# A conditional and regularized approach for large-scale spatiotemporal wind power forecasting

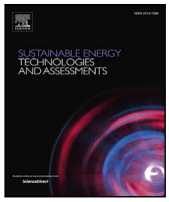
Simon Camal, Robin Girard, Maxime Fortin, Augustin Touron, Laurent Dubus

Sustainable Energy Technologies and Assessments 65 (2024) 103743

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

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



# A conditional and regularized approach for large-scale spatiotemporal wind power forecasting

Simon Camal [a,*], Robin Girard [a], Maxime Fortin [b], Augustin Touron [b], Laurent Dubus [b]

[a] MINES Paris - PSL University, Research Centre PERSEE, Sophia Antipolis, France
[b] RTE - PES/DRD/DS, Paris, France

ARTICLE INFO

*Keywords:*
Wind power forecasting
High dimension
Lasso
Interpretability
Transmission System Operator

ABSTRACT

Large-scale wind power forecasting at intraday horizons of minutes-ahead up to hours-ahead is essential to secure important operations in transmission systems. It is clear that recent information collected about neighboring sites improve the predictive performance of autoregressive models. At the scale of a region or of a country, regularization or feature selection are needed to mitigate the high dimensionality of the autoregressive model. Unconditional approaches of regularization have shown limited added value compared to benchmark models in the context of wind power forecasting. This work proposes an intraday wind power forecasting method that predicts the production of any wind farm in the control area of a Transmission System Operator (TSO), taking into account the information collected from other wind farms. The method combines feature selection, regularization and local-learning via conditioning on recent production levels or on expected weather conditions. Improvements in Root Mean Squared Error (RMSE) with respect to other models, evaluated on a dataset with a large number of wind farms are comprised between 4% (10-min horizon) and 11% (3-h horizon). Interpretability of the forecasting model is demonstrated via an analysis of the model coefficients and a discussion of the performance in a challenging situation, namely a wind front.

## Notation

Let $\mathcal{S}$ be a set, then the cardinality of this set is denoted by $|\mathcal{S}|$. Superscript indices, e.g. $.^{(s)}$ in $\mathbf{y}^{(s)}$, define that the production vector is associated to the site of index $s$ in the set $\mathcal{S}$. Estimated regression coefficients are denoted by $\hat{\boldsymbol{\beta}}$.

## Introduction

The increasing penetration of wind power in power systems and electricity markets leads to the need of operational wind power forecasting at different temporal horizons and geographical scales. At the level of a single wind farm, forecasting wind power is a mature discipline on prediction horizons ranging from the next minutes to the next days [1]. Many power system applications such as the management of local grid constraints or the dispatch of flexibility in a renewable virtual power plant require simultaneous predictions of multiple wind farms at large scale, e.g. an entire region or country. A first simple approach consists in applying meteorological models which convert Numerical Weather Prediction (NWP) models into power predictions thanks to calibrated power curves of wind turbines. However, statistical forecasting models are known to outperform meteorological models in the context of intraday wind power forecasting [2]. A surge in Machine Learning based approaches has been observed in recent years for the next-minute to next-hour prediction of wind power [3]. At large scale, both statistical and Machine Learning models face the curse of dimensionality as the number of explanatory variables grows linearly with the number of explanatory sites or the number of variables considered per site.

In terms of remote information, an essential input for short-term wind power forecasting models at multiple sites is recent measurements collected from neighboring wind farms, that act as 'virtual power sensors' and have proved to improve forecasting performance up to a few hours ahead [2]. A state-of-the art approach further developed in this work is an Auto-Regressive Spatio-Temporal (ARST) model where recent measurements of the wind farms neighboring the target site are used as explanatory variables following [4]. The feature matrix $\mathbf{X}_k^{(s)}$ collects in (1) the $N$ power measurements data points of the target production site $s$ and its neighbors resulting in a set of sites $\mathcal{S}$ indexed by $\{(1), \dots, (s), \dots, (|\mathcal{S}|)\}$. A set of time lags $\mathcal{L}$ indexed by $\{1, \dots, l, \dots, |\mathcal{L}|\}$ is applied to the most recent available measurements

* Corresponding author.
*E-mail addresses:* simon.camal@minesparis.psl.eu (S. Camal), robin.girard@minesparis.psl.eu (R. Girard), maxime.fortin@rte-france.com (M. Fortin), augustin.touron@rte-france.com (A. Touron), laurent.dubus@rte-france.com (L. Dubus).

https://doi.org/10.1016/j.seta.2024.103743
Received 10 August 2023; Received in revised form 7 March 2024; Accepted 12 March 2024
Available online 18 March 2024

**Nomenclature**

**Abbreviations**

| | |
|---|---|
| AR,ARST,VAR | Auto-Regressive, id. Spatio-Temporal, Vector Auto-Regressive |
| ccf | Cross-correlation function |
| NWP | Numerical Weather Predictions |
| TSO | Transmission System Operator |

**Parameters**

| | |
|---|---|
| $\sigma$ | Conditional bandwidth |
| $\theta$ | Regularization parameter of the $\ell_1$-norm of regression coefficients |
| $p$ | Number of explanatory variables without regularization |
| $p_\theta$ | Number of explanatory variables with regularization |
| c | Number of conditioning variables |
| M | Number of kernel centers |

**Sets**

| | |
|---|---|
| $\mathcal{K}$ | Set of prediction horizons |
| $\mathcal{L}$ | Set of temporal lags |
| $\mathcal{P}$ | Set of wind power sites in the portfolio |
| $\mathcal{S}$ | Set of explanatory wind power sites in the forecasting model |

**Variables**

| | |
|---|---|
| $\beta$ | Regression coefficients |
| $\mathbf{W}$ | Matrix of conditioning weights |
| $\mathbf{X}$ | Matrix of explanatory variables |
| $\mathbf{y}$ | Vector of normalized power production of a Wind farm |
| $\mathbf{z}$ | Vector of conditioning variables |
| $\rho^{\epsilon}(.)$ | Correlation in errors of auto-regressive models |
| $\rho^{y}(.)$ | Correlation in production of different wind farms |

of each site considering the forecasting lead-time $k$. This leads to a number of $p$ distinct explanatory variables $p = 1 + |S||\mathcal{L}|$ comprising the addition of a constant intercept.

$$\mathbf{X}_k^{(s)} = \begin{bmatrix} 1 & \mathbf{y}_1^{(1)} & \cdots & \mathbf{y}_1^{(s)} & \cdots & \mathbf{y}^{(|S|)} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & \mathbf{y}_t^{(1)} & \cdots & \mathbf{y}_t^{(s)} & \cdots & \mathbf{y}_t^{(|S|)} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & \mathbf{y}_N^{(1)} & \cdots & \mathbf{y}_N^{(s)} & \cdots & \mathbf{y}_N^{(|S|)} \end{bmatrix} \in \mathbb{R}^{N \mathrm{x} p} \tag{1}$$

where $\mathbf{y}_t^{(s)} = \left[ y_{t-k}^{(s)} \cdots y_{t-k-l}^{(s)} \cdots y_{t-k-|\mathcal{L}|}^{(s)} \right] \in \mathbb{R}^{|\mathcal{L}|}, \forall t \in [1, \ldots, N]$

Least-square minimization leads to the well-known analytical solution of regression parameters $\hat{\beta}_k^{(s)} = \left( \mathbf{X}_k^{(s)T} \mathbf{X}_k^{(s)} \right)^{-1} \left( \mathbf{X}_k^{(s)T} \mathbf{y}_k^{(s)} \right)$. When $|S|$ is large (order of hundreds of sites), the computation and inversion of the matrix $\mathbf{X}_k^{(s)T} \mathbf{X}_k^{(s)} \in \mathbb{R}^{p \mathrm{x} p}$ become quickly intractable with standard matrix multiplication and inversion techniques.

