Smart4RES collaborative analytics for renewable energy forecasting: federated learning and data markets

Ricardo Bessa (ricardo.j.bessa@inesctec.pt) | INESC TEC

IEA TASK 51 Workshop "State of the Art and Research Gaps"

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

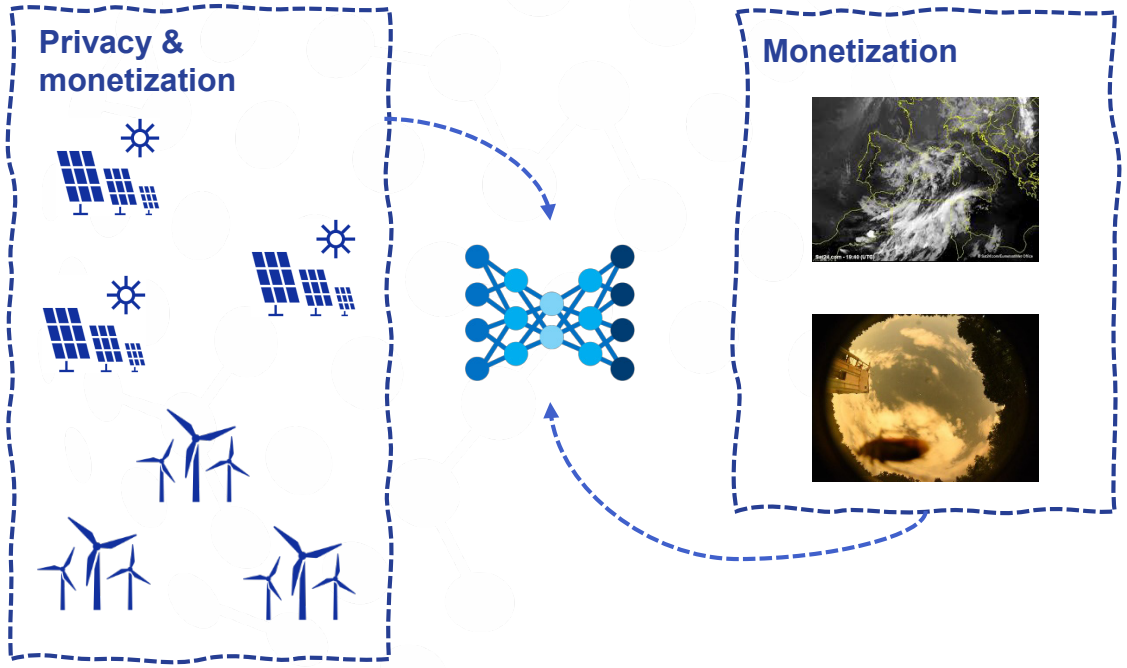
Motivation

Increasing volume of geographically distributed data

Main Barriers

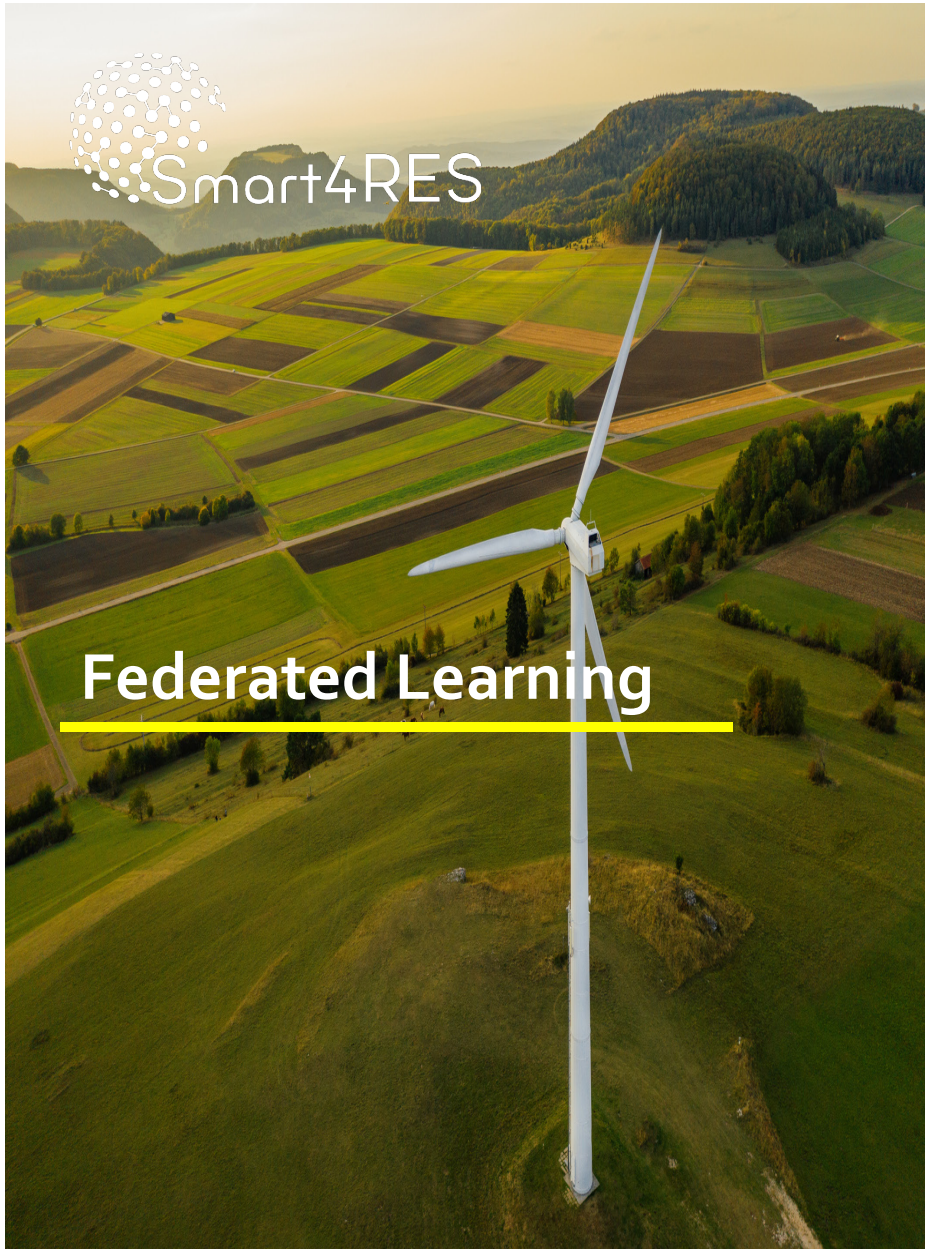
- Data **privacy** and **confidentiality**
- Lack of **monetary and non-monetary incentives** for sharing data
- Lack of **business cases** for collaborative analytics

