

STOCHASTIC SIMULATION OF PREDICTIVE SPACE-TIME SCENARIOS OF WIND SPEED USING OBSERVATIONS AND PHYSICAL MODELS

BY JULIE BESSAC^{*}, EMIL CONSTANTINESCU^{*,†} AND MIHAI ANITESCU^{*,†}

Argonne National Laboratory^{} and The University of Chicago[†]*

We propose a statistical space-time model for the prediction of atmospheric wind speed based on deterministic numerical weather predictions and historical measurements. We consider a Gaussian multivariate space-time framework that combines multiple sources of past physical model outputs and measurements along with model predictions in order to produce a probabilistic wind speed forecast within the prediction window. We illustrate this strategy on ground wind speed forecast for several months in 2012 for a region near the Great Lakes in the US. The results show that (i) the prediction is improved in the mean-squared sense relatively to the numerical forecasts as well as (ii) in probabilistic scores. Moreover, (iii) the samples are shown to produce realistic wind scenarios based on the sample spectrum.

1. Introduction . In this study we propose a statistical space-time model for the prediction of atmospheric wind speed based on numerical weather predictions and historical measurements. We focus on a region around lake Michigan in the US; however, the framework proposed here is not specific to it. The wind speed predictions are based on deterministic numerical weather prediction (NWP) model outputs in a framework that integrates past dependence between observational measurements and the NWP model outputs. The aim of this work is to improve the wind speed forecasts provided by the NWP model based on the past relation, that is modeled linearly, between measurements and NWP forecasts.

Atmospheric surface wind prediction is very important for the energy, agricultural, and security sectors and it has known a considerable development for the past years. Several components of the wind field can be predicted: the zonal and meridional components, see [16, 26]; wind speed, see [6, 25] and wind direction, [2]. Recent works on wind speed and wind power statistical prediction focus on the generation of predictive scenarios that enables to account for the prediction error, see [20, 19]. However, few criteria of quality assessment of scenarios have been proposed; multivariate

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: sample, L^AT_EX 2 ϵ

(multiple time-step ahead and/or space or ensemble forecasts) criteria can be used but do not account for the nature of time trajectory of scenarios. In [19] an event-based criterion is proposed to assess the quality of scenarios to reproduce wind events and to compare scenarios from different models.

In general predicting wind speed based on measurements and physical model outputs results in multivariate space-time processes that are typically inhomogeneous due to the presence of different types of data. Multivariate space-time modeling has been an area of intense research focus in the past decades, see [10] for a review on bivariate geo-statistical modeling and see [3], where hierarchical Bayesian modeling is discussed for multiple dependent datasets.

Combining multiple sources of data is an increasing field of research due to the large variety of sources of data available nowadays. Data fusion is also part of multivariate modeling and various statistical models have been proposed. Some of the recent studies include the following. In [11] a Bayesian hierarchical model is built to combine model outputs and observed measurements to provide spatial prediction for chemical species. A hidden process is used to represent the unobserved 'true' concentration of sulfur dioxide and the sources of data are affine transformations of this 'true' process. A similar approach was previously used in a space-time context for multiple measurements of snow water equivalent data in [8]. In [17], a hierarchical model based on Spatial Random Effects model is built to combine several outputs of regional climate model output in a spatial framework. In [4], a space-time hierarchical Bayesian model is proposed to fuse measurements and model outputs of air quality data with an extension of a downscaling model introduced some years ago. A hierarchical approach to multivariate spatial modeling and prediction is developed in [22, 23], where the specification of the conditional and a marginal distribution is made instead of specifying the joint distribution, which involves cross-covariances. Indeed the modeling of multivariate covariance structure is challenging and is still an on-going research area, see [1, 12, 5].

Forecast uncertainty can be accounted through ensemble forecasts; however, this strategy is known to be often uncalibrated and under-dispersive. In the context of improving numerical forecasts, statistical methods have been proposed to provide probabilistic forecasts; such methods post-process the single or ensemble forecasts and tend to address the issue of bias and dispersion. These methods are known as model output statistics (MOS) and ensemble model output statistics (EMOS) and are used to identify shortcomings of the raw ensemble from past measurement-forecast pairs. In [21], a finite mixture model named Bayesian model averaging (BMA) is intro-

duced for producing probabilistic forecasts based on ensemble forecasts. In [13], a regression model between the measurements and the members of the ensemble forecast is proposed as a post-processing statistical tool. The assessment of multivariate predictive distribution has been discussed in [15], where tools to assess calibration and sharpness of the predictive distributions are investigated.

In this paper, we propose a bivariate space-time Gaussian process to improve forecasts from an NWP model. The forecasts of wind speed are combined with historical measurements data and provide scenarios of prediction. A particularly important aspect of our model is to account for the space-time dependence between the two datasets. To the best of our knowledge, this dependence is not accounted by the MOS methods proposed in the literature for wind speed. Moreover, in this work we consider a framework where future information from the NWP is used, whereas common MOS methods work with contemporaneous information in space and time. The model is specified in a hierarchical way in order to avoid the characterization of the full space-time bivariate covariance. This specification initially proposed in [22, 23] in a spatial context, is here extended to a space-time modeling.

The paper is organized as follow, in Section 2 we introduce the modeling context and the model. In Section 3, we describe the two sources of data that are used and combined. In Section 4, the model is validated on different months of the year and the quality of space-time prediction at one out-of-sample station is assessed. We highlight the improvements in terms of forecasting accuracy of the proposed model with respect to the NWP forecasts. We conclude with presenting general improvements made by the model with respect to the NWP data in Section 5 and highlight some perspectives to improve the shortcomings of the models.

2. A statistical model for NWP model outputs. In this section, we introduce a Gaussian modeling framework that embeds the space-time dependence between measured observations and NWP model forecasts. In [22, 23], a model is built to conveniently combine spatial data, here we extend this framework to a space-time context.

2.1. Modeling objectives. The modeling context is the following, let us assume that both measured observations Y_{Obs} and NWP forecasts Y_{NWP} are available from time t_1 to time t_k . In the sequel, the term “observations” refers to the observational measurements. Observations are available at J_0 locations $S = \{s_1, \dots, s_{J_0}\}$ and NWP forecasts are available over a grid that covers these stations. The NWP model is run every day for a period of h hours, time can be written in terms of blocks of length h . Henceforth, we

consider a time-window of $h = 24$ hours, and let us denote by b_i the i -th time block of length h , $b_i = \{t_{k_i}, \dots, t_{k_i+h-1}\}$.

The objective here is to predict the measurements Y_{Obs} between time t_{k_K} and t_{k_K+h-1} at the stations $S = \{s_1, \dots, s_{J_0}\}$ and possibly at locations $\{s_{J_0+1}, \dots, s_J\}$ where no historical measurements are recorded, from NWP forecasts that are available between t_{k_K} and t_{k_K+h-1} . This can be summarized by:

$$(2.1) \quad \left(\begin{array}{c} y_{\text{Obs}}^a(b_1; s_1, \dots, s_{J_0}) \\ y_{\text{Obs}}^a(b_2; s_1, \dots, s_{J_0}) \\ \vdots \\ y_{\text{Obs}}^a(b_K; s_1, \dots, s_{J_0}) \\ y_{\text{Obs}}^u(t_{k_K}; s_1, \dots, s_{J_0}, s_{J_0+1}, \dots, s_J) \\ \vdots \\ y_{\text{Obs}}^u(t_{k_K+h-1}; s_1, \dots, s_{J_0}, s_{J_0+1}, \dots, s_J) \end{array} \right) \text{ and } \left(\begin{array}{c} y_{\text{NWP}}^a(b_1; s_1, \dots, s_{J_0}) \\ y_{\text{NWP}}^a(b_2; s_1, \dots, s_{J_0}) \\ \vdots \\ y_{\text{NWP}}^a(b_K; s_1, \dots, s_{J_0}) \\ y_{\text{NWP}}^a(t_{k_K}; s_1, \dots, s_{J_0}, s_{J_0+1}, \dots, s_J) \\ \vdots \\ y_{\text{NWP}}^a(t_{k_K+h-1}; s_1, \dots, s_{J_0}, s_{J_0+1}, \dots, s_J) \end{array} \right),$$

where the super-script ‘‘a’’ stands for available and ‘‘u’’ for unavailable quantities.