An appealing multivariate approach for wind power spatio-temporal forecasting is Vector Auto-Regression (VAR). Sparsity is a desired property of the VAR model in order to ensure tractability, as the number of parameters increases with the square of the number of explanatory wind farms. In [5], a sparse model is achieved by conditional selection of coefficient pairs according to a statistical test on the significance of spatio-temporal correlations. A more direct approach to obtain sparsity consists in applying Lasso to a VAR model [6], solved by coordinated cyclic gradient descent. Adding correlation constraints to the optimization problem formed by a sparsity-controlled approach [7] is found to increase performance on spatio-temporal wind power forecasting compared to a sparsity-control approach.

Formulating the autoregressive model as a local regression permits to take into account the non-linear effects of uncertainties in explanatory variables and wind power production, as done in [8] for a single Wind farm. Further, variability in wind speed and consequently in wind power production over temporal scales (from sub-hourly to annual) should be addressed by the prediction model. At sub-hourly scales, wind speed variability due mostly to boundary layer turbulence can be modeled by downscaling NWP using Model Output Statistics [9], which increases predictions of accuracy of wind speed and wind power production at a single site. A second option consists in an online formulation of the autoregressive model to handle the non-stationarity of the wind production process [10]. Similarly, the spatio-temporal model of [6] combines Lasso regularization with an online formulation which updates iteratively an exponential weight on recent measurements. Wind power forecasts at such sub-hourly resolution facilitate the scheduling of balancing reserve [11] or flexibility levers for local grid constraints [12]. Multi-resolution methods have been proposed [13,14] to ensure consistency between forecasts at low and high resolution (e.g. the hourly mean of minute-resolution forecast equals the hourly-resolution forecast). This avoids discrepancies in sequential decisions in the operation of power systems such as a hours-ahead economic dispatch and a minutes-ahead balancing decision [14]. Variations of wind over longer periods e.g. months, seasons or years may force power system operators to modify their unit commitment or economic dispatch decisions, and should therefore also be predicted. A simple way to capture these slow variations in wind speed and wind direction is to train the regression model over a sliding window [15]. An alternative is to derive a seasonal autoregressive model, cf. [16] in the context of offshore wind speed predictions.

Machine Learning (ML) models can also tackle the problem at hand, starting by Random Forest (RF) trained on selected features from the different neighboring farms and integrated in an ensemble forecasting approach [17]. Spatio-temporal ML wind forecasting methods have also been proposed, especially based on graphs as in [18] or in [19] for minutes to hours-ahead wind speed forecasting at multiple offshore sites. In the latter, a two-stage approach models first spatial dependencies by a Graph Neural Network (GNN), then a Fast-Fourier Transformer creates predictors of temporal variability at the site level that are used to update the GNN nodes. Interestingly, a two-stage method is also employed in the probabilistic spatio-temporal approach proposed by [20], but here ML is used to predict temporal variability at the farm level and a flexible copula (DVINE) captures asymmetrical spatial dependencies in the residuals of the temporal models.

The main purpose of this work is to propose a conditional ARST forecasting approach of wind power production at the scale of a large region or country and intraday horizons. VAR approaches are appealing for the problem at hand but are not investigated here because we argue that the proposed ARST model compares favorably to a VAR model in terms of simplicity of the approach compared to the potentially lower forecasting performance. The original contributions of this paper are the following:

1. *A scalable intraday ARST Wind power forecasting model thanks to the selection of explanatory sites and an efficient computation of the design matrix*. Existing spatio-temporal wind power forecasting methods focus mostly on next-minutes ahead exploiting past power measurements [6,7], but are not easily extendable to hours-ahead forecasting where predicted weather conditions should be considered.

2. *A local learning approach is proposed to condition the forecasting model on intraday features such as predicted weather conditions*. The local learning approach on predicted wind conditions reduces forecasting error compared to benchmarks. Using wind direction to define a priori sparse hierarchical structures in the covariance of wind power production at multiple sites [21] is difficult when there is significant uncertainty about wind direction.
3. *Results are evaluated on two real-world case studies encompassing wind farms across an entire country (France and Australia)*. The improved performance is therefore representative of the heterogeneity in spatio-temporal situations encountered at the level of a region or a country, in terms of weather conditions and geographical distribution of Wind farms. This contrasts with previous case studies focusing on simulated wind profiles [20,21], small regions [21] or countries [6].

## Methodology

The methodology of this work is articulated in three steps:

1. *Pre-processing time series and computing the design matrix*. Periods with missing values, abnormal values or detected power curtailment are discarded from every production time series. The pre-processed series are used to efficiently construct the design matrix (cf. Section "Efficient computation of the design matrix").
2. *Selection of explanatory variables*. is performed in Section "Selection of explanatory sites" in order to focus learning on relevant sites and temporal lags, before formulating the regularized regression model, conditioned by recent production levels or weather forecasts.
3. Analysis of results on two real-world datasets

In what follows, in order to reduce notational clutter, superscripts attached to the target site $.^{(s)}$ are omitted wherever possible.

### *Preprocessing large-scale wind power data*

A filtering method is employed on wind power time series to (1) treat periods where the available power capacity is inferior to the installed capacity and (2) detect long periods with zero production that cannot be associated to lack of wind. Details on the method, based on the detection of boxes in the production signal of a wind farm, can be found in Appendix A.

### *Efficient computation of the design matrix*

The large design matrix ($\mathbf{X}^T\mathbf{X}$) is computed by iterating over the row and column indices of $\mathbf{X}$ associated with the production of all sites at different temporal lags, see formulation in Appendix A.

### *Characterizing the spatio-temporal propagation of forecasting error*

In this work it is assumed that the forecasting model does not dispose of the location of the individual farms, but of the following information to characterize neighboring sites and the weather condition:

- the indices of the $S \setminus s$ nearest geographical neighbors of a target site $s$, where $S \subset \mathcal{P}$, with $\mathcal{P}$ the set of sites in the entire portfolio in the TSO area.
- the most recent NWP available and downscaled at the level of each farm

The assumption that the forecasting model has no access to farm locations can be justified by data privacy constraints or cybersecurity protocols, e.g. in the case where the forecasting model is operated by a third party and communicated to the TSO. Without access to farm locations, spatial dependencies cannot be modeled explicitly from geophysical information. However, this paper considers an autoregressive method with explanatory variables. With such a method, information on physical locations, topology or orography can only be used to select candidate informative sites.

