granularity of market from hourly to 15-mins), new needs imposed by technology (e.g. if having storage to operate jointly with wind farms, the temporal resolution of forecasts should be higher), etc. However, some of these new forecast products may also come from novel ideas and development from the R&D side, since profiting of novel data sources, new computational capabilities, etc. Clear examples relate to the use of sky imagers in solar power forecasting, see e.g. [4], allowing to produce high-resolution forecasts, as well as the implementation and solving of LES-type models on GPUs [5], also allowing for very high resolution weather forecasting. In the present report, the main emphasis is placed on short-term forecasts with lead times ranging between a few minutes ahead and several days ahead, in line with the majority of developments over the last few decades. It is to be noted, however, that increasingly focus is also given to longer term forecasts at the monthly and seasonal scales [6], which may be useful for hedging, adequacy studies, etc. Obviously this should eventually have an impact of potential new forecast products. Finally, it should be noted that, further than forecast product mainly, the discussion in this report is to extend to novel business models in relation with renewable energy forecasting, e.g., in connection with collaborative and distributed learning, privacy-preserving learning and data markets.

[1] See the ECMWF page on the user guide to ECMWF forecast products at: https://www.ecmwf.int/en/about/media-centre/focus/user-guide-ecmwf-forecast-products

Driven by those trends and new opportunities, the aim of this document is to uncover

- what kind of technological and regulatory developments may drive the need for new forecasting products,
- the need for new forecast products as expressed by forecast users, and
- the possibilities on the R&D side to bring new forecast products (and novel business models) to forecast users.

The document is structured as following. Section 2 gives a clear introduction to the concept of forecast product and its role in renewable energy forecasting. Section 3 concentrates on recent regulatory developments that may motivate new forecast products. Similarly, Section 4 discuses recent technological evolution that may motive new forecast products while Section 5 covers the evolution in operation practice. Section 6 looks at the user side, and how they may request themselves new forecast products. Sections 7 takes the point of view of recent R&D developments in forecasting that may lead to new forecast products and novel business models in renewable energy forecasting. Finally, Section 8 gathers a set of conclusions and perspectives for future developments. This report also relies on an extensive set of references, for the readers for further dig into the topics covered. In that sense, it may also be seen as a survey paper on the topic of forecast products for renewable energy forecasting.

# 2 What is a forecast product?

The fundamental aim of forecasting is to express and communicate information about future events: forecasts ought to be *informative* (most likely, based on the decision problem at hand) and concern specific *events*. Some of the basic concepts in forecasting have been theorized over the period 1960-2010, when for instance looking at the nature of goodness of weather forecasting [7] and various ways to verify forecasts, or when aiming to find the best forecasting techniques based on available data, decision problem at hand, etc. [8]. Forecasts are rarely used without a specific purpose in mind, i.e., they are inputs to specific decision-problems. The nature and presentation of the forecasts is then to be a function of what the forecaster has to offer (based on data, models, etc.) and what the forecast user needs for decision-making. In the case where forecasts are to be used for a broad range of decision problems, they are made as generic as possible, as for the example of broad-audience weather forecasts. However, they could be made even more suitable and informative after dedicated interaction between the forecaster and the forecast user. Such interaction is to result in the definition of a forecast product.

Let us consider a set of very simple examples:

- The operator of a storage unit is performing basic arbitrage in an electricity market, i.e., based on the idea that, when the price is higher than average energy is released (and sold), while when the price is lower than average, energy is stored (and hence bought). The event of interest is then *"whether the price is above, or below, average"*
- A retailer is to buy the necessary amount of energy from the wholesale electricity market for his pool of consumers. Consequently, the event of interest is *"how much is the electricity demand for that the pool of consumers"*

Already here, one notices that those two simple problems relate to different types of events. In the storage arbitrage case, the forecasts are for a binary variable (above/below), while for the retailer case, the forecasts are for a continuous variable (energy consumption value).

Considering a decision-problem-oriented approach, forecast products are to be seen as an interface between the forecaster and the forecast user. Indeed, the forecaster uses relevant data streams, models and forecasting approaches, own expertise, as well as visualization tools, to eventually produce the right information that can be used by the forecast used as input to decision-making. This is illustrated in Figure 1. The forecast product should then somewhat encapsulate all the work and expertise of the forecaster in a format that makes it useful to the decision-maker. In additional to usefulness, other attributes may be sought after, e.g. simplicity, user-friendliness, etc.

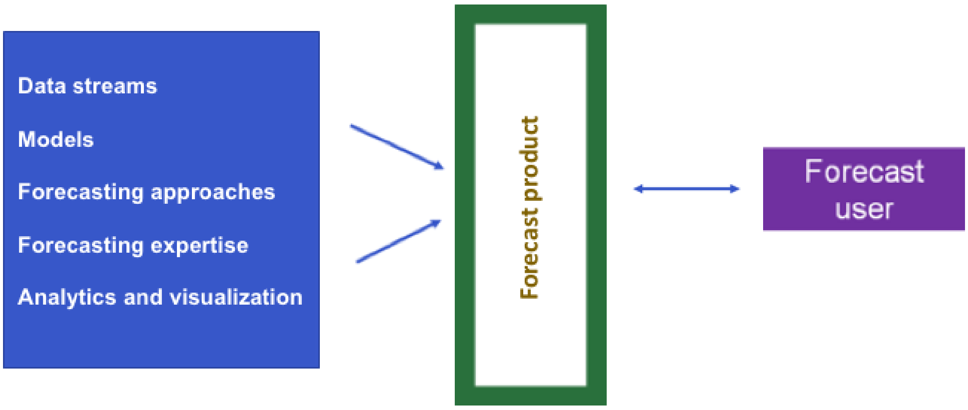


Figure 1: Forecast products are to be seen as the interface between forecaster and forecast user, within a decision-problem-oriented approach.

Back to the storage arbitrage problem, a forecast product is to acknowledge the characteristics of the decision problem at hand, as well as the way the problems is to be solved by the forecast user. For instance, if the forecast user employs deterministic optimization approach as basis to decision-making, forecasts for above/below should certainly also be deterministic. In contrast, if the forecast user employs stochastic optimization instead, forecasts should take a probabilistic format.

Today, when thinking of forecast products within renewable energy forecasting, it is of utmost importance to describe the type of event of interest, which will be summarized by the variable(s) of interest, lead times, spatial information (given location, grid, portfolio aggregation), etc. Similarly, the informativeness of the forecasts is to be defined: should it be probabilistic or deterministic, inform of dependencies between variables or not, etc.? Those aspects should be kept in mind when browsing through the following sections.

# 3 Regulatory and market developments

Owing to the large-scale integration of renewable energy generation capacities, the decentralization of energy system, the liberalization of electricity markets and the more proactive role of energy consumers, the market and regulatory framework for energy systems is changing substantially. Many of those changes motivate the development of new forecast products as discussed in the following. We mainly focus on the existing energy and ancillary service markets, as well as the recent push towards more decentralized electricity markets, e.g. in a peer-to-peer and community-based framework.

## 3.1 Energy and ancillary services

The observed evolution of electricity markets in Europe shows three salient technical characteristics: (1) a higher *volatility* due to the growing influence of actors with limited predictability such as variable renewable energy, demand response or interconnection transfers [9]; (2) a shift towards exchanges *closer to delivery*, evidenced by the increase of 10 TWh in the intra-day volumes of EPEX between 2017 and 2018 [10], and by the closure of balancing energy Ancillary Services (AS) foreseen in the last hour before delivery by the Electricity Balancing Guideline of the European Commission [11]; (3) *a wider range of markets* including tenders for balancing AS (Frequency Containment Reserve (FCR), automatic Frequency Restoration Reserve (aFRR), Replacement Reserve (RR)).

Volatility impacts the price of imbalances on the energy market, which is an important factor for the optimal offer of RES production. With the single price settlement being adopted in Europe for pricing imbalances, renewable producers who bid on the energy market should anticipate the system imbalance to decide whether to deviate from their day-ahead schedule. If the deviation helps reduce the system imbalance they may earn additional profit, otherwise they should minimize their expected deviations. Deterministic forecasting of the imbalance price has been proposed [12], but probabilistic predictions should be favoured given the uncertainty levels associated to high volatility. In some market conditions volatility is so high that it may be hard to achieve a well-performing price prediction. Additionally to the existing approach of a probabilistic forecast of the imbalance signal or volume [13], an alternative forecasting product consists in issuing a probabilistic classification of expected imbalance regimes.

The evolution of markets closer to delivery creates more interest on the prediction of intra-day prices. Intra-day markets in Europe differ vastly in terms of procurement (continuous or sessions) and liquidity. An interesting challenge for intra-day price forecasting is to propose a probabilistic forecasting product of intra-day energy prices (see for example [14]) which is generic enough to be useful to intra-day decision-making in different market configurations including continuous markets [9].

RES power plants have started to participate in the provision of AS in several European countries [15]. This *widens the range of markets* where RES actors can bid. However given the requests of high level of reliability and minimum bid size formulated by system operators for these services [16], aggregations of multiple RES plants (possibly mixing energy sources) within Virtual Power Plants (VPP) may be necessary to reduce the variability of production and secure enough reserve capacity [17]. In this context, the prediction of extremely reliable levels of the VPP production is needed and should include coherent forecasts over the hierarchy levels of the VPP (from a single plant to the total aggregation) in order to optimize bidding and dispatch. A last need of forecasting product to guarantee the technical feasibility of AS provision is the prediction of the available active power (AAP), i.e. the estimated maximum production level under curtailment for reserve provision. The prediction of the AAP is established for offshore wind farms [18], but still offers rooms for improvement [19] and should be developed for onshore wind farms with complex terrains and large photovoltaic plants.

