

# Smart4RES

## Literature Highlights

Project Newsletter #8

March 2022



### EDITO

The vision of Smart4RES is to bring a holistic approach, covering the whole model and value chain related to RES forecasting, from weather forecasting up to end-use applications. To achieve this, Smart4RES proposes disruptive research ideas with the aim of breaking new ground rather than making incremental improvements. The knowledge created so far forms the basis for the next generation of modelling and short-term forecasting tools of weather-based RES power production and related decision making.

A sneak peek of the research articles published is presented in this 8th edition of the project Newsletter.

---

### In brief, the articles published so far addressed the following bottlenecks:

Need of new NWP products adapted to RES use cases	3
Decaying commercial value of forecast products	8
Need for models adequate in high-RES integration scenarios and digitalisation of the energy sector	10
Lack of open data and of price-based incentives to share data	11



## Table of content

<b>Need of new NWP products adapted to RES use cases</b>	<b>3</b>
Uncertainty of shortwave cloud radiative impact due to the parameterization of liquid cloud optical properties (Météo-France, JAMES)	<b>3</b>
Cloud height measurement by a network of all-sky imagers (DLR, Atmos. Meas. Tech. Discuss.)	4
A Hybrid Solar Irradiance Nowcasting Approach (DLR, Solar RRL)	5
Applying self-supervised learning for semantic cloud segmentation of all-sky images (DLR, Atmos. Meas. Tech. Discuss.)	6
Measurement of diffuse and plane of array irradiance by a combination of a pyranometer and an all-sky imager (DLR, Solar Energy)	7
<b>Decaying commercial value of forecast products</b>	<b>8</b>
Online forecast reconciliation in wind power prediction (DTU, Electric Power Systems Research)	8
Forecasting Conditional Extreme Quantiles for Wind Energy (INESC TEC, Electric Power Systems Research)	9
<b>Need for models adequate in high-RES integration scenarios and digitalisation of the energy sector</b>	<b>10</b>
Prescriptive Trees for Integrated Forecasting and Optimization Applied in Trading of Renewable Energy (ARMINES, IEEE Transactions on Power Systems)	10
<b>Lack of open data and of price-based incentives to share data</b>	<b>11</b>
Towards Data Markets in Renewable Energy Forecasting (INESC TEC & DTU, IEEE Transactions on Sustainable Energy)	11
A critical overview of privacy-preserving approaches for collaborative forecasting (INESC TEC & DTU, International Journal of Forecasting)	13
Privacy-preserving Distributed Learning for Renewable Energy Forecasting (INESC TEC & DTU, IEEE Transactions on Sustainable Energy)	14
Regression markets and application to energy forecasting (DTU, TOP)	15

## Need of new NWP products adapted to RES use cases

### Uncertainty of shortwave cloud radiative impact due to the parameterization of liquid cloud optical properties

E. Jahangir, Q. Libois, F. Couveux, B. Vié, D. Saint-Martin. JAMES. 2021.

Climate predictions differ a lot from one model to another, and the difficulty to simulate how clouds will behave in a warmer world is largely responsible for that. The radiative effect of clouds depends on the size of the individual droplets forming a cloud, a quantity that is not explicitly represented in atmospheric models. In this study, Météo-France investigates how not accounting for the detailed droplet size distribution affects the capability of climate models to reliably predict the radiative impact of clouds. By assuming a variety of droplet size distributions in a set of simulations, it has been observed that apparently similar clouds (i.e. containing the same amount of condensed water) can have very different radiative impacts depending on the assumed distribution. This is primarily attributed to the estimation of the effective radius of cloud droplets, a key quantity that drives cloud radiative properties. Differences up to 20% are observed on critical quantities such as fluxes at top-of-atmosphere and at the surface. The absorption of radiation within clouds is also significantly altered. The impact on the global estimate of the shortwave cloud radiative effect (a measure of how much clouds cool the Earth by reflecting solar radiation) is around 13%, which highlights the need to improve the representation of the microphysical characteristics of clouds to run more reliable weather and climate predictions.

#### What's ahead?

This study was restricted to the shortwave impact of liquid clouds. To have a global picture of the impact of droplet size distribution assumptions on cloud radiative impact, it should be extended to ice clouds and to the longwave range (LW). These new parameterizations, which are available [here](#), will be tested in the operational model of Météo-France, AROME, to investigate whether this improves the solar irradiance forecasts, and consequently PV production forecasts.