This paper proposes mitigations to capture the spatial dependencies that are implicitly present in the available data without access to the physical location of the farms. First, the model is built based on the information collected from geographically neighboring sites. Among all neighbors, some sites are likely to have low correlation with the target site, because e.g. away from prevailing wind directions experienced by the target site. This is why a second alternative mitigation collects information from sites that are the closest in terms of a distance proxy $d_{\text{proxy}}$, which corresponds to the inverse of the linear correlation between the production at target site $s$ and other sites $j \in P$, $d_{\text{proxy}}(j,s) = 1/\rho^{y}(\mathbf{y}^{(j)}, \mathbf{y}^{(s)}) \quad \forall j, s \in \mathcal{P}$. These sites do not necessarily coincide with geographical neighbors.

The two site selection methods presented above are not directly linked to the performance of statistical learning. Consequently, we test an alternative site selection of subset sites based on the analysis of the correlation between the training errors $\epsilon^{AR}$ of an AR model applied to the target site and potential explanatory sites. Following Girard and Allard [2], the maximum correlation between these errors characterizes the spatio-temporal dependencies in forecasting error for a given horizon difference $k$, for any pair of sites $j, s \in \mathcal{P}$: $\lceil \rho^{\epsilon}_{j,s} \rceil(k) = \max_{|k_2-k_1|=k} \rho^{\epsilon}\left(\left(\epsilon^{AR,(j)}_{t+k_1}\right)_{t\in N}, \left(\epsilon^{AR,(s)}_{t+k_2}\right)_{t\in N}\right)$. Sites that have low correlation of AR errors with the target site at the requested horizon are likely to experience distinct conditions from the target site and therefore convey limited information. The maximum correlation of errors between wind farms of the pool $\lceil \rho^{\epsilon} \rceil(k)$ for two target sites is illustrated in the Supplementary Material.

### *Selection of explanatory sites*

Once pre-processed, the dataset is reduced for each target site $s$ in order to obtain a tractable subset where past production from the target site is complemented by a past production from selected explanatory sites, resulting in a subset of sites $S \subset \mathcal{P}$. The implemented selection methods are listed below.

The explanatory sites consist of the set of sites $S \setminus s$ selected according to the following methods:

- **Geographical neighbors.** Set formed by the geographical neighbors of the target site.
- **Proxy of distance by correlation in production.** Set of sites that have the smallest proxy of the distance to the target site.
- **Maximum correlation of AR forecasting error.** Set of sites that show the smallest distance in terms of correlation in the error of the AR forecasting model with the target site. The distance $d_{\text{AR-errors}}$ is the inverse of the maximum correlation of AR errors between the target site $s$ and another site $j$. Note that this distance depends of the prediction horizon.
- **Average cross-correlation with lagged productions at distant sites** The explanatory sites consist of the set of sites $S \setminus s$ that show the smallest distance in terms of the cross-correlation function (ccf) of their lagged production level with the production of target sites, as expressed in (2). More specifically, the distance is the inverse of the average ccf for lags greater or equal to the prediction horizon.

$$\text{ccf}_{j,s}(k,\ell) = \rho((y^{(j)}_{t-\ell})_{t\in\{1,\dots,N\}}, (y^{(s)})_{t\in\{1,\dots,N\}}), \quad \ell \geq k, \forall k \in \mathcal{K} \tag{2}$$

$$d^{\text{ccf}}_{j,s}(k) = \frac{1}{\frac{1}{|\mathcal{L}|-k}\sum_{\ell\in\mathcal{L},\ell\geq k} \text{ccf}_{j,s}(k,\ell)}, \quad \forall k \in \mathcal{K} \tag{3}$$

*Conditioning the ARST model by intraday features*

The ARST model is now conditioned by intraday features. Conditioning is implemented by adding local weights to the linear regression model [22]. The following feature variables have been considered as potential conditioning variables at intraday horizons:

- most recent power measurement at runtime and target site
- wind speed forecast at runtime and target site
- wind direction forecast at runtime and target site

The conditioning is applied to one or to a combination of the variables listed above, resulting in a conditioning vector $\mathbf{z} \in \mathbb{R}^c$ where $c$ is the number of conditioning variables. Conditioning weights are derived in (4) from Gaussian kernels of width $\sigma$ and centered on a series of conditioning points $\mathbf{z}_i, i \in [1, \dots, M]$, forming a set of $M$ kernels.

$$w(\mathbf{z}_i, \mathbf{z}_t) = \exp(-(\mathbf{z}_t - \mathbf{z}_i)^2/(2\sigma^2)), \quad i \in [1, \dots, M] \tag{4}$$

In what follows, indices associated to the target site $s$ are dropped to avoid notational clutter. Local regression coefficients $\hat{\boldsymbol{\beta}}_{k,i}$ are derived by weighted least-square minimization for each lead-time $k$ and conditioning index $i$ over training points $t \in [1, \dots, N]$: $\hat{\boldsymbol{\beta}}_{k,i} = \text{argmin}_{\beta} \sum_{t\in[1,N]} w(\mathbf{z}_i, \mathbf{z}_t)(y_{t+k} - \mathbf{X}_{k,t}^T \boldsymbol{\beta})^2$.

The analytic solution $\hat{\boldsymbol{\beta}}_{k,i} = (\mathbf{X}_k^T \mathbf{W}_i \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{W}_i \mathbf{y}$ involves a weighted design matrix for each conditioning index $\mathbf{W}_i(\mathbf{z})$ as follows:

$$\mathbf{W}_i(\mathbf{z}) = \begin{pmatrix} w(\mathbf{z}_i, \mathbf{z}_1) & \dots & 0 \\ \vdots & w(\mathbf{z}_i, \mathbf{z}_t) & \vdots \\ 0 & \dots & w(\mathbf{z}_i, \mathbf{z}_N) \end{pmatrix}, \quad \forall i \in [1, \dots, M]$$

Finally production forecasts of the target site consist in the interpolation of local predictions associated to the different kernels as a function of their conditional weights at runtime $w(\mathbf{z}_i, \mathbf{z}_t), i = [1, \dots, M]$: $\hat{y}^s_{t+k|t}(\mathbf{z}_t) = 1/\sum_{i=1}^M w(\mathbf{z}_i, \mathbf{z}_t). \sum_{i=1}^M w(\mathbf{z}_i, \mathbf{z}_t) \hat{\boldsymbol{\beta}}_{k,i} \mathbf{X}_{k,t}$.