In this context the model is trained on the following available pairs:

$$\left((y_{\text{Obs}}^a(b_1; S), y_{\text{NWP}}^a(b_1; S)), (y_{\text{Obs}}^a(b_2; S), y_{\text{NWP}}^a(b_2; S)), \dots, (y_{\text{Obs}}^a(b_K; S), y_{\text{NWP}}^a(b_K; S)) \right),$$

and the prediction is made from $y_{\text{NWP}}^a(b_{K+1}; S, s_{J_0+1}, \dots, s_J)$ to estimate $y_{\text{Obs}}^u(b_{K+1}; S, s_{J_0+1}, \dots, s_J)$, where $b_{K+1} = \{t_{k_K}, \dots, t_{k_K+h-1}\}$. Each day of $h = 24$ hours, the WRF model is run independently from the previous day because WRF is initialized from a reanalysis or assimilated dataset.

In a probabilistic sense, we aim to compute

$$(2.2)$$

$$p(y_{\text{Obs}}^u(b_{K+1}) | y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) =$$

$$(2.3)$$

$$\int p(y_{\text{Obs}}^u(b_{K+1}), \theta | y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_1), \dots, y_{\text{Obs}}^a(b_K), y_{\text{NWP}}^a(b_1), \dots, y_{\text{NWP}}^a(b_K)) \, d\theta$$

where θ is a random set of model parameters, blocks $b_{1:K}$ are available and b_{K+1} is a predicted block, and the spatial components are suppressed for brevity. Note that b_{K+1} is not necessarily a block coming right after b_K , but rather a day that is not observed. To simplify the computation of (2.3) we now make several assumptions. First, we assume (i) that we have approximate independence of $y_{\text{Obs}}^u(b_{K+1})$ on $y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})$ conditional on

$y_{\text{NWP}}^a(b_{K+1})$. In hierarchical models, as ours, which has NWP predictions as its first layer and the observation sites as the second layer, it is common to make the assumption that random variables on the second layer are independent conditional on the realizations of the ones in the first layer, see [9]. This is exactly correct if the additional randomness occurs from the noise of different unrelated sensors. In our case, as we are considering the error of NWP models, the difference between prediction and observations most likely occurs due to features not modeled by NWP. They may be the use of lower resolution or models that have been obtained by some level of space-time homogenization of the physics of the model considered. In this case, the difference is the modeling of subscale noise, which can be assumed to have short temporal correlation scales, see [18]. Moreover, our use of 24h temporal blocks as opposed to every time index would strengthen the validity of approximate conditional independence on NWP simulations of wind realizations at observation sites. The independence of $y_{\text{Obs}}^u(b_{K+1})$ on $y_{\text{NWP}}^a(b_{1:K})$ conditional on $y_{\text{NWP}}^a(b_{K+1})$ may also be a good approximation given the short temporal correlation scales of subscale noise discussed above.

As a result, assumption i implies that the integrand in (2.3) can be approximated as

$$\begin{aligned}
 & p(y_{\text{Obs}}^u(b_{K+1}), \theta | y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) \\
 &= p(y_{\text{Obs}}^u(b_{K+1}) | \theta, y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) p(\theta | y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) \\
 (2.4) \quad & \approx p(y_{\text{Obs}}^u(b_{K+1}) | \theta, y_{\text{NWP}}^a(b_{K+1})) p(\theta | y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K}))
 \end{aligned}$$

In this study, we assume that θ^* can be obtained (ii) maximizing the likelihood

$$\begin{aligned}
 (2.5) \quad & \theta^* = \operatorname{argmax}_{\theta} \mathcal{L}(\theta; y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) = \operatorname{argmax}_{\theta} p(\theta | y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) .
 \end{aligned}$$

With assumptions ($i-ii$), and thus using (2.4) in (2.3) we obtain that

$$\begin{aligned}
 (2.6) \quad & \int p(y_{\text{Obs}}^u(b_{K+1}), \theta | y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_1), \dots, y_{\text{Obs}}^a(b_K), y_{\text{NWP}}^a(b_1), \dots, y_{\text{NWP}}^a(b_K)) \, d\theta \approx \\
 (2.7) \quad & p(y_{\text{Obs}}^u(b_{K+1}) | \theta^*, y_{\text{NWP}}^a(b_{K+1}))
 \end{aligned}$$

In what follows we consider multivariate normal distributions for (2.7). In our approach we have found it productive to model statistically the output

of NWP itself. The way this can be thought of is that NWP is a noisy realization of a latent underlying process NWP^V (which models the evolution of spatially averaged quantities). Then NWP conditional on this NWP^V we then assume to be independent for two different temporal blocks – all temporal correlation between successive blocks is due to NWP^V itself. The same reasoning from above now applies by replacing NWP^V in that discussion. Note that we never forecast the NWP output using the statistical model we develop, we only forecast its relationship to the observations. Thus, there is no need to model explicitly the temporal correlation between different blocks of NWP as long as a sample is produced by the WRF model by a (for the purpose of this paper) black box mechanism which emulates the correct interblock correlation by its relationship to NWP^V . Moreover, if such an assumption does not hold completely, it can only lead to more conservative forecasts. Another way our approach can be thought of is as a regression approach with noisy NWP predictors and the observational unit being one temporal block over the entire geographical area. As a result, the likelihood in (2.5) factorizes in product form for different temporal blocks.

To summarize our approach, for a given statistical model, we first estimate θ^* from the available data (model and observations) using (2.5), then, using (2.7) a predictive distribution is obtained by conditioning only on the NWP predictions for the same temporal block and plugging-in the maximum likelihood estimate θ^* :

$$(2.8) \quad p(y_{\text{Obs}}^u(b_{K+1}) | y_{\text{NWP}}^a(b_{K+1}), y_{\text{Obs}}^a(b_{1:K}), y_{\text{NWP}}^a(b_{1:K})) \approx p(y_{\text{Obs}}^u(b_{K+1}) | \theta^*, y_{\text{NWP}}^a(b_{K+1}))$$

These choices are motivated by computational tractability, by the fact that we assume that the information missed by NWP is a subscale type information, which, as mentioned above, is assumed to have short time correlations conditional on NWP realizations, and by the fact that we do not forecast NWP itself, but rather the relationship between NWP and observations. In Section 2.2 we review a hierarchical approach for Gaussian processes and in Section 2.3 we present the model used for the mean and covariance functions that introduce the parametrization θ .

2.2. The hierarchical bivariate model. Gaussian processes are chosen for their convenience in expressing conditional distributions. As done in other studies, we use the wind speed data directly without any transformation [14, 11, 23]. Moreover, we have not observed a significant departure from normality within the data sets used for this study. Square-root or Box-Cox transformations can be used to preprocess the data, but we do not expect

that will influence the modeling choices. We write the joint distribution of the process $(Y_{\text{Obs}}, Y_{\text{NWP}})$ as:

$$(2.9) \quad \begin{pmatrix} Y_{\text{Obs}} \\ Y_{\text{NWP}} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_{\text{Obs}} \\ \mu_{\text{NWP}} \end{pmatrix}, \begin{pmatrix} \Sigma_{\text{Obs}} & \Sigma_{\text{Obs},\text{NWP}} \\ \Sigma_{\text{Obs},\text{NWP}}^T & \Sigma_{\text{NWP}} \end{pmatrix} \right).$$