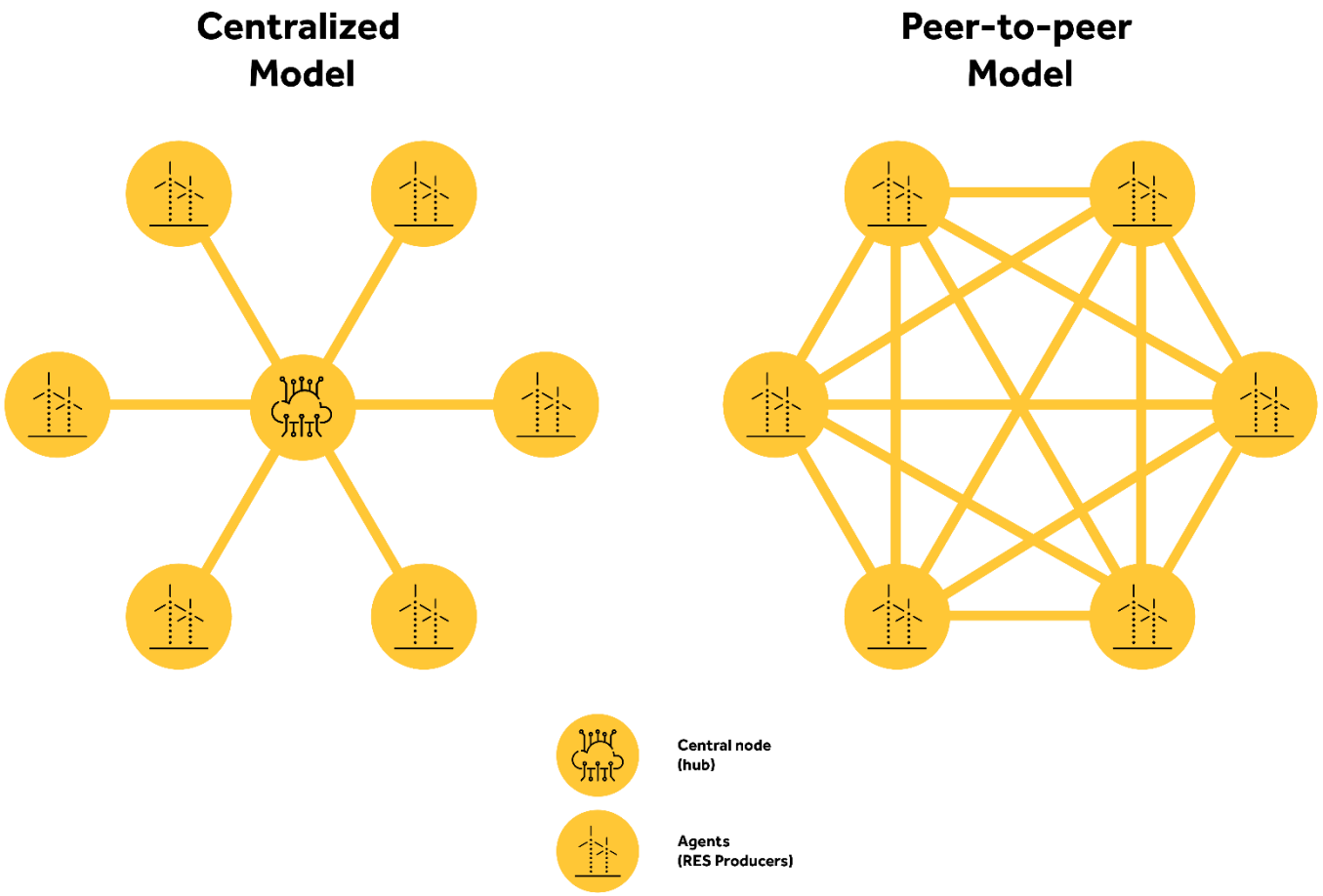
Collaborative analytics & RES forecasting



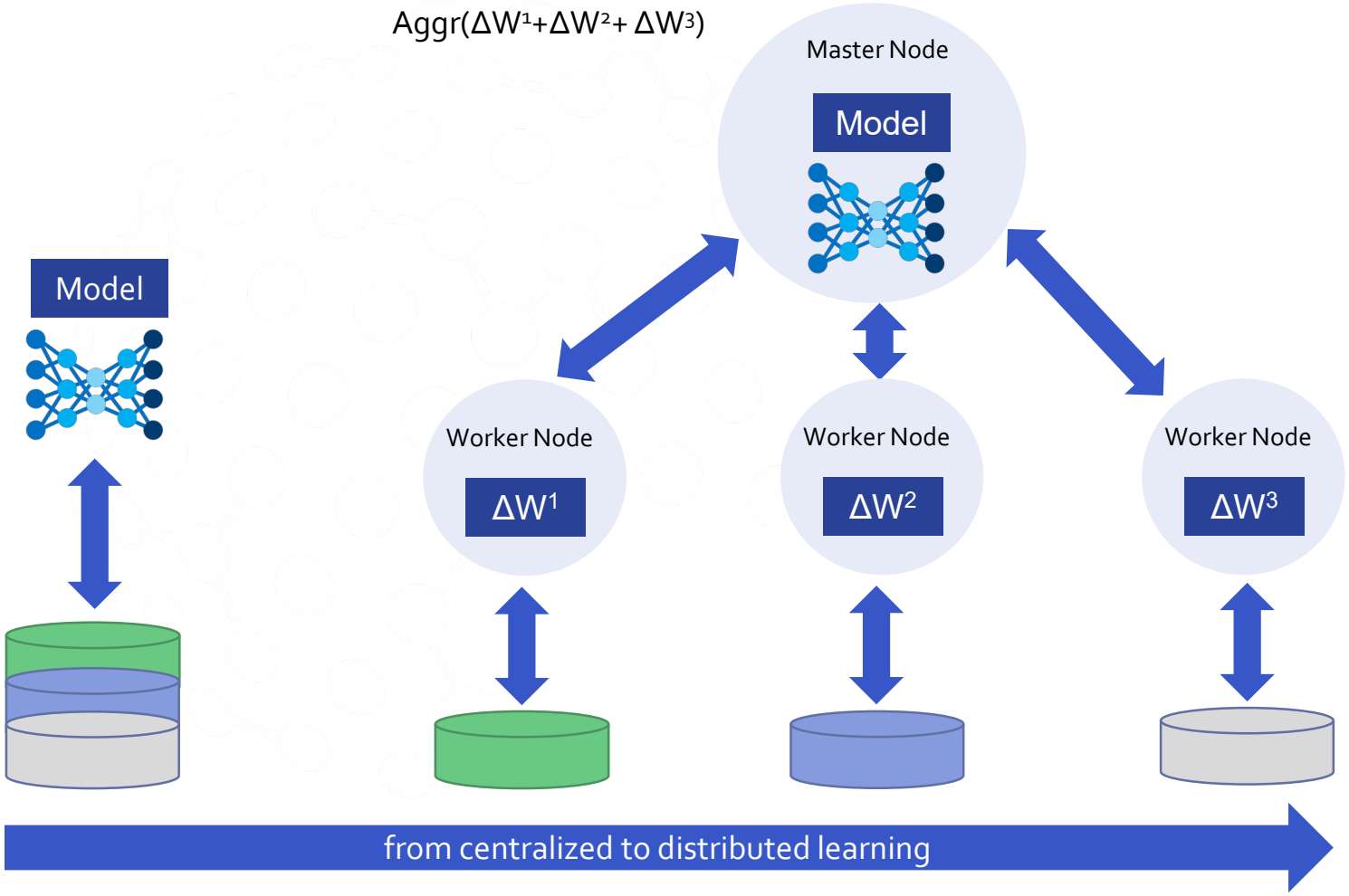
- **Federated learning models with data from different owners / data sources**
  - Privacy-preserving protocols
  - Different data exchange schemes