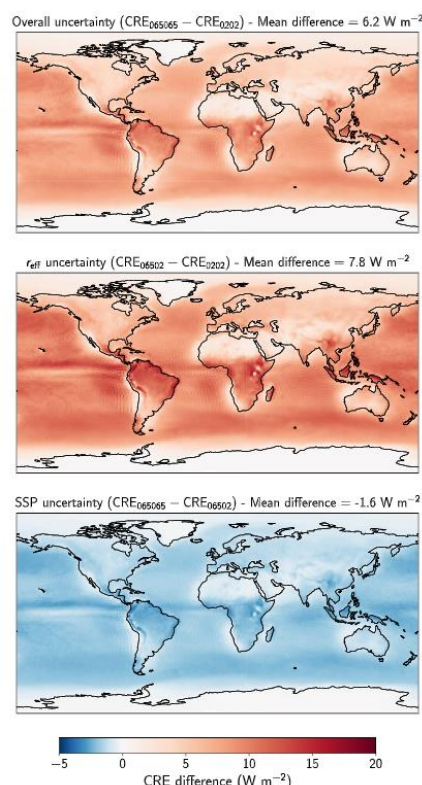


Figure 1 Global difference maps of the 2000–2004 average cloud radiative effect (CRE) computed from outputs of the CNRM-CM6-1 amipsimulation for various choices of cloud single scattering properties (SSPs) and reff parameterizations.

Read the article

## Cloud height measurement by a network of all-sky imagers

Blum, N. B., Nouri, B., Wilbert, S., Schmidt, T., Lünsdorf, O., Stührenberg, J., Heinemann, D., Kazantzidis, A., and Pitz-Paal, R. *Atmos. Meas. Tech. Discuss.*, 2021.

Cloud base height (CBH) is an important parameter for many applications including solar irradiance forecasting for the next seconds to hours ahead (nowcasting). To nowcast solar irradiance, systems based on all-sky imagers (ASIs), cameras monitoring the entire sky dome above their point of installation, have been demonstrated. Accurate knowledge of the CBH is required to nowcast the spatial distribution of solar irradiance around the ASI's location at a resolution down to 5 m. In this article, DLR demonstrates experimentally that measuring the CBH with a network of ASIs allows to enhance the accuracy, in comparison to the individual ASI pairs.

The ASI network-based estimation of the CBH aims to combine the measurements of the CBH from ASI pairs arranged in proximity and organized in a network. Conditional probabilities are modeled from historic CBH measurements received from ASI pairs and a reference ceilometer.

The influence of camera distance on the accuracy of ASI-based estimation of the CBH is decisive. For low clouds (CBH < 4 km), small camera distances were found to lead to most accurate measurements. Under these conditions, deviations were found to increase steadily with camera distance as described in the literature. For higher clouds (especially for CBH > 8 km) larger camera distances were found to increase accuracy.

Depending on the meteorological conditions on-site, the study provides suggestions to select a camera distance for an ASI pair with a stereoscopic estimation of the CBH. The presented approach for merging the measurements of ASI pairs in an ASI network combined the favored properties of the individual ASI pairs.

### What's ahead?

- A trade-off between the CBH accuracy and coverage or costs must be found for ASI networks.
- The presented approach can be transferred to other sites using the conditional probabilities of the CBH found at the Oldenburg site. Found distributions may then be extended to include more frequent observations of high clouds.
- The approach will be enhanced by first extending the utilised statistics of the measured CBH with data from other sites at which a combination of ASI pair and ceilometer is available. Such an extended data set will additionally allow to use more elaborate statistical methods, including neural networks.
- A procedure for the generation of irradiance nowcasts based on the whole ASI network, utilising the method to estimate the CBH has also been implemented and will be published soon.

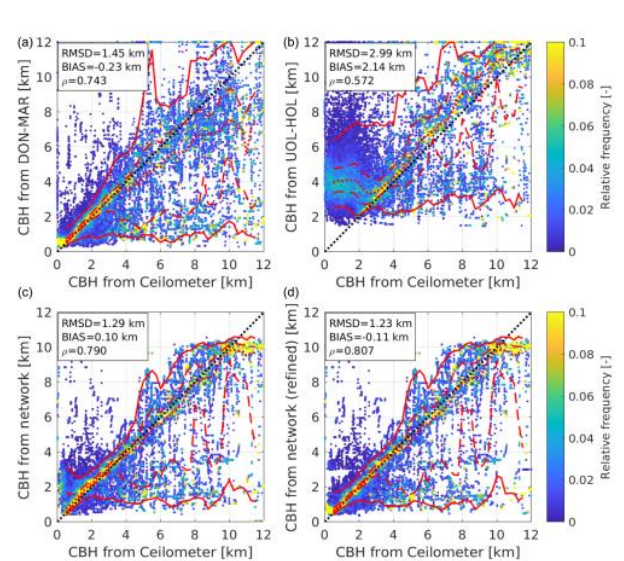


Figure 2 Relative frequency of ASI-based CBH estimation for a given CBH from the ceilometer. An evaluation of two of the ASI pairs, DON-MAR (a) and UOL-HOL (b), with respective camera distances of 0.8 and 5.7 km, and from the ASI network, without (c) and with refinements (d), is shown.

