

Economic Impacts of Wind Covariance Estimation on Power Grid Operations

Cosmin G. Petra, Victor M. Zavala, Elias D. Nino-Ruiz, and Mihai Anitescu

Preprint ANL/MCS-P5148-0614

Abstract—We study the impact of capturing spatiotemporal correlations between multiple wind supply points on economic dispatch procedures. Using a simple dispatch model, we first show analytically that over/underestimation of correlation leads to positive and negative biases of dispatch cost, respectively. A rigorous, large-scale computational study for the State of Illinois transmission grid with real topology and physical constraints reveals similar conclusions. For this study, we use the Rao-Blackwell-Ledoit-Wolf estimator to approximate the wind covariance matrix from a small number of wind samples generated with the numerical weather prediction model WRF and we use the covariance information to generate a large number of wind scenarios. The resulting stochastic dispatch problems are solved by using the interior-point solver PIPS-IPM on the BlueGene/Q (Mira) supercomputer at Argonne National Laboratory. We find that strong and persistent biases result from neglecting correlation information and indicate to the need to design a market that coordinates weather forecasts and uncertainty characterizations.

Index Terms—covariance, correlation, spatiotemporal, estimation, uncertainty, wind power, dispatch

I. INTRODUCTION

Achieving efficient grid operations in the presence of intermittent renewable power is a challenge because these supply sources follow complex spatiotemporal patterns (see Figure 1) that extend over wide geographical regions (e.g., tens to hundreds of kilometers) and long periods of time (i.e., hours to days). Reserve allocation procedures therefore can be ineffective, and more adaptive and systematic approaches based on stochastic and robust optimization techniques are needed.

Stochastic and robust optimization techniques rely on uncertainty characterizations. Correlation (or covariance) information, in particular, is key because this guides forecast aggregation/disaggregation procedures and because it is needed to characterize risk in dispatch cost and revenues of market players. For instance, if the supply of a wind farm in a region is uncorrelated from that in another region, these can be forecasted independently without affecting dispatch cost. When correlations exist, however, one would expect that using independent forecasts will introduce errors in the uncertainty characterization and this will bias dispatch cost and shift incentives of the players (wind power suppliers, suppliers, and consumers). This situation was hypothesized in the stochastic market setting of Pritchard and coworkers [9] (see Section 5). Before the present study,

C. G. Petra, V. M. Zavala, and M. Anitescu are with the Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. E-mail: {petra, vzavala, anitescu}@mcs.anl.gov.

E. D. Nino-Ruiz is with the Department of Computer Science, Virginia Tech, Blacksburg, VA 24061. E-mail: enino@vt.edu

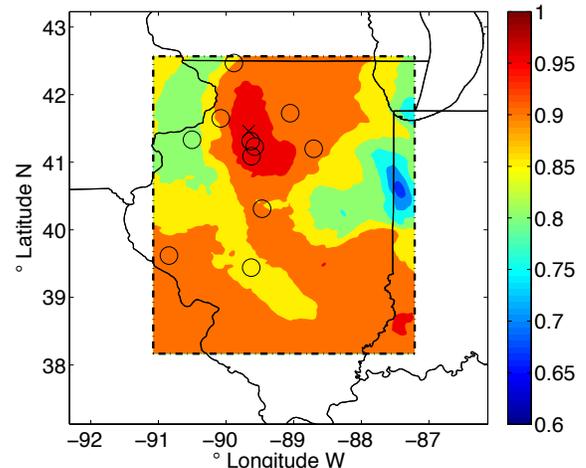


Fig. 1. Spatial correlation for wind speed in the State of Illinois.

however, no evidence existed to show that this is indeed the case and how large the cost bias would be.

From a market implementation point of view, determining long-range correlations is a challenge because wind farm or solar power plant owners might not be willing to share their forecasting procedures and their site information with other markets players and the ISO. Consequently, they might prefer to construct their own forecasts and uncertainty levels, possibly neglecting correlations with other sites. Owners might also need to ignore long-range correlations because of computational limitations faced by their forecasting vendors. Note also that keeping forecasting information confidential provides a mechanism for manipulation under a stochastic market setting (e.g., a supplier overestimates its uncertainty). Computational challenges arise because properly resolving the space-time resolution of numerical weather prediction (NWP) systems requires significant computational power [2]. NWP systems are extremely computationally intensive, and few computing sites exist in the world that can obtain forecasts that accurately capture both short-range conditions at the supply site and long-range behavior. In other words, limits exist on the resolution of uncertainty characterizations achievable, and these limits lead to ambiguity.

Properly designing decentralized markets that factor in uncertainty in weather-driven supply is necessary, but this requires significantly more complex information exchange mechanisms between the ISO and market players compared with existing deterministic settings [9], [16]. To design such information exchange mechanisms, one needs to understand the effects that certain information has on performance. In this work, we study the effect of long-range correlation information on dispatch cost

in the presence of wind power. We first present an analytic example to prove that correlation between suppliers output can positively or negatively bias dispatch cost (depending on the correlation direction). We then perform a detailed computational study using a stochastic economic dispatch setting in the Illinois transmission system. We use validated wind speed ensemble forecasts obtained with the NWP system WRF that are propagated through typical wind power curves in order to obtain wind power ensembles. We use the Rao-Blackwell-Ledoit-Wolf estimator to generate scenarios from the limited number of WRF samples available, and we demonstrate the efficacy of this estimator. Our computational study reveals that dispatch cost biases that would scale up to an order of hundreds of millions of dollars a year (if similar correlation patterns across the year hold) can be introduced by ignoring long-range correlation. We also show that confidence levels of dispatch cost differ significantly from the actual ones when correlations are neglected. In our study, the confidence intervals when ignoring correlations were narrower, thus underestimating the number scenarios necessary to close the gap. Our study thus indicates that, as hypothesized in the stochastic market setting of [9], centralized forecasts that can properly account for correlations are superior to localized ones when used for constructing wind power bids by suppliers in markets with significant wind power penetration.

The paper is structured as follows. In Section II we present a motivating example to illustrate the effect of correlation information on dispatch cost. In Section III we present a detailed computational study using data for the state of Illinois transmission system. This section describes the dispatch model, the scenario generation procedure, the covariance estimator, and the numerical results. Section IV presents concluding remarks.

II. MOTIVATING ANALYTICAL EXAMPLE

Consider a single-node system with three suppliers and one demand. The first two suppliers, G_1 and G_2 , have uncertain power output and the outputs follow Gaussian distributions, $\mathcal{N}(w_1, \sigma_1)$ and $\mathcal{N}(w_2, \sigma_2)$. We define $\rho \in [-1, 1]$ as the correlation coefficient and assume that both suppliers have a cost p_w . The third supplier, G_3 , is assumed to be deterministic; this supplies power at cost p_{th} with $p_{th} > p_w$ and has infinite capacity. The demand quantity, d , is assumed to be deterministic and inelastic.