The positive-definiteness of block matrices is generally difficult to ensure when specifying the three blocks in (2.9) independently. Therefore, in order to avoid the specification of the full covariance in (2.9), we follow the hierarchical conditional modeling proposed by [22, 23] and we model $(Y_{\text{Obs}}|Y_{\text{NWP}})$ and (Y_{NWP}) , where $(Y_{\text{Obs}}|Y_{\text{NWP}})$ stands for the conditional distribution of Y_{Obs} given Y_{NWP} . When $(Y_{\text{Obs}}, Y_{\text{NWP}})$ is a Gaussian process, $(Y_{\text{Obs}}|Y_{\text{NWP}})$ and (Y_{NWP}) follow a Gaussian distribution, then only first and second order structures are to be specified. Consequently the model is described by the following distributions:

$$(2.10) \quad (Y_{\text{Obs}}|Y_{\text{NWP}}) \sim \mathcal{N}(\mu_{\text{Obs}|NWP}, \Sigma_{\text{Obs}|NWP})$$

a linear dependence between Y_{Obs} and Y_{NWP} agrees reasonably with the data analysis, so we choose the following dependence:

$$(2.11) \quad \mu_{\text{Obs}|NWP} = \text{E}(Y_{\text{Obs}}|Y_{\text{NWP}}) = \mu + \Lambda Y_{\text{NWP}},$$

and

$$(2.12) \quad Y_{\text{NWP}} \sim \mathcal{N}(\mu_{\text{NWP}}, \Sigma_{\text{NWP}}).$$

Finally, from these equations, we express the full joint distribution given by (2.9) as:

$$(2.13) \quad \begin{pmatrix} Y_{\text{Obs}} \\ Y_{\text{NWP}} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu + \Lambda \mu_{\text{NWP}} \\ \mu_{\text{NWP}} \end{pmatrix}, \begin{pmatrix} \Sigma_{\text{Obs}|NWP} + \Lambda \Sigma_{\text{NWP}} \Lambda^T & \Lambda \Sigma_{\text{NWP}} \\ (\Lambda \Sigma_{\text{NWP}})^T & \Sigma_{\text{NWP}} \end{pmatrix} \right).$$

2.3. Statistical model. In order to provide space-time prediction and also to ensure model parsimony, we propose a parameterization in space and time of the involved quantities such that the first and second order structures of the conditional and the marginal distributions defined by (2.10) and (2.12) are specified following an exploratory analysis of the datasets.

2.3.1. *Marginal mean structure of (Y_{NWP}) .* The empirical mean function of Y_{NWP} exhibits spatial patterns associated to the geographical coordinates but also to several parameters of the NWP model due to the large water mass of the lake. (especially, the land-use, that is a categorical variable that represents the type of land used in the parameterization of the NWP model). Time-periodic effects are present in the first order structure of Y_{NWP} and are accounted through harmonics of different frequencies. In Figure 7, these spatial and temporal patterns are plotted. We write:

$$(2.14) \quad \begin{aligned} E(Y_{\text{NWP}}(t, s)) = & \left(\beta_0 + \beta_1 \cos\left(\frac{2\pi t}{24}\right) + \beta_2 \sin\left(\frac{2\pi t}{24}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \right. \\ & \left. \beta_4 \sin\left(\frac{2\pi t}{12}\right) + \beta_5 \cos\left(\frac{2\pi t}{8}\right) + \beta_6 \sin\left(\frac{2\pi t}{8}\right) \right) \\ & \left(\alpha_0^{\text{LU}(s)} + \alpha_1(s) \right), \end{aligned}$$

where t is measured in hour, $\text{LU}(s)$ is an integer that represents the land-use associated to station s used in the model. $(\alpha_l^i)_{l=1, \dots, n}$, with n the number of possible land-use, $(\alpha_1(j))_{j=1, \dots, J_0}$ and $(\beta_k)_{k=0, \dots, 6}$ are real numbers.

2.3.2. *Marginal covariance structure of (Y_{NWP}) .* The block-structure of the space-time covariance of the data suggests expressing wind speed at each station as a linear transformation of an unobserved common signal with added noise, see the top panels of Figure 4. Intuitively we can think of this common signal as an average flow over the studied region. The wind speed at each site is a linear transformation of this average flow. The temporal dynamics of the unobserved signal is modeled with a squared exponential covariance. The following structure is used:

$$Y(b_i, s_j) = L_{s_j} Y_0(b_i) + \epsilon_{s_j}(b_i),$$

where b_i is a temporal window of $h = 24$ lags and s_j the spatial location, L_{s_j} is a $h \times h$ -matrix. The various ϵ_{s_j} are assumed independent from each other and from Y_0 . This model is inspired in part by an earlier study [7], where the L operators were used to represent a known functional relation. In our case, L_{s_j} is a parameterized matrix that is inferred from the data.

The overall covariance of Y has the following structure where:

$$(2.15) \quad \text{cov}(Y(\cdot, s_i), Y(\cdot, s_j)) = (L_{s_i} K_0 L_{s_j}^T) + \delta_{i-j} K_{s_i},$$

where δ stands for the Kronecker symbol and for $j \in \{1, \dots, J_0\}$, the $h \times h$ -matrices K_{s_j} are written as:

$$K_{s_j}[l, k] = a_{s_j} \exp(-b_{s_j} (|t_k - t_l|)^2) + \delta_{k-1} c_{s_j},$$

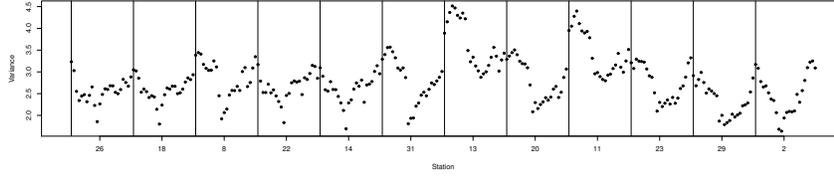


Fig 1: Variance of NWP outputs at every hour and each station of the sub-region C2

and

$$K_0[l, k] = a_0 \exp(-b_0(|t_k - t_l|)^2) + \delta_{k-l}c_0,$$

Following the data analysis, the $h \times h$ -matrices L_{s_j} are parametrized as tridiagonal matrices. Given the study of the variance in Figure 1, the diagonal and off-diagonal quantities are modeled with a quadratic dependence in time and spatially dependent coefficients. The diagonal, sub-diagonal and super-diagonal of the matrix L_{s_j} are written as:

$$\begin{aligned} L_{s_j}[i, i] &= (1 + a_1 \text{Lat}(s_j) + a_2 \text{Long}(s_j)) + (1 + a_3 \text{Lat}(s_j) + a_4 \text{Long}(s_j)) \times i + \\ &\quad (1 + a_5 \text{Lat}(s_j) + a_6 \text{Long}(s_j)) \times i^2, \\ L_{s_j}[i, i - 1] &= (1 + a_7 \text{Lat}(s_j) + a_8 \text{Long}(s_j)) + (1 + a_9 \text{Lat}(s_j) + a_{10} \text{Long}(s_j)) \times i + \\ &\quad (1 + a_{11} \text{Lat}(s_j) + a_{12} \text{Long}(s_j)) \times i^2, \\ L_{s_j}[i, i + 1] &= (1 + a_{13} \text{Lat}(s_j) + a_{14} \text{Long}(s_j)) + (1 + a_{15} \text{Lat}(s_j) + a_{16} \text{Long}(s_j)) \times i + \\ &\quad (1 + a_{17} \text{Lat}(s_j) + a_{18} \text{Long}(s_j)) \times i^2, \end{aligned}$$

for $i \in \{1, \dots, h\}$. We work in relatively small areas and use distances in latitude and longitude here and for the rest of this work.