[Read the article](#)

## A Hybrid Solar Irradiance Nowcasting Approach: Combining All Sky Imager Systems and Persistence Irradiance Models for Increased Accuracy

Bijan Nouri, Niklas Blum, Stefan Wilbert, Luis F. Zarzalejo. *Solar RRL*, 2021.

Accuracy being the decisive factor for the effective use of nowcasts, in this second study, DLR aims to increase the nowcast accuracy by combining real-time-capable hybrid ASI and persistence nowcasting approaches, which persist with the prevailing irradiance conditions, while maintaining the spatial coverage and resolution obtained by the ASI system. This hybrid approach combines the strengths while reducing the respective weaknesses of both approaches. The validation procedure is conducted with eight distinct reference stations distributed over an area of roughly 1 km<sup>2</sup>. The benchmark analysis showed that:

- Both the purely persistence- and the ASI-based approaches are clearly outperformed by the hybrid approach over all LTs.
- The performance of the nowcasting approaches also depends strongly on the conditions at hand: the persistence approach shows the best performance for low variability conditions, whereas it is outperformed by the ASIs during highly variable conditions.
- Validation results show reductions of the root mean square deviation of up to 12% due to the hybrid approach.

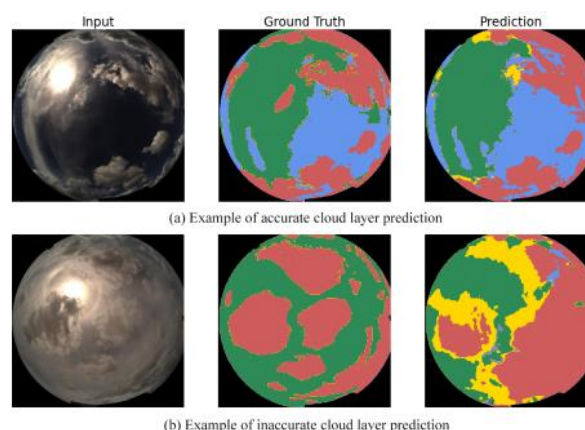


Figure 3 Positive and negative prediction examples compared to ground truth for a given input from the validation set (blue: sky; red: low layer; yellow: mid-layer; green: high-layer)

### What's ahead?

- Dense networks of ASIs with overlapping viewing angles can lead to further notable performance enhancements of ASI-based nowcasts. Such networks can also increase significantly the nowcasts' horizon while maintaining the same resolution.
- Future research will evaluate in detail the nowcasts potential to improve the performance of electrical grids and power plants under variable conditions, despite the nowcasts uncertainties.

[Read the article](#)

### ***Applying self-supervised learning for semantic cloud segmentation of all-sky images***

*Fabel, Y., Nouri, B., Wilbert, S., Blum, N., Triebel, R., Hasenbalg, M., Pascal, K., Luis, Z., & Pitz-Paal, R. Atmos. Meas. Tech. Discuss, 2022.*

Semantic segmentation of ground-based all-sky images (ASIs) can provide high-resolution cloud coverage information of distinct cloud types, applicable for meteorology-, climatology- and solar-energy-related applications. In this article, DLR and its partners present an approach to pretrain deep neural networks for ground-based sky observation using raw image data. This approach allows to exploit many more data than in purely supervised training and thus increase the model's performance. Indeed, results show that these networks can be trained more effectively without the need for labeling thousands of images by hand. For cloud segmentation, this offers the possibility for simultaneously detecting and distinguishing clouds with associated properties in ASIs using deep learning.

In the first step, about 300 000 ASIs are used in two different pretext tasks for pretraining. One of them pursues an image reconstruction approach. The other one is based on the DeepCluster model, an iterative procedure of clustering and classifying the neural network output. In the second step, the model is fine-tuned on a small labeled dataset of 770 ASIs, of which 616 are used for training and 154 for validation. For each of them, a ground truth mask was created that classifies each pixel into clear sky or a low-layer, mid-layer or high-layer cloud.

- By comparing the results from DLR's pretrained models with the ones from ImageNet or random initialization, the benefits of self-supervised learning were shown. Considering only pixel accuracy, the pretrained models reach over 85 %, compared to 82.1 % (ImageNet initialization) and 78.3 % (random initialization).
- The most significant advantage becomes evident when regarding the distinction of cloud classes. For more challenging cloud types such as mid- or high layer clouds, precision of the self-supervised models is often 10–20 percentage points higher.

#### ***What's ahead?***

- More studies on other pretext tasks or other methods exploiting raw image data could further improve the cloud classification.
- Future models could be trained using large datasets of multiple cameras at different sites, potentially capable of generalizing well on any camera.