By construction, one can deduce that as much cheap power as possible should be produced. If this does not satisfy all demand, then G_3 will be dispatched to fulfill the remaining demand. Consequently, the negative dispatch cost is

$$c_d = \mathbb{E} [p_w \min(X_1 + X_2, d) + p_{th} \max(d - X_1 - X_2, 0)]. \quad (1)$$

To show the dependence $c_d = c_d(\rho)$, we write (1) as follows:

$$\begin{aligned} c_d &= \mathbb{E} [p_w d + p_w \min(X_1 + X_2 - d, 0) \\ &\quad + p_{th} \max(d - X_1 - X_2, 0)] \\ &= p_w d + \mathbb{E} [-p_w \max(d - X_1 - X_2, 0) \\ &\quad + p_{th} \max(d - X_1 - X_2, 0)] \\ &= p_w d + \mathbb{E} [(p_{th} - p_w) \max(d - X_1 - X_2, 0)] \\ &= p_w d + (p_{th} - p_w) \mathbb{E} [d - (X_1 + X_2) | X_1 + X_2 \leq d]. \end{aligned} \quad (2)$$

Here, $\mathbb{E}[X|Y]$ denotes the expectation of X conditional on event Y . Furthermore, since the random variable $X = X_1 + X_2$ is normally distributed, $X \sim \mathcal{N}(\mu, \sigma)$, where

$$\begin{aligned} \sigma &= \sigma(\rho) \\ &= \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2} \text{ and } \mu = w_1 + w_2, \end{aligned} \quad (3)$$

c_d can be expressed

$$\begin{aligned} c_d &= p_w d + (p_{th} - p_w) \mathbb{E} [d - X | X \leq d] \\ &= p_w d + (p_{th} - p_w) \cdot d \cdot \Phi(d, \sigma) - (p_{th} - p_w) \mathbb{E} [X | X \leq d], \end{aligned} \quad (4)$$

where Φ is the cumulative density function of X ,

$$\Phi(x, \sigma) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right]. \quad (5)$$

We denote the probability density function of Y by $\phi(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$. One can show (see Lemma 1 from the Appendix) that

$$\mathbb{E} [X | X \leq d] = -\sigma^2 \phi(d, \sigma) + \mu \Phi(d, \sigma). \quad (6)$$

Combining this equation with (4), we obtain as the

$$c_d(\sigma) = p_w d + (p_{th} - p_w) ((d - \mu)\Phi(d, \sigma) + \sigma^2 \phi(d, \sigma)). \quad (7)$$

The dependence of the dispatch cost c_d on the correlation parameter ρ becomes clear and is illustrated in Figure II. We also have the following result:

Proposition 1: The dispatch cost $c_d(\rho)$ is a strictly increasing function of the correlation parameter ρ .

Proof: See the Appendix.

This result shows that there can be a positive or negative bias (depending on the correlation direction) between dispatch cost values computed from different correlation structures (different uncertainty characterizations). This can be interpreted as a dispatch cost bias introduced by an error in the correlation coefficient resulting from computational or market implementation limitations.

With increasing ρ , the probability distribution of the total available wind $X = X_1 + X_2$ exhibits an increasing variance (see (3)). Consequently, it is less likely to satisfy demand using the cheap uncertain supply and more likely to dispatch the more expensive supply, thus resulting in a higher cost. For instance, if two suppliers are fully positively correlated ($\rho = 1$), a low output of one will result in a low output of the other. In the extreme case of full negative correlation ($\rho = -1$) cheap uncertain supply will most likely be used because if the output of one wind farm drops, then this implies that the output of the other one increases, so at least one wind farm is always active. From Figure II we note also that, at low and high demands, correlation has less effect on the dispatch cost.

From a purely dispatch cost perspective, one can argue that an inaccurate forecasting system that overestimates and underestimates the correlations may lead to the same expected dispatch cost in the long run, compared with that obtained with an ideal forecasting system. While we do not refute this possibility, we point out that inaccurate forecasting also

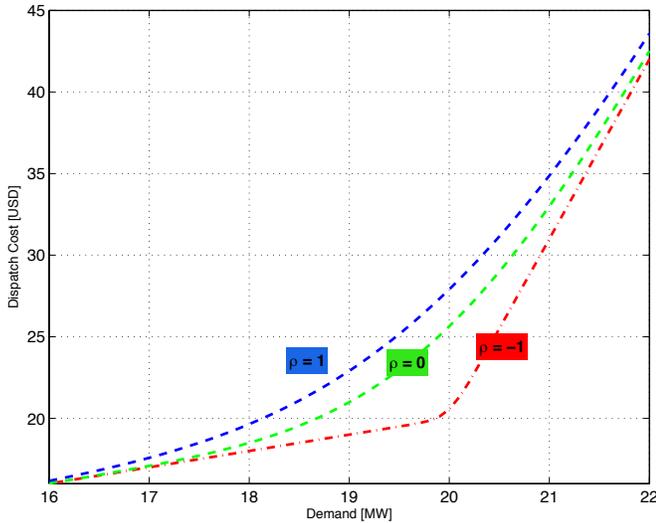


Fig. 2. Dispatch cost c_d as a function of demand d for negative correlation ($\rho = -1$), no correlation ($\rho = 0$), and positive correlation ($\rho = 1$).

results in gaps in prices and revenues and, hence in *market inefficiencies*. These inefficiencies will bias incentives toward certain players in unpredictable ways that might not average out in the long run.

Errors in correlation between uncertain suppliers can occur for various reasons. For example, the owners of the wind farms might submit their own forecasts and scenarios, which ISO will treat independently when dispatching. Such a situation is discussed in [10]. Alternatively, correlations or the covariance matrix may be poorly estimated in the NWP systems because of the prohibitive computational cost of accurate estimators. Hence, we advocate the use of ISO-centric weather forecasting systems of atmospheric conditions while mapping to wind power forecasts (that capture local and wind turbine effects) can be done internally by wind farm owners in order to prevent disclosure of information. The numerical simulations of the next section indicate that savings of more than \$10,000 per dispatch period can be obtained for full-scale power grids, such as the one of the State of Illinois. The savings over one year can reach a hundred million dollars, assuming a comparable amount of savings per dispatch period in the rest of the year.