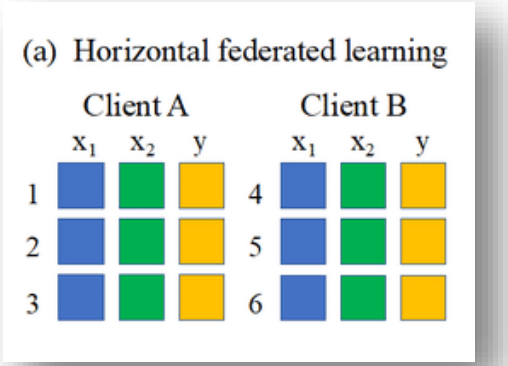
- **Algorithmic solutions for data markets**: data price as a function of use case specific value

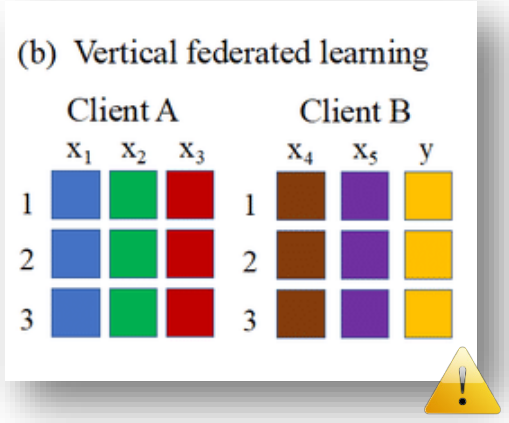

Smart4RES
Federated Learning


Centralized Model
Peer-to-peer Model
Central node (hub)
Agents (RES Producers)

Federated learning concept



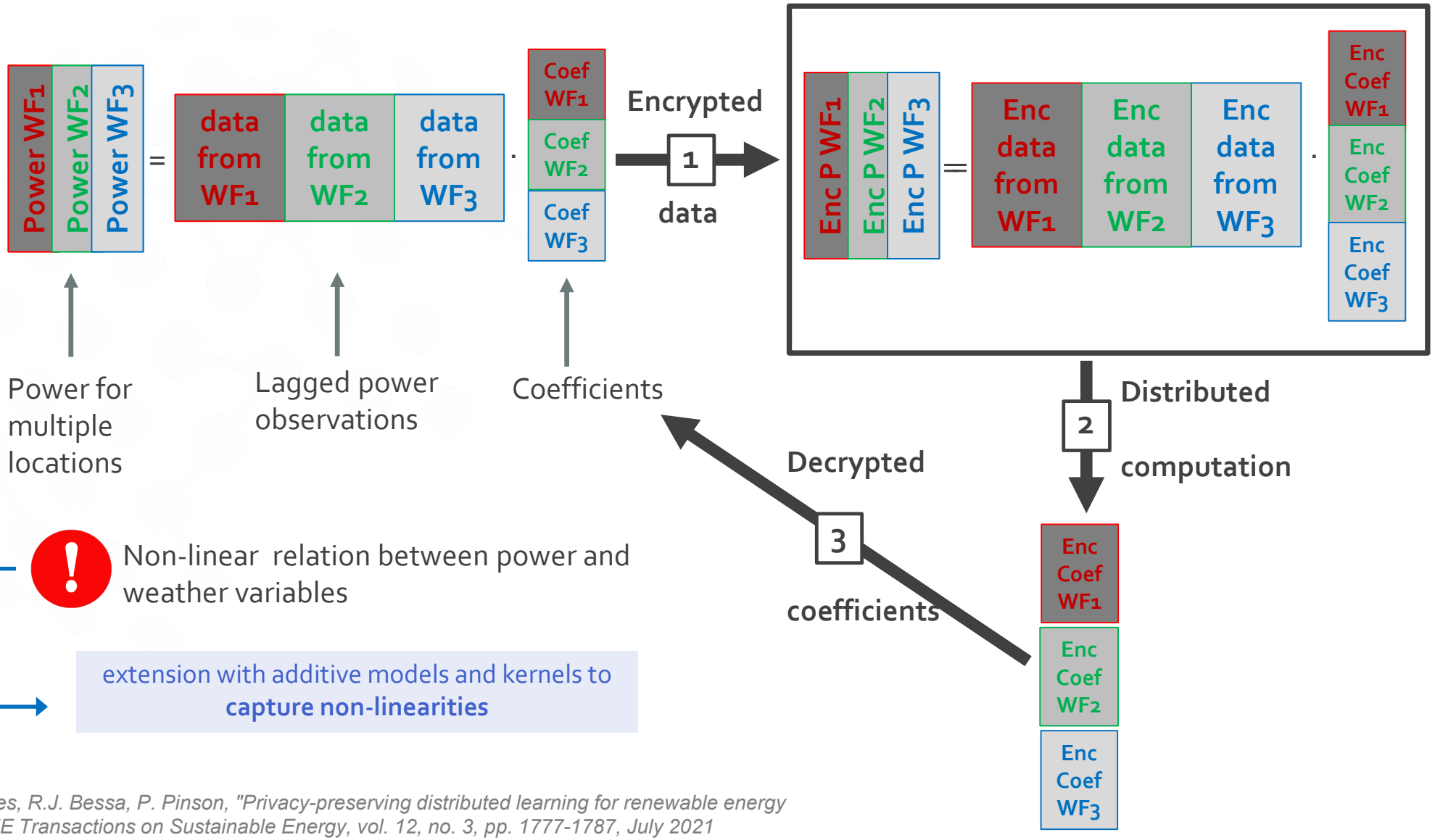




Focus of this presentation

Source: Chen, S., et al. (2021). FL-QSAR: a federated learning-based QSAR prototype for collaborative drug discovery. Bioinformatics, 36(22-23), 5492-5498.

Privacy-preserving federated learning: RES forecasting



Non-linear relation between power and weather variables

extension with additive models and kernels to **capture non-linearities**

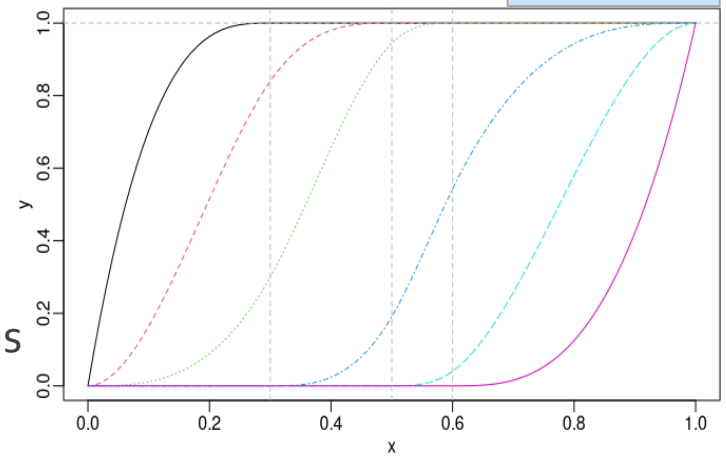
Ref: C. Gonçalves, R.J. Bessa, P. Pinson, "Privacy-preserving distributed learning for renewable energy forecasting," IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 1777-1787, July 2021