2.3.3. Conditional mean structure of $(Y_{\text{Obs}}|Y_{\text{NWP}})$. Scatterplots of observations and model outputs suggest that a linear dependence between the variables is reasonable. In [22], several configurations of the transition matrix Λ are proposed depending on its use. For instance, a transition matrix from atmospheric pressure to wind speed is derived from geostrophic equations in [23]. The observations exhibit daily and half-daily periodicity (with various intensities depending on the month of the year) and spatial patterns, see Figure 3, which are accounting in the following. However, the relation between the two datasets does not exhibit significant time-dependence that requires a time-varying dependence. We use spatial and temporal neighbors

to explain the observed wind speed. The land-use (LU) is included in the transition matrix, since it defines different behaviors in the NWP model data. We choose the following transition between the two datasets:

$E(Y_{\text{Obs}}(t, s)|Y_{\text{NWP}}) = \mu(t, s) + (\Lambda Y_{\text{NWP}})(t, s)$, with

$$\mu(t, s) = \left(\beta_0 + \beta_1 \cos\left(\frac{2\pi t}{24}\right) + \beta_2 \sin\left(\frac{2\pi t}{24}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \beta_4 \sin\left(\frac{2\pi t}{12}\right) \right) \left(1 + \beta_5 \text{Lat}(s) + \beta_6 \text{Long}(s) \right),$$

$$(\Lambda Y_{\text{NWP}})(t, s) = \sum_{i=1}^h \alpha^{(\text{LU}(s))}(|t - t_i|) \sum_{k=1}^3 f_k(\Delta \text{Lat}, \Delta \text{Long})(s, s_k) Y_{\text{NWP}}(t_i, s_k), \quad t_1 \leq t \leq t_h,$$

where

- $\alpha^{(\cdot)}(\cdot)$ are temporal weights, parameterized according to $\alpha^{(l)}(\Delta t) = \theta_0^{(l)} \exp(-\theta_1^{(l)}|\Delta t|) + \theta_2^{(l)}$, for the time-difference Δt in $\{0, \dots, h-1\}$; the integer $l \in \{1, \dots, n\}$ is the land-use value of the closest grid point of s ; $\alpha^{(l)}(0) = 1$ for identifiability purpose,
- f_k are linear functions of the differences in latitude and in longitude $\Delta \text{Lat}(s_i, s_j) = |\text{Lat}(s_i) - \text{Lat}(s_j)|$ and $\Delta \text{Long}(s_i, s_j) = |\text{Long}(s_i) - \text{Long}(s_j)|$,
- s_1, s_2, s_3 nearest spatial neighbor grid points of s selected according to the radial distance, but other distances are possible. Moreover, for simplicity we consider here nearest neighbors, but other choices of predictors can be made such as upwind stations for instance.

2.3.4. Conditional covariance structure of $(Y_{\text{Obs}}|Y_{\text{NWP}})$. The analysis of the empirical conditional covariance suggests the use of the parametric shape proposed in (2.15), with a different set of parameters.

2.4. Estimation of the parameters. Maximum likelihood is chosen to estimate the parameters. The likelihood of the model for the observed dataset $y_{\text{Obs}}(t_1, \dots, t_T; s_1, \dots, s_{J_0}), y_{\text{NWP}}(t_1, \dots, t_T; s_1, \dots, s_{J_0})$ is written as:

$$\begin{aligned} \mathcal{L}(\theta; y_{\text{Obs}}(t_1, \dots, t_T; s_1, \dots, s_{J_0}), y_{\text{NWP}}(t_1, \dots, t_T; s_1, \dots, s_{J_0})) \\ = p_{\theta}(y_{\text{Obs}}(t_1, \dots, t_T; s_1, \dots, s_{J_0}), y_{\text{NWP}}(t_1, \dots, t_T; s_1, \dots, s_{J_0})) \\ = p_{\theta}(y_{\text{NWP}}(t_1, \dots, t_T; s_1, \dots, s_{J_0})) p_{\theta}(y_{\text{Obs}}(t_1, \dots, t_T; s_1, \dots, s_{J_0}) | y_{\text{NWP}}(t_1, \dots, t_T; s_1, \dots, s_{J_0})). \end{aligned}$$

This is the particular instantiation of (2.5).

Each day, the WRF model is run independently from the previous day so we consider statistical independence between each day, which leads to the

following product:

$$\begin{aligned} p_\theta(y_{NWP}(t_1, \dots, t_T; s_1, \dots, s_{J_0})) &= \prod_{i=1}^K p_\theta(y_{NWP}(t_{k_i}; \mathbf{S}), \dots, y_{NWP}(t_{k_i+23}; \mathbf{S})) \\ &= \prod_{i=1}^K p_\theta(y_{NWP}(b_i; \mathbf{S})) \end{aligned}$$

where $\mathbf{S} = \{s_1, \dots, s_{J_0}\}$ and $\{b_1, \dots, b_K\} = \{t_1, \dots, t_{24}, t_{25}, \dots, t_T\}$ with $b_i = \{t_{k_i}, \dots, t_{k_i+23}\}$. For each $i \in \{1, \dots, K\}$:

$$\begin{aligned} p_\theta(y_{NWP}(b_i; \mathbf{S})) &= \\ &= \frac{1}{\sqrt{(2\pi)^{J_0} \det(\Sigma_{NWP})}} \exp\left(-\frac{1}{2}(y_{NWP}(b_i; \mathbf{S}) - \mu_{NWP})^T \Sigma_{NWP}^{-1} (y_{NWP}(b_i; \mathbf{S}) - \mu_{NWP})\right), \end{aligned}$$

where μ_{NWP} and Σ_{NWP} are the parametric mean and covariance expressed in (2.14) and (2.15). The log-likelihood associated to the marginal of Y_{NWP} is then expressed as:

$$\begin{aligned} &\log(p_\theta(y_{NWP}(t_1, \dots, t_T; s_1, \dots, s_{J_0}))) \\ &= -\frac{1}{2} \sum_{i=1}^K \left(\log((2\pi)^{J_0}) + \log(\det(\Sigma_{NWP})) + (y_{NWP}(b_i; \mathbf{S}) - \mu_{NWP})^T \Sigma_{NWP}^{-1} (y_{NWP}(b_i; \mathbf{S}) - \mu_{NWP}) \right). \end{aligned}$$

Similarly the conditional distribution is written as:

$$\begin{aligned} &\log(p_\theta(y_{Obs}(t_1, \dots, t_T; s_1, \dots, s_{J_0}) | y_{NWP}(t_1, \dots, t_T; s_1, \dots, s_{J_0}))) \\ &= -\frac{1}{2} \sum_{i=1}^K \left(\log((2\pi)^{J_0}) + \log(\det(\Sigma_{Obs|NWP})) \right. \\ &\quad \left. + (y_{Obs}(b_i; \mathbf{S}) - \mu - \Lambda y_{NWP}(b_i; \mathbf{S}))^T \Sigma_{Obs|NWP}^{-1} (y_{Obs}(b_i; \mathbf{S}) - \mu - \Lambda y_{NWP}(b_i; \mathbf{S})) \right). \end{aligned}$$

In practice, a preliminary least square estimation of the parameters is realized between the empirical and parametric first and second order structures of Y_{Obs} and Y_{NWP} . These estimates are used as initial conditions of the maximum likelihood procedure.

2.5. *Kriging.* Space-time predictions of Y_{Obs} from Y_{NWP} are obtained from the kriging equations, see [28], with the mean and covariance defined by (2.13). For t_0 in $\{t_{k+1}, \dots, t_{k+h}\}$ and s_0 in $\{1, \dots, J_0, J_0 + 1, \dots, J\}$ defined in (2.1),

$$(2.16) \quad (Y_{Obs}(t_0; s_0) | Y_{NWP}(t_{k+1}, \dots, t_{k+h}; 1, \dots, J_0, J_0 + 1, \dots, J)) \sim \mathcal{N}(\hat{\mu}_{Obs}(t_0; s_0), \hat{\Sigma}_{Obs}(t_0; s_0))$$

with

(2.17a)

$$\hat{\mu}_{Obs}(t_0; s_0) = (\mu + \Lambda\mu_{NWP})(t_0; s_0) + c_0^T \Sigma_{NWP}^{-1}((b_K; 1, \dots, J); (b_K; 1, \dots, J))(Y_{NWP} - \mu_{NWP})(b_K; 1, \dots, J),$$

(2.17b)

$$\hat{\Sigma}_{Obs}(t_0; s_0) = \Sigma_{Obs}((t_0; s_0); (t_0; s_0)) + c_0^T \Sigma_{NWP}^{-1}((b_K; 1, \dots, J); (b_K; 1, \dots, J))c_0, \\ c_0 = \Sigma_{Obs, NWP}((t_0; s_0); (b_K; 1, \dots, J)).$$

The distribution (2.16) is used to generate the scenarios of prediction of wind speed in Section 4. This is in fact the predictive distribution presented in (2.8).

3. Wind data. In order to improve forecasts from the considered numerical model, two sources of data are combined, one source consists of ground measurements and the other is the Weather Research and Forecast (WRF) model outputs. The measurement data are recorded across an irregular network and at each observational station, we pick the closest gridded point of NWP outputs. This results in the two datasets having the same number of spatial locations; however, the proposed model is not restricted to this spatial layout and can handle for datasets with different numbers of stations. In the following, the time series of the two datasets are filtered in time by a moving average process over a window of one hour to remove small scale effects and focus on a larger temporal scale and they are picked every hour.