III. DETAILED COMPUTATIONAL STUDY

In this section, we present a detailed computational framework to analyze the effects of correlation on dispatch cost and prices. The framework comprises of a stochastic dispatch model, numerical weather prediction, and covariance estimation and scenario generation. Our setting includes realistic data for the Illinois transmission system, and we assume a wind adoption scenario of 17%.

A. Stochastic Dispatch Model

In our analysis we use the stochastic dispatch model from [10] that captures forward and spot components. The model assumes that the market is cleared at the current time using predictions of the uncertain conditions at the next time period.

When uncertainty is realized, the suppliers are allowed to sell additional power or buyback under the realized conditions. The formulation is given by:

$$\begin{aligned} \min_{x_i, X_i(\omega)} & \\ & \sum_{i \in \mathcal{G}} (p_i x_i + \mathbb{E}_\omega [p_i^+ (X_i(\omega) - x_i)_+ - p_i^- (X_i(\omega) - x_i)_-]) \end{aligned} \quad (8a)$$

s.t.

$$\tau_n(f) + \sum_{i \in \mathcal{G}(n)} x_i = d_n, \quad n \in \mathcal{N} \quad (8b)$$

$$\tau_n(F(\omega)) - \tau_n(f) + \sum_{i \in \mathcal{G}(n)} (X_i(\omega) - x_i) = 0, \quad n \in \mathcal{N}, \omega \in \Omega \quad (8c)$$

$$f, F(\omega) \in \mathcal{U}, \omega \in \Omega \quad (8d)$$

$$(x_i, X_i(\omega)) \in \mathcal{C}_i(\omega), \quad i \in \mathcal{G}, \omega \in \Omega \quad (8e)$$

Here, \mathcal{N} denotes the set of nodes (buses) and \mathcal{L} the set of transmission lines. The set of all suppliers is denoted by \mathcal{G} . Subsets $\mathcal{G}(n)$ denote the set of players connected to node n . The forward dispatched quantities for players are x_i , and the spot quantities under scenario ω are $X_i(\omega)$. The forward power flow through line $\ell \in \mathcal{L}$ is denoted by f_ℓ , and f denotes the vector of all line flows. Similarly, $F(\omega)$ denotes the vector of line flows $F_\ell(\omega)$ for each scenario ω . The demand is assumed to be deterministic and inelastic and is represented by d_n , $n \in \mathcal{N}$.

The scenarios ω characterize the randomness in the model due to unpredictable capacities and are mathematically expressed as random vectors defined on some probability space (Ω, \mathcal{F}, P) . The expectation \mathbb{E}_ω is taken with respect to the measure P . In practice, one considers a finite approximation of Ω obtained through sampling.

The objective consists of minimizing the forward dispatch cost $\sum_{i \in \mathcal{G}} p_i x_i$ plus the expected adjustment or recourse dispatch cost $\sum_{i \in \mathcal{G}} \mathbb{E}_\omega [p_i^+ (X_i(\omega) - x_i)_+ - p_i^- (X_i(\omega) - x_i)_-]$ specific to individual scenarios. Here $[y]_+ = \max\{y, 0\}$ and $[y]_- = \max\{-y, 0\}$. The coefficients p_i denote the bid price, and p_i^+ and p_i^- are price bids for real-time corrections of the generators. A supplier i asks $p_i^+ > p_i$ to sell additional power or asks $p_i^- < p_i$ to buy power from the system (e.g., reduces output). In our model we have used $p_i^- = 0.8p_i$ and $p_i^+ = 1.2p_i$.

The decision variables are the dispatch quantities for each generator, x_i , $X_i(\omega)$ and the power flows f_ℓ and $F(\omega)$. The forward dispatches x_i are “ahead” decisions that accounts for randomness; the spot redispatches $X_i(\omega)$ represent “real-time” decisions that are appropriate corrections once an individual realization ω of the randomness is observed.

Function $\tau_n(\cdot)$ is a mapping of the flow vector to the node n . We denote by $\nu_1(n)$ the inflow lines into node $n \in \mathcal{N}$ and by $\nu_0(n)$ the outflow lines. Equation (8b) describes the power flow through a node $n \in \mathcal{N}$ which is the sum of power $\tau_n(f) = \sum_{l \in \nu_1(n)} f_l - \sum_{l \in \nu_0(n)} f_l$ imported via the transmission lines to node n and power $\sum_{i \in \mathcal{G}(n)} x_i$ produced at node n . Equation (8c) is the second-stage correspondent of (8b), enforcing power flow balance at each node for each scenario ω . It is shown in [10] that the multipliers associated with this “residual” formulation (not with the simpler equivalent form,

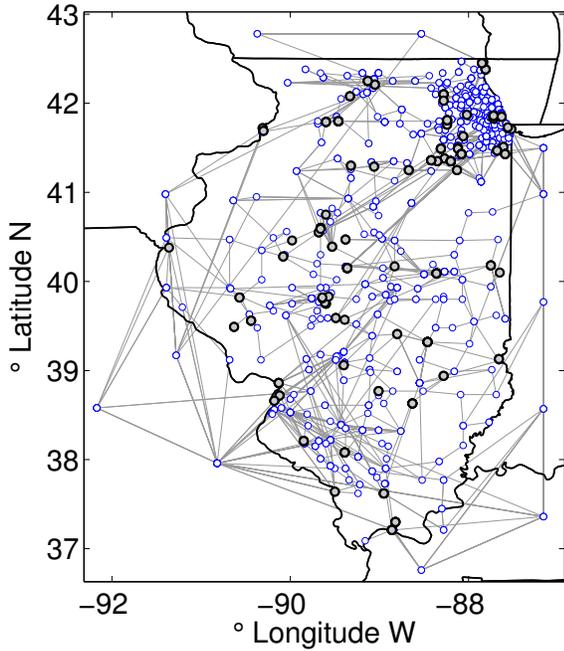


Fig. 3. Illinois transmission system.

$\tau_n(F(\omega)) + \sum_{i \in \mathcal{G}(n)} X_i(\omega) = 0$) gives the clearing prices to be used in the spot (or real-time) market. Equation (8d) represents maximum flow constraints on individual lines, with \mathcal{U} usually being a polyhedron.

Equation (8e) expresses constraints on supply that come from technological limits of the generators (such as maximum/minimum capacity, limited ramp-up/down power on short notice) and intermittent availability of energy of some generators. In our model, \mathcal{C}_i is deterministic for thermal generators (natural gas, coal, heater oil and nuclear) and given by $\mathcal{C}_i = \{(x_i, X_i) : x_i, X_i \in [m_i, M_i], |x_i - X_i| \leq r_i\}$, expressing capacity and ramp constraints. The uncertain output is modeled by $\mathcal{C}_i(\omega) = \{(x_i, X_i(\omega)) : x_i \in [0, M_i], X_i(\omega) \in [0, W_i(\omega)]\}$, showing that the forward dispatch x_i can be allowed to reach maximum installed capacity and the spot dispatch $X_i(\omega)$ can only be allowed to reach maximum power generated under scenario ω . We consider ramp constraints with only one time step.