[Read the article](#)



## Measurement of diffuse and plane of array irradiance by a combination of a pyranometer and an all-sky imager

N. Blum, S. Wilbert, B. Nouri, J. Lezaca, D. Huckebrink, A. Kazantzidis, D. Heinemann, L. Zarzalejo, M. Jiménez, R. Pitz-Paal. *Solar Energy* 2022.

Accurate measurements of global tilted irradiance (GTI) and diffuse horizontal irradiance (DHI) are crucial for many solar energy applications. GTI is required to predict the output of a photovoltaic (PV) generation or non-concentrating solar-thermal collectors. DHI is used to derive the rear side plane of array irradiance that is used by bifacial PV installations. Combined with GHI, DHI is used to derive the direct irradiance which is useful for tracked or concentrating collectors. However, measurement equipment known so far is costly and/or limited in its scope.

To tackle the shortcomings associated with GTI and DHI measurements, the authors suggest a comparably inexpensive and robust setup of a thermopile pyranometer, combined with an all-sky imager (ASI), to measure GTI and DHI. This combination allows to avoid radiometric calibrations of the camera. Together with the inclusion of commonly

available image meta data, it allows to check, correct and self-calibrate major influences present in the measurement. Validity of the approach is assured by tests at two distinct sites. The study concludes that for GTI, the method is more accurate than solar tracker-based transposition.

### What's ahead?

- The method will be tested at several further sites and benchmarked with other methods to derive diffuse and direct irradiance that do not need a solar tracker.
- The method has been demonstrated for its application in all sky imager-based forecasting systems and a corresponding study will be published soon.
- The method could improve solar resource assessment and solar plant modelling noticeably while also enabling nowcasting and reducing investment and maintenance costs.



Figure 4 Setup at CIEMAT's PSA (top row): Solar tracker with pyrheliometer and shaded pyranometer (left), pyranometers in 3 tilted planes (center) and ASI (right). The pyranometer

[Read the article](#)

## Decaying commercial value of forecast products

### Online forecast reconciliation in wind power prediction

Chiara Di Modica, Pierre Pinson, Souhaib Ben Taieb. *Electric Power Systems Research*, 2021

Among modern challenges of the electric power sector, ensuring coherency of forecasts among various agents and at various aggregation levels has recently attracted attention. A number of reconciliation approaches have been proposed, from both game-theoretical and statistical points of view. However, most of these approaches make unrealistic unbiasedness assumptions and overlook the fact that the underlying stochastic processes may be nonstationary.

In this article a data-driven approach to forecast reconciliation is introduced in a multivariate regression framework. The forecast reconciliation relies on a multivariate least squares estimator to determine the regression parameters for each level of the forecast hierarchy. The main interest of that approach is that it eventually allows for online forecast reconciliation, hence allowing to adapt to nonstationarity in the underlying stochastic processes. A proof of reconciliation by design is also provided. Even trained on specific past data, the approach proposed by DTU allows to reconcile any new forecasts out of sample. Overall, this approach allows for adaptive learning with exponential forgetting.

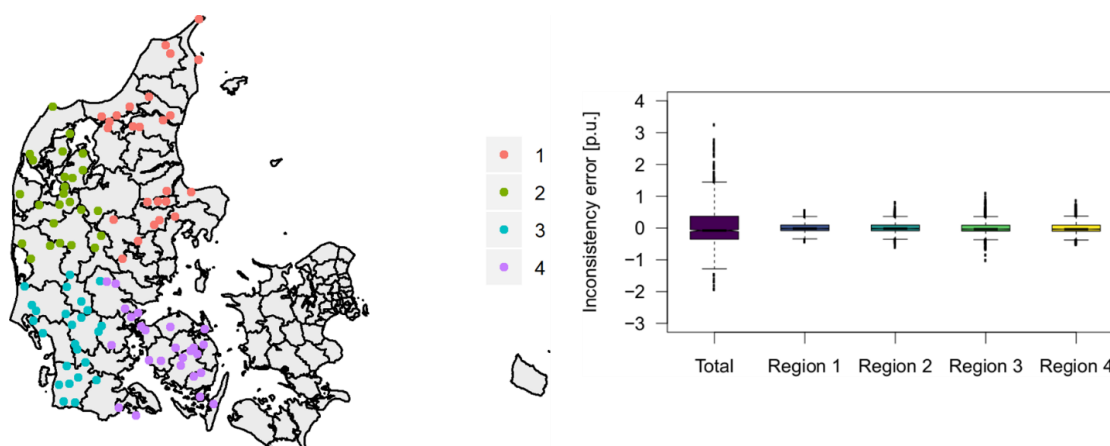


Figure 5 100 Danish sites selected from the complete Danish wind power dataset, divided into 4 hypothetical regions. 15-min wind generation forecasts are generated for each site, each region, and the whole country independently, and the inconsistency error is displayed for the upper levels of the hierarchy over a randomly chosen period of 2 weeks.