3.1. Direct observations. Observational data are extracted from the Automated Surface Observing System (ASOS) network, they are available at <ftp://ftp.ncdc.noaa.gov/pub/data/asos-onemin>. The network of collecting stations covers the US territory. The studied data are 1-minute data selected over the states of Wisconsin, Illinois, Indiana and Michigan, see Fig. 2. The measured wind speed is discretized in integer knots (on knot is about 0.5 m/s). We do not apply any additional treatment to account for this discretization since the data are filtered over a window of 1 hour, see [25] for discussion about the discretization of wind speed. The orography of this region is simple and flat; however, the presence of Lake Michigan has strong impacts on wind conditions. Several months are investigated and reveal different behaviors, especially periodicities differ from winter to spring and summer months. In the sequel for homogeneity purpose the dataset is sub-divided in 3 spatial clusters depicted in Fig. 2. A spatial clustering

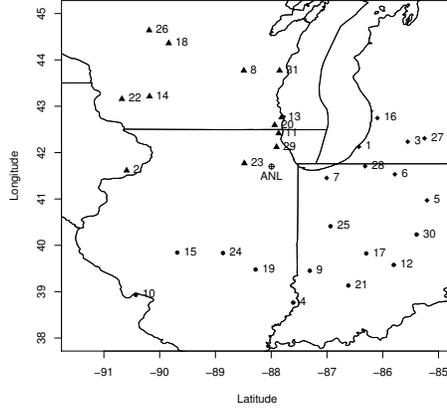


Fig 2: Map of the considered area. Clusters are respectively denoted as C_1 for points represented by \bullet , stations of C_2 are represented by \blacktriangle and C_3 by \blacklozenge . The station represented by \oplus is Argonne National Laboratory, IL, this station will be used for validation in the following.

is performed on wind speed in order to distinguish among different average regional weather conditions. This is a proxy for different NWP forecast behaviors. These three clusters are treated independently hereafter.

3.2. Numerical weather prediction data. State-of-the-art NWP forecasts are generated by using WRF v3.6 ([24]) which is a state-of-the-art numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. WRF has a comprehensive description of the atmospheric physics that includes cloud parameterization, land-surface models, atmosphere-ocean coupling, and broad radiation models. The terrain resolution can go up to 30 seconds of a degree (less than 1 km^2). The NWP forecasts are initialized using North American Regional Reanalysis fields. The NWP forecasts are initialized using North American Regional Reanalysis fields data set that covers the North American continent (160W-20W; 10N-80N) with a resolution of 10 minutes of a degree, 29 pressure levels (1000-100 hPa, excluding the surface), every three hours from 1979 until present. Simulations are started every day during January, May and August 2012 and cover the continental U.S. on a grid of $25 \times 25 \text{ Km}$ with a time resolution of 10 minutes.

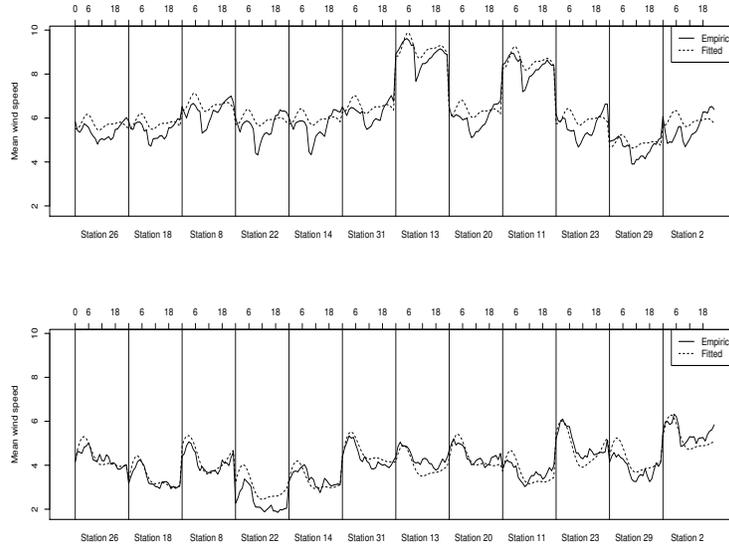


Fig 3: Empirical and fitted parametric mean of wind speed at each hour of a day and at each station in the sub-region C2 in January. Vertical lines separate each station, within each of these windows, each hour of the day is considered. Top panel: mean of Y_{NWP} , bottom panel: mean of Y_{Obs} .

4. Results. In this section, in a first time we make some analysis of the estimated parameters and in a second time we explore qualitatively and quantitatively the capacity of the model to provide accurate forecasts. Three different months of the year (January, May and August) are considered and are studied independently in order to investigate the model performance under different conditions. For each month, the model is trained on two thirds of the month and validated on the remaining third. The two training thirds are rolled over all the possible permutations.

4.1. Analysis of the estimated parameters. In this section, we investigate the maximum likelihood estimation of the mean and covariance of the process. In a first step, the empirical mean and covariance are compared to the fitted parametric ones proposed in Section 2. The mean of the process (Y_{Obs}, Y_{NWP}) is depicted on Figure 3; for each station, the mean at each hour of the day is plotted. The structure of the estimated mean of the two processes is accurately reproduced in terms of temporal and spatial patterns. In Figure 4, the empirical and fitted space-time correlation are plotted. A

great part of the structure is captured by the proposed parametric shapes; however the global shapes tend to be smoothed by the parametric models. The non-separability between space and time that is visible on the empirical off-diagonal blocks is not entirely captured by the parametric model on the top panels. The analysis of the matrices L_s , that are involved in the covariance model (2.15), reveals different configurations given the sub-region and the period of the year, which can be expected since these operators can be interpreted as a linear projector of a process that is common to all the stations. Average air flows are different according the season and the location, the dependence from a common process that would contain this information is likely to differ in space and in time across the year.

The matrix Λ , which appears in both mean and covariance components, is of importance since it links the NWP forecasts to the objective predictive quantities. The analysis of Λ reveals that the intensity of temporal dependence varies with the land-use; however, the temporal persistence is curtailed to a few hours across the different land-uses.

In a second time, the uncertainty associated to the estimation of the parameters is accounted for. Following [27], samples from a normal distribution, with the mean given by the maximum likelihood estimated parameters and the covariance given by the inverse of the hessian of the log-likelihood, are generated. In Figure 5, the maximum likelihood estimation of the parameters and the associated samples are plotted. The parameters that present the most estimation variance are several parameters a_i that appear in the matrices L in Sub-Section 2.3.2. In the parameters of $\mu_{Obs|NWP}$, parameters with a high estimation variance are the ones associated to μ defined in Sub-Section 2.3.3. In these cases, a lack of data in the estimation of these specific parameters may cause this high estimation variance. Notice that the improvement in the predictive variance is inferior to 5% when the uncertainty is accounted in the generation of the predictive scenarios in comparison with the predictive variance when the uncertainty on the parameters is not accounted.

4.2. *Assessment of the quality of the predictive model.* In this part, samples are generated from the predictive distribution defined by Equation (2.16) and are called scenarios or samples in the following. The mean of these samples can be used as a point-wise prediction but the objective here is to embed the uncertainty associated with the prediction by working with samples from the predictive distribution.