Our model was set up for the State of Illinois power grid which comprises 2,522 lines, 1,908 buses, 870 load buses and 225 generators. The topology of the network is presented in Figure 3. To obtain a large wind power installed capacity (approximately 17%), we needed to create synthetic wind farms in addition to the existing ones. The synthetic farms were chosen to replace some existing coal or gas generators. This approach was taken specifically to avoid possible network congestion that would limit the amount of real wind adoption. In addition, we replaced only thermal generators that were mirrored by other (usually identically) generators in order to ensure that enough thermal generation was available to satisfy demand in the low-wind scenarios. The generation cost for the wind farm was set to 5\$/MW, the lowest across all generators.

B. Wind Scenario Generation

Wind direction and speed samples required for our study are obtained from WRF. The WRF model [12] is a state-of-the-art numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. WRF is the result of a multi-agency and university effort to build a highly parallelizable code that can run across scales ranging from large-eddy to global simulations. The comprehensive description of the atmospheric physics includes cloud parameterization, land-surface models, atmosphere-ocean coupling, and broad radiation models.

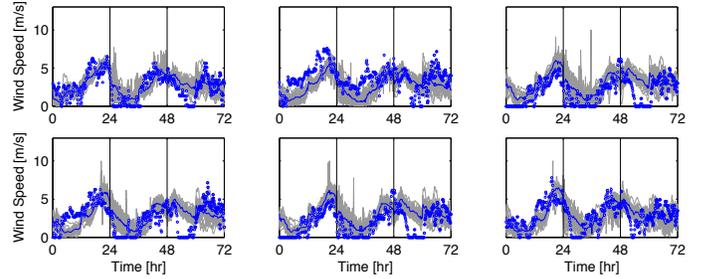


Fig. 4. Wind speed realizations for 6 wind farm locations in Illinois and observations (dots) at nearest meteorological stations. Vertical lines represent beginning of day (12:00 AM).

We set up a computational nested domain structure for WRF including a high-resolution sector that covers the State of Illinois and two additional domains of larger coverage but lower resolution that provide boundary conditions for the nested domains [2]. This is illustrated in Figure 5. The initial conditions from the assimilated state (also known as reanalyzed state) of the atmosphere are randomly perturbed and propagated in time through the nonlinear NWP model to obtain a set of ensembles that describe possible trajectories of the atmospheric conditions. The computational cost of this procedure is significant. Computing a single ensemble for Illinois over a 24-hour time window takes 6 hours of wall-clock time running on 32 processors. Because of these limitations, we computed only 30 ensembles for June 4, 2006 [2]. The ensembles have been validated using observations obtained from the National Climatic Data Center. The ensembles for six different wind farm locations are shown in Figure 4.

Each ensemble provides the components of the wind velocity that are transformed to wind speed. This gives a 3D field in geographical coordinates (latitude, longitude, and elevation) evolving over time where the field points match the discretization mesh in the inner domain. The wind farm locations, however, do not match the discretization mesh. In addition, the typical hub height used for wind farms (80 meters) may also not match the WRF vertical layers. To remedy this issue, we use linear and bilinear interpolation to compute wind speeds at the farm locations from the WRF ensembles, therefore obtaining a set of 30 ensembles for the speeds at the desired 3D coordinates. We then compute the sample mean \bar{x} , sample covariance \mathbf{S} , and the Rao-Blackwell-Ledoit-Wolf (RBLW) estimator $\hat{\Sigma}_{RBLW}$ as described in the next section. Using this approach, we can compute many wind speed scenarios by sampling a multivariate Gaussian with mean \bar{x} and estimated covariance matrix

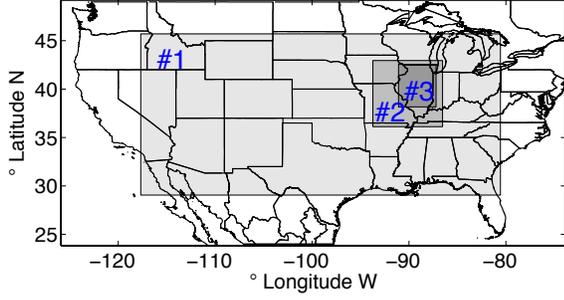


Fig. 5. WRF simulation domain with resolutions of #1 - 32 km², #2 - 6 km², and #3 - 2 km².

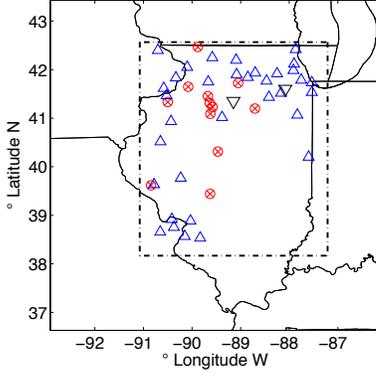


Fig. 6. Locations of subset of wind farms (red) and meteorological stations (blue).

$\widehat{\Sigma}_{RBLW}$.

Our computational framework is not restricted to the use of WRF and RBLW covariance estimators and can accommodate scenarios generated with different scenario generation techniques for wind speed, assuming that these can handle the large spatial coverage presented in our computational study. Other alternative scenario generation techniques include copula transformation [14], [15], probabilistic models [13], and time-recursive estimation of covariance matrix [8].

C. Covariance Estimation Methods and Validation

The simplest estimator for the covariance using the WRF samples would be the sample covariance matrix. In our setting, however, this matrix is rank-deficient and therefore noninvertible since the number of WRF samples is smaller than the number of the wind farms. Because of this, we cannot factorize it and sample from it. In addition, the sample covariance estimator can be unreliable when few samples are available. Consequently, we investigate the use of other estimators and compare their performance.

Mathematically, the problem of covariance estimation can be expressed as the problem of estimating the matrix $\Sigma \in \mathbb{R}^{p \times p}$ of a random vector $\mathbf{y} \in \mathbb{R}^p$ based on a set of realizations or samples $\{\mathbf{x}_i\}_{i=1}^n$, for the case when n (number of samples) is smaller than p (dimension of uncertainty space). In our case, p is the number of wind supply points. Let $\bar{\mathbf{x}} = \frac{1}{n} \cdot \sum_{i=1}^n \mathbf{x}_i \in \mathbb{R}^p$ denote the sample mean and $\mathbf{S} = \frac{1}{n} \cdot \Delta \mathbf{X} \cdot \Delta \mathbf{X}^T \in \mathbb{R}^{p \times p}$ denote the sample covariance matrix, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in$

$\mathbb{R}^{p \times n}$ and $\Delta \mathbf{X} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}] \in \mathbb{R}^{p \times n}$.