### What's ahead?

- The approach may be readily used for multi-step ahead forecasts and day-ahead forecasting
- The forecast reconciliation problem is seen as centralized, but it could be readily distributed using e.g. ADMM and the likes, since consisting of a convex optimization problem.
- Similarly, sparsification was not considered here, while it may be clearly of interest to minimize the number of alterations to forecasts in the reconciliation process
- Other types of models may be thought of in a multivariate regression framework.
- Although the model was restricted to the linear setting, as done in all reconciliation literature, one could generalize it to the nonlinear setting, that would make it possible to account for conditional effects (e.g. from weather conditions and prevailing wind direction) as well as regime-switching, either explicitly or by the use of an adaptive forgetting factor scheme.

[Read the article](#)



### **Forecasting Conditional Extreme Quantiles for Wind Energy**

*Carla Gonçalves, Laura Cavalcante, Margarida Brito, Ricardo J. Bessa, João Gama. Electric Power Systems Research, 2021.*

The growing integration of renewable energy sources (RES) brings new challenges to system operators and market players and robust forecasting models are fundamental for handling their variability and uncertainty. This fomented a growing interest in RES probabilistic forecasting techniques and its integration in decision-aid under risk.

Probabilistic forecasting of distribution tails (i.e., quantiles below .05 and above .95) is challenging for nonparametric approaches since data for extreme events are scarce. A poor forecast of extreme quantiles can have a high impact in various power system decision-aid problems (e.g., reserve capacity setting, dynamic line rating, RES market trading). An alternative approach, more robust to data sparsity, is extreme value theory (EVT), which uses parametric functions for modelling distribution's tails.

In this work, INESC TEC apply conditional EVT estimators for truncated variables to historical data. The parametric function parameters are conditioned by covariates such as wind speed or direction from a numerical weather predictions grid.

Three major benefits are provided by this work:

- Covariates are used to produce conditional forecasts of quantiles without any limitation in the number of variables.
- The parametric EVT-based estimator can be combined with any nonparametric model (artificial neural networks, GBT, random forests, etc.) without any major modification.
- Sharp and calibrated forecasts that avoid over- and under-estimation of risk.

The results for a wind power plant located in Galicia, Spain, show that the proposed method outperforms state-of-the-art methods in terms of quantile score. The proposed method can be transposed to other use cases in the energy sector, such as risk management in portfolio's future returns and study grid resilience to adverse weather events.

#### **What's ahead?**

- Inclusion of information from weather ensembles, as additional covariates, in order to exploit its capability to capture extreme events with a physically-based approach;
- Generalization of the proposed method to other energy-related time series, e.g., solar power and electricity price;
- New proper scoring rules are needed to evaluate the forecasting skill of extreme (rare) events.

[Read the article](#)

## Need for models adequate in high-RES integration scenarios and digitalisation of the energy sector

### Prescriptive Trees for Integrated Forecasting and Optimization Applied in Trading of Renewable Energy

Akylas Stratigakos, Simon Camal, Andrea Michiorri, Georges Kariniotakis. *IEEE Transactions on Power Systems*, 2022.

Almost every aspect of power systems, including real-time operation, investment planning, and electricity markets, is subject to uncertainty stemming from multiple sources, such as stochastic renewable production and unknown market conditions. Facilitated by the influx of data associated with the transition towards the smart grid era, recent years have seen a burgeoning development of energy analytics tools that enable stakeholders to derive decisions in the presence of contextual information. Data-driven decision-making comprises a sequential process with two components, (energy) forecasting and optimization. The first involves forecasting uncertain quantities, e.g., renewable production or market prices, at a future time interval and in a form that ranges from deterministic (point) forecasts to probabilistic forecasts, scenarios (trajectories) and uncertainty sets. Next, these forecasts are used as inputs in an optimization problem to derive optimal decisions (prescriptions).

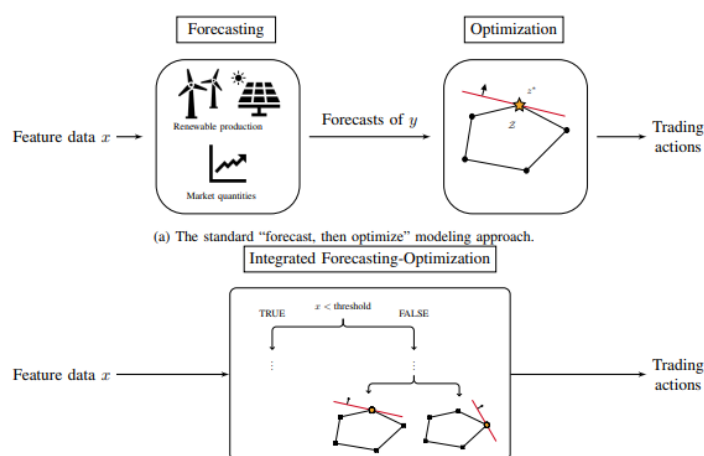


Figure 6 Deriving trading decisions from contextual information. In both cases the objective is uncertain and must be inferred