4.2.1. *Qualitative exploration of the predictions.* In a first step, as a visual assessment of the prediction, we investigate observed time series and

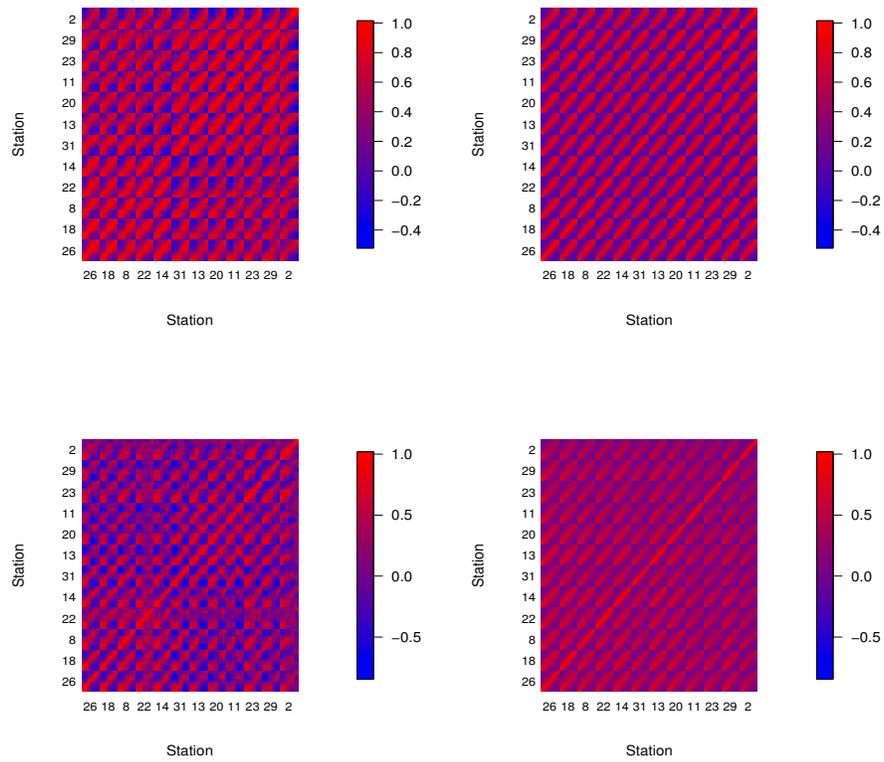


Fig 4: Empirical (left column) and fitted parametric (right-column) space-time correlation estimated in January 2012 in the sub-region C2. Top panels: NWP data and NWP parametric model, bottom panels: empirical conditional correlation and associated fitted ones.

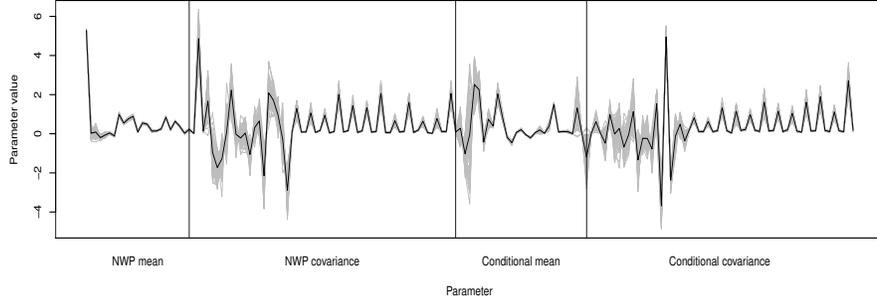


Fig 5: Maximum likelihood estimation of the parameters of the model and associated Gaussian samples. Vertical lines separate the different sets of parameters, from left to right are the parameters of μ_{NWP} , Σ_{NWP} , $\mu_{Obs|NWP}$ and $\Sigma_{Obs|NWP}$.

generated predictive scenarios for a part of the months of January and August, see Figure 6. Measured wind speed, that is to be predicted, is plotted as a reference to evaluate the accuracy of the prediction. NWP wind forecasts are also plotted as they are predictors and as a target to be improved with respect to the measurements. For both months under display, the global trend of the measured time series is well captured by the predictive mean and by the scenarios. The predictive samples cover the measurements that are to be predicted (see left panels) and the predictive mean realizes, most of the time, an improvement with respect to the NWP forecasts. Moreover, each sample has a temporal dynamics consistent with the observed temporal behavior (see right panels). The improvement of the proposed prediction is more visible in August (bottom panels), this is likely due to the periodic components that are stronger in this period of the year and that are well captured by the model, see also Figure 3 described below. Furthermore, the spread of the scenarios is more important in January than in August, this is likely due to the fact that wind speed has more variability in winter as illustrated in the observed variances in Table 1, which makes it less predictable. We note that the scenarios are not spreading at the end of each prediction window as observed in the literature, this is due to the fact that the NWP predictors are available over the entire prediction window and such spread increase is not obvious in the model-measurement discrepancy.

In Figure 7, mean wind speed at each hour of the day is depicted for the measurements, NWP forecasts and forecasts from the model at a station that

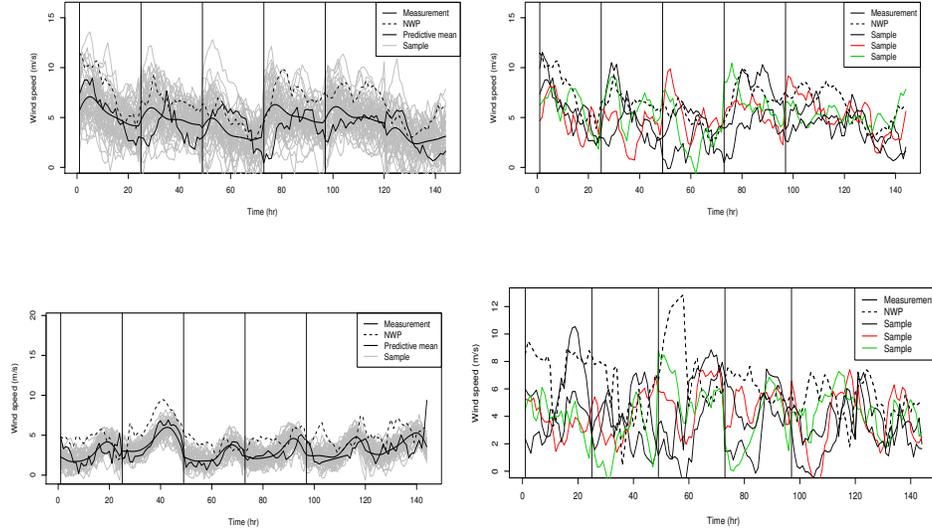


Fig 6: Time series of wind speed at the station with the median RMSE. January 2012 (top) - August 2012 (bottom) for six days. Left panels: 50 predictive samples are plotted, right panel: 3 samples are plotted.

has the median RMSE in sub-region C2. The temporal evolution of the mean differs from the measurements to the NWP data; however, the proposed model is able to compensate for this discrepancy well, which is also visible in time series of Figure 6. In Figure 8 we show the mean wind speed at each station in August. The mean is estimated for the measurements, the NWP forecasts and the predictions from our model. The NWP forecasts show a higher mean than the measurements, especially around the lake, likely due to the parameterization of the NWP model. The proposed predictive model is able to correct this over-estimation and provide a mean consistent with the measured one. Moreover, we note that the spatial structure of measurement is well captured by our forecasts.

The variance of the processes is shown on Fig 9, the proposed model also corrects in space and time the variances that are not well captured by the NWP model. Stations identified as 7 and 9 are near the lake and the variance present in the NWP forecast is consistent with the mean overprediction. In general, the space-time correlation of the NWP forecasts and the model predictions are relatively consistent with the space-time correlation of measurements, which is reflected in part in the performance metrics considered

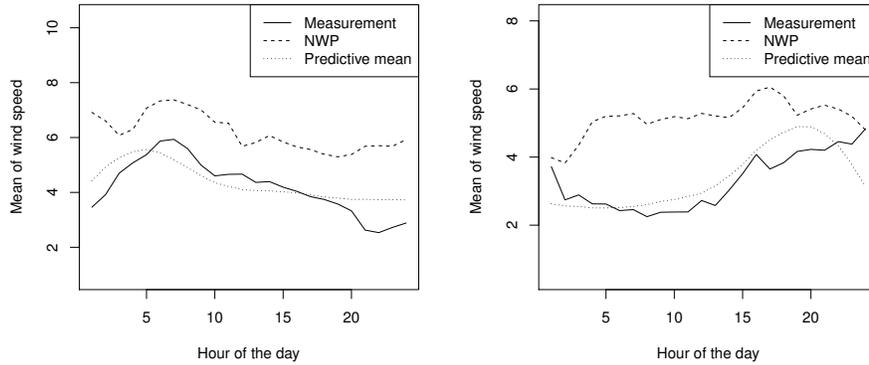


Fig 7: Mean wind speed at each hour of the day in January (left) and August (right), the quantities are plotted for the station with the median RMSE in sub-region C2.