Covariance estimators for the case $p > n$ use regularizations of the sample covariance matrix \mathbf{S} to overcome the rank deficiency. One of the most common regularizations is the perturbation of the covariance matrix by a multiple of the identity, which leads to shrinkage estimators of the type

$$\widehat{\Sigma} = \alpha \cdot \mathbf{I} + \beta \cdot \mathbf{S}, \quad (9)$$

where $\mathbf{I} \in \mathbb{R}^{p \times p}$ is the identity matrix and α and β are regularization parameters or weights chosen to minimize the estimation error $\|\widehat{\Sigma} - \Sigma\|$, where $\|\cdot\|$ usually refers to the Frobenius norm. Since the real covariance matrix Σ is unknown, different techniques are used to compute statistical estimates for α and β .

Ledoit and Wolf [3] propose the estimator $\widehat{\Sigma}_{LW}$ given by

$$\widehat{\Sigma}_{LW} = \frac{b^2}{d^2} \cdot m \cdot \mathbf{I} + \frac{a^2}{d^2} \cdot \mathbf{S},$$

where

$$m = \text{tr}(\mathbf{S} \cdot \mathbf{I})/p, \quad d^2 = \text{tr}([\mathbf{S} - m \cdot \mathbf{I}] \cdot [\mathbf{S} - m \cdot \mathbf{I}]^t)/p, \quad (10)$$

$$b^2 = \min \left(\frac{1}{n} \cdot \sum_{k=1}^n \frac{\text{tr}([\mathbf{x}_k \cdot \mathbf{x}_k^t - \mathbf{S}] \cdot [\mathbf{x}_k \cdot \mathbf{x}_k^t - \mathbf{S}]^t)}{p}, d^2 \right), \quad (11)$$

and $a^2 = d^2 - b^2$. Here, $\text{tr}(A)$ denotes the trace of matrix A .

Chen et al. [1] propose a refinement of the Ledoit-Wolf estimator, called the Rao-Blackwell Ledoit-Wolf estimator, which we denote by $\widehat{\Sigma}_{RBLW}$, that has better asymptotical properties (as $n \rightarrow \infty$) than $\widehat{\Sigma}_{LW}$. This estimator is given by

$$\widehat{\Sigma}_{RBLW} = \rho_{RBLW} \cdot \mathbf{I} + (1 - \rho_{RBLW}) \cdot \mathbf{S} \quad (12a)$$

$$\rho_{RBLW} = \min \left(\frac{\frac{n-2}{n} \cdot \text{tr}(\mathbf{S}^2) + \text{tr}^2(\mathbf{S})}{(n+2) \cdot [\text{tr}(\mathbf{S}^2) - \frac{\text{tr}^2(\mathbf{S})}{p}]}, 1 \right). \quad (12b)$$

A second estimator proposed in [1] is the oracle approximating shrinkage (OAS) estimator, which rely on a iterative procedure to provably converge to

$$\widehat{\Sigma}_{OAS} = \rho_{OAS} \cdot \mathbf{I} + (1 - \rho_{OAS}) \cdot \mathbf{S} \quad (13a)$$

$$\rho_{OAS} = \min \left(\frac{(1 - 2/p) \cdot \text{tr}(\mathbf{S}^2) + \text{tr}^2(\mathbf{S})}{(n + 1 - 2/p) [\text{tr}(\mathbf{S}^2) - \text{tr}^2(\mathbf{S})/p]}, 1 \right) \quad (13b)$$

The performance of these estimators is highly dependent on the data set. For example, in [1] $\widehat{\Sigma}_{OAS}$ performs better than $\widehat{\Sigma}_{RBLW}$ and $\widehat{\Sigma}_{LW}$ for data coming from a fractional Brownian motion, but all three estimators perform comparably for Gaussian AR(1) processes (for small values of n).

We demonstrate the performance of the three estimators using the lightweight quasi-geostrophic (QG) model, which is representative of a realistic atmospheric or oceanic data assimilating system [11]. We consider the QG model on the 2D Cartesian domain $\Omega = [0, 1] \times [0, 1]$ to estimate the covariance between 4×4 grid at 40 consecutive times ($p=160$). The QG

model describes the motion of a fluid and is mathematically expressed as

$$q_t = -\psi_x - \epsilon \cdot \mathcal{J}(\psi, q) - A \cdot \Delta^3 \psi + 2 \cdot \pi \cdot \sin(2 \cdot \pi \cdot y), \quad (14)$$

where q is the potential vorticity, ψ is the stream function, Δ is the Laplacian operator, $\mathcal{J}(\psi, q) = \psi_x \cdot q_y - \psi_y \cdot q_x$ is the Jacobian, and x and y are the horizontal and vertical components, respectively. The coefficients A and ϵ are set to 2×10^{-12} and $A = 10^{-5}$, respectively.

The model uses homogeneous boundary conditions $\partial\Omega = \partial\psi = \partial q = 0$ and initial conditions of the form

$$\begin{aligned} \psi_{i,j} = & \sin(\alpha + \beta \cdot 4 \cdot y_i \cdot x_j) + \cos(\alpha + \beta \cdot 2 \cdot y_i \cdot x_j) \\ & + \sin(\alpha + \beta \cdot 2 \cdot y_i \cdot x_j) \cdot \cos(\alpha + \beta \cdot 4 \cdot y_i \cdot x_j), \end{aligned} \quad (15)$$

where (x_i, y_j) , $1 \leq i, j \leq 64$, are the discretization points. Parameters α and β describe the shift and amplitude waves, respectively.

The samples are built by picking four points from the discretization of Ω at 40 consecutive times and randomly perturbing the initial conditions using $\alpha = 1 + |\mu_1|$ and $\beta = \mu_2$, where μ_1 and μ_2 are uniformly distributed random variables, $\mu_1 \sim \mathcal{U}(0, 10^{-4})$ and $\mu_2 \sim \mathcal{U}(0, 10^{-2})$. Each initial condition is propagated in time providing a sample associated with the 160 points of interest.