Regarding AS market conditions, the transition from regulated mechanisms to markets based on short-term tenders is recent and still ongoing, so developments on AS price forecasting are limited compared to energy markets. A specific challenge scarcely tackled in the state of the art is the prediction of the acceptance probability of the AS bid [20], in the case of renewable-based AS bids. Lastly, new forecasting products for intra-day reserve prices are expected when intra-day markets for reserve activation (e.g. aFRR, FRR, RR) will emerge in Europe before 2025. Given that reserve markets are being standardized in Europe, a standard forecasting product for intra-day reserve prices may be a realistic perspective.

Finally, it is an open question to consider the new forecasting price products identified above as predictions of the price level, or of the spread price between the price of interest and a reference price on a second market (for instance spread between day-ahead energy price and day-ahead reserve price, or spread between intra-day reserve price and intra-day energy price). The spread price has two advantages: it is a synthetic indicator for the arbitrage between two markets, and can be directly integrated into bidding strategies, therefore streamlining decisions when compared to forecasting products for each market price.

## 3.2 Towards community-based and peer-to-peer electricity markets

The concept of prosumers, i.e. those in the power system that both can produce and consume (typically, a household with solar PV production units), has been around for quite some time. It appeared and gained interest with the increased deployment of solar PV production units on rooftops, hence mainly relating to PV. This is while, actually, the concept of renewable energy communities has been there quite longer: it can be traced back to the deployment of wind turbines in Denmark in the 1960s and 1970s, for which locals had stakes in the project. Conceptually, those energy communities both consumed and produced electric energy. Today, there is a strong push towards adapting electricity markets to better accommodate those prosumers and energy communities, since allowing to unleash some need flexibility for renewable energy integration while also sharing and spreading the investment burden and risk. In practice, this may translate to accelerating the decentralization of energy systems, while also democratizing electricity (and more generally energy) markets based on novel business models [21]. This trend is there in recent works performed within academia, R&D developments, new offering and operational practice in industry, as well as initiatives at the political and regulatory levels (e.g., with the EU Clean Energy Package). This development was also helped by the recent focus on blockchain (and distributed ledger technologies more generally) as potential backbone of future smart grids and electricity markets [22]. A recent review of approaches to peer-to-peer markets can be found in [23] while an alternative approach involving online matching is described in [24].

Necessarily, the design of electricity markets with more focus on peer-to-peer and community-based setup will generate new needs for forecasting products. Such markets will most likely require higher-resolution forecasts (say, with 5-minute resolution) since they will be closer to real-time while involving more local and volatile generation and consumption. They also require to look at more localized forecasts, for single households or groups of households. The forecasts are to involve net load, i.e., the difference between consumption and local generation. In practice it means that it combines several underlying variables to be predicted, as for the case of hybrid power plants for instance.

# 4 Evolution in technology and impact on forecasting needs

On the technology evolution side, some of the most important developments that may impact the need for forecast products include storage and hybrid power systems. Typically one expects that this calls for forecasts with higher temporal resolution (to obtain storage operation policies), combines wind and solar power generation forecasts for hybrid power plants, etc. Considering the special case of wind farms (especially, offshore), it is also clear that as the turbines get bigger and go higher, having wind forecasts at a single level may not be optimal, and one should look at forecasting full profiles. Some of those aspects are reviewed in the following.

## 4.1 Storage

The use of energy storage technologies in combination with RES power plants introduce two main requirements for forecasting systems: (a) modelling of the temporal and spatial dependency structure of forecast uncertainty; (b) high temporal resolution forecasts (i.e., $<$5-min) for model predictive control. For instance, focusing on the temporal structure of uncertainty, it is intuitive that the need for storage, was well as its optimal operation, will be different if one systematically over- (resp. under-) predict renewable energy generation, or if forecast errors alternate between being positive and negative. In the first case, one would have to continuously charge (resp. discharge) the storage for a sustained period of time – hence requiring large storage to cope. In the second case instead, one would alternatively charge and discharge the storage at each and every time step, making that the eventual storage capacity required is much smaller. This information about the temporal dependency within forecast uncertainty can be summarized within the error auto-correlation function for instance.

The study presented in [25] showed that the storage capacity rating grows non-linearly with the forecast errors auto-correlation and the expected lifetime is influenced linearly and ranges from 2 to 17 years for low and high values of error auto-correlation. The charge/discharge power rating is only marginally influenced by the error auto-correlation and is mainly affected by the error standard deviation. These results have two major implications:

- *Operational domain*: The forecast uncertainty should include the temporal dependency of forecast errors since it impacts the state-of-charge equation and the degradation function of the battery (usually included as a term in the objective function). Small-scale geographically distributed DER (storage and RES technologies) can be combined to compose a federated virtual power plant (see [26] for more details) and, in this case, the spatial-temporal dependency structure is a fundamental requirement in the uncertainty forecast.

- *Investment planning*: Auto-correlation and standard deviation of forecast errors should be considered when sizing energy storage, e.g., discarding auto-correlation can underestimate the storage capacity by a factor of about seven [27]. The forecast skill maps proposed in the FP7 EU project SafeWind (2008-2012) can be adapted to provide useful information for storage investment planning, such as spatial and/or temporal dependency of forecast errors and sharpness (linked to standard deviation) of forecast uncertainty, considering topographical effects or weather phenomena [28].

The simultaneous provision of multiple services (peak shaving, frequency containment reserve, etc.) by storage units is economically attractive and requires two temporal layers, day-ahead operational planning and real-time control [29]. The real-time control operates in a second-to-second basis but, for some services and when combined with RES power plants, it can be used to either compensate the deviation between market offer and measured electrical energy in a fixed time window or ensure a net energy around zero after a short period (e.g., 5-min). In this case, a high temporal resolution forecast, considering also a high frequency of forecast updates, is a valuable input to improve the tracking performance and minimize deviations.

## 4.2 Hybrid power plants

Hybrid power plants considered here combine different energy sources (e.g. Wind, Photovoltaics, Run-of-River Hydro) in a single physical location. A forecasting product of the combined power generation can take the form of probability density functions or ensembles (see next paragraph). Forecasts should reflect the correlations between energy sources which are expected to vary largely with hour of day and seasons [30]. Concerning their probabilistic performance, forecasts should display a level of uncertainty that is coherent with the uncertainty resulting from the combination of multiple sources. Innovative forecasting products for hybrid power plants may include other combinations of energy sources than the widespread wind-solar combination and consider a wider range of horizons, e.g. seasonal, day-ahead and intra-day horizons [31].

The ending of feed-in tariff support schemes for RES is opening the opportunity for hybrid power plants with wind and solar technologies, which can deliver a high number of equivalent full load hours (i.e. annual energy supply divided total number of hours in a year), hence yielding an attractive levelized cost of energy (LCOE) when compared to individual generation from wind or solar. This concept requires a forecasting product that takes the form of a unique (i.e., wind plus solar energy) power generation ensemble that considers the spatial-temporal dependency structure of wind speed and irradiance forecasts [32].

Different renewable energy sources can also be aggregated within a VPP operating distinct production sites. The correlation level between energy sources is expected to be low as a function of energy sources employed [33] and of the distance between sites. In this context, probabilistic forecasts of VPP production, which can be density forecasts, trajectories [34] or uncertainty sets constitute a new forecasting product. A hierarchical forecast of production from the plant level up to the total aggregated portfolio of the VPP is valuable for many applications, for instance to optimize the dispatch of a flexibility bid offered by the VPP.

## 4.3 Specifics of renewable energy technologies

Presently, each wind and solar power plant transmits large volumes of field-data (from SCADA systems, additional sensors and inverters) that are very valuable for O&M system and need to be combined with weather data (measurements and forecasts). In comparison to wind turbines, research for PV power plants is still in an early stage and some faults are not covered [35]. In this context, RES forecasting can be used to assess performance loss in both wind turbines and PV panels and is particularly relevant to detect soiling effects or, in the opposite direction, can benefit from information collected by sensors that detect soiling [36].

On the other hand, maintenance of offshore wind turbines is also becoming a high priority [37]. New forecasting products are proposed in [38] for assessing safety conditions during crew transfers to and from offshore installations. The final product is a probabilistic forecast of vessel motion up to five days ahead that can take the shape of temporal trajectories or an "access score" (i.e., categorical variable: good, fair, poor, no-go) that facilitates communication to decision-makers.

A tendency towards larger sizes of renewable plants is observed, especially for PV plants and offshore wind farms. In this context, the O&M challenges mentioned above must be addressed by forecasting tools that are scalable to large-size plants, for which the volume of data increases but also possibly the diversity of problems to be treated. In such large plants, heterogeneity in weather conditions can be observed across the panels or turbines composing the plant. New forecasting tools are expected to estimate the production of such large PV plants and in particular estimate at very short-term their available power [39]. Turbine-level wind forecasts have also been found to be valuable on large wind farms [40]. Finally, wind turbine technology developments have led to higher power extraction thanks e.g. to higher hub heights and rotor dimensions. In the context of new wind farms and re-powering of existing farms, a change in the wind power curve is expected, which drives the need for forecasting products adapted to new farm configurations (e.g. prediction of production at different profiles and higher hub heights) in order to predict correctly the expected future production levels and associated uncertainty [33].

As available land for RES production tends to reduce, new locations of RES power plants are actively considered, such as floating PV on hydro-power reservoirs or floating offshore wind turbines. These particular geographical locations have not been considered in past forecasting activities, therefore new forecasting products could consist in probabilistic forecasts adapted to such floating plants, which experience specific weather conditions. Finally, small to medium run-of-river hydro receives increasing attention from the forecasting community because of its possible provision of flexibility and variability levels which can be seasonally high (e.g. close to mountains). A probabilistic forecast for such small/medium run-of-river hydro power plants is a novel forecasting product.