*Efficient regularization for large-scale wind power forecasting*

The second methodological contribution of this work consists in the evaluation of the added value of sparse models obtained by the addition of a regularization penalty to promote sparsity in the learning pattern of the ARST model. The first penalty employed is the well-known Lasso penalty that penalizes in (5) the $\ell$-1 norm of regression coefficients $\|\hat{\beta}\|_1 = \sum_{d=1}^D |\hat{\beta}_d|$. A constraint ensures that the penalty is lower or equal to a threshold $\theta$, tuned by cross-validation as further discussed in the Case Study, cf. Section "Model configuration". The aim is to assess empirically the merits of the Lasso regularization when applied to a high-dimensional and conditioned problem, specifically:

- Is the ARST problem considering all potential explanatory sites performing better with regularization?
- When regularization is applied to methods implementing a site selection or conditioning, is it beneficial to the forecasting performance and to the interpretation of the model?
- What is the sensitivity of forecasting performance to the bandwidth of conditioning kernels?

$$\hat{\boldsymbol{\beta}}_{k,i} = \underset{\beta}{\text{argmin}} \sum_{t\in\{1,\dots,N\}} w_i (y_{t+k} - \mathbf{X}_{k,t}^T \boldsymbol{\beta})^2 \tag{5}$$

$$s.t. \sum_{d=1}^{p} |\beta_d| \leq \theta$$

## Case study

*Dataset description*

The method is evaluated on two wind power datasets covering large geographical scales:

1. **Dataset 1**: A large-scale dataset is composed of a subset $\mathcal{P} = 747$ sites of the entire pool of wind power sites in France. This subset corresponds to an ensemble of sites for which the preprocessed time series do not exhibit large periods of zero production or reduced available capacities. The forecasting model disposes of the historical time series, installed capacity and NWP wind forecasts from ARPEGE-MeteoFrance downscaled to each site.
2. **Dataset 2**: Open dataset with ten wind farms covering different regions in Australia, namely the dataset associated to the Gefcom 2014 Wind Track [23], where wind power time series and NWP downscaled to each site are available. We adopt here the same forecasting configuration than in [24] in order to compare with the results obtained by their Lasso-VAR model on this dataset.

*Model configuration*

The forecasting model is applied to derive forecast for all sites in the portfolio of $\mathcal{P}$ sites is selected. The proposed ARST model is compared with the following benchmarks:

- AR model, unconditioned and conditioned to the most recent power measurement at target site
- unconditioned Random Forest (RF) model, implemented with the R library *ranger*, chosen for its robustness and known level-field performance in hours-ahead wind power forecasting [3].
- For Dataset 1, operational multivariate wind power forecasting model implemented at RTE. This benchmark model is the first step of the wind power forecasting method Préole developed by RTE. It consists of an initial estimate of the production of each wind farm based on calibrated wind turbine power curves and NWP from ARPEGE-MeteoFrance. The second step of Préole, not tested here, corrects the initial estimation as a function of recent power observations and of the autocorrelation in forecasting errors.
- For Dataset 2, the Lasso-VAR developed and evaluated on this dataset by [24]

The configuration parameters of the forecasting model for the present case study are listed in the upper section of Table 1. Three conditioning approaches are tested: conditioning on last power measurements for the AR, conditioning on wind directions and on wind speeds for the ARST model. Conditioning parameters (kernel centers $\mathbf{z}_i$ and bandwidth $\sigma$), identical for both datasets, are presented in the lower section of Table 1. Positions of the conditioning points correspond to $M = 10$ points uniformly placed along the interval of possible values.

Finally, regularization is applied to the ARST model, in its unconditioned and conditioned variants. The threshold value $\theta$ is tuned by cross-validation on the following set of thresholds, applying a progressively higher constraint on the $\ell_1$-norm of the unconditioned model: $\theta \in \{\sum_{d=1}^{p_\theta} |\hat{\beta}_d|, \quad p_\theta = (p/10, p/50, p/100)\}$

*Evaluation and analysis of model behavior*

The forecasting performance is evaluated in terms of bias, normalized MAE and RMSE over the range of prediction horizons. The normalization is done by diving the score by the installed capacity of each site [25]. In order to assess the results over the entire pool, we present the average RMSE of the pool but also the distribution of score over the sites, approximated by its median and quartiles. The added value of the ARST model is quantified by deriving the RMSE improvement compared to an unconditional AR model, using the same lags of recent production at the target site.

**Table 1**
Forecasting parameters.

| Config. parameter | Dataset 1 (France) | Dataset 2 (Australia) |
|---|---|---|
| Training window | Fixed | Sliding, 1 month increment |
| Training start/end | 2018-01/2018-12 | 2012-01/2012-12 → 2012-11/2013-10 |
| Testing | 2019-01/2019-12 | 2013-01 → 2013-11 |
| Horizons | 10 min to 3 h | 1 h to 6 h |
| Temporal resolution | 10 min | 1 h |
| Temporal lags | (10, 20, 30 min, 1 h, 2 h, 4 h) | (1 h to 6 h) |
| Number of evaluated sites | 492 | 10 |
| **Condit. Parameters** | **Centers ($\mathbf{z}$)** | **Bandwidth ($\sigma$)** |
| Last power [Capacity factor] | {0.1, 0.2, …, 0.9 } | {0.02,0.03,…,0.08} |
| Wind direction [° ] | {15°,53°,91°, …,360° } | {15°,45°,90°,360° } |
| Wind speed [m/s] | {0,3.33,6.66, …,30 } | {1,2,3,4} |

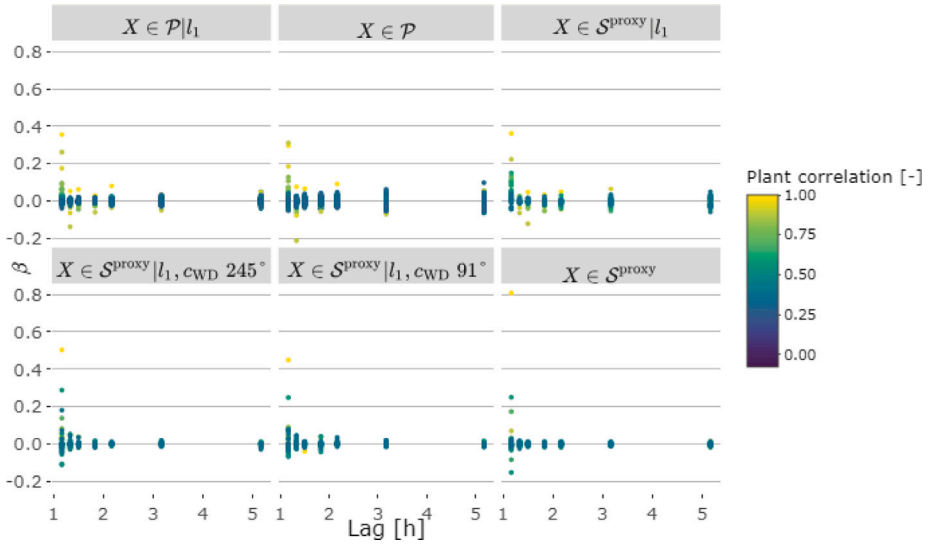



**Fig. 1.** Regression coefficients of the ARST model variants at 1h horizon for target site '1000', as a function of the lag added to the last available measurement. The color scale indicates the level of linear correlation between the target site and the explanatory site. $X \in \mathcal{P}|l_1$ (resp. $X \in \mathcal{P}$): ARST with all sites in portfolio $\mathcal{P}$ with (resp. without) $l_1$ regularization; $X \in \mathcal{S}^{\text{proxy}}|l_1$ (resp. $X \in \mathcal{S}^{\text{proxy}}$): ARST with sites selected by distance proxy with (resp. without) $l_1$ regularization; $c_{WD}$ 245° (resp 91°) indicates conditioning by wind direction centered on position 245° (resp 91°).