The quality of the estimators is inferred based on the percentage relative improvement in average loss norm (PRIAL) that describes how much an estimator $\hat{\Sigma}$ improves the estimation of Σ with respect to \mathbf{S} . PRIAL is defined as

$$\delta(\Sigma, \mathbf{S}, \hat{\Sigma}) = \frac{\mathbb{E}[\|\Sigma - \mathbf{S}\|^2] - \mathbb{E}[\|\Sigma - \hat{\Sigma}\|^2]}{\mathbb{E}[\|\Sigma - \mathbf{S}\|^2]}. \quad (16)$$

For this estimator, we note that larger is better, that S itself would give a value of 0 for δ , whereas a perfect estimator would result in $\delta = 1$. The true covariance matrix Σ is evaluated using 200 samples. We evaluate the estimators using $n = 25$ and $n = 40$ samples. The PRIAL norms are shown in Table I. We note that $\hat{\Sigma}_{RBLW}$ offers the best improvement of the estimation over the sample covariance matrix and is more robust for different number of samples. In Figure 7 we show the structure of Σ and $\hat{\Sigma}_{RBLW}$ ($n = 25$ samples) and note that RBLW estimator preserves the structure of the true covariance Σ .

TABLE I
PRIAL NORMS FOR THE ESTIMATORS FOR THE QG MODEL USING $n = 25$
AND $n = 40$ SAMPLES.

n	$\delta(\Sigma, \mathbf{S}, \hat{\Sigma}_{LW})$	$\delta(\Sigma, \mathbf{S}, \hat{\Sigma}_{RBLW})$	$\delta(\Sigma, \mathbf{S}, \hat{\Sigma}_{OAS})$
25	26%	66%	62%
40	35%	64%	61%

D. Integrative Numerical Study

To perform our benchmarks, we consider two strategies:

- **(Corr):** This strategy computes the forward dispatch solution using scenarios. This is done by using a distribution with covariance matrix $\hat{\Sigma}_{RBLW}$. In other words, this

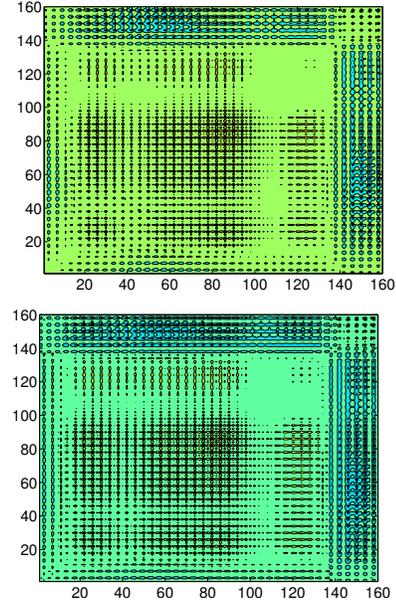


Fig. 7. True covariance matrix (top) and Rao-Blackwell-Ledoit-Wolf estimator (bottom) for the sample size $n = 25$.

strategy assumes that the ISO has correct spatiotemporal information.

- **(Indep):** This strategy computes the forward dispatch solution using scenarios that do not capture correlation. This is done by using a distribution with diagonal covariance matrix $\text{diag}(\hat{\Sigma}_{RBLW})$. We recall that the use of independent scenarios corresponds to the case in which each player submits to the ISO a set of scenarios created based on its own forecast, leaving the ISO without information about the correlation between players.

For both cases, we define as the **predicted cost** as the dispatch cost obtained from the solution of the corresponding dispatch problem. We then fix the ahead decisions to evaluate cost at a new set of scenarios obtained by sampling the distribution capturing correlation. We denote this cost as the **realized cost**.

We devised the simulations to also reveal the effect of dispatch cost and the number of scenarios used. For this purpose we have solved the dispatch model with $S = \{4, 8, 16, 32, 64, 128, 256\}$ scenarios. Because the dispatch cost is a random variable (it is an estimator), we computed error bands shown in Figure 8. We generated 256 batches of S scenarios for each $S = \{4, 8, 16, 32, 64, 128, 256\}$, solved the dispatch model for each batch, and computed the mean and the standard deviation of the dispatch cost.

The sizes of the resulting optimization problems range from 14,635 decision variables and 12,884 constraints for $S = 4$ to 763,579 decision variables and 704,372 constraints for $S = 256$. As can be seen, the size of the problems is significant. To solve these problems, we exploit the well-known “dual-block angular” structure using our parallel solver PIPS-IPM [4], [5], [6], [7]. To compute the mean and standard deviation of the dispatch cost for each $S = \{4, 8, 16, 32, 64, 128, 256\}$ we solve 256 instances in parallel, each instance using S parallel processes. In these experiments we used “Intrepid” IBM BG/P and “Mira” IBM BG/Q supercomputers at Argonne National

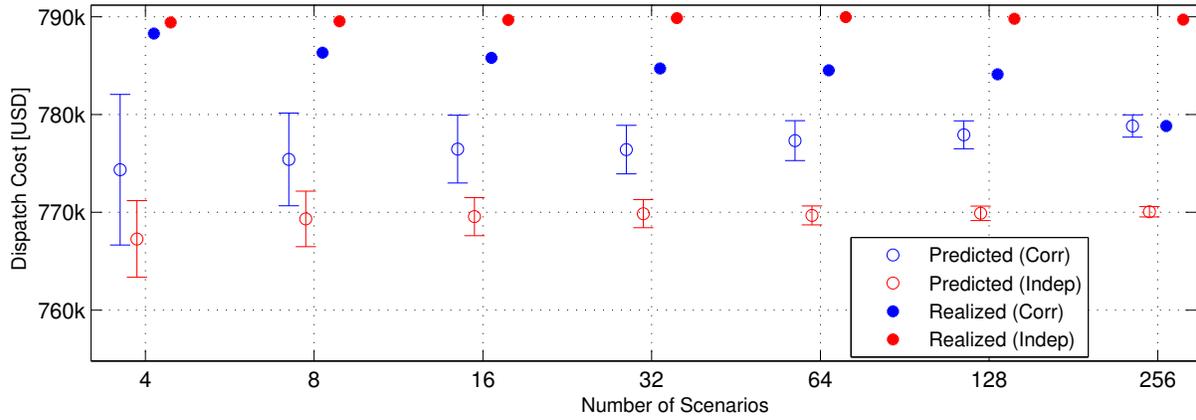


Fig. 8. Mean and 95% confidence intervals for dispatch cost predicted with correlated and independent scenarios. We also show “realized” dispatch costs, corresponding to costs obtained by implementing the predicted dispatch decisions under the correlated scenarios sampling scheme.