Forecasting models learn by minimizing a loss function that stands as a proxy for task-specific costs (e.g., trading, scheduling) without considering the downstream optimization, which in practice creates a performance bottleneck and obscures the impact of data on decisions. The paper presented here proposes an integrated forecasting and optimization approach to improve prescriptive performance in renewable trading applications.

To validate this modeling approach, ARMINES consider two case studies related to trading renewable energy. First, the authors examine trading in a day-ahead market and propose strategies that balance optimal trading decisions and predictive accuracy. An average increase in aggregate profit of 3.82% and 0.62% for trading in a day-ahead market under a single- and dual-price balancing mechanism has been observed. In an additional case study that included a generic storage in the dual-price setting, a 3.07% profit increase was shown. The empirical results demonstrate improved prescriptive performance compared to solutions derived under the standard stochastic optimization framework. Valuable insights on how explanatory data impact optimization performance, and how this impact evolves under different market designs, are further provided in this article.

#### What's ahead?

Future work could focus on learning in an adaptive (online) setting and enhancing model interpretability.

[Read the article](#)

## Lack of open data and of price-based incentives to share data

### Towards Data Markets in Renewable Energy Forecasting

Carla Gonçalves, Pierre Pinson, Ricardo Bessa. *IEEE Transactions on Sustainable Energy*, 2021

Large amount of data is being collected from geographically distributed renewable energy sources such as wind turbines and photovoltaic panels and sensors (e.g., pyranometers). These data include power generation and weather measurements like air temperature, wind speed and direction, irradiation, etc. These time-series data can improve renewable energy sources forecasting skill for different time horizons (e.g. hours-ahead, day-ahead) and consequently the revenue from electricity market players. However, data owners/RES agents may be unwilling to share their data, even if privacy is ensured, due to a form of prisoner's dilemma: all could benefit from data sharing, but in practice no one is willing to do because they are most likely competitors in the same electricity market. Smart4RES proposal hence consists of a data marketplace, to incentivize collaboration between different data owners through the monetization of data.