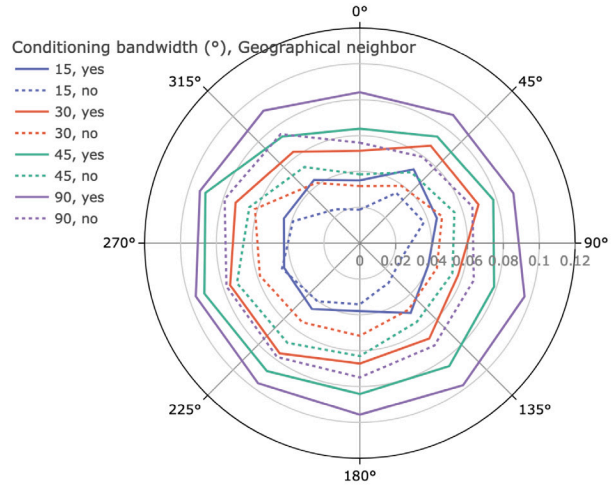
## Results

### *Effect of conditioning and regularization on the regression model*

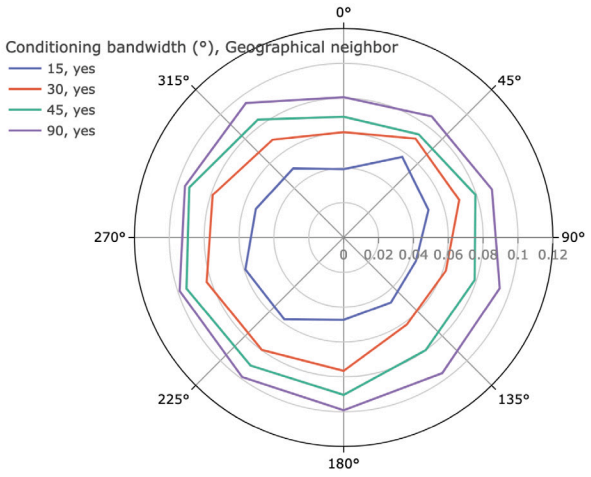
The evaluation starts by an interpretation of regression coefficients. Fig. 1 shows the regression coefficients derived for a 1-h horizon at a specific target site in Dataset 1 (results are similar for other target sites). The ARST models without site selection (top-left and top-center) spread coefficients across all available sites in $\mathcal{P}$. Non-zero coefficients are mostly associated to recent lags below 1h as expected. The application of $\ell_1$-regularization reduces homogeneously low coefficients towards zero, which corresponds to a known behavior of the lasso [22]. The inefficiency of the ARST models without feature selection is demonstrated by the fact that sites with the highest absolute value of coefficients are not the closest in terms of proxy distance to the target site (distance proxy values of 0.4–0.6). In other words, the ARST models without feature selection exhibit a significant bias due to their inability to treat the high level of collinearity in the feature space. In contrast, the ARST models implementing feature selection by retaining the closest sites in terms of distance proxy ($X \in \mathcal{S}^{\text{proxy}}$) show high absolute values of coefficients at close sites. The model without conditioning and lasso derives high coefficients to the most recent lagged production ($\leq$ 1h) of sites close to the target site. Adding conditioning and Lasso to this model brings a twofold modification of regression coefficients: (1) the weight given to lagged production of the target site (right side of x-axis) varies as a function of the wind direction, with a maximum value for the wind direction conditioning point of 15°; (2) absolute values of coefficients are reduced and negative coefficients are close to zero in contrast with the non-regularized model.

Mean absolute values of coefficients obtained for the different conditioning centers (wind directions) for two conditional ARST methods, selecting explanatory sites from the geographical neighbors or according to the distance proxy are shown in Fig. 2. Regression coefficients associated to sites that are geographical neighbors are similar in both methods. For explanatory sites that are not neighbors, the method selecting sites according to the distance proxy obtains higher weights in westerly directions, which is one of the main prevailing wind direction for onshore farms in France (Atlantic ocean). Coefficients obtained by this method combine a relatively direction-independent information from geographical neighbors with direction-specific information from sites outside the neighborhood.

Finally, the conditioning bandwidth influences the selection of most explicative sites as shown in Fig. 3. The number of unique explanatory sites associated to significant regression coefficients decreases with increasing values of conditional bandwidth (wind direction), which indicates higher selectivity. We note that selectivity is more pronounced on non-neighboring sites for the method with sites selected according to the distance proxy.




(a) Site selection based on distance proxy




(b) Site selection based on geographical neighbors

**Fig. 2.** Mean absolute coefficients of conditioned and regularized method for the different conditioning wind direction centers, for two different site selection methods. Horizon = 3h (30 randomly selected target sites). Coefficients of explanatory sites that are (resp. are not) geographical neighbors are represented by solid (resp. dashed) lines.

*Forecasting performance*

*Analysis of forecasts during a wind front*

A relevant question for spatio-temporal forecasting is to know whether the proposed model helps predict production during a wind front where forecasting errors are high due to phasing errors (e.g. delay between the observed production increase and the prediction) and amplitude errors (e.g. error in wind speed magnitude during a plateau of high wind speed near the rated wind speed of a turbine). The prediction obtained by the proposed model during such a wind front with a forecasting horizon of 1h is illustrated in green in Fig. 4, for a particular target site. The model exploits local information such as the lagged observed production of the explanatory site which has the highest regression coefficient (in gray in the Figure), whereas the benchmark AR model ignoring this type of information predicts with a phase error during the period of increasing production and amplitude errors during the plateau of maximum production. If the geographical position of the different sites is available, this analysis can be generalized beyond this specific situation to estimate the spatio-temporal propagation of prediction errors following the method proposed in [2].

*Performance of combined Lasso and site selection*

The proposed ARST model is now evaluated on the two datasets, for every site in each portfolio. The mean RMSE obtained for Dataset 1 across sites for the different methods is shown in Table 2. The proposed ARST model with selection of sites by distance proxy, conditioned by wind direction and regularized by lasso shows the best performance across all models and all horizons. The conditioned AR model on last production level brings very close result to the unconditional AR. The site selection by distance proxy (linear correlations in production) achieves lower mean RMSE than the other site selection methods, namely based on AR errors, geographical neighbors and cross-correlation functions for all horizons. At the short horizon of $k = 10$ min, conditioning seems to have little added value as the most recent observation of the target site is the most useful feature. Conditioning on wind direction appears to be more effective than conditioning on wind speed. This may reflect the fact that collecting information from sites upstream or experiencing similar wind directions is more informative than knowing expected wind magnitude at neighboring sites, which is indirectly reflected in the neighbors' recent production measurements.

Improvement vs the AR model is observed for bandwidth values of $\sigma \geq 45°$, whereas conditioning with a bandwidth of 15° overfits and consequently has degraded performance, even below the unconditional AR reference. Conversely, it is observed that conditioning with a bandwidth of 360° has similar performance than the unconditional ARST, which is expected because having a bandwidth equal to the entire range of possible values for the conditioning variables leads to zero conditioning. Interestingly, conditioning seems to improve performance only when associated with Lasso regularization and with a bandwidth of 90°. The choice of a relatively sparse model ($p_\theta = p/10$) leads to the best median improvement up to 11% at 3h horizon. For further horizons starting from 5–6 h (not shown here), the benchmark model Préole presents better performance than the present model because the NWP information becomes dominant over recent production lags and therefore the spatio-temporal approach has less interest.