# 5 Evolution of approaches to power system operation

During the last years, the use of uncertainty forecasts for setting reserve requirements has been the most successful use case and, in one way or another, transmission system operators (TSOs) are already using information from predictive distributions for this problem [41]. This section describes some recent challenges in system operation for power systems with near-100% RES, as well as new processes and needs from industry that, in most cases, can motivate the development of innovative forecasting products.

## 5.1 Operational management of the grid constraints

Transmission and distribution system operators (TSOs and DSOs), with the help of load and RES forecasts, are implementing operational management practices, which means moving from reactive to proactive management, in order to handle grid technical constraints and solve/mitigate congestion and voltage issues, e.g. to avoid line disconnection and further cascading failures. The set of possible decisions are generators re-dispatching, load shedding, demand response, phase-shifter transformer tap positions and grid topology reconfiguration [42]. The activation of this flexibility can be planned ahead in order to guarantee sufficient capacity to handle the predicted technical issues. The main goal is to operate the grid close to its limits in order to postpone grid reinforcements and maximize RES integration.

In this context, it is important to underline the following quote from [43] *"... TSOs are reluctant to move to shorter time periods. They are concerned that the risk of failing to contract the required level of reserves increases with a shorter window to source these reserves. TSOs can have some reservations about a dynamic, daily adjusted procurement volume since this requires an additional probabilistic assessment of the forecast errors and ramps of the next day."* This means that days-ahead RES power forecasts will remain a core product for TSOs in the operational planning of the power system. It is important to note that short notification times for demand response might mean a high price for using this resource, thus a trade-off between "level of uncertainty" and flexibility price must be considered in the decision-making processes [44].

Some examples are the *Apogée* framework where RTE (the French TSO) integrated in the Energy Management Systems (EMS) two days-ahead predictive tools for grid state and remedial actions (preventive and corrective) [45] and the Distribution Management System (DMS) of Enedis (the French DSO) that was enhanced with functions to forecast the grid operation constraints and solve problems with different actions (i.e., voltage regulation, generation curtailment, flexibility offers) [46]. The following requirements are present in both use cases: (a) provide fast decision-aid to human operators (see for an example of interpretability requirements by human operators [47]); (b) reduce the volume of information (including the number of uncertainties) that is displayed in the dispatch center. For new forecasting products, this means that they should contribute to reduce the computational time associated to decisions and visualization of information is a key factor to reduce human operator stress. In other words, forecasting products should generate high-level and integrated (e.g. with decision-aid tools) information. The inclusion of the spatial dependency structure in the forecasting product is mandatory in grid power flow analysis and the temporal dependency can be a requirement if inter-temporal constraints are considered, such as the number of tap position changes in on-load tap changer (OLTC), the maximum number of hours with line overload and storage state-of-charge (although the operation of this resource by TSOs and DSOs remains a "grey area" in the regulatory framework).

## 5.2 Inertia and frequency control in isolated power systems

Typically, isolated power systems are more vulnerable to disturbances comparatively to interconnected systems, since extreme frequency excursions are observed after the occurrence of network major power imbalances, which is due to their limited control and to their low amount of synchronous inertia. The growing integration of converter-interfaced RES (CI-RES), such as wind and solar, in isolated power systems is introducing challenges in power system management due to RES variability and uncertainty [48]. Moreover, the CI-RES generation units are connected to the grid through power electronic interfaces and, without additional control features, they lack the inherent ability of synchronous machines to provide voltage and frequency control [48]. Thus, the large-scale integration of non-dispatchable CI-RES raises major challenges to system operators, since it leads to the reduction of conventional thermal-based synchronous machines, decreasing the frequency regulation capabilities and online operational reserves, jeopardizing system stability. This might force system operators to curtail RES-based generation units to ensure adequate spinning reserve capability [49]. In order to cope with this problem, the establishment of new connection requirements has been developed for island power systems, requiring CI-RES to actively participate in the provision of regulation services, similarly to conventional power plants [50]. Another solution that has been identified and exploited is the installation of fast power-frequency regulation solutions (e.g., battery energy storage), acting as a complementary resource in the provision of power-frequency regulation [51].

The Australian energy market operator (AEMO) studied the impact of a N-1 contingency event over the system dynamic stability of two distinct cases: network operation without and with a large amount of wind and PV generation. For the first case, conventional unit commitment (UC)/economic dispatch (ED) algorithms scheduled a large amount of synchronous machines, and thus a large amount of synchronous inertia available in the system, given that the maximum post-contingency frequency deviation was well contained between the acceptable frequency operational band (49.5–50.5Hz) [52]. On the other hand, for the second case, fewer synchronous machines were scheduled by conventional UC/EC algorithms. Hence, for the same contingency event, the corresponding maximum frequency deviation rapidly breaches its constraint (49.5 Hz). This leads to the conclusion that the synchronous inertia inadequacy should be considered in the UC/ED problem with a large amount of non-dispatchable CI-RES generation.

RES variability is also critical in isolated power systems, e.g. PV power fluctuations can reach up to ±30% of the rated capacity per 10 seconds, ±70% per minute, increasing to

±80% per 10 minutes [53]. These large fluctuations may lead to major power imbalances that can compromise the network frequency stability [48] and, thus, some network operators of isolated systems impose ramp-rate limits to the PV power output [54]. Imposing such limits prevents RES from being exploited in an efficient way and may largely limit the revenue of the owners.

So, beyond local control strategies, innovative centralized control approaches are also required, for both preventive operational planning and real-time operation support. In this sense, load and RES generation forecasts play a key role in ensuring the network frequency transient stability. They are providential for the network operator to preventively determine an adequate UC/ED solution in operating scenarios with large share of RES generation. Regarding the preventive operational planning, one-day-ahead load and RES forecasts are required (with a 15-min temporal resolution), serving as an input for the UC/ED problem to find a generation dispatch solution for the next day. In such forecasts, beyond the mean value, additional statistics are important, such as the standard deviation, minimum and maximum value.

Moreover, load and RES intra-day forecasts (i.e., 15-min resolution, time horizon up to 24 hours and updated every 6-hour) are required to provide the opportunity for the system operator to apply corrective measures if necessary. For these requirements, the combination of large-eddy simulations (LES) and the computational power offered by graphics processing units (GPUs) enable operational NWP with high temporal (e.g. 30-sec) and spatial resolution [55]. From these high-resolution forecasts, it is possible to derive new forecasting products, such as variability indices for weather variables. Finally, in order to support the network real-time operation, power forecasts with a very-short-time resolution (i.e., 10 seconds to 1 minute) are required, with a lead time of 1-2 hours or, in other words, the system operator should be informed 1 or 2 hours in advance about the amplitude of power variability.

## 5.3 System operation under extreme/emergency scenarios

As mentioned in [56], a possible evolution for the future control and dispatch centers is to increase the level of automation, while human operators will intervene solely in abnormal grid operating states. This means that data science based algorithms will provide automated decision-making in scenarios covered by a high volume of historical data, while human operators handle scenarios with scarce data. This division of tasks imposes different requirements in terms of data sources for uncertainty forecasting products. As discussed in the International Energy Agency (IEA) Task 36 (Forecasting for Wind Power) [18], statistically-based uncertainty forecasts (i.e., without the use of meteorological ensembles) are unable to accurately forecast extreme weather events and are highly dependent on available historical data. Therefore, to better plan system operation under extreme scenarios the forecasting products should be based on meteorological ensembles, where data science has an important role to play in the post-processing of ensemble members and the conversion to power forecasts.

It is important to underline the fact that weather-dependent events are one of the major causes for disturbances and contingencies in electric power systems [57] and, with the support of forecasts, system operators can plan hours-ahead short-term resilience measures, e.g. ensure islanding operation capabilities [58]. In terms of forecasting products, this requires an accurate (location and severity) and event-driven representation of the information. Ideas already studied for public communication of weather uncertainty [59] can be borrowed by the energy sector (but tailored-made for specific events) and event-based evaluation framework adopted to evaluate the quality of the products [60].

Presently, outage forecasting tools (e.g. Elenia Oy, EDP, Enedis) are already using NWP data to predict the number and location (e.g. region, substation, feeder) of grid faults for a time horizon ranging between three hours and three days, supporting a better management and position of repair crews [61–63]. However, state-of-the-art methodologies are mainly based on deterministic NWP with a numerical representation of variables and data are handled by traditional machine learning methods with known difficulties in dealing with rare events and performing causality discovery. Future forecasting products should use weather ensembles as an input, with an event or categorical representation of weather variables, and be able to find causal relations to allow for fast decision-making and interpretability.

## 5.4 Dynamic line rating (DLR)

DLR is a technology aiming at exploiting measurements of conductors' thermal state in order to modify dynamically the thermal rating of an overhead line. Due to variable environmental conditions (e.g. high wind, cold temperature) it is possible to increase the real-time allowable current, which enables high integration levels of RES and peak load management and leads to investment deferral [64]. By combining NWP with machine learning it is possible to produce probabilistic forecasts of overhead line ampacity for the next hours and days [65], which is a forecasting product itself. State-of-the-art results and discussions with end-users suggest that the use of low quantiles (below 5%) is necessary, e.g., according to [66], an accurate estimation and use of quantiles below 1% can lead to overall lower reserve costs and frequency of incidents. As there is no direct measurement of line rating (it is estimated instead), in addition to the forecasting product itself, the evaluation of the accuracy and the use of ampacity "measurements" in statistical trainings is an important issue to be further considered in research and forecast product development.