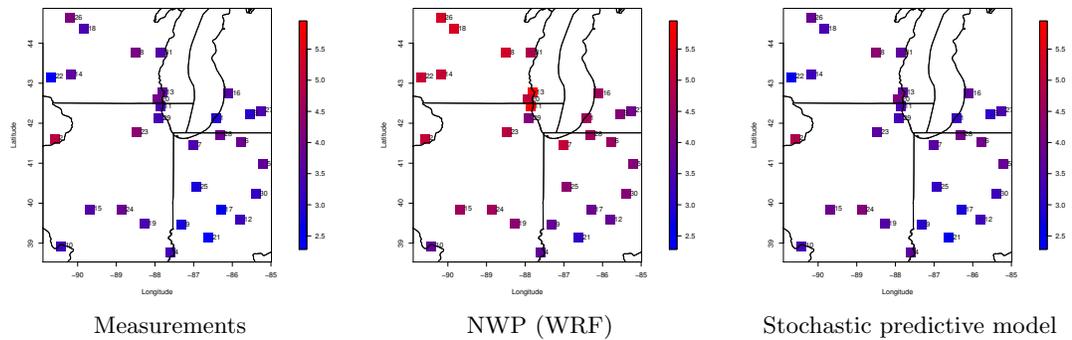


Fig 8: Mean wind speed at each station of the studied area in August.

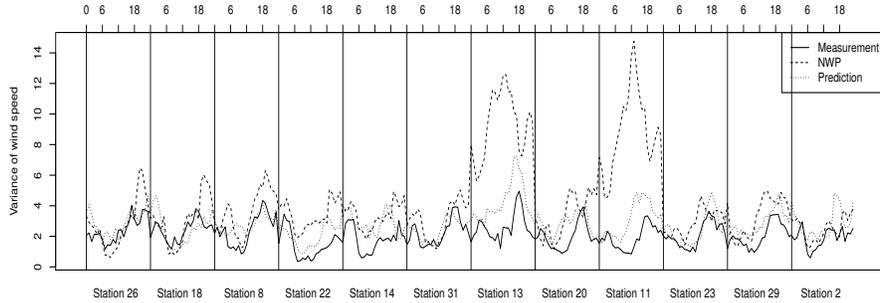


Fig 9: Variance of the 3 processes (measurements, NWP outputs and prediction from the proposed model) at each station and each hour of the day in August in sub-region C2

below.

The spectral content of the scenarios and of the observations is estimated and depicted in Fig. 10, the average spectrum of the estimated spectrum on each sample is also plotted. The estimated spectra of the scenarios cover most of the spectrum of the observations. The overall shape of the estimated spectrum and of the average spectrum indicate a robust agreement, especially in August where small frequencies are accurately captured. In this and other spectral estimates, the spectral content at high frequency is sometimes slightly overpredicted; we believe this is due to the fact that the forecasts do not attempt to correct for discontinuities at the boundaries between temporal blocks. Nevertheless, the features of the spectrum of the measurements appear well captured by our model. Therefore, our model appears to be quite appropriate as a realistic wind scenario generator.

4.2.2. *Quantitative assessment of the quality of the predictions.* In a second step, we study general metrics to assess quantitatively the overall improvement of the model in comparison with the WRF model outputs, see Table 1. In this paper, we study general metrics since there are no specific user-application here; however we expect similar performances when using specific metrics. The root mean square error (RMSE) is computed for the predictive mean of the proposed distribution and for the NWP forecasts. We consider also the energy score (ES) which represents a multivariate generalization of the continuous ranked probability score (CRPS) (see [15, 19]). This metric is an omnibus metric that enables to compare ensemble forecasts and scenarios with point-wise prediction; it is computed on predictive

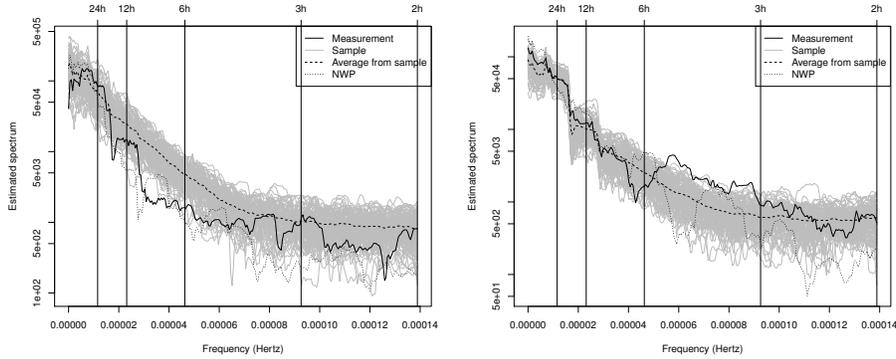


Fig 10: Estimated spectrum in January (left) and August (right) for the station with the median RMSE in sub-region C2.

samples and on NWP forecasts. The energy score is a proper scoring rule, the lower is the energy score the better is the proposed forecast.

On the sub-region C_2 , the model shows the greatest improvement in terms of RMSE and energy score, this is likely due to the presence of Lake Michigan. Indeed, the NWP embeds this presence through the lake mask and land-use but this may be overestimated in comparison with the behaviors of the observations. The improvement of RMSE is more significant in May and August, this is likely due to the periodic components that are well captured by the model as said earlier. The energy score clearly favors the proposed model in comparison to the WRF outputs. The mean of the observations is well captured by the prediction made with the model, the variance is sometimes over-estimated depending the sub-region and the period of the year, but most of the variances are well reproduced.

4.3. *Validation at Argonne station.* We use an independent wind speed dataset collected at Argonne National Laboratory in order to further validate the model. The wind speed at Argonne is predicted by the model fitted on the cluster C_2 , without using this information in the model training. These data are obtained from the weather tower, and are available at the URL <http://www.atmos.anl.gov/ANLMET/>. These data are quality controlled, available every 15 minutes, and each measure is the average of the last 15 minutes-interval data. Predictive scenarios are generated from the space-time prediction distribution (2.16). The RMSE and energy score is available in Table 1. The quality of prediction is very sensitive to the land-use of the NWP predictor, and a spurious land-use may lead to a poor quality forecast

Model	RMSE	Energy score		Mean of Y_{Obs}	Variance of Y_{Obs}
NWP (Jan. 2012, C_1)	1.85	48	Measurements	4.6	5.31
Model (Jan. 2012, C_1)	1.65 (10.9%)	30	Samples	4.58	5.48
NWP (May 2012, C_1)	2.97	77	Measurements	2.87	2.78
Model (May 2012, C_1)	1.9 (36%)	35	Samples	3.14	5.8
NWP (Aug. 2012, C_1)	1.73	44	Measurements	2.49	2
Model (Aug. 2012, C_1)	1.13 (34.9%)	21	Samples	2.55	2.39
NWP (Jan. 2012, C_2)	2.55	66	Measurements	4.32	5.05
Model (Jan. 2012, C_2)	1.81 (29%)	33	Samples	4.55	6.37
NWP (May 2012, C_2)	3.29	85	Measurements	3.56	4.47
Model (May 2012, C_2)	1.86 (43.4%)	35	Samples	3.71	3.83
NWP (Aug. 2012, C_2)	1.9	48	Measurements	2.29	2.48
Model (Aug. 2012, C_2)	1.31 (40.4%)	21	Samples	2.39	2.68
NWP (Jan. 2012, C_3)	2.05	53	Measurements	4.29	5.04
Model (Jan. 2012, C_3)	1.82 (20%)	35	Samples	4.32	9.48
NWP (May 2012, C_3)	2.34	61	Measurements	3.48	3.36
Model (May 2012, C_3)	1.85 (21%)	34	Samples	3.39	6.44
NWP (Aug. 2012, C_3)	1.72	44	Measurements	2.3	2.22
Model (Aug. 2012, C_3)	1.22 (28.9%)	22	Samples	2.31	2.15
NWP (Jan. 2012, ANL)	2.33	60	Measurements	3.5	3.88
Model (Jan. 2012, ANL)	1.59 (31.6%)	31	Samples	3.63	7.17
NWP (May 2012, ANL)	3	78	Measurements	2.87	3.03
Model (May 2012, ANL)	2.06 (31.3%)	38	Samples	3.62	5.47
NWP (Aug. 2012, ANL)	2.02	51	Measurements	2.04	1.09
Model (Aug. 2012, ANL)	1.08 (46.7%)	19	Samples	2.24	2.5

TABLE 1

Statistics and metrics are given for the station that represents the median RMSE in each cluster denoted as C_i , for $i = 1, 2, 3$. They are evaluated on the concerned month for time prediction. Associated to the model RMSE is the percentage of improvement of the model with respect to the NWP data. ANL refers to the station at Argonne National Laboratory that is predicted in space and time from the model trained on the cluster C_2 .