Privacy-preserving federated learning: RES forecasting

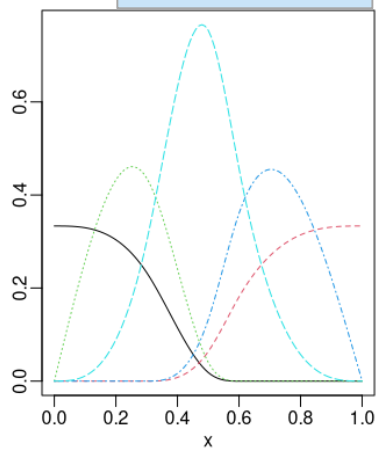
each variable is transformed in k variables

$$X_1 \rightarrow f_1(X_1), f_2(X_1), \dots, f_k(X_1)$$

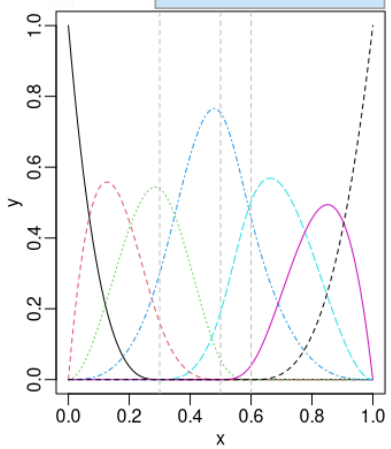
I-splines



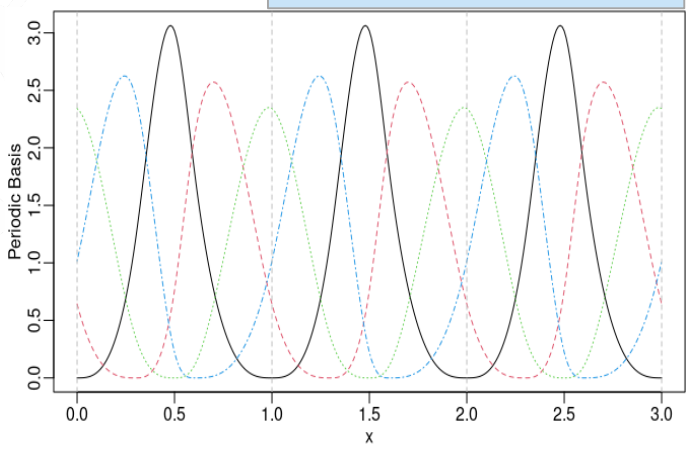
Nc-splines



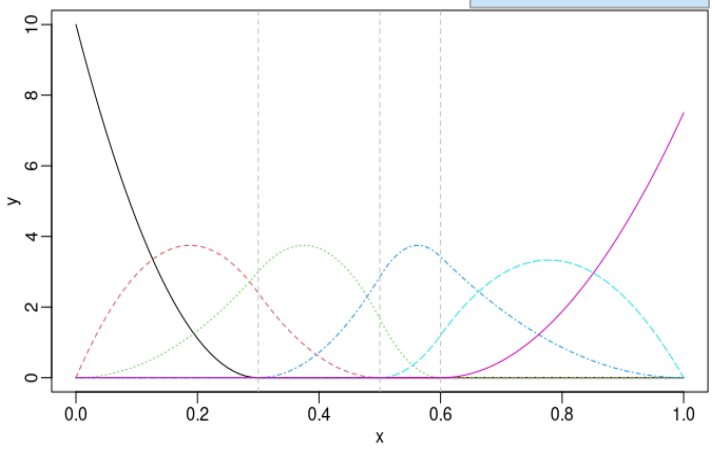
B-splines



periodic M-splines



M-splines



Supported by our privacy-preserving protocol

Privacy-preserving federated learning: RES forecasting

Each observation is transformed to the distance relative to the others

$$y|\mathbf{x} = \sum_t K(\|\mathbf{x} - \mathbf{x}_t\|)\, B_t \quad \rightarrow \quad \hat{B} = \mathbf{K}^{-1} y$$

Negative Distance Kernel (NDK)

$$K(\|\mathbf{x} - \mathbf{x}_t\|) = -\|\mathbf{x} - \mathbf{x}_t\|^2$$

$$\begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ x_{1,3} & x_{2,3} \end{pmatrix} \rightarrow \mathbf{K} = \begin{pmatrix} \cancel{\exp}(-[(x_{1,1}-x_{1,1})^2 + (x_{2,1}-x_{2,1})^2]/(\cancel{2\sigma})) & \exp(-[(x_{1,1}-x_{1,2})^2 + (x_{2,1}-x_{2,2})^2]/(2\sigma)) & \\ \exp(-[(x_{1,2}-x_{1,1})^2 + (x_{2,2}-x_{2,1})^2]/(2\sigma)) & \exp(-[(x_{1,2}-x_{1,2})^2 + (x_{2,2}-x_{2,2})^2]/(2\sigma)) & \dots \\ \exp(-[(x_{1,3}-x_{1,1})^2 + (x_{2,3}-x_{2,1})^2]/(2\sigma)) & \exp(-[(x_{1,3}-x_{1,3})^2 + (x_{2,3}-x_{2,3})^2]/(2\sigma)) & \end{pmatrix}$$

$$K = K_{A1} + K_{A2}$$

K_{A1} Encrypted
K_{A2}
Y_{A1}
Y_{A2} data

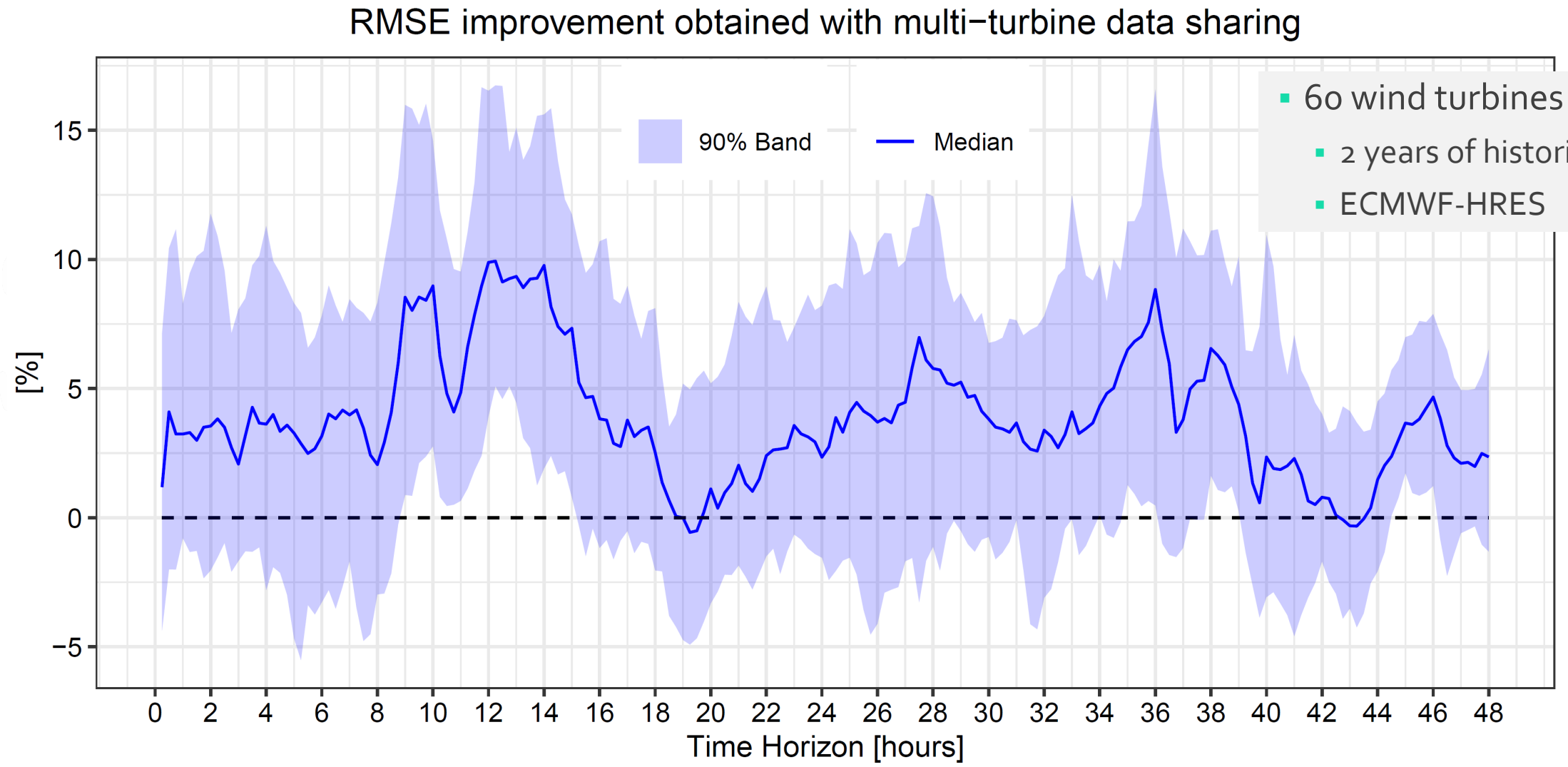
$\rightarrow$ $\mathbf{M}K_{A1}$, $\mathbf{M}K_{A2}$, $\mathbf{M}Y_{A1}$, $\mathbf{M}Y_{A2}$ $\rightarrow$