In Dataset 2, only a reduced set of wind farms are monitored in the Australian territory. In this context, no site selection is performed, i.e. all sites are considered as explanatory. Detailed results are reported in Appendix B and in the Supplementary Material. Interestingly here, conditioning on predicted wind speed leads to better results in terms of bias, RMSE and MAE than conditioning on predicted wind direction. With a reduced number of neighboring farms as 'virtual sensors', conditioning on wind speed compensates the limited information on power regimes at the vicinity of the target site by a non-linear proxy between wind speed and expected production. This contribution becomes secondary in Dataset 1, where the identification of significantly explicative sites is assisted by conditioning on expected wind direction.

## Conclusions

This paper presented an auto-regressive spatio-temporal method for the prediction of Wind power production at a regional or national scale and for horizons of minutes to hours ahead.

The site selection approach of retaining the sites with the highest linear correlation in production improves the RMSE up to 5% compared to a case where the model considers the 50 nearest geographical neighbors (Dataset 1). Conditioning increases forecasting performance in different spatio-temporal contexts. In Dataset 1 with a dense network of wind farms (50 neighbors for each target site), conditioning by predicted wind direction decreases RMSE if the conditioning bandwidth is large enough (90°) in order to reach a balance between site selectivity and diversity of weather conditions. In Dataset 2 with a low density of neighboring farms, conditioning by expected wind speed brings the highest improvement, as it helps capture better the varying effect of wind on power at the target site and at the small number of potentially

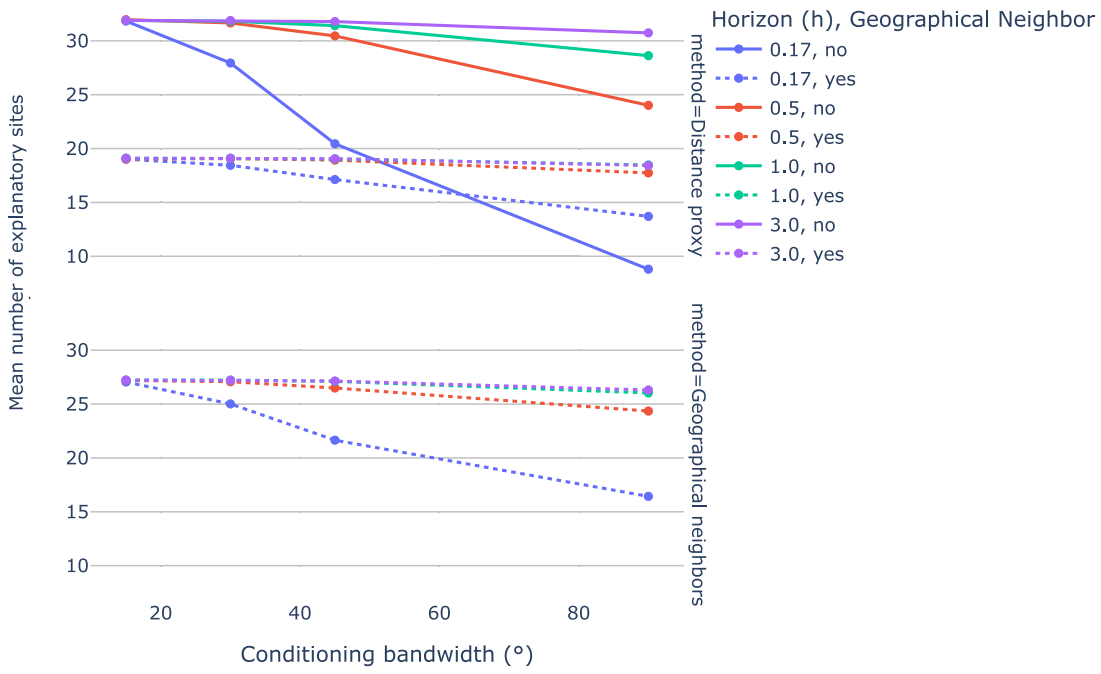



**Fig. 3.** Mean number of unique explanatory sites for which the absolute regression coefficient is significantly positive (larger than median values of non-zero coefficients), as a function of the conditioning bandwidth on wind direction. Top: explanatory sites selected based on distance proxy, Bottom: explanatory sites as geographical neighbors.

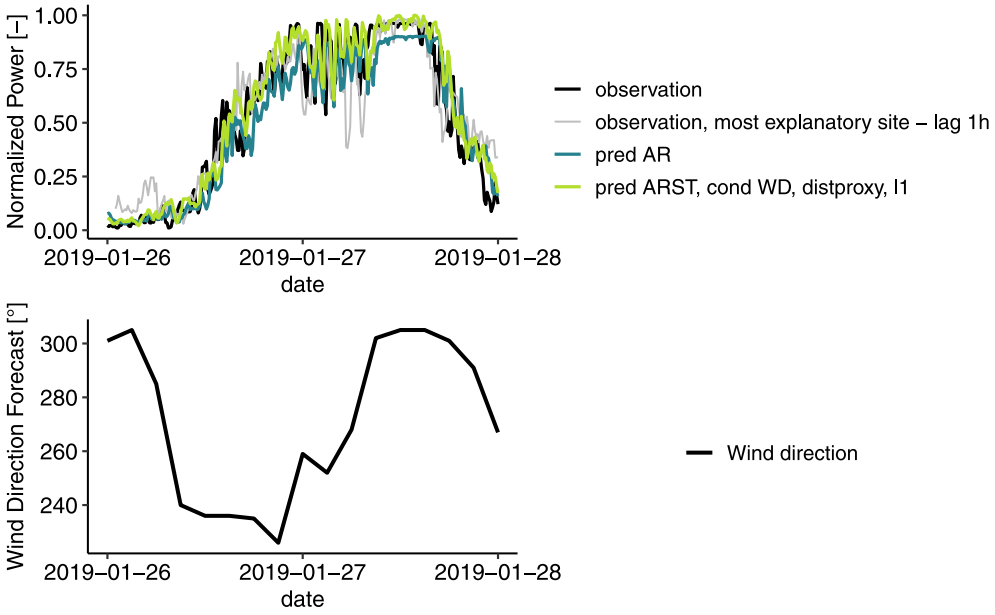



**Fig. 4.** Prediction (1h forecasting horizon) during a wind front with the proposed conditional regularized ARST model in green, compared with an AR model in blue, and observed production in black. The bottom figure shows the predicted wind direction at the target site.

informative neighboring sites. The regularization by LASSO enables to improve the RMSE of the ARST model by 10% at 2-h horizon compared to an unconditional AR model ignoring off-site information.