The work in [66] suggests the following requirements for DLR forecasting products: (a) use of conditional parametric functions for tail's modelling, where the full probability distribution is not needed; (b) forecasting value (and not just accuracy) should be considered to select different DRL forecasts since small differences the quantile scoring rule values can be translated to big differences in system operation cost.

## 5.5 Inertia forecasting

In power systems with a high penetration of RES generators connected via power converters, the inertia provided by synchronous generators is reduced. System operators need to estimate the expected level of inertia in order to avoid instabilities and prepare mitigation actions such as an increase of inertia or decrease of maximum possible loss. Uncertainties in the estimation of inertia have recently increased, leading to the need of accurate point forecasts or probabilistic forecasts of inertia in a low-inertia power system [67].

Inertia originally consisted in two main contributors: transmission-system-connected synchronous generators and generation and demand units with synchronous rotating masses embedded in distribution systems. The inertia contribution of transmission-system-connected synchronous generator units has so far been predicted using a physical model [68, 69]. This physical model relies upon forecasts of the online statuses of the individual generators. However, in systems operated by independent system operators, forecasts of generators' statuses have appeared to be inaccurate during periods with low energy prices [68]. Moreover, (accurate) forecasts of the generators' statuses are not always available to system operators if energy markets are operated by independent market operators. This asks for models to accurately forecast which generators are expected to be online as a result of energy market trading. To forecast the embedded units' inertia contribution, existing forecast models only have a simple linear structure [69]. This is driven by the small data set of estimates of embedded units' inertia contribution that is available, as these contributions can only be accurately estimated during infrequent, large disturbances in demand or supply. Moreover, existing forecast models provide only point forecasts of expected inertia to the system operator [67].

The characteristics of inertial response are quickly evolving over time. First of all, the amount of synchronous generator units is decreasing. This changes the generation mix that is online and increases the relative importance of the contribution of embedded units, which is more variable and uncertain. Second, although not widely deployed in practice yet, power-electronic-interfaced wind generators and storage units can provide emulated inertial response, so called virtual inertia, if they are adequately controlled. Although the objective of this emulated inertial response is to mimic the inertial response of synchronous generators, its characteristics differ as it is typically provided by more variable and uncertain energy buffers. These evolutions in inertial response call for inertia forecast models that are able to deal with non-stationarities and are probabilistic in nature to give system operators insight into the uncertainty of the system inertia that will be available [67].

# 6 What do forecast users think about new forecast products?

End-users will be very different, from TSOs, DSOs, plant operators to smart cities, smart buildings even 'smart households'. Some of them will need **explainability** (interpretable models, physics-inspired models) but others will 'just' need **robustness** (no knowledge nor understanding of forecasts, no 'corrective' action).

A difference can also be found in the literature between **distributed** algorithms and **decentralized** algorithms depending on who are the end-users:

- distributed algorithms: one central server ('master') and 'contractual' agents = need for **privacy**?

- decentralized algorithms: no central server but collaborations between different agents = need for **incentives**?

The International Energy Agency Wind Task 36 on Forecasting has recently conducted a questionnaire asking power system stakeholders about their current use of forecasting solutions [70]. It appeared that the majority of respondents (around 70% of 24 persons representing system operators, traders, R&D, power producers) are aware of probabilistic forecasts but only about 20% of them do use probabilistic forecasts while operating. This is not explained by a lack of staff or IT solutions, but 54% agreed to state that they 'fear that speculative planning may result in a loss'. Therefore the applicability of probabilistic forecasting in the industry seems to be due in parts to a lack of knowledge of the utility of such forecasting solutions for energy trading or grid management.

The respondents of the questionnaire developed by Smart4RES [71] have indicated the properties of the new forecasting products they are the most interested in. The background of the 12 respondents is as follows, reusing some of the categories proposed in [70]: 3 Energy Service Organizations, 1 weather forecasting provider, 1 Power Management Company, 7 R&D organizations. Results are shown in Table 1. Explainability and robustness are the most voted properties, however the small number of answers limits the interpretation of these results. In the IEA survey, respondents indicated that they did not use directly numerical weather predictions [70]. Interestingly, in the Smart4RES questionnaire, a weather service provider spontaneously responded and showed interest in power system applications, showing that the existing gap between weather predictions and RES applications can be filled. Furthermore, a DSO indicated that they use weather predictions for the mitigation of extreme phenomena. In this context, they identified that the predictions they use were limited in their ability to predict localized and sudden storm events. Lastly, respondents mentioned specific needs on the topic of RES forecasting:

- R&D respondents mentioned needs that are in line with their subjects of interests, for instance multi-time scale forecasting, scenarios modelling the dependence between wind and hydro power plants, or the access to data on RES plants and power systems.
- Energy Service Organizations look forward to forecasts which are useful to tackle the challenges they face such as high deviations in RES and load forecasts, EV load forecasts or flexibility potential.
- The need for forecasts that operate in real-time and/or are issued at very-short-term, in a nowcasting approach.

| Properties of new forecasting products | Number of respondents interested |
|---|---|
| Improve the explainability of forecasting | 7 |
| Ensure robustness of forecasting | 7 |
| Provide incentives to collaborate | 5 |
| Address high dimensionality or high resolution | 5 |
| Preserve Privacy | 3 |

Table 1: Interest to properties of new forecasting products from respondents to Smart4RES questionnaire

A very good overview of the link between forecasting and the use of forecasts in practice was recently published in [18]. The interested reader is then referred to that publication for an extensive coverage of the topic.

# 7 Prospects from the R&D side[2]

Considering the needs expressed by forecast users, since the beginning of the development of forecasting methodologies for renewable energy applications, many on the R&D sides (in academia, research institutes as well as small and large companies) have brought in novel ideas and concepts that were readily answering the needs of forecast users. They have additionally proposed some novel forecasting concepts and products, which were eventually adopted by forecast users. In this section, we review some of the latest directions within R&D related to renewable energy forecasting, also underlying their prospects in terms of forecast improvement and impact on business models. Those topics include hierarchical forecasting, missing data, novel probabilistic forecasting products, collaborative analytics, as well as data markets.

## 7.1 Hierarchical forecasting

With the expansion of renewable energy generation sources, there will be many more production sites, scattered all over territories, from household solar panels to large wind farms. This shall have a huge impact on the way electric power networks are exploited and on the forecasting products needed to do so. In particular, such an evolution calls for an increasing development of what is usually known in the literature as hierarchical forecasting.

Hierarchical forecasting involves multiple time series that are hierarchically organized and can be aggregated at several levels. It requires not only good prediction accuracy at each level of the hierarchy but also coherency between levels. In the case of an electric power network, different agents may need forecasts at different aggregation levels of the grid, all those forecasts having to be coherent. While ensuring consistency, hierarchical forecasting can also be a way of improving the forecasts involved in the hierarchy, by taking into account the possible dependencies between the series.

Common original methods in hierarchical forecasting were split between bottom-up versus top-down approaches, or compromised on a middle-out approach. In the past few years a third class of methods has thrived: the reconciliation approach, which can come along with a combination step or not. In [72] the authors propose a new approach which provides optimal forecasts that are better than forecasts produced by either a top-down or a bottom-up approach. They independently forecast all series at all levels of the hierarchy, then use a regression model to optimally combine and reconcile these forecasts.They show that the resulting revised forecasts add up appropriately across the hierarchy, are unbiased and have minimum variance under some simple assumptions. However, their method

[2]Note that, since this part is focused on recent research focus within forecasting for renewable energy applications, the text may be more technical and more difficult to digest by a broad audience.

require the initial forecasts to be unbiased, which precludes regularized estimators like LASSO or ridge regression ones for initial forecasts. The reconciliation of load forecasts in distribution grids proposed in [73] showed that the error on the aggregated forecast is lowered significantly by the reconciliation (10% relative improvement in this case), but not base forecasts (at most 0.1% relative improvement). Authors note though that higher improvements for base forecasts may be obtained with a reconciliation approach formulated conditionally on the error of base forecasters. In [74] the authors introduce a game-theoretically optimal reconciliation method, which is guaranteed to only improve any given set of forecasts. This opens up new possibilities in constructing the initial forecasts, which no longer have to be unbiased. Unlike [72], this new approach does not address the goal of sharing information between hierarchical levels but only reconciliation. A separate procedure is required to share information.

Those reconciliation methods initiate a new path in hierarchical forecasting but only deal with forecasting the mean of each time series of the hierarchy. This contrasts with the shift in the literature towards probabilistic forecasting. To the best of our knowledge, the first method which aimed at providing probabilistic forecasts for hierarchical time series was proposed in the $34^{th}$ International Conference on Machine Learning (in 2017), see [75]. Generating probabilistic forecasts for hierarchical time series is challenging: in addition to computing entire distributions, which might be very different throughout the hierarchy, because of the hierarchical structure, it requires catching the dependencies between the distributions, marginal predictive distributions being not enough. The authors propose an algorithm to compute the conditional predictive cumulative distribution function for all series in the hierarchy. In particular they compute the joint distributions using copulas. An updated version of this work can also be found in [76].

Since then, works on hierarchical forecasting have still focused mainly on forecasting the mean. We can cite [77] which proposes a new reconciliation approach that incorporates the information from a full co-variance matrix of forecast errors in obtaining a set of coherent forecasts. This can be used to construct prediction intervals but assuming the forecast errors are normally distributed and conditionally on the base forecasts to be unbiased. In [78] this approach is updated by adding non-negativity constraints to ensure that the coherent forecasts are strictly non-negative. In [79] the authors use machine learning techniques to allow for non-linear combination of the base forecasts. In [80], [81] and [82] the issue of online hierarchical forecasting is tackled.