$$\hat{B} = [\mathbf{M}(K_{A1} + K_{A2})]^{-1}[\mathbf{M}Y_{A1}, \mathbf{M}Y_{A2}]$$

Supported by our privacy-preserving protocol

$$\mathbf{M}K_{A1} + \mathbf{M}K_{A2} = \mathbf{M}(K_{A1} + K_{A2})$$

Numerical results: Improvement w.r.t. single owner data



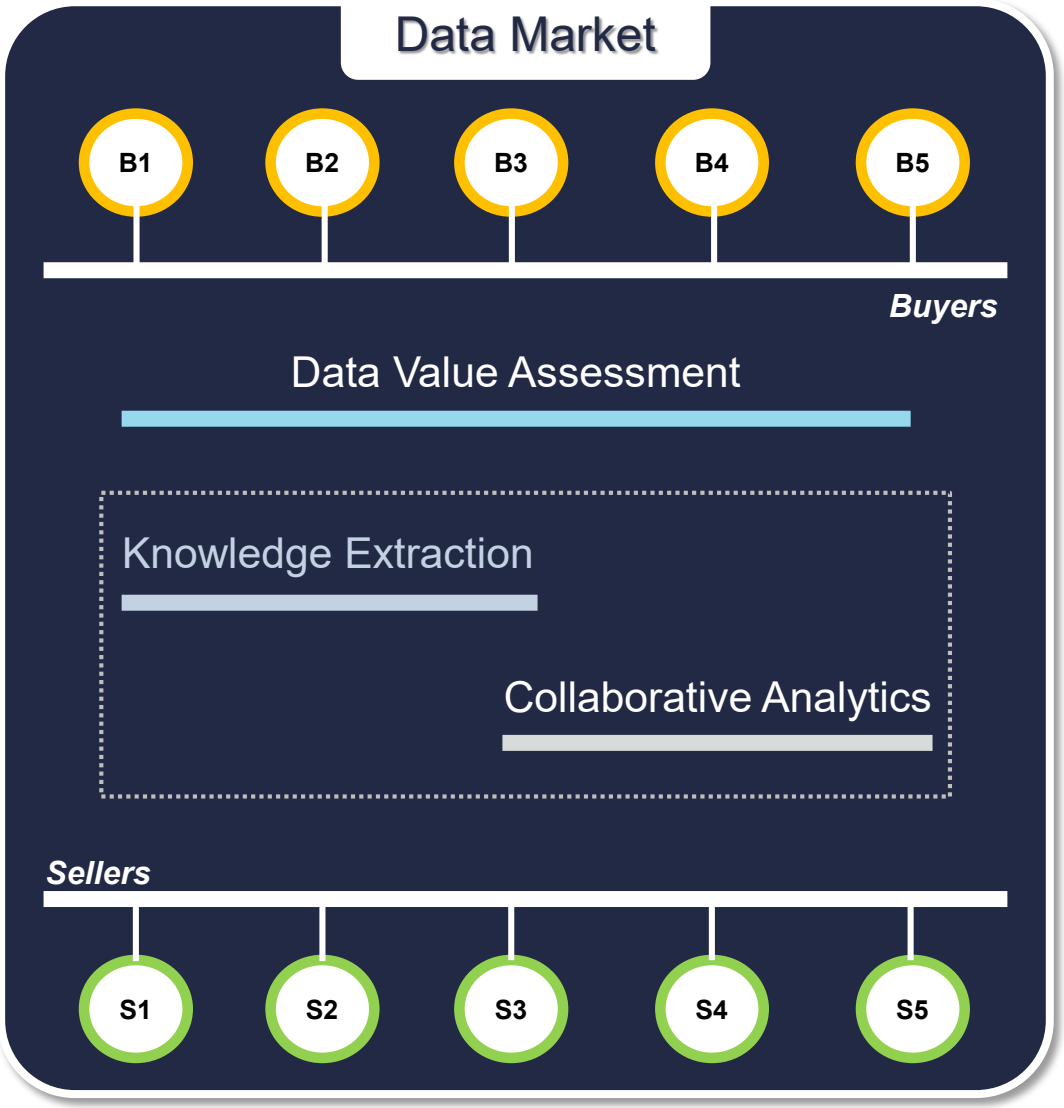
- 60 wind turbines
  - 2 years of historical data
  - ECMWF-HRES

Numerical results: Splines and kernel performance

ms-splines over Gradient Boosting Trees (no privacy)

NDK over Gradient Boosting Trees (no privacy)


Smart4RES
Data Markets


Data Market
B1
B2
B3
B4
B5
Buyers
Data Value Assessment
Knowledge Extraction
Collaborative Analytics
Sellers
S1
S2
S3
S4
S5