Laboratory. The Intrepid supercomputer has 40 racks with a total 40,960 nodes and a 3D torus high-performance interconnect; each BG/P node has a 850 MHz quad-core PowerPC processor and 2 GB of RAM. Mira is the replacement for Intrepid and consists of 48 racks, each of 1024 nodes, and a 5D torus interconnect. Each BG/Q node has 16 PowerPC A2 cores operating at 1600 Mhz and 16 GB of memory. In our simulations we have used up to 16,384 nodes on each system (for the largest run corresponding to $S=256$). On Mira, the total execution times for solving 256 batches in parallel are between 6 minutes (for $S = 4$) and 8 minutes (for $S = 256$). The slight increase in the execution times with S is primarily caused by I/O overhead, the optimization solution times remaining relatively constant (a bit more than 5 minutes for $S = 4$ and almost 6 minutes for $S = 256$). Serial times required to perform all simulations would be on the order of days.

Our results are summarized in Figure 8. From here, we can make the following observations:

- A positive dispatch cost gap exists between the realized dispatch costs of the (Corr) and the (Indep) strategies of about 10,000 USD/hr or 1.5%. This can add up to 100 million USD/yr. This gap should be interpreted as the error induced by the suboptimality of the forward decision due to the use of an incorrect probability distribution. Furthermore, the gap does not close as the number of scenario increases, being consistent with the analytical dispatch model of Section II.
- The (Indep) and the (Corr) strategies exhibit different error bands, and the bands converge at significantly different rates. For the (Corr) strategy, the error bands are small for 64 or more scenarios. In particular, the standard deviation of the dispatch cost is 0.45% and 0.36% when using 128 and 256 scenarios, respectively. This suggests that $\mathcal{O}(10^2)$ scenarios offer a good approximate dispatch cost even for a large number of wind farms. The (Indep) strategy underestimates the amount of scenarios needed.
- The predicted and realized costs converge for the (Corr) strategy while the gap does not close for the (Indep) strategy because the (Indep) strategy uses the wrong probability distribution. The gap is nearly 20,000 USD/hr which can

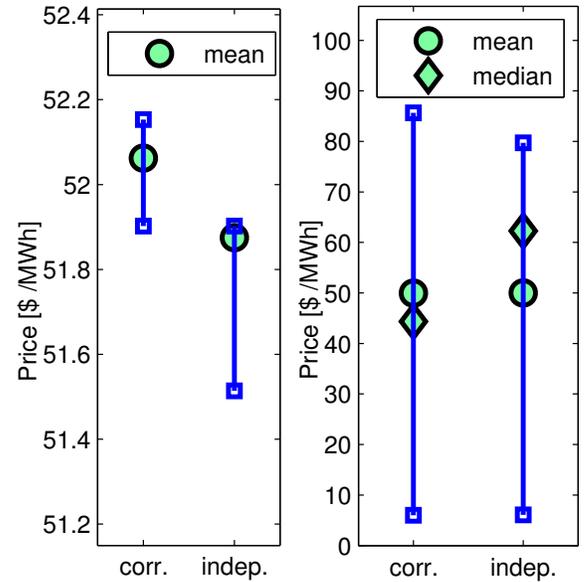


Fig. 9. Forward and spot prices with confidence intervals.

translate to up to 200 million USD/yr. This gap should be interpreted as a predictive error induced by improperly characterizing the uncertainty of the atmospheric conditions (ignoring correlations).

- The (Corr) strategy correctly predicts the dispatch cost as the scenarios are increased, as expected.

We also present in Figure 9 the distribution of the locational marginal prices at a certain network bus computed with the two sampling strategies. In particular, we observe that the (Corr) ahead prices are consistently larger than the (Indep) ahead prices, a result related to the gap in the dispatch cost. Also, looking at the real-time prices, one can see that while the two strategies give the same mean, the distribution of the prices is different, with the (Indep) prices being negatively skewed.

Our results indicate that a coordinated weather forecasting capability is necessary to create proper uncertainty characterizations of atmospheric conditions. Note that this does not imply that the ISO would perform wind power forecasts for

the suppliers because it would be impossible to capture local conditions. The suppliers can still be responsible of translating scenarios of meteorological conditions that capture long-range correlations into wind power supply scenarios using private knowledge of their own system at local conditions.

IV. CONCLUSIONS

We have demonstrated that neglecting correlations between multiple wind supply points can result in strong biases of dispatch cost. Our conclusions are drawn from a detailed study that incorporates high-resolution wind speed forecasts from numerical weather prediction models, covariance estimation techniques, and stochastic dispatch models. We believe that the results are of relevance as they suggest that a coordinated uncertainty characterization procedure is needed to capture long-range correlation.

APPENDIX

Proof of Proposition 1

We start with the following technical result.

Lemma 1: $\mathbb{E}[X|X \leq d] = -\sigma^2\phi(d) + \mu\Phi(d)$.

Proof: By direct computation we have

$$\begin{aligned} \mathbb{E}[X|X \leq d] &= \int_{-\infty}^d x\phi(x)dx = \\ &= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^d \left(-\frac{x-\mu}{\sigma^2} - \frac{\mu}{\sigma^2}\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= -\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \Big|_{-\infty}^d \\ &\quad + \frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^d \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= -\sigma^2\phi(d) + \mu\Phi(d). \end{aligned}$$

We are now prove Proposition 1.

Proof of Proposition 1:

We first show that $c_d(\sigma)$ given by (7) is strictly increasing. For this let $p(x, \sigma) = -\frac{(x-\mu)^2}{2\sigma^2}$; therefore $\phi(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(p(x, \sigma))$. Observe that

$$\frac{d}{dx}p(x, \sigma) = -\frac{(x-\mu)}{\sigma^2}, \quad (17a)$$

$$\frac{d}{d\sigma}p(x, \sigma) = +\frac{(x-\mu)^2}{\sigma^3} \quad (17b)$$

$$\frac{d}{dx}\phi(x, \sigma) = \phi(x, \sigma) \cdot \frac{d}{dx}p(x, \sigma) \quad (17c)$$

$$\begin{aligned} \frac{d}{d\sigma}\phi(x, \sigma) &= \phi(x, \sigma) \cdot \frac{d}{d\sigma}p(x, \sigma) - \frac{1}{\sigma^2\sqrt{2\pi}} \exp(p(x, \sigma)) \\ &= \frac{(x-\mu)^2}{\sigma^3} \cdot \phi(x, \sigma) - \frac{1}{\sigma}\phi(x, \sigma). \end{aligned} \quad (17d)$$