This work opens several perspectives for large-scale intraday wind power forecasting. A first improvement would consist in deriving seamless predictions from a single model over multiple horizons as done for PV by [26]. A second perspective consists in reformulating the conditional model in an online learning setting in order to improve scalability and adaptability to weather cycles or composition of the wind farm portfolio. Finally, a third perspective consists in building upon the resilient forecasting model proposed by [27], where lasso regression is reformulated as a robust regression where the missingness variables at test time is modeled by sets of binary variables [28]. The decision rules corresponding to missing data could be adapted to account for

**Table 2**
RMSE of the different methods, mean across target sites, as a function of lead-times. Best performance per horizon is indicated in bold.

| Method, Conditioning \ Forecasting horizon | k= 0.17h | k= 0.5h | k= 1h | k= 3h |
|---|---|---|---|---|
| AR, production $\sigma = 0.02$ | **4.98** | 8.27 | 10.50 | 15.13 |
| AR, production $\sigma = 0.03$ | 5.21 | 8.81 | 11.25 | 16.15 |
| AR, production $\sigma = 0.04$ | 5.21 | 8.80 | 11.24 | 16.14 |
| AR, production $\sigma = 0.07$ | 5.20 | 8.80 | 11.23 | 16.13 |
| AR, production $\sigma = 0.08$ | 5.20 | 8.80 | 11.23 | 16.13 |
| ARST-AR errors, unconditional | 5.18 | 8.61 | 10.95 | 15.28 |
| ARST-proxy, WD $\sigma = 15^{\circ}$ | 5.39 | 9.11 | 11.58 | 16.56 |
| ARST-proxy, WD $\sigma = 360^{\circ}$ | 5.01 | 8.22 | 10.31 | 14.75 |
| ARST-proxy, WD $\sigma = 45^{\circ}$ | 5.07 | 8.39 | 10.57 | 15.23 |
| ARST-proxy, WD $\sigma = 90^{\circ}$ | 5.02 | 8.26 | 10.37 | 14.88 |
| ARST-proxy-$\ell_1$, WD $\sigma = 90^{\circ}$ | **4.98** | **7.34** | **9.17** | **13.22** |
| ARST-ccf-$\ell_1$, WD $\sigma = 90^{\circ}$ | 5.07 | 8.30 | 10.59 | 15.24 |
| ARST-proxy, WS $\sigma = 2$ m/s | 5.16 | 8.36 | 10.40 | 14.93 |
| ARST-proxy, WS $\sigma = 3$ m/s | 5.11 | 8.27 | 10.28 | 14.62 |
| ARST-proxy, WS $\sigma = 4$ m/s | 5.10 | 8.24 | 10.24 | 14.52 |
| ARST-proxy, unconditional | 5.18 | 8.39 | 10.47 | 14.89 |
| ARST-proxy-$\ell_1$, unconditional | 4.99 | 8.19 | 10.27 | 14.70 |
| ARST-neighbors, unconditional | 5.00 | 8.29 | 10.56 | 15.40 |
| Benchmark,unconditional | 5.50 | 9.09 | 12.50 | 16.18 |
| Persistence,unconditional | 5.31 | 9.09 | 11.70 | 17.19 |
| RF,unconditional | 5.30 | 8.88 | 11.34 | 16.31 |

curtailment situations observed at target site and neighboring sites, e.g. by sets of curtailment levels (from 0% to 100%).

## CRediT authorship contribution statement

**Simon Camal:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Robin Girard:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization. **Maxime Fortin:** Writing – review & editing, Resources, Funding acquisition, Data curation, Conceptualization. **Augustin Touron:** Writing – review & editing, Resources, Funding acquisition, Data curation, Conceptualization. **Laurent Dubus:** Supervision, Resources, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data for Dataset 1 is confidential. Dataset 2 is an open dataset.

## Acknowledgments

The authors would like to thank RTE for the provision of wind power production data and benchmark forecasts derived from their operational model Préole, and Météo France for the access to NWP predictions.

## Appendix A

This appendix presents details on filtering time series and computation of the design matrix of the spatio-temporal regression method.

**Algorithm 1** Iterative computation of the design matrix

```
i ← 0
for s ∈ S do
    for k ∈ {1, ..., L} do
        j ← 0
        for s' ∈ S do
            for k' ∈ {1, ..., L} do
                (X^T X)_ij = Σ_{t∈{1,...,N}} y^(s)_{t−k} . y^(s')_{t−k'}
                j ← j + 1
            end for
        end for
    end for
end for
```

*Filtering wind power time series*

At each timestep $t$, the algorithm computes the rectangular box $b_t$, defined in (A.1) as the area above the observed normalized production $y_t$ up to the theoretical maximum (here 1), for the past period until the most recent timestep $t'$ with production higher than the observed production $y_t$. This filtering approach is computationally fast and versatile as it detects both potential partial curtailments and long idle periods without production.

$$b_t = 1_{\{y_{t-1} \le y_t\}}.(1 - y_t).(t - t'), \quad \forall t \in \{1, \dots, N\}$$
$$\text{with} \quad t' = \max_{1 \le i < t} \{i, y_i \ge y_t\} \tag{A.1}$$

*Computation of the design matrix*

The design matrix is computed in Algorithm 1 by iterating through indices of its rows and columns, denominated hereafter $i$ and $j$ respectively.

## Appendix B

This Appendix collects results obtained on Dataset 2 (Gefcom 2014, Australia). Table B.1 shows the average RMSE across sites of the regularized ARST model conditioned on either predicted wind direction

**Table B.1**
Mean RMSE of spatio-temporal forecasting models, across sites of Dataset 2 as a function of the horizon. Best model in **green** and second best in **orange**.

| Method | $k = 1h$ | $k = 2h$ | $k = 3h$ | $k = 4h$ | $k = 5h$ | $k = 6h$ |
|---|---|---|---|---|---|---|
| ARST-$\ell_1$, WD $\sigma = 15^\circ$ | 10.06 | 14.98 | 18.04 | 20.29 | 22.05 | 22.79 |
| ARST-$\ell_1$, WD $\sigma = 30^\circ$ | 9.98 | 14.87 | 17.94 | 20.22 | 22.01 | 23.47 |
| ARST-$\ell_1$, WD $\sigma = 45^\circ$ | 9.98 | 14.89 | 17.98 | 20.31 | 22.15 | 23.64 |
| ARST-$\ell_1$, WD $\sigma = 90^\circ$ | 9.93 | 14.89 | 18.07 | 20.46 | 22.37 | 23.95 |
| ARST-$\ell_1$, WS $\sigma = 1$ m/s | **9.63** | **13.49** | **15.33** | **16.35** | **16.94** | **17.32** |
| ARST-$\ell_1$, WS $\sigma = 2$ m/s | **9.63** | **13.82** | **16.02** | **17.36** | **18.21** | **18.79** |
| ARST-$\ell_1$, WS $\sigma = 3$ m/s | **9.73** | 14.22 | 16.77 | 18.47 | 19.66 | 20.53 |
| ARST-$\ell_1$, WS $\sigma = 4$ m/s | 9.80 | 14.49 | 17.28 | 19.22 | 20.66 | 21.77 |
| ARST-$\ell_1$, WS/WD $\sigma = \{2$ m/s, $30^\circ\}$ | 9.89 | 14.53 | 17.29 | 19.26 | 20.68 | 21.80 |
| ARST-$\ell_1$, WS/WD $\sigma = \{3$ m/s, $15^\circ\}$ | 9.95 | 14.61 | 17.39 | 19.36 | 20.79 | 21.90 |
| ARST-$\ell_1$, WS/WD $\sigma = \{3$ m/s, $30^\circ\}$ | 9.84 | 14.50 | 17.28 | 19.25 | 20.69 | 21.80 |
| ARST-$\ell_1$, WS/WD $\sigma = \{3$ m/s, $45^\circ\}$ | 9.86 | 14.54 | 17.38 | 19.41 | 20.94 | 22.16 |
| ARST-$\ell_1$, WS/WD $\sigma = \{3$ m/s, $90^\circ\}$ | 9.90 | 14.71 | 17.69 | 19.87 | 21.56 | 22.92 |
| ARST-$\ell_1$, WS/WD $\sigma = \{4$ m/s, $30^\circ\}$ | 9.86 | 14.52 | 17.33 | 19.35 | 20.86 | 22.06 |
| PP-Lasso-VAR [24] | 9.85 | 14.46 | 17.29 | 19.38 | 21.04 | 22.39 |
| Analogs [24] | 10.48 | 15.52 | 18.89 | 21.45 | 23.46 | 25.15 |

or wind speed. Additional information is reported in the Supplementary Material as follows:

1. a figure showing average biases and MAE across sites of the regularized and conditioned model on wind speed, depending on horizon and conditional bandwidth
2. a figure showing the evolution of regression coefficients as a function of training sliding windows.

## Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.seta.2024.103743.

## References


[1] Kariniotakis G. Renewable energy forecasting - From models to applications. Woodhead publishing series in energy, Woodhead Publishing Series in Energy - Elsevier; 2017.

[2] Girard R, Allard D. Spatio-temporal propagation of wind power prediction errors. Wind Energy 2013;(August 2012). http://dx.doi.org/10.1002/we.1527.

[3] Tawn R, Browell J. A review of very short-term wind and solar power forecasting. Renew Sustain Energy Rev 2022;153:111758. http://dx.doi.org/10.1016/j.rser.2021.111758.

[4] Agoua XG, Girard R, Kariniotakis G. Short-term spatio-temporal forecasting of photovoltaic power production. IEEE Trans Sustain Energy 2018;9(2):538–46. http://dx.doi.org/10.1109/TSTE.2017.2747765.

[5] Dowell J, Pinson P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. IEEE Trans Smart Grid 2016;7(2):763–70. http://dx.doi.org/10.1109/TSG.2015.2424078.

[6] Messner JW, Pinson P. Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. Int J Forecast 2019;35(4):1485–98. http://dx.doi.org/10.1016/j.ijforecast.2018.02.001.

[7] Zhao Y, Ye L, Pinson P, Tang Y, Lu P. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. IEEE Trans Power Syst 2018;33(5):5029–40. http://dx.doi.org/10.1109/TPWRS.2018.2794450.

[8] Pinson P, Nielsen HA, Madsen H, Nielsen TS. Local linear regression with adaptive orthogonal fitting for the wind power application. Stat Comput 2008;18(1):59–71. http://dx.doi.org/10.1007/s11222-007-9038-7.

[9] Dupré A, Drobinski P, Alonzo B, Badosa J, Briard C, Plougonven R. Sub-hourly forecasting of wind speed and wind energy. Renew Energy 2020;145:2373–9. http://dx.doi.org/10.1016/j.renene.2019.07.161.

[10] Pierrot A, Pinson P. Adaptive generalized logit-normal distributions for wind power short-term forecasting. In: 2021 IEEE Madrid PowerTech, PowerTech 2021 - Conference Proceedings. (1). 2021, http://dx.doi.org/10.1109/PowerTech46648.2021.9494900, arXiv:2012.08910.

[11] Hosseini SA, Toubeau J-F, Amjady N, Vallée F. Day-ahead wind power temporal distribution forecasting with high resolution. IEEE Trans Power Syst 2023;1–11. http://dx.doi.org/10.1109/TPWRS.2023.3295915.

[12] Bessa RJ, Moaidi F, Viana J, Andrade JR. Uncertainty-aware procurement of flexibilities for electrical grid operational planning. IEEE Trans Sustain Energy 2023;1–14. http://dx.doi.org/10.1109/TSTE.2023.3305865.

[13] Nejati M, Amjady N, Zareipour H. A new multi-resolution closed-loop wind power forecasting method. IEEE Trans Sustain Energy 2023;14(4):2079–91. http://dx.doi.org/10.1109/TSTE.2023.3259939.

[14] Wang C, Pinson P, Wang Y. Seamless and multi-resolution energy forecasting. 2023, arXiv:2401.05413.

[15] Goncalves C. Renewable energy forecasting – extreme quantiles, data privacy and monetization [Ph.D. thesis], Doctoral Program in Applied Mathematics of the Univer- sities of Minho, Aveiro and Porto; 2021, Available at https://repositorio-aberto.up.pt/bitstream/10216/135580/2/487920.pdf.

[16] Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM. Energy 2021;227:120492. http://dx.doi.org/10.1016/j.energy.2021.120492.

[17] De Caro F, De Stefani J, Vaccaro A, Bontempi G. DAFT-E: Feature-based multivariate and multi-step-ahead wind power forecasting. IEEE Trans Sustain Energy 2022;13(2):1199–209. http://dx.doi.org/10.1109/TSTE.2021.3130949.

[18] Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 2019;10(2):670–81. http://dx.doi.org/10.1109/TSTE.2018.2844102.

[19] degaard Bentsen LO, Warakagoda ND, Stenbro R, Engelstad P. Spatio-temporal wind speed forecasting using graph networks and novel transformer architectures. Appl Energy 2023;333:120565. http://dx.doi.org/10.1016/j.apenergy.2022.120565.

[20] Arrieta-Prieto M, Schell KR. Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model. Int J Forecast 2022;38(1):300–20. http://dx.doi.org/10.1016/j.ijforecast.2021.05.013.

[21] Liu Y, Davanloo Tajbakhsh S, Conejo AJ. Spatiotemporal wind forecasting by learning a hierarchically sparse inverse covariance matrix using wind directions. Int J Forecast 2021;37(2):812–24. http://dx.doi.org/10.1016/j.ijforecast.2020.09.009.

[22] Hastie, Tibshirani. The elements of statistical learning. In: Springer text in statistics. 2013, http://dx.doi.org/10.1007/978-1-4419-9863-7-941.

[23] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. Int J Forecast 2016;32(3):896–913. http://dx.doi.org/10.1016/j.ijforecast.2016.02.001.

[24] Goncalves C, Bessa RJ, Pinson P. Privacy-preserving distributed learning for renewable energy forecasting. IEEE Trans Sustain Energy 2021;12(3):1777–87. http://dx.doi.org/10.1109/TSTE.2021.3065117.

[25] Messner JW, Pinson P, Browell J, Bjerregård MB, Schicker I. Evaluation of wind power forecasts—An up-to-date view. Wind Energy 2020;23(6):1461–81. http://dx.doi.org/10.1002/we.2497.

[26] Carriere T, Vernay C, Pitaval S, Kariniotakis G. A novel approach for seamless probabilistic photovoltaic power forecasting covering multiple time frames. IEEE Trans Smart Grid 2020;11(3):2281–92. http://dx.doi.org/10.1109/TSG.2019.2951288.

[27] Stratigakos A, Andrianesis P, Michiorri A, Kariniotakis G. Towards resilient energy forecasting: A robust optimization approach. IEEE Trans Smart Grid 2023;1. http://dx.doi.org/10.1109/TSG.2023.3272379.

[28] Xu H, Caramanis C, Mannor S. Robust regression and lasso. IEEE Trans Inf Theor 2010;56(7):3561–74. http://dx.doi.org/10.1109/TIT.2010.2048503.