Data market concept for RES forecasting

Buyers

Objective: Improve forecasting skill

Payment depends on the **gain** obtained by using market **sellers'** data

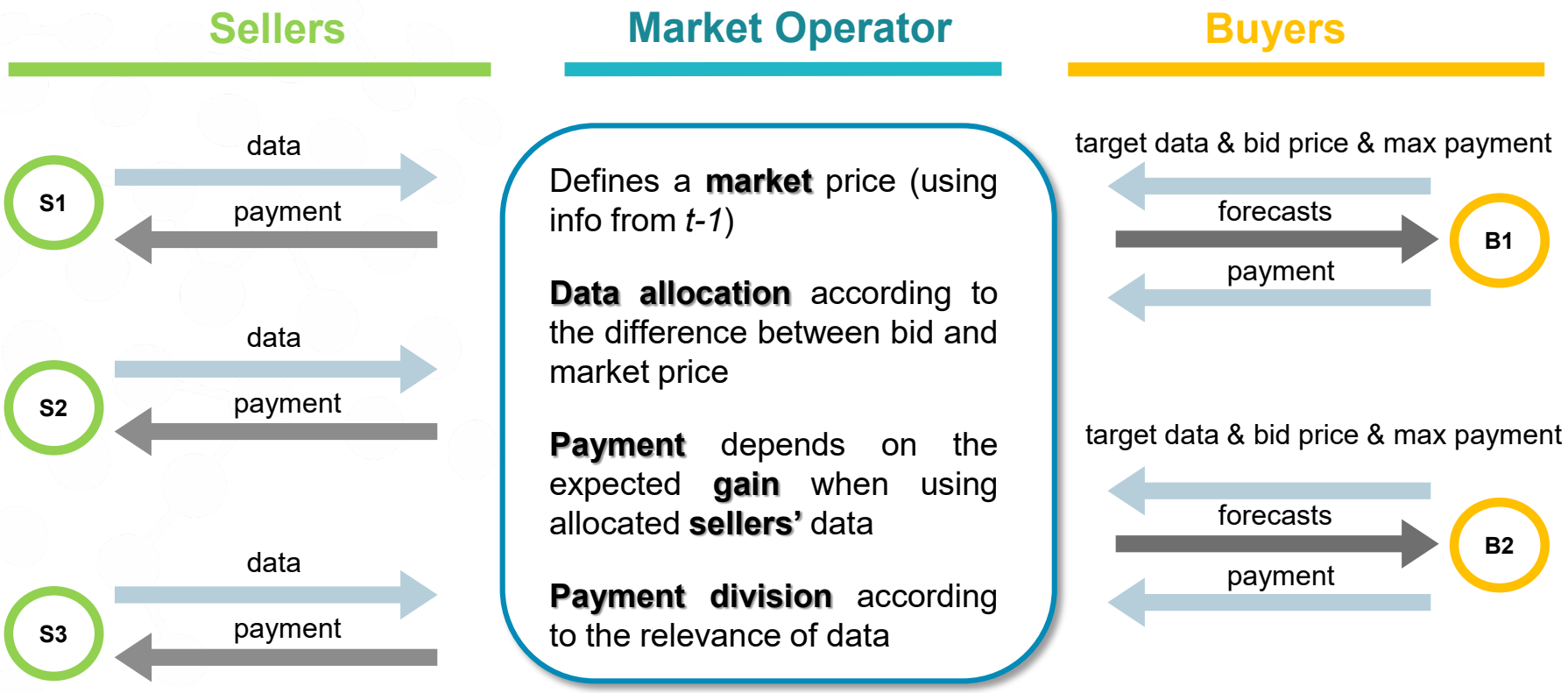
Data value found through collaborative analytics

Sellers

Objective: Monetize their data

Revenue depends on the actual contribution to **Buyers** forecast skill

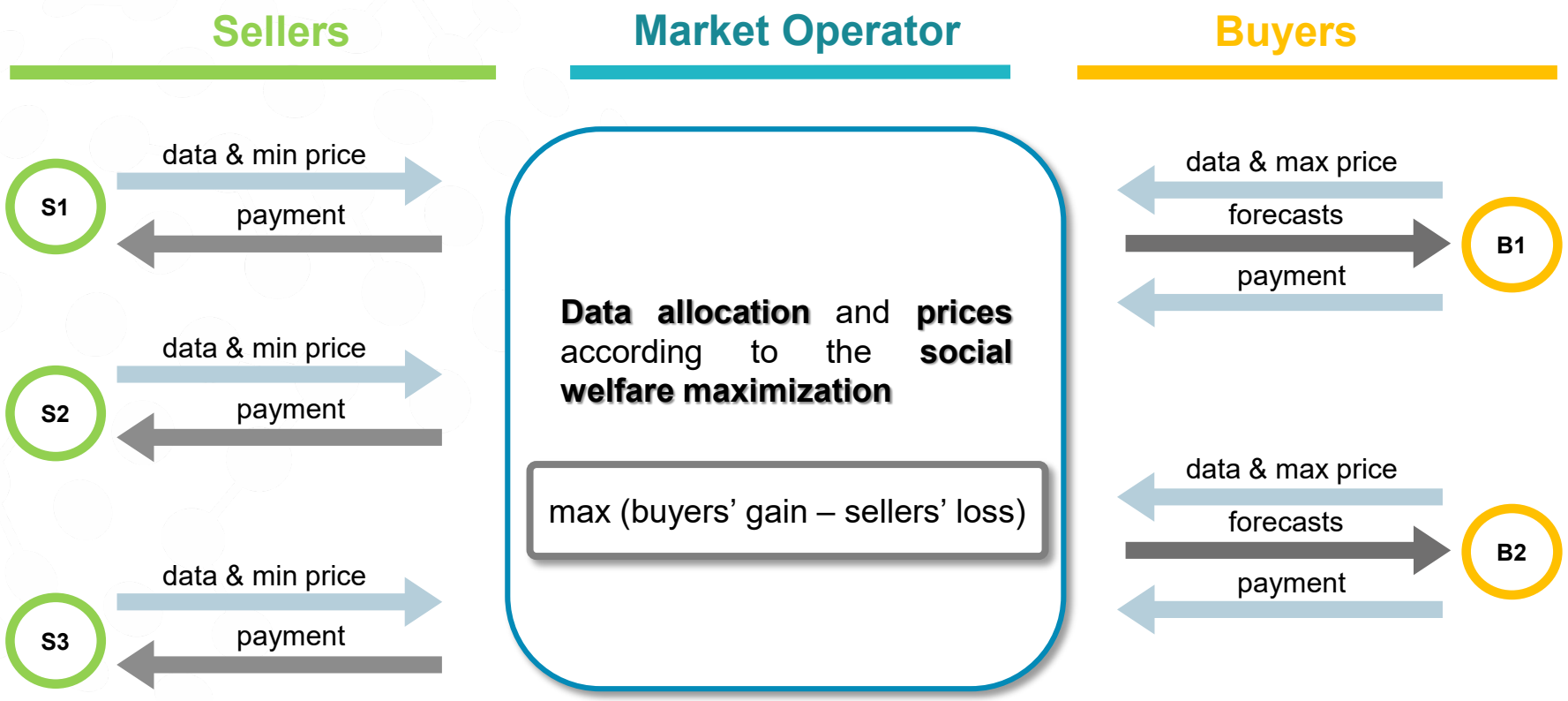
Data market: No-regret mechanism



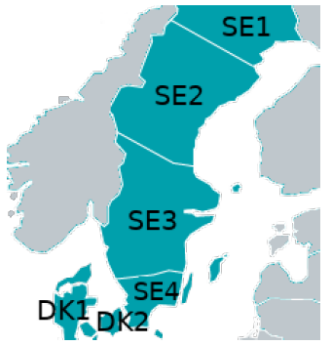
Sellers' loss when sharing their data is not considered

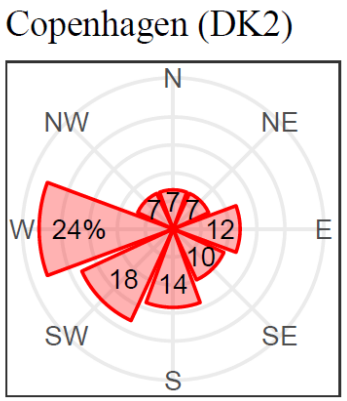
Goncalves, C., Pinson, P., Bessa, R. J. (2021). Towards data markets in renewable energy forecasting. IEEE Transactions on Sustainable Energy, 12(1), 533-542