## 7.2 Missing data

As much as data collection, transmission and storage has improved tremendously over the last few decades, there are still situations and periods for which data may be missing. While the issue of missing data may not be apparent at first, actually in practice it may significantly affect the learning (for model parameter tuning) and forecasting processes. In addition, the way the data is missing, e.g. at random or not, for sparse short periods or long ones, etc. may require different approaches to be accommodated. This is why we discuss these aspects in detail in the following.

A widely-used nomenclature for missing data mechanisms is that from [83]. It divides them into three types, depending on the relationship between the missingness and the value of the observations:

- MCAR – Missing Completely At Random: when the probability of a data point being missing is completely independent of any variables in the dataset.
- MAR – Missing At Random: when the missingness in one variable is independent of its own value but does depend on the value of another.
- MNAR – Missing Non At Random: when the missingness in one variable depends on its own value.

A subsequent part of the literature only considers the first two "simple" mechanisms and struggles for the harder, yet prevalent, MNAR case. Indeed when data is missing for renewable energy generation, we are in a case of MNAR, with missing patterns related to the forecast output. For instance, in [84] the authors studied a dataset of 30 European wind farms, with two years of 10-minute resolution data. Three main sources of missing data were brought out, which are

- Data missing in the raw time series due to sensor measurements;
- Missing periods due to site-wide maintenance works;
- Curtailments.

While the first source of missing data may be considered at random, the other two may not, since planned maintenance activities are often scheduled for times with lower wind speed, and wind farm sites are more likely to be curtailed close to rated power from grid restraints limiting power flows. Nevertheless, the rates of missing values computed by the authors for this dataset show that the main source of missing values is the first one.

The simplest method to deal with missing data is the *complete case analysis*, where any data points with partial information are discarded. The first inconvenience of this method is that it will produce biased results if data is not MCAR, because the remaining data is no longer a random sample from the underlying distribution. This may be counteracted by weighting the remaining data using missing probabilities but this does not prevent reduction in size of the dataset. This is not an alternative in high dimension, as "One of the ironies of working with Big Data is that missing data play an ever more significant role" [85]. For example, considering a data set of 5 variables where a single value is missing with probability 0.01, 95% of the observations are complete, i.e., we have values for all of the 5 variables; with 300 variables, only 5% of the observations are now complete. For a wind farm of 100 wind turbines, this would mean that about 63% of the recorded observations would be complete. But as we use lagged observations for forecasting, one missing value shall lead to the cancellation of several observations instead of one, and to an even worse proportion of data being kept.

Another popular method to deal with missing data is *imputation*. It consists in providing a value to replace the missing one. The simplest way of doing it is taking the mean of the variable, referred to as (single) mean imputation. More sophisticated imputation methods include regression imputation, stochastic regression imputation, linear and spline interpolation, Last Observation Carried Forward, Next Observation Carried Backward in the case of time series, among others. The popular practice of imputing with the mean of the variable on the observed entries is known to have serious drawbacks, as it distorts the joint and marginal distributions of the data, see [86]. But mean imputation has never really been studied when the aim is to predict an output. Abundant literature addresses the missing data issue in an inferential framework, but very few studies are dedicated to supervised learning settings. For instance, in the scenario 1 of [84], the authors studied the case of data missing in the training set, when the task is forecasting the wind power generation of 10 sites, on a 2.5-hour ahead horizon, with VAR models. They observed that for a 11.65% rate of missing data, the complete case analysis leads to a worsening of forecasts of 19%, while a simple single mean imputation allows them to keep a worsening of only 1.27%. This kind of unexpected result might be brought together with a striking theoretical result obtained in [87]: while single mean imputation is bad for estimation, it is not bad for prediction. This is, to the best of our knowledge, the first result justifying this convenient practice of handling missing values.

The third way of dealing with missing data is *likelihood-based methods*. They rely on the well-known Expectation-Maximization algorithm which was first proposed in [88]. The EM algorithm iteratively finds the maximum likelihood estimator of the data distribution alternating between missing data imputation given the observed data, and complete data maximum likelihood estimation. In this case, the imputation of missing data is not the goal of the procedure but a side effect of each iteration. An online version of the EM algorithm was first derived in [89], where the parameters are updated at each observation by using the gradient of the incomplete data likelihood, weighted by the complete data Fisher information matrix. In [90] the authors proposed a new approach more directly connected to the usual EM algorithm, which did not rely on integration with respect to the complete-data distribution. The resulting algorithm is thus simpler and suitable for conditional (regression) models. With this kind of algorithms, not only the issue of missing data is tackled while forecasting, but their online version also makes it possible to forecast in high-dimensional settings.

Finally, a sound link has to be made with the hierarchical forecasting topic, when obviously information coming from other levels of a hierarchy might be of help while filling or handling missing values to the purpose of forecasting.

## 7.3 New probabilistic forecasting products

Looking at it from the R&D side, forecasts are to be thought of in a probabilistic framework, even if, eventually, those will be communicated in a simplified manner to the practitioner and forecast user, e.g. as single-value forecasts (given the loss function and decision problem at hand), intervals, ramp forecasts, etc. [18]. As hinted by [91], for the case of renewable energy generation, the complete forecasting problem may consist in predicting the joint density for all lead times and locations considered (possibly also the different variables e.g. wind, solar and load). However, in practice, one may need to revert to simpler versions of that problem – for instance forecasting of the marginal densities for each lead time and locations, individually, or focusing on certain pre-defined events like ramps [92]. Very few works have focused on informing about the complete joint densities. Joint densities would in that case give the complete information about what will happen for each location, time and renewable energy type involved, as well as dependencies. For instance, it may summarize the uncertainty in future generation from two wind farms individually, while also telling about the potential correlation in generation increasing or decreasing simultaneously at those locations. To describe those joint densities in a practical and pragmatic manner, different forecast products can be considered, like scenarios (also referred to as ensemble forecasts in meteorological forecasting) which are a conventional input to a wide range of decision problems under uncertainty, or simultaneous prediction intervals [93].

The research on how to optimally accommodate probabilistic forecast information in decision problems has advanced tremendously over the last decade, with a wealth of optimization approaches being considered, and with emphasis placed on a broad range on application areas (e.g., active distribution grid management [94], chance-constrained optimal power flow [95]). This has eventually triggered the need to think of novel probabilistic forecast products, which would be specifically tailored to certain decision problems and approaches to solve them. For instance in [96], the authors aim at obtaining uncertainty sets in a data-driven manner, as input to robust optimization problem in electricity markets. This idea of readily obtaining uncertainty sets in a data-driven manner was previously proposed by [97] and [98] for the case of ellipsoidal and polyhedral multivariate uncertainty regions, respectively.

## 7.4 Collaborative learning and forecasting

Many works in the literature have revealed that there would be benefits in using information in the vicinity of the sites of interest when forecasting their future power generation. The benefits may be in the form of increased forecast accuracy, but also of a decrease in the frequency of large/extreme forecast errors [99]. Such information may come from meteorological stations, remote sensing devices (sky imagers, weather radars, etc.), other wind farms and solar parks, and alternative weather forecasts over neighboring grid points. Many examples can be found, e.g. with focus on directly looking at a set of locations and all relevant dependencies [100], and going towards higher dimensions and online learning setups [101].

When looking at such proposals to improve forecast accuracy, one somewhat assumes that the data are to be shared by all agents involved. In practice, however, it is rarely the case that agents of the power system (wind farm operators, solar parks operators, system operators, portfolio managers, traders, etc.) are willing to share their private data and information. Reasons for that behaviour include competition in the electricity markets, secrecy regarding the performance of installed assets, etc.

Adapting to such a situation requires new business models in renewable energy forecasting, to find ways to reap the benefits from distributed collection and storage of data, possibly without actively sharing the data themselves. Some proposals were recently pushed forward towards collaborative learning and forecasting, specifically for renewable energy forecasting applications - see [102] and [103] for instance. In both cases, the authors explore alternative setups based on distributed optimization to yield this collaborative approach. In addition, for the case of [103], special emphasis is placed on lighting the computational burden and adapting to slow changes in the underlying process dynamics within an online learning environment. A natural criticism to those methods is that, even if the data are not actively shared, some could design attack models (based on inverse optimization or Bayesian inference) permitting to eventually retrieve the original data of the agents. This issue has therefore triggered additional research within privacy-preserving analytics, i.e., with the aim to protect the privacy of those agents engaged in collaborative learning and forecasting. For a recent overview of methods and challenges related to privacy-preserving analytics, and application to forecasting, the reader is referred to [104]. Finally, if looking broader than renewable energy forecasting applications only, this trend towards collaborative and privacy-preserving analytics is in line with more general developments in the field of machine learning, towards federated learning [105], following the term proposed by Google scientists in 2016.

## 7.5 Data and forecasts becoming products themselves?

The quantity and quality of data accessible to each RES agent are reflected in their forecasting performance, which may then directly impact their imbalance costs in electricity markets. The current ways to obtain such data for a RES plant is primarily through historical data and local measurements, which have technological and economic limitations due to the size and location of the plant [106, 107]. Considering the correlation of weather data among sites within certain geographical distances, it becomes apparent that sharing data among these RES agents can improve their forecasting capabilities, as discussed in the above.