The derivative of Φ with respect to σ can be computed as

follows:

$$\begin{aligned} &\frac{d}{d\sigma}\Phi(d, \sigma) \\ &= \frac{d}{d\sigma} \int_{-\infty}^d \phi(p(x, \sigma))dx \\ &= \int_{-\infty}^d \frac{d}{d\sigma}\phi(p(x, \sigma))dx \\ &\stackrel{(17d)}{=} \int_{-\infty}^d \frac{(x-\mu)^2}{\sigma^3} \cdot \phi(p(x, \sigma))dx - \frac{1}{\sigma} \int_{-\infty}^d \phi(p(x, \sigma))dx \\ &= -\int_{-\infty}^d \phi(p(x, \sigma)) \cdot \left(-\frac{x-\mu}{\sigma^2}\right) \cdot \frac{x-\mu}{\sigma} dx - \frac{1}{\sigma}\Phi(d, \sigma) \\ &\stackrel{(17a)}{=} -\int_{-\infty}^d \phi(p(x, \sigma)) \cdot \frac{d}{dx}p(x, \sigma) \cdot \frac{x-\mu}{\sigma} dx - \frac{1}{\sigma}\Phi(d, \sigma) \\ &\stackrel{(17c)}{=} -\int_{-\infty}^d \frac{d}{dx}\phi(p(x, \sigma)) \cdot \frac{x-\mu}{\sigma} dx - \frac{1}{\sigma}\Phi(d, \sigma) \\ &= -\phi(p(x, \sigma)) \cdot \frac{x-\mu}{\sigma} \Big|_{-\infty}^d \\ &\quad + \frac{1}{\sigma} \int_{-\infty}^d \phi(p(x, \sigma))dx - \frac{1}{\sigma}\Phi(d, \sigma) \\ &= -\frac{d-\mu}{\sigma} \cdot \phi(p(d, \sigma)). \end{aligned} \quad (18)$$

By differentiating (7) and using (17d) and (18), we compute

$$\begin{aligned} \frac{d}{d\sigma}c_d(\sigma) &= (p_{th} - p_w) \left((d-\mu) \frac{d}{d\sigma}\Phi(d, \sigma) + 2\sigma\phi(d, \sigma) \right. \\ &\quad \left. + \sigma^2 \frac{d}{d\sigma}\phi(d, \sigma) \right) \\ &= (p_{th} - p_w) \left(-\frac{(d-\mu)^2}{\sigma} \cdot \phi(d, \sigma) + 2\sigma\phi(d, \sigma) \right. \\ &\quad \left. + \frac{(d-\mu)^2}{\sigma} \cdot \phi(d, \sigma) - \sigma\phi(d, \sigma) \right) \\ &= \sigma(p_{th} - p_w)\phi(d, \sigma) > 0, \end{aligned} \quad (19)$$

which shows that $c_d(\sigma)$ is strictly increasing. Since $\sigma(\rho) = \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}$ is also a strictly increasing function of ρ , we conclude that c_d is also strictly increasing function of the correlation coefficient ρ . ■

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, under Contract No. DE-AC02-06CH11357. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. Computing time on *Intrepid* was granted by a 2012-2013 DOE INCITE award, ‘‘Optimization of Complex Energy Systems under Uncertainty,’’ PI Mihai Anitescu, Co-PI Cosmin G. Petra.

REFERENCES

- [1] YILUN CHEN, A. WIESEL, AND A.O. HERO, *Shrinkage estimation of high dimensional covariance matrices*, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, 2009, pp. 2937–2940.
- [2] E.M. CONSTANTINESCU, V.M. ZAVALA, M. ROCKLIN, SANGMIN LEE, AND M. ANITESCUS, *A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation*, IEEE Transactions on Power Systems, 26 (2011), pp. 431–441.

- [3] OLIVIER LEDOIT AND MICHAEL WOLF, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis, 88 (2004), pp. 365 – 411.
- [4] MILES LUBIN, COSMIN G. PETRA, AND MIHAI ANITESCU, *The parallel solution of dense saddle-point linear systems arising in stochastic programming*, Optimization Methods and Software, 27 (2012), pp. 845–864.
- [5] MILES LUBIN, COSMIN G. PETRA, MIHAI ANITESCU, AND VICTOR ZAVALA, *Scalable stochastic optimization of complex energy systems*, in Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC '11, New York, USA, 2011, ACM, pp. 64:1–64:64.
- [6] COSMIN G. PETRA AND MIHAI ANITESCU, *A preconditioning technique for Schur complement systems arising in stochastic optimization*, Computational Optimization and Applications, 52 (2012), pp. 315–344.
- [7] COSMIN G. PETRA, OLAF SCHENK, MILES LUBIN, AND KLAUS GÄRTNER, *An augmented incomplete factorization approach for computing the Schur complement in stochastic optimization*, accepted to SIAM Journal on Scientific Computing (2014).
- [8] P. PINSON, H. MADSEN, H. A. NIELSEN, G. PAPAETHYMIU, AND B. KLCKL, *From probabilistic forecasts to statistical scenarios of short-term wind power production*, Wind Energy, 12 (2009), pp. 51–62.
- [9] GEOFFREY PRITCHARD, GOLBON ZAKERI, AND ANDREW PHILPOTT, *A single-settlement, energy-only electric power market for unpredictable and intermittent participants*, Operations Research, 58 (2010), pp. 1210–1219.
- [10] ———, *A single-settlement, energy-only electric power market for unpredictable and intermittent participants*, Operations Research, 58 (2010), pp. 1210–1219.
- [11] PAVEL SAKOV AND PETER R. OKE, *A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters*, Tellus A, 60 (2008), pp. 361–371.
- [12] W.C. SKAMAROCK, J.B. KLEMP, J. DUDHIA, D.O. GILL, D.M. BARKER, M.G. DUDA, X.-Y. HUANG, W. WANG, AND J.G. POWERS, *A description of the Advanced Research WRF version 3*, Tech. Report Tech Notes-475+ STR, NCAR, 2008.
- [13] J. SUMAILI, H. KEKO, V. MIRANDA, ZHI ZHOU, A. BOTTERUD, AND JIANHUI WANG, *Finding representative wind power scenarios and their probabilities for stochastic models*, in 16th International Conference on Intelligent System Application to Power Systems, Sept 2011, pp. 1–6.
- [14] J. TASTU, P. PINSON, AND H. MADSEN, *Space-time trajectories of wind power generation: Parameterized precision matrices under a Gaussian copula approach*, in Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension, in press, 2014.
- [15] M. WYTOCK AND J.Z. KOLTER, *Large-scale probabilistic forecasting in energy systems using sparse Gaussian conditional random fields*, in 2013 IEEE 52nd Annual Conference on Decision and Control, Dec 2013, pp. 1019–1024.
- [16] V. M. ZAVALA, J. BIRGE, AND M. ANITESCU, *A stochastic market clearing formulation with consistent pricing properties*, draft (2014).

(To be removed before publication) 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.