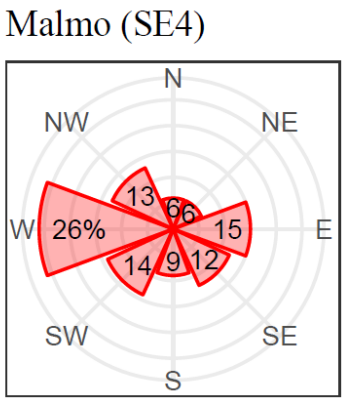
Data market: Social welfare maximization



Numerical results



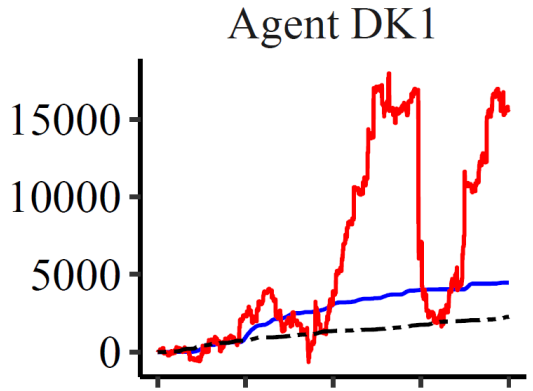




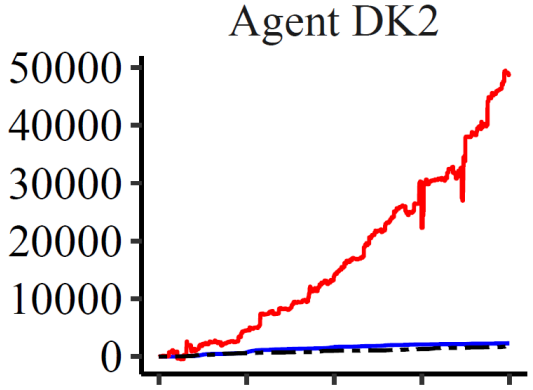
Results in: *Goncalves, C., Pinson, P., Bessa, R. J. (2021). Towards data markets in renewable energy forecasting. IEEE Transactions on Sustainable Energy, 12(1), 533-542*

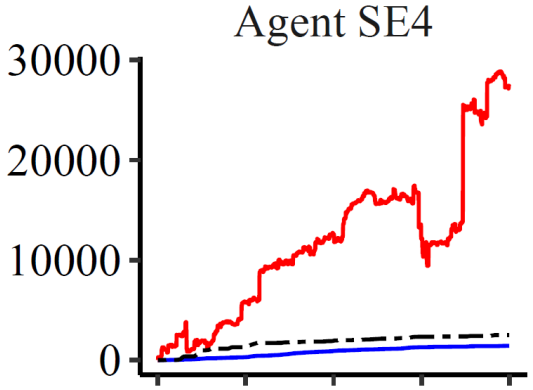
Cumulative Data Market Revenue — Cumulative Extra Revenue from Electricity Market — Cumulative Payment

Agent with highest reward for data sharing

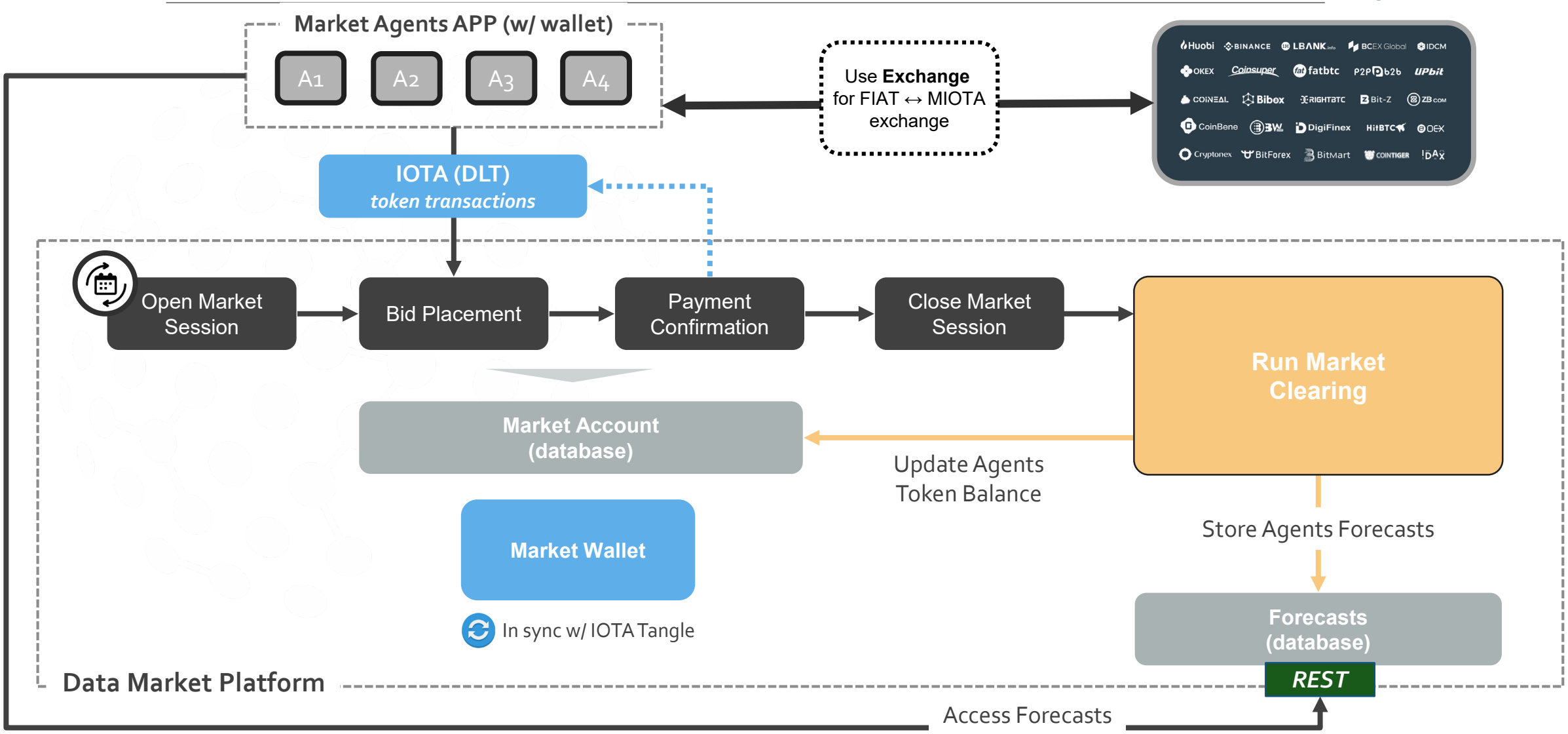


Agents that maximize electricity market revenue due to “extra” data





Data market IOTA prototype



Smart4RES data markets portfolio

IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, VOL. 12, NO. 1, JANUARY 2021 533

Towards Data Markets in Renewable Energy Forecasting

Carla Gonçalves, Pierre Pinson, *Fellow, IEEE*, and Ricardo J. Bessa, *Senior Member, IEEE*