One big challenge of data sharing is the preservation of privacy. In a competitive market, agents tend to be cautious about giving away any private information to their competitors. Here, we assume the forecasting accuracy is a non-decreasing function of data. In other words, obtaining more data can only improve an agent's forecasting accuracy or at least maintain an agent's forecasting accuracy at the same level. This may lead to the following outcomes:

1. The agent with more information can refine their bids in the electricity market to achieve lower regulation costs and higher revenues.
2. As the overall forecasting accuracy increases for a significant number of RES agents, the real-time imbalance between the total actual generation and the total forecast generation is likely to reduce, resulting in lower regulation unit costs for all agents.

The first outcome brings revenue to the agents who gain additional data, and the second outcome benefits all RES agents. From the perspective of an agent who has data that can be shared with others, the first outcome is revenue neutral, and the second is revenue positive. It promises tremendous potential for setting up a data platform for agents to collaborate and achieve higher mutual profits. By putting a price on the data shared on this platform, it allows an agent to compare their valuation of privacy and their increased profit, thus helping them make decisions on whether to share their data with other agents or not.

The rapid advancements in information and communication technologies in recent years have made information sharing a very simple task that can be done at almost zero cost. However, as people become more aware of the value of data, there has been an increasing interest in reforming data markets to expand their scope and participation [108]. Based on the market architecture, the current state-of-the-art data platforms can be put into four categories: trade, forward auction, reverse auction, and double auction.

In a data trade, only one data seller and one data buyer enter the market at a time [109]. Through negotiation, the seller and the buyer reach a deal on the price and the quantity/quality of the traded data. In this market setup, the traded data bring additional profit to the buyer, while the seller gets a share of that profit through the trading process itself [110]. Each trade can take place between different seller-buyer pairs as long as the trades are independent of each other.

In a data forward auction, there is a single data seller, e.g. a data collector, and multiple data buyers who submit bids to purchase data from the seller. This monopolist data seller is the leader of the market, and they can either choose the buyer that bids the highest price [111] or create different versions of their data and sell them to multiple buyers at different prices depending on the versions selected by the buyers [112, 113].

In a data reverse auction, there is a single data buyer and multiple data sellers. The monopolistic data buyer makes offers to purchase data from the sellers. Under a limited budget, the data buyer only selects the data that would maximize the buyer's utility [114, 115]. The sellers, on the other hand, only accept offers that outweigh their loss of privacy as a result of data sales [116, 117]. In an environment where the sellers' data are correlated, one agent's decision to sell their data may cause privacy leakage for other agents, thus reducing the value of their data [118]. As a result, some agents who initially place a higher value on their privacy than the offer they receive may eventually decide to accept the offer just because others have sold their data.

Finally, in a data double auction, there are multiple data sellers and multiple data buyers. This market usually requires a third party market operator to purchase data from the sellers and sell data to the buyers. This market operator can either be a for-profit data vendor [119, 120] or a non-profit 'marketplace' [121]. An iterative auction mechanism is shown in [122] to be able to achieve the maximized social surplus in a setup where the sellers, the buyers, and the market operator are all profit-driven.

In the context of an electricity wholesale market, RES agents can trade their information with each other, which makes them both buyers and sellers. This adds another level of complexity into the design of the data market because an agent not only has a privacy value on their own data, they also have to take into account the possible outcomes of other agents acquiring more data as outlined above, which might in turn affect the agent's own profits. Using the basic framework proposed in [121], a data platform was formulated in [123], where wind producers can sell and purchase data from other wind producers to achieve better forecasts. The transactions are conducted through a centralized market operator who collects a portion of the increased profits from each buyer and redistributes them among the sellers, while ensuring the data privacy of all participants.

A basic assumption made in [123] is that wind producers are price takers and that their bidding policy of the wind generators cannot impact imbalance prices [124]. This means that only the first aforementioned outcome as a result of improved forecasting capabilities has been considered, where the traded data can only benefit the buyers in the electricity market. An immediate extension of this work is to incorporate the second outcome into the model as well, allowing the better collective forecasting capabilities to lower the imbalance prices for all agents, which offers additional financial incentives for selling data. Other important research topics relating to the establishment of a data market include the maximization of social welfare in a competitive environment, and the formulation of a fair and sustainable profit allocation strategy among the participants.

Looking ahead, data trading among RES agents may lead to additional scenarios besides the two obvious outcomes mentioned above. For example, in current forward electricity wholesale markets, generation bids are only accepted based on their prices and quantities without any consideration given to their reliability. As the RES penetration continues to grow, maintaining the power network stability will become more challenging given the RES variability, so it is plausible that the system operator might start prioritizing RES producers with higher forecasting accuracy when accepting RES bids in order to reduce the real-time balancing efforts and costs. The interaction between the data market and the electricity market under those circumstances will be a very interesting area of future research.

# 8 Conclusions and perspectives

Forecasting of renewable energy generation is often referred to as a mature research field, with well-established methods, views on input data, interface with forecast users and decision problems, etc. This has been confirmed by the growing literature on the topic, as well as operational and commercial offering of forecasting solutions.

However, as technology as well as market and regulatory framework evolve very fast (and possibly in a non-negligible manner), this induces a new push towards looking at potential advances in terms of forecasting. The most common objective naturally is to improve forecast accuracy and quality, but it is also to rethink forecast products in a problem-oriented manner. Our objective here was to look at the need for new and enhanced forecast products from various perspectives, e.g. as motivated by various developments within technology, market, regulation, operational practice etc. but also as seen from the point of view of forecast users and R&D.

Looking at it from a broader perspective, it is clear that new and enhanced forecast products will be developed and offered in an operational context in the coming period, for instance with focus on higher resolution and new power system operational practice. More generally though, we certainly see the possibility for substantial changes in business models related to forecasting. Some changes in business models may for instance aim to yield optimal value from all the relevant data being collected, but not shared. It may additionally focus on the consistency of forecasts at various granularity and aggregation levels. Those changes in business models actually go beyond the case of renewable energy forecasting only – they are a general trend supported by the digitalization of our economies and societies.

It is the aim of the EU project Smart4RES to explore the possibilities of both new products and new business models in renewable energy forecasting in the coming, both in terms of novel developments, but also in shortening the time to market.

# References

[1] R. Inman, H. Pedro, and C. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013.

[2] G. Kariniotakis, *Renewable Energy Forecasting: From Models to Applications*. Amsterdam (The Netherlands): Elsevier, 2017.

[3] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *WIREs ENergy and Environemnt*, vol. 9, p. e365, 2020.

[4] W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881–2893, 2011.

[5] J. Schalkwijk, H. Jonker, A. Siebesma, and E. V. Meijgaard, "Weather forecasting using gpu-based large-eddy simulations," *Bulletin of the American Meteorological Society*, vol. 96, no. 5, pp. 715–723, 2015.

[6] A. Orlov, J. Sillmann, and I. Vigo, "Better seasonal forecasts for the renewable energy industry," *Nature Energy*, vol. 5, pp. 108–110, 2020.

[7] A. Murphy, "What is a good forecast? an essay on the nature of goodness in weather forecasting," *Weather and Forecasting*, vol. 8, no. 2, pp. 281–293, 1993.

[8] J. C. Chambers, S. K. Mullick, and D. D. Smith, "How to choose the right forecasting technique," *Harvard Business Review*, vol. 49, pp. 45–71, 1971.

[9] B. Uniejewski, G. Marcjasz, and R. Weron, "Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO," *International Journal of Forecasting*, vol. 35, no. 2, pp. 1533–1547, 2019.

[10] EPEX SPOT, "Annual Report 2018," tech. rep., 2018.

[11] The European Commission, "Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing," *Official Journal of the European Union*, vol. 2017, no. November, pp. 312/6 – 312/53, 2017.

[12] I. Dimoulkas, M. Amelin, and M. R. Hesamzadeh, "Forecasting balancing market prices using hidden markov models," in *2016 13th International Conference on the European Energy Market (EEM)*, pp. 1–5, 2016.

[13] J. Bottieau, L. Hubert, Z. De Greve, F. Vallee, and J.-F. Toubeau, "Very-Short-Term Probabilistic Forecasting for a Risk-Aware Participation in the Single Price Imbalance Settlement," *IEEE Transactions on Power Systems*, vol. 35, pp. 1218–1230, Mar. 2020.

[14] J. R. Andrade, J. Filipe, M. Reis, and R. J. Bessa, "Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model," *Sustainability (Switzerland)*, vol. 9, no. 11, pp. 1–29, 2017.

[15] IRENA, "Innovative Ancillary Services: innovation Landscape Brief," p. 24, 2019.

[16] J. Strahlhoff, A. Liebelt, S. Siegl, and S. Camal, “Development and Application of KPIs for the Evaluation of the Control Reserve Supply by a Cross-border Renewable Virtual Power Plant,” in *Proceedings of the 49th Annual Meeting of the German Informatics Society: 50 years of the German Informatics Society - Computer Science for Society, INFORMATICS 2019*, 2019.

[17] S. Camal, A. Michiorri, and G. Kariniotakis, “Optimal Offer of Automatic Frequency Restoration Reserve from a Combined PV/Wind Virtual Power Plant,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6155–6170, 2018.

[18] R. J. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B. M. Hodge, U. Cali, and G. Kariniotakis, “Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry,” *Energies*, vol. 10, no. 9, pp. 1–48, 2017.

[19] T. Göçmen, G. Giebel, P.-E. Réthoré, and J. P. Murcia Leon, “Uncertainty Quantification of the Real-Time Reserves for Offshore Wind Power Plants,” in *Proceedings of the 15th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*, 2017.

[20] M. Merten, F. Rücker, I. Schoeneberger, and D. U. Sauer, “Automatic frequency restoration reserve market prediction: Methodology and comparison of various approaches,” *Applied Energy*, vol. 268, p. 114978, 2020.

[21] Y. Parag and B. Sovacool, “Electricity market design for the prosumer era,” *Nature Energy*, vol. 1, pp. 1–6, 2016.