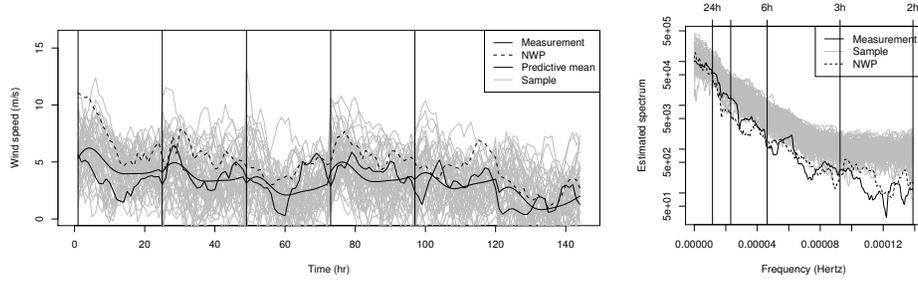


Fig 11: Time series at Argonne in January 2012 and estimated spectrum.

from the model. In Fig. 11, predictive scenarios and measured wind speed at Argonne are plotted, the quality of prediction is good but not as accurate as when only time prediction is made as in Sub-section 4.3. Moreover, wind speed from Argonne is not extracted from the ASOS dataset, which may also lead to some discrepancy in the forecast, due to different recording process.

5. Conclusions. We have introduced a statistical space-time modeling framework for the prediction of atmospheric wind speed based on deterministic numerical weather predictions and historical measurements. We have used a Gaussian multivariate space-time process that combines multiple sources of past physical model outputs and measurements along with model predictions to forecast wind speed at observations sites and at a validation (independent site). We applied this strategy on ground wind speed forecast for a region near the Great Lakes in the US. The results show that the prediction is improved in the mean-squared sense as well as in probabilistic scores. Moreover, the samples are shown to produce realistic wind scenarios based on the sample spectrum. The proposed model enables to correct the first and second order space-time structure of the numerical forecasts to match the structure of the measurements.

Acknowledgements. This work was supported by the US Department of Energy through contract no. DE-AC02-06CH11357. We thank Prof. Michael Stein for comments on multiple versions of this manuscript.

References.

- [1] Apanasovich, T. V. and Genton, M. G.: Cross-covariance functions for multivariate random fields based on latent dimensions, *Biometrika*, 97, 15–30, 2010.
- [2] Bao, L., Gneiting, T., Gritmit, E. P., Guttorp, P., and Raftery, A. E.: Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction, *Monthly Weather Review*, 138, 1811–1821, 2010.
- [3] Berliner, M.: Hierarchical Bayesian modeling in the environmental sciences, *AStA Advances in Statistical Analysis*, 2, 2000.
- [4] Berrocal, V. J., Gelfand, A. E., and Holland, D. M.: Space-Time Data fusion Under Error in Computer Model Output: An Application to Modeling Air Quality, *Biometrics*, 68, 837–848, 2012.
- [5] Bourotte, M., Allard, D., and Porcu, E.: A flexible class of non-separable cross-covariance functions for multivariate space-time data, Preprint, 2015.
- [6] Brown, B. G., Katz, R. W., and Murphy, A. H.: Time series models to simulate and forecast wind speed and wind power, *Journal of climate and applied meteorology*, 23, 1184–1195, 1984.
- [7] Constantinescu, E. and Anitescu, M.: Physics-based covariance models for Gaussian processes with multiple outputs, *International Journal for Uncertainty Quantification*, 3, 47–71, 2013.
- [8] Cowles, M. K., Zimmerman, D. L., Christ, A., and McGinnis, D. L.: Combining snow water equivalent data from multiple sources to estimate spatio-temporal trends and compare measurement systems, *Journal of agricultural, biological, and environmental statistics*, 7, 536–557, 2002.
- [9] Cressie, N. and Wikle, C. K.: *Statistics for spatio-temporal data*, John Wiley & Sons, 2011.
- [10] Fanshawe, T. R. and Diggle, P. J.: Bivariate geostatistical modelling: a review and an application to spatial variation in radon concentrations, *Environmental and ecological statistics*, 19, 139–160, 2012.
- [11] Fuentes, M. and Raftery, A. E.: Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models, *Biometrics*, 61, 36–45, 2005.
- [12] Genton, M. G. and Kleiber, W.: Cross-covariance functions for multivariate geostatistics, *Stat Sci*, 2014.
- [13] Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, *Monthly Weather Review*, 133, 1098–1118, 2005.
- [14] Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E.: Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method, *Journal of the American Statistical Association*, 101, 968–979, 2006.
- [15] Gneiting, T., Stanberry, L. I., Gritmit, E. P., Held, L., and Johnson, N. A.: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds, *Test*, 17, 211–235, 2008.
- [16] Hering, A. S. and Genton, M. G.: Powering up with space-time wind forecasting, *Journal of the American Statistical Association*, 105, 92–104, 2010.
- [17] Kang, E. L. ., Cressie, N., and Sain, S. R.: Combining outputs from the North American regional climate change assessment program by using a Bayesian hierarchical model, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61, 291–313, 2012.

- [18] Majda, A. and Wang, X.: Nonlinear dynamics and statistical theories for basic geophysical flows, Cambridge University Press, 2006.
- [19] Pinson, P. and Girard, R.: Evaluating the quality of scenarios of short-term wind power generation, *Applied Energy*, 96, 12–20, 2012.
- [20] Pinson, P., Christensen, L. E. A., Madsen, H., Sorensen, P. E., Donovan, M. H., and E., J. L.: Regime-switching modelling of the fluctuations of offshore wind generation, *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 2327–2347, 2008.
- [21] Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review*, 133, 1155–1174, 2005.
- [22] Royle, J. and Berliner, L.: A hierarchical approach to multivariate spatial modeling and prediction, *Journal of Agricultural, Biological, and Environmental Statistics*, pp. 29–56, 1999.
- [23] Royle, J., Berliner, L., Wikle, C., and Milliff, R.: A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea, in: *Case Studies in Bayesian Statistics*, pp. 367–382, Springer, 1999.
- [24] Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X.-Y., Wang, W., and Powers, J.: A description of the Advanced Research WRF Version 3, Tech. Rep. Tech Notes-475+ STR, NCAR, 2008.
- [25] Sloughter, J. M. L., Gneiting, T., and Raftery, A. E.: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, *Journal of the American Statistical Association*, 105, 25–35, 2010.
- [26] Sloughter, J. M. L., Gneiting, T., and Raftery, A. E.: Probabilistic wind vector forecasting using ensembles and Bayesian model averaging, *Monthly Weather Review*, 141, 2107–2119, 2013.
- [27] Stein, M. L.: Spatial interpolation of high-frequency monitoring data, *The Annals of Applied Statistics*, pp. 272–291, 2009.
- [28] Stein, M. L.: *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media, 2012.

Government License The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

ADDRESS OF THE FIRST AND SECOND AUTHORS
ARGONNE NATIONAL LABORATORY
9700 SOUTH CASS AVENUE
LEMONT, IL, 60439-4844
E-MAIL: jbessac@anl.gov
emconsta@mcs.anl.gov

ADDRESS OF THE THIRD AUTHOR
ARGONNE NATIONAL LABORATORY
9700 SOUTH CASS AVENUE
LEMONT, IL, 60439-4844
E-MAIL: anitescu@mcs.anl.gov