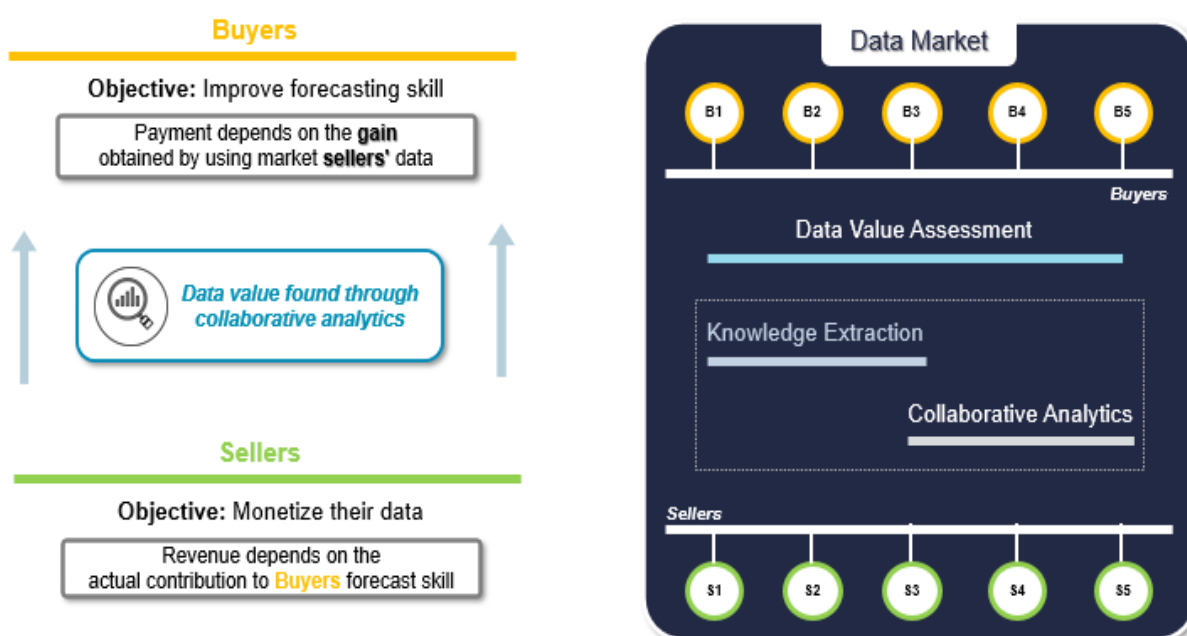


Figure 7 Data market concept  
© INESC

In this article, it is proposed to consider a marketplace where RES agents sell data (historical power production, NWP, etc), purchase forecasts and pay according to resulting forecasting accuracy. This avoids the confidentiality problem of sharing raw data directly. Cooperation between sellers is done through a market operator who receives all agents data and prepares forecast. To do so, the authors proposed to adapt an existing auction mechanism to the case of RES forecasting data. The algorithmic approach accommodates the temporal nature of the data, i.e., lagged time-series act as covariates and models are updated continuously using a sliding window.

Synthetic data was used in a controlled case study where it was possible to confirm the correct allocation of revenue across sellers by the market operator, as well as the fact that buyers who did not benefit from the forecasts of others did not pay for such forecasts. A test case with Nordpool wind energy data is presented to illustrate and assess the effectiveness of such data

markets. All agents (or data owners) are shown to benefit in terms of higher revenue resulting from the combination of electricity and data markets. The results support the idea that data markets can be a viable solution to promote data exchange between RES agents and contribute to reducing system imbalance costs.

### *What's ahead?*

- The loss of RES agents when sharing their data should be considered when defining the data price. A seller sharing data with its competitors expects a compensation for the potential impact on its business.
- Some improvements are required when using a sliding-window approach to reduce the computational effort, allowing the market operator to update the model weights through online learning whenever the variables in the data market remain the same.
- Address data-privacy issues.
- Develop peer-to-peer data trading schemes (i.e., without a central node as market operator) for prosumers in local energy communities, in such a way that data sellers can set their own data price.

[Read the article](#)

### ***A critical overview of privacy-preserving approaches for collaborative forecasting***

*Carla Gonçalves, Ricardo J. Bessa, Pierre Pinson. International Journal of Forecasting, 2021.*

In the same vein than the precedent article, the present article focuses in data privacy-preserving methods. Indeed, due to personal data protection questions, RES agents/data owners might be unwilling to share their data, which increases the interest in collaborative privacy-preserving forecasting.

The authors conducted a review of the state-of-the-art in statistical methods for collaborative forecasting with privacy-preserving approaches. This work is not restricted to a simple overview of the existing methods, and it performs a critical evaluation of said methods, from a mathematical and numerical point of view, namely when applied to the Vector AutoRegressive (VAR) model. The VAR model has been widely used to forecast variables that may have different data owners and is appropriate to update very short-term forecasts (e.g. from 15 minutes to 6 hours ahead) with recent data but also in short-term electricity price forecasting.

With this paper, the authors intend to show gaps and downsides of current methods and to present insights for further improvements towards fully privacy-preserving VAR forecasting methods.

The existing state-of-the-art privacy-preserving techniques are divided into three groups: data transformation, secure multi-party computations, and decomposition methods. The analysis shows that state-of-the-art techniques have limitations in preserving data privacy:

- Data transformation requires a trade-off between privacy and accuracy
- Secure multi-party computations either result in computationally demanding techniques or do not fully preserve privacy in VAR models
- Decomposition-based methods rely on iterative processes and after a number of iterations, the agents have enough information to recover private data.

[Read the article](#)



## Privacy-preserving Distributed Learning for Renewable Energy Forecasting

Carla Gonçalves, Ricardo Bessa, Pierre Pinson. *IEEE Transactions on Sustainable Energy*, 2021.

To tackle the privacy issue and methods limitations mentioned earlier, this paper formulates a novel privacy-preserving framework that combines data transformation techniques with the alternating direction method of multipliers (ADMM). The method proposed focus on privacy-preserving protocols for very short-term forecasting with the Vector Autoregressive (VAR)

model. This approach allows not only to estimate the model in a distributed fashion but also to protect data privacy, coefficients and covariance matrix. Besides, asynchronous communication between peers is addressed in the model fitting, making it possible to update the forecast model based on information from a subset of agents and improve the computational efficiency of the proposed model. The mathematical formulation is flexible enough to be applied in two different collaboration schemes: centralized communication with a neutral node and peer-to-peer communication, and in a way that original data cannot be recovered by central node or peers.