[22] M. Andoni, V. Robu, and D. Flynn, “Crypto-control your own energy supply,” *Nature*, vol. 548, p. 158, 2017.

[23] T. Sousa, T. Soares, P. Pinson, F. Moret, T. Baroche, and E. Sorin, “Peer-to-peer and community-based markets: A comprehensive review,” *Renewable Sustainable Energy Reviews*, vol. 104, pp. 367–378, 2019.

[24] H. Esch, F. Moret, P. Pinson, and A. Marin Radoszynski, “Online matching and preferences in future electricity markets,” in *Proceedings of the Nineteenth Yale Workshop on Adaptive and Learning Systems*, 2019.

[25] A. Michiorri, J. Lugaro, N. Siebert, R. Girard, and G. Kariniotakis, “Storage sizing for grid connected hybrid wind and storage powerplants taking into account forecast errors autocorrelation,” *Renewable Energy*, vol. 117, pp. 380–392, Mar. 2018.

[26] T. Morstyn, N. Farrell, S. J. Darby, and M. D. McCulloch, “Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants,” *Nature*, vol. 3, pp. 94–101, 2018.

[27] P. Haessig, B. Multon, H. B. Ahmed, S. Lascaud, and P. Bondon, “Energy storage sizing for wind power: impact of the autocorrelation of day-ahead forecast errors,” *Wind Energy*, vol. 18, pp. 43–57, Jan. 2015.

[28] J. S. Rodrigo, L. F. Paredes, R. Girard, G. Kariniotakis, K. Laquaine, N. Stoffels, and L. von Bremen, *Renewable Energy Forecasting: From Models to Applications*, ch. The role of predictability in the investment phase of wind farms, pp. 341–357. Woodhead Publishing Series in Energy, 2017.

[29] E. Namor, F. Sossan, R. Cherkaoui, and M. Paolone, "Control of battery storage systems for the simultaneous provision of multiple services," *IEEE Transactions on Smart Grid*, vol. 10, pp. 2799–2808, May 2019.

[30] M. Kay and ARENA, "Forecasting and Characterising Grid Connected Solar Energy and Developing Synergies with Wind," tech. rep., 2015.

[31] K. Engeland, M. Borga, J. D. Creutin, B. François, M. H. Ramos, and J. P. Vidal, "Space-time variability of climate variables and intermittent renewable electricity production – A review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 600–617, 2017.

[32] S. Alessandrini and T. McCandless, "The schaake shuffle technique to combine solar and wind power probabilistic forecasting," *Energies*, vol. 13, p. 2503, 2020.

[33] M. Koivisto, P. Maule, N. Cutululis, and P. Sorensen, "Effects of Wind Power Technology Development on Large-scale VRE Generation Variability," in *Proceedings of the 13th IEEE PowerTech Milano 2019: Leading Innovation for Energy Transition*, June 2019.

[34] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, and L. Badesa, "Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications," *Applied Energy*, vol. 242, pp. 1396–1406, May 2019.

[35] A. Colli, "Failure mode and effect analysis for photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 804–809, Oct. 2015.

[36] S. Costa, A. Diniz, and L. L. Kazmerski, "Solar energy dust and soiling R&D progress: Literature review update for 2016," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2504–2536, 2018.

[37] B. P. Lau, E. M. Ma, and M. Pecht, "Review of offshore wind turbine failures and fault prognostic methods," in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012)*, May 2012.

[38] C. Gilbert, J. Browell, and D. McMillan, "Probabilistic access forecasting for improved offshore operations," *International Journal of Forecasting*, In Press, 2020.

[39] Reinhard Mackensen, Y.-M. Saint-Drenan, D. Jost, R. Fritz, N. Asanalieva, M. Widdel, and M. Hahler, *Regelenergie durch Wind- und Photovoltaikparks Abschlussbericht.* 2017.

[40] J. Browell, C. Gilbert, and D. McMillan, "Use of turbine-level data for improved wind power forecasting," in *2017 IEEE Manchester PowerTech*, pp. 1–6, IEEE, June 2017.

[41] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Lazaro, and N. Detlefsen, "Methodologies to determine operating reserves due to increased wind power," *IEEE Transactions on Sustainable Energy*, vol. 3, pp. 713–723, Oct. 2012.

[42] R. Brazier, L. Cunha, P. Hermans, G. de Jong, T. Knop, M. Lallemand, M. Merkel, A. S. Risnes, M. de la Torre Rodríguez, and P. de Wit, "An integrated approach to active system management," tech. rep., CEDEC, E.DSO for Smart Grids, Eurelectric, GEODE, ENTSO-E, Apr. 2019.

[43] T. Schittekatte, V. Reif, and L. Meeus, "The EU electricity network codes," tech. rep., European University Institute, 2019.

[44] N. Marinho and Y. Phulpin, "Economics of distributed flexibilities for the grid considering local short-term uncertainties," in *2015 IEEE Eindhoven PowerTech*, (Eindhoven, Netherlands), 2015.

[45] J.-Y. Astic, G. Bareux, T. Buhagiar, M. Dussartre, N. Omont, P. de Longeaux, Y. Jacquemart, B. Meyer, B. Peyron, and L. Saludjian, "Control center designs: New functions and challenges for the transmission system operator," *IEEE Power and Energy Magazine*, vol. 16, pp. 57–66, Feb. 2018.

[46] H. Morais, C. Paris, O. Carré, M. Carlier, and B. Bouzigon, "Levers optimization in short-term operational planning for real distribution systems," in *25th International Conference on Electricity Distribution*, (Madrid, Spain), June 2019.

[47] A. Marot, B. Donnot, S. Tazi, and P. Panciatici, "Expert system for topological remedial action discoveryin smart grids," in *2018 Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion*, (Dubrovnik, Croatia), Nov. 2018.

[48] H. Vasconcelos, C. Moreira, A. Madureira, J. P. Lopes, and V. Miranda, "Advanced control solutions for operating isolated power systems: Examining the Portuguese islands," *IEEE Electrification Magazine*, vol. 3, pp. 25–35, Mar. 2015.

[49] B. Kroposki, B. Johnson, Y. Zhang, V. Gevorgian, P. Denholm, B.-M. Hodge, and B. Hannegan, "Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy," *IEEE Power and Energy Magazine*, vol. 15, pp. 61–73, Mar. 2017.

[50] E. Rodrigues, G. J. Osório, R. Godina, A. W. Bizuayehu, J. M. Lujano-Rojas, and J. P. S. Catalão, "Grid code reinforcements for deeper renewable generation in insular energy systems," *Renewable and Sustainable Energy Reviews*, pp. 163–177, Jan. 2016.

[51] G. N. Psarros, E. G. Karamanou, and S. A. Papathanassiou, "Feasibility analysis of centralized storage facilities in isolated grids," *IEEE Transactions on Sustainable Energy*, vol. 9, pp. 1822–1832, Oct. 2018.

[52] H. Gu, R. Yan, and T. K. Saha, "Minimum synchronous inertia requirement of renewable power systems," *IEEE Transactions on Power Systems*, vol. 33, pp. 1533–1543, Mar. 2018.

[53] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "Power output fluctuations in large scale PV plants: One year observations with one second resolution and a derived analytic model," *Progress in Photovoltaics*, vol. 19, pp. 218–227, Mar. 2011.

[54] "Madeira island grid code," tech. rep., Região Autónoma da Madeira. Decreto Regulamentar Regional n.º 8/2019/M, 2019.

[55] C. Gilbert, J. W. Messner, P. Pinson, P.-J. Trombe, R. Verzijlbergh, P. van Dorp, and H. Jonker, "Statistical post-processing of turbulence-resolving weather forecasts for offshore wind power forecasting," *Wind Energy*, vol. 23, pp. 884–897, Apr. 2020.

[56] A. M. Prostejovsky, C. Brosinsky, K. Heussen, D. Westermann, J. Kreusel, and M. Marinelli, "The future role of human operators in highly automated electric power systems," *Electric Power Systems Research*, vol. 175, p. 105883, Oct. 2019.

[57] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience ofpower systems: Impacts and possible mitigation strategies," *Electric Power Systems Research*, vol. 217, pp. 259–270, 2015.

[58] R. J. Santos, A. Neves, A. Araújo, P. Castro, F. Guerra, C. Gouveia, and J. Damásio, "Project SENSIBLE's results from MV/LV coordinated island operation in a distribution grid," in *25th International Conference on Electricity Distribution*, (Madrid, Spain), June 2019.

[59] S. L. Joslyn and R. M. Nichols, "Probability or frequency? Expressing forecast uncertainty in public weather forecasts," *Meteorological Applications*, vol. 16, pp. 309–314, Aug. 2009.

[60] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, Aug. 2012.

[61] B. Almeida, G. Faria, T. Soares, R. J. Santos, J. Ferreira Pinto, T. Santos, I. Preto, and C. Monteiro, ""outage forecast" – a real application of machine learning on grid operation management strategies," in *25th International Conference on Electricity Distribution*, (Madrid, Spain), June 2019.

[62] S. Vähäkuopus, H. Paananen, L. Anttila, and T. Kupila, "Predicting the impacts of the major disturbances for better resource management and situational awareness," in *25th International Conference on Electricity Distribution*, (Madrid, Spain), June 2019.

[63] O. Faivre, P. Cochet, J. Mérigeault, and S. Folleville, "Optimizing network replacement with AI," in *25th International Conference on Electricity Distribution*, (Madrid, Spain), June 2019.