***Abstract*—Geographically distributed wind turbines, photovoltaic panels and sensors (e.g., pyranometers) produce large volumes of data that can be used to improve renewable energy sources (RES) forecasting skill. However, data owners 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 do. Our proposal hence consists of a data marketplace, to incentivize collaboration between different data owners through the monetization of data. We adapt here an existing auction mechanism to the case of RES forecasting data. It accommodates the temporal nature of the data, i.e., lagged time-series act as covariates and models are updated continuously using a sliding window. A test case with 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.**

***Index Terms*—Collaborative forecasting, data marketplace, data pricing, renewable energy, electricity market.**

α_t^*	Nominal level which minimizes $C_t^{\uparrow/\downarrow}$.
$F_{i,t}^{-1}(\alpha_t^*)$	Forecasted conditional quantile for nominal level α_t^*.
$\hat{\psi}_t^{\uparrow}, \hat{\psi}_t^{\downarrow}$	Forecasted upward / downward regulation price.
$\hat{p}_t^{\uparrow}, \hat{p}_t^{\downarrow}$	Probability of up/downward regulation at time t.
N	Number of RES power agents.
$\mathcal{A}$	Overall set of power plants, $\mathcal{A} = \{1, \ldots, N\}$.
T	Number of historical records.
H	Length of the time horizon.
$x_{i,t}$	Power measurements for RES agent i at time t.
$\hat{q}_{\alpha_t^*}^i$	Forecasted quantile α_t^* for site $i \in \mathcal{A}$ at time t.
$\hat{\beta}_{\cdot}^{\cdot}$	Linear quantile regression coefficients.
$\mathbf{x}_i^{\mathrm{S}}$ (or $\mathbf{x}_i^{\mathrm{B}}$)	Data from seller (or buyer) i.
$\mathbf{X}^{\mathrm{S}}$	Data from all sellers, $\mathbf{X}^{\mathrm{S}} = [\mathbf{x}_1^{\mathrm{S}}, \ldots, \mathbf{x}_N^{\mathrm{S}}] \in \mathbb{R}^{T \times N}$.
$\mathcal{M}_i$	Forecasting model for power production of agent i.
$\mathcal{G}_i$	Gain function for buyer i.
μ_i	Private valuation for each unit gain.
b_i	Public bid price (buyer i is willing to pay $b_i < \mu_i$).

Monetizing Customer Load Data for an Energy Retailer: A Cooperative Game Approach

Liyang Han*, Jalal Kazempour*, Pierre Pinson†
*Department of Electrical Engineering, †Department of Technology, Management and Economics
Technical University of Denmark
Kongens Lyngby, Denmark
liyha@elektro.dtu.dk, seykaz@elektro.dtu.dk, ppin@dtu.dk

***Abstract*—When energy customers schedule loads ahead of time, this information, if acquired by their energy retailer, can improve the retailer's load forecasts. Better forecasts lead to wholesale purchase decisions that are likely to result in lower energy imbalance costs, and thus higher profits for the retailer. Therefore, this paper monetizes the value of the customer schedulable load data by quantifying the retailer's profit gain from adjusting the wholesale purchase based on such data. Using a cooperative game theoretic approach, the retailer translates their increased profit in expectation into the value of cooperation, and redistributes a portion of it among the customers as monetary incentives for them to continue providing their load data. Through case studies, this paper demonstrates the significance of the additional profit for the retailer from using the proposed framework, and evaluates the long-term monetary benefits to the customers based on different payoff allocation methods.**

***Index Terms*—data monetization, energy market, cooperative game theory, newsvendor model, probabilistic forecasting**

generation [4] and controllability from smart appliances and energy storage systems are fueling the energy customers' interest in becoming active participants in the market operation [5]. Recent literature on this topic heavily focuses on the creation of distributed energy markets, which generally allow customers to share or trade energy among themselves [6]. Recognizing that the way customers control and schedule their energy resources holds important information that can help energy retailers to improve their demand forecast, we identify customer load data as another commodity that creates value, and thus has the potential to be monetized in the energy market.

A common method of collecting energy customer load data is through the use of smart meters. Privacy is a major concern despite research effort in developing privacy-preserving techniques to process customer data [7]. The lack of fairness

Electric Power Systems Research 212 (2022) 108442

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Trading data for wind power forecasting: A regression market with lasso regularization

Liyang Han[a,*], Pierre Pinson[b], Jalal Kazempour[a,*]
[a] Department of Wind and Energy Systems, Technical University of Denmark, Kgs. Lyngby, Denmark
[b] Department of Technology, Management and Economics, Technical University of Denmark, Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
Data market
Lasso
Linear regression
Time series
Wind power forecasting

ABSTRACT

This paper proposes a regression market for wind agents to monetize data traded among themselves for wind power forecasting. Existing literature on data markets often treats data disclosure as a binary choice or modulates the data quality based on the mismatch between the offer and bid prices. As a result, the market disadvantages either the data sellers due to the overestimation of their willingness to disclose data, or the data buyers due to the lack of useful data being provided. Our proposed regression market determines the data payment based on the least absolute shrinkage and selection operator (lasso), which not only provides the data buyer with a means for selecting useful features, but also enables each data seller to individualize the threshold for data payment. Using both synthetic data and real-world wind data, the case studies demonstrate a reduction in the overall losses for wind agents who buy data, as well as additional financial benefits to those who sell data.

TOP
https://doi.org/10.1007/s11750-022-00631-7

ORIGINAL PAPER

Regression markets and application to energy forecasting

Pierre Pinson[1] · Liyang Han[2] · Jalal Kazempour[2]

Received: 7 October 2021 / Accepted: 7 March 2022

Abstract
Energy forecasting has attracted enormous attention over the last few decades, with novel proposals related to the use of heterogeneous data sources, probabilistic forecasting, online learning, etc. A key aspect that emerged is that learning and forecasting may highly benefit from distributed data, though not only in the geographical sense. That is, various agents collect and own data that may be useful to others. In contrast to recent proposals that look into distributed and privacy-preserving learning (incentive-free), we explore here a framework called regression markets. There, agents aiming to improve their forecasts post a regression task, for which other agents may contribute by sharing their data for their features and get monetarily rewarded for it. The market design is for regression models that are linear in their parameters, and possibly separable, with estimation performed based on either batch or online learning. Both in-sample and out-of-sample aspects are considered, with markets for fitting models in-sample, and then for improving genuine forecasts out-of-sample. Such regression markets rely on recent concepts within interpretability of machine learning approaches and cooperative game theory, with Shapley additive explanations. Besides introducing the market design and proving its desirable properties, application results are shown based on simulation studies (to highlight the salient features of the proposal) and with real-world case studies.

Keywords Energy forecasting · Data markets · Mechanism design · Regression · Estimation

Mathematics Subject Classifications 62F99 · 62J99 · 68T05 · 91B26 · 62M20

Smart4RES

Concluding

Acknowledgements:

Carla Gonçalves, INESC TEC

Pierre Pinson, Imperial College London

- Advantage: with federated learning & spatial-temporal models **we only have 1 model to maintain**
- **Value of spatial-temporal information** (power, NWP) is a **proven result**
- **Different mechanisms** are possible **for data markets**
  - Challenge: **data value for the seller**
  - To reach full potential, we need to **go beyond the forecasting use cases**
- **Data sharing** in the RES sector **aligned with EU initiatives** like GAIA-X and Common Data Spaces