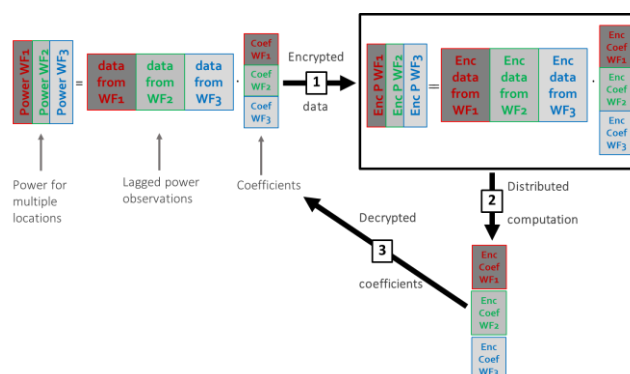


Figure 8 VARn(p) model with privacy  
© INESC

centralized communication with a neutral node and peer-to-peer communication, and in a way that original data cannot be recovered by central node or peers.

The results for a solar energy dataset show that the proposed method is robust to privacy breaches and communication failures and delivers a forecast skill comparable to a model without privacy protection and outperformed a state-of-the-art method based on analog search.

### What's ahead?

- Investigate uncertainty forecasting and application to non-linear models (and consequently longer lead times): the proposed privacy-preserving protocol can be applied to non-linear regression by extending the additive model structure to a multivariate setting or by local linear smoothing.
- The extension to other non-linear multivariate models, such as long short-term memory networks and variants, requires further researcher and significant changes in the protocol.

[Read the article](#)

## Regression markets and application to energy forecasting

Pierre Pinson, Liyang Han, Jalal Kazempour. Accepted in TOP (under edition)

For many operational problems of the energy system, it is assumed that data can be shared and centralized for the purpose of solving the analytics task at hand. In other words, various agents collect and own data that may be useful to others. However, in practice, it is rarely the case that the agents are willing to freely share their data.

In contrast to recent proposals that look into distributed and privacy-preserving learning (incentive-free), DTU explores a framework called regression markets. There, agents aiming to improve their forecasts post a regression task (that is widely used as a basis for energy forecasting) for which other agents may contribute by sharing their data for their features and get monetarily rewarded for it. Such regression markets rely on recent concepts within interpretability of machine learning approaches and cooperative game theory. In this generic regression market framework, willingness to pay for the buyer and willingness to sell for the sellers are considered.

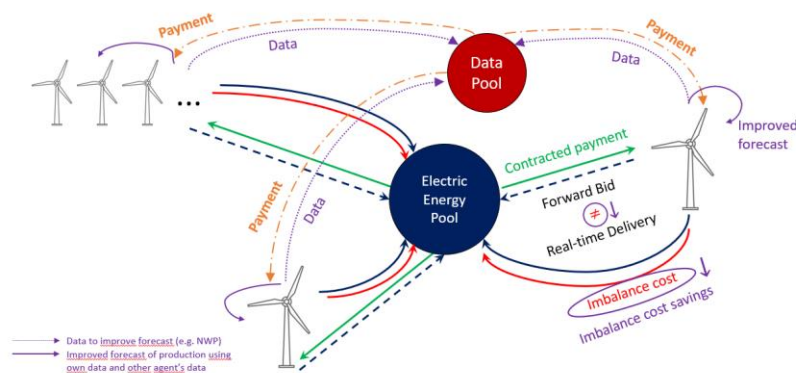


Figure 9 RES wholesale market: many agents in a centralised model  
© DTU

In this article, the agents and preliminaries regarding regression tasks are described before introducing the regression market mechanisms proposed by DTU. The overall concept is presented for both batch and online setups, with a description of feature valuation and allocation policies. The extension to the out-of-sample regression and forecasting case is also covered. The properties of the regression market mechanisms are finally presented and proven, based on a set of simulation studies for a broad range of models and cases. Application to real-world forecasting case-studies, with both mean and quantile forecasting problems are discussed at the end.

Overall, the proposed regression market can be applied to not only linear models, but also models of higher order and quantile regression. This framework is compatible with online learning without violating the market properties.

### What's ahead?

- The concepts and key elements of the approach should be extended to the case of other analytics tasks, e.g. classification, filtering, etc., and to the nonlinear case
- The properties of the various regression markets may be further studied, for instance in a regret analysis framework, to provide some interesting bounds and potential fairness implications.

[Read the article](#)

**To get more insights, DTU invites you to read the following conference papers:**

- Liyang Han, Jalal Kazempour, Pierre Pinson. Monetizing Customer Load Data for an Energy Retailer: A Cooperative Game Approach. 2021 IEEE Madrid PowerTech, IEEE, Jun 2021, Madrid, Spain. eprint 2012.05519.

[Download the e-print in arXiv](#)

- Amandine Pierrot and Pierre Pinson. Adaptive Generalized Logit-Normal Distributions for Wind Power Short-Term Forecasting. 2021 IEEE Madrid PowerTech, IEEE, Jun 2021, Madrid, Spain. eprint 2012.08910.

[Download the e-print in arXiv](#)

In this article an approach is proposed to improve both **point and probabilistic forecasts**. It introduces batch and online learning as well.



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