[64] B. Meyer, J.-Y. Astic, P. Meyer, F.-X. Sardou, C. Poumarede, N. Couturier, M. Fontaine, C. Lemaitre, J. Maeght, and C. Straub, "Power transmission technologies and solutions: The latest advances at RTE, the French transmission system operator," *IEEE Power and Energy Magazine*, vol. 18, pp. 43–52, Mar. 2020.

[65] A. Michiorri, H.-M. Nguyen, S. Alessandrini, J. B. Bremnes, S. Dierer, E. Ferrero, B.-E. Nygaard, P. Pinson, N. Thomaidis, and S. Uski, "Forecasting for dynamic line rating," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1713–1730, Dec. 2015.

[66] R. Dupin, L. Cavalcante, R. J. Bessa, G. Kariniotakis, and A. Michiorri, "Extreme quantiles dynamic line rating forecasts and application on network operation," *Energies*, vol. 13, p. 3090, June 2020.

[67] E. Heylen, G. Strbac, and F. Teng, "Challenges and opportunities of inertia estimation and forecasting in low-inertia power systems." Aug. 2020.

[68] P. Du and J. Matevosyan, "Forecast System Inertia Condition and Its Impact to Integrate More Renewables," *IEEE Transactions on Smart Grid*, vol. 9, pp. 1531–1533, Mar. 2018.

[69] D. Wilson, J. Yu, N. Al-Ashwal, B. Heimisson, and V. Terzija, "Measuring effective area inertia to determine fast-acting frequency response requirements," *International Journal of Electrical Power & Energy Systems*, vol. 113, pp. 1–8, Dec. 2019.

[70] G. Giebel, J. Cline, H. Frank, W. Shaw, P. Pinson, B.-M. Hodge, G. Kariniotakis, J. Madsen, and C. Möhrlen, "Wind power forecasting: IEA Wind Task 36 & future research issues," *Journal of Physics: Conference Series*, vol. 753, p. Article 032042, Sept. 2016.

[71] Smart4RES, "Smart4RES - Questionnaire on the use and needs of forecasting for renewable energy applications," 2020. https://ec.europa.eu/eusurvey/runner/02027011-b975-7860-1c35-4037eb359764.

[72] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, "Optimal combination forecasts for hierarchical time series," *Computational Statistics and Data Analysis*, vol. 55, pp. 2579–2589, Mar. 2011.

[73] L. Nespoli, V. Medici, K. Lopatichki, and F. Sossan, "Hierarchical demand forecasting benchmark for the distribution grid," *Electric Power Systems Research*, vol. 189, p. 106755, 2020.

[74] T. van Erven and J. Cugliari, *Modeling and Stochastic Learning for Forecasting in High Dimensions*, ch. Game-Theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts, pp. 297–317. Springer International Publishing, 2015.

[75] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Coherent probabilistic forecasts for hierarchical time series," in *Proceedings of the 34th International Conference on Machine Learning (ICML-2017)*, 2017.

[76] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical probabilistic forecasting of electricity demand with smart meter data," *Journal of the American Statistical Association*, vol. 00, no. 0, pp. 1–17, 2020.

[77] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization," *Journal of the American Statistical Association*, vol. 114, no. 526, pp. 804–819, 2019.

[78] S. L. Wickramasuriya, B. A. Turlach, and R. J. Hyndman, "Optimal non-negative forecast reconciliation," *Statistics and Computing*, 2020.

[79] E. Spiliotis, M. Abolghasemi, R. J. Hyndman, F. Petropoulos, and V. Assimakopoulos, "Hierarchical forecast reconciliation with machine learning." June 2020.

[80] M. Huard, R. Garnier, and G. Stoltz, "Hierarchical robust aggregation of sales forecasts at aggregated levels in e-commerce, based on exponential smoothing and holt's linear trend method." June 2020.

[81] M. Brégère and M. Huard, "Online hierarchichal forecasting for power consumption data." Mar. 2020.

[82] C. Di Modica, P. Pinson, and S. Ben Taieb, "Online forecast reconciliation in wind power prediction," *Electric Power Systems Research*, vol. 190, no. 106637, 2021.

[83] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[84] R. Tawn, J. Browell, and I. Dinwoodie, "Missing data in wind farm time series: Properties and effect on forecasts," in *Proceedings of the 21st Power Systems Computation Conference (PSCC-2020)*, 2020.

[85] Z. Zhu, T. Wang, and R. J. Samworth, "High-dimensional principal component analysis with heterogeneous missingness." July 2019.

[86] R. Little and D. Rubin, *Statistical Analysis with Missing Data, 3rd edition*. Wiley, 2019.

[87] J. Josse, N. Prost, E. Scornet, and G. Varoquaux, "On the consistency of supervised learning with missing values." July 2020.

[88] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[89] D. M. Titterington, "Recursive parameter estimation using incomplete data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 2, pp. 257–267, 1984.

[90] O. Cappé and E. Moulines, "On-line expectation-maximization algorithm for latent data models," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 71, no. 3, pp. 593–613, 2009.

[91] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[92] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting ramps of wind power production with numerical weather prediction ensembles," *Wind Energy*, vol. 16, no. 1, pp. 51–63, 2013.

[93] R. J. Bessa, "From marginal to simultaneous prediction intervals of wind power," in *18th International Conference on Intelligent System Application to Power Systems (ISAP)*, 2015.

[94] T. Soares, R. J. Bessa, P. Pinson, and H. Morais, "Active distribution grid management based on robust AC optimal power flow," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6229–6241, 2017.

[95] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Review*, vol. 56, no. 3, pp. 461–495, 2014.

[96] A. Velloso, A. Street, D. Pozo, J. M. Arroyo, and N. G. Cobos, "Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 958–969, 2020.

[97] F. Golestaneh, P. Pinson, R. Azizipanah-Abarghooee, and H. Gooi, "Ellipsoidal prediction regions for multivariate uncertainty characterization," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4519–4530, 2018.

[98] F. Golestaneh, P. Pinson, , and H. Gooi, "Polyhedral predictive regions for power system applications," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 693–704, 2019.

[99] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 480–489, 2014.

[100] J. Dowell and P. Pinson, "Very-short-term wind power probabilistic forecasts by sparse vector autoregression," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 763–770, 2016.

[101] J. Messner and P. Pinson, "Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1485–1498, 2019.

[102] Y. Zhang and J. Wang, "A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5714–5726, 2018.

[103] B. Sommer, P. Pinson, J. Messner, and D. Obst, "Online distributed learning in wind power forecasting," *International Journal of Forecasting*, vol. 37, no. 1, pp. 205–223, 2021.

[104] C. Goncalves, R. Bessa, and P. Pinson, "A critical overview of privacy-preserving approaches for collaborative forecasting," *International Journal of Forecasting*, 2020. available online.

[105] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. Jan. 2019, no. 12, 2019.

[106] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[107] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762–777, 2014.

[108] D. Bergemann and A. Bonatti, "Markets for information: An introduction," *Annual Review of Economics*, vol. 11, pp. 85–107, 2019.

[109] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A survey on big data market: Pricing, trading and protection," *IEEE Access*, vol. 6, pp. 15132–15154, 2018.

[110] M. Babaioff, R. Kleinberg, and R. Paes Leme, "Optimal mechanisms for selling information," *Proceedings of the ACM Conference on Electronic Commerce*, pp. 92–109, 2012.

[111] K. Bimpikis, D. Crapis, and A. Tahbaz-Salehi, "Information sale and competition," *Management Science*, vol. 65, no. 6, pp. 2646–2664, 2019.

[112] H. Yu and M. Zhang, "Data pricing strategy based on data quality," *Computers and Industrial Engineering*, vol. 112, pp. 1–10, 2017.

[113] D. Bergemann, A. Bonatti, and A. Smolin, "The design and price of information," *American Economic Review*, vol. 108, no. 1, pp. 1–48, 2018.

[114] Y. Chen, N. Immorlica, B. Lucier, V. Syrgkanis, and J. Ziani, "Optimal data acquisition for statistical estimation," *ACM EC 2018 - Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 27–44, 2018.

[115] Y. Chen and S. Zheng, "Prior-free data acquisition for accurate statistical estimation," *ACM EC 2019 - Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 659–677, 2019.

[116] R. Montes, W. Sand-Zantman, and T. Valletti, "The value of personal information in online markets with endogenous privacy," *Management Science*, vol. 65, no. 3, pp. 1342–1362, 2019.

[117] J. P. Choi, D. S. Jeon, and B. C. Kim, "Privacy and personal data collection with information externalities," *Journal of Public Economics*, vol. 173, pp. 113–124, 2019.

[118] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Too much data: Prices and inefficiencies in data markets," vol. 23, no. 3, p. 2019, 2019.

[119] Y. Jiao, P. Wang, D. Niyato, M. Abu Alsheikh, and S. Feng, "Profit maximization auction and data management in big data markets," *IEEE Wireless Communications and Networking Conference, WCNC*, 2017.

[120] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "An online pricing mechanism for mobile crowdsensing data markets," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[121] A. Agarwal, M. Dahleh, and T. Sarkar, "A marketplace for data: An algorithmic solution," *ACM EC 2019 - Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 701–726, 2019.

[122] X. Cao, Y. Chen, and K. J. Ray Liu, "Data Trading with Multiple Owners, Collectors, and Users: An Iterative Auction Mechanism," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 2, pp. 268–281, 2017.

[123] C. Goncalves, P. Pinson, and R. J. Bessa, "Towards Data Markets in Renewable Energy Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 3029, no. c, pp. 1–1, 2020.

[124] P. Pinson, C. Chevallier, and G. N. Kariniotakis, “Trading wind generation from short-term probabilistic forecasts of wind power,” *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1148–1156, 2007.