2023-07-16

Day-ahead estimation of renewable generation uncertainty set for more efficient market clearing

Day-Ahead Estimation of Renewable Generation Uncertainty Set for More Efficient Market Clearing

Alexander Wasilkoff
Division of Systems Engineering
Boston University
Boston, MA, U.S.
wasilkof@bu.edu

Panagiotis Andrianesis
Division of Systems Engineering
Boston University
Boston, MA, U.S.
panosa@bu.edu

Michael Caramanis
Division of Systems Engineering
Boston University
Boston, MA, U.S.
mcaraman@bu.edu

***Abstract*—Wholesale Day Ahead Markets (DAMs) dominated by renewable generation can improve social surplus by driving unit commitment (UC) and security constrained economic dispatch (SCED) in a novel systematic uncertainty aware manner. A new robust uncertainty aware DAM model can anticipate worst-case output of renewable generation and use a system reliability constraint involving an uncertainty set of renewable capacity realizations. The uncertainty set yields unique worst-case renewable capacities, which can be written as linear functions of a tentative, most-recent, SCED. A few repetitions of the UC-SCED converge to the optimal DAM clearing under the new system reliability constraint, endogenously determining reserve requirements, and potentially prices for energy and reserves. This paper focuses on constructing the uncertainty set essential to this new uncertainty aware Market paradigm and illustrates it on a realistic size Balancing Area with hundreds of wind farms.**

***Index Terms*—Power Markets, Robustness, Uncertain Systems, Wind Energy Integration.**

I. Introduction

Wholesale Day Ahead Markets (DAMs) in renewable generation dominated Balancing Areas are increasingly challenged by uncertainty and volatility. At the same time, the evolution of demand to include flexible storage-like loads, such as distributed energy resources, provides new opportunities to use versatile and spatiotemporally granular dynamic prices to coordinate renewable generation with schedulable flexible demand. Revision of DAM that can take advantage of flexible demand requires (i) systematic modeling of renewable generation uncertainty based on detailed forecasts of their available capacity, and (ii) pricing of associated requisite reserves to elicit efficient demand response to renewable uncertainty.

Forecasts are routinely generated by sophisticated vendors for multiple locations looking up to two days ahead on a 5-minute time scale. Forecasts are re-estimated during the day enabling a better understanding of how uncertainty evolves as a function of the look ahead horizon. We argue that systematic quantification of renewable generation uncertainty anticipated at DAM gate closure can increase social surplus by improving unit commitment (UC) and security constrained economic dispatch (SCED) so as to decrease overall costs. To this end, we reimagine a new robust uncertainty aware DAM model that anticipates worst-case realizations of renewable generation and translates it to more efficient UC and Energy and Reserves schedules. The resulting DAM model is augmented by a new hourly system reliability constraint, relying on the proposed uncertainty set to be made robust and overcome computationally intractable stochastic optimization approaches. We argue that an appropriately constructed uncertainty set can anticipate worst-case renewable generation expressed as linear functions of a tentative, most-recent, SCED. A few repetitions of the mixed integer linear programming (MILP) problem are sufficient for the requisite iterative market clearing to converge to the optimal UC-SCED solution and provide energy and reserve prices consistent with optimal demand response.

This paper focuses on the construction of the aforementioned uncertainty set represented by a hyper-ellipsoid (positive definite quadratic form) with desired properties, namely, the determination of a unique renewable generation worst case obtainable in closed form from the uncertainty set parameters. The literature contains various alternatives surveyed below, which we believe do not dominate our approach. Ref. [1] employs a two-stage adaptive robust method for solving the UC problem, where the final solution remains feasible for all random variable realizations residing in the uncertainty set represented by uncertainty budget constraints. It does not account for pairwise correlations across wind farms rendering it possibly unacceptable by individual wind farms. Ref. [2] uses principal component analysis (PCA) with kernel smoothing to extract uncertainty information for the construction of polytope uncertainty sets, but it lacks transparency on how individual wind farms contribute to the system worst case. Ref. [3] derives ellipsoidal uncertainty sets for wind generation and provides a complex feedback mechanism to leverage knowledge of historical wind speed observations. It asserts that ellipsoidal uncertainty sets perform better than competing box sets and budget uncertainty sets demonstrating a higher ratio of probabilistic guarantee to uncertainty set volume. Nevertheless, it is applied to a system with only a few wind farms and does not investigate, as we do, issues that arise with singular covariance matrices. Ref. [4] proposes a data-driven uncertainty set used for Robust Optimal UC exploring temporal and spatial correlations outperforming box and budget uncertainty sets; its use of temporal correlations and historical data might complement our approach.

Research supported by ARPA-E's PERFORM grant.

Our proposed uncertainty set construction addresses the undesirable singularity of the individual renewable generation covariance matrix by reducing dimensionality through aggregation into clusters. For a specific hour, the resulting cluster uncertainty is captured by a non-singular positive definite covariance matrix estimated from forecast data. The uncertainty set is then instantiated by (i) its ellipsoidal shape corresponding to the quadratic form associated with the inverse covariance matrix, and (ii) its size represented by the ellipsoid's radius calibrated by a probabilistic guarantee that the uncertainty set includes renewable capacity realizations. Given the uncertainty set's shape and size, a unique worst case of cluster outputs is then estimated, ratios of worst case to scheduled capacity are calculated and assigned to the individual wind farms aggregated to each cluster. Numerical illustrations are provided on sanitized data from about 200 wind farms of a large U.S. Balancing Area. Our main contribution involves the construction of an uncertainty set that enables robust modeling of a system reliability constraint, representing an endogenous determination of reserves imposed by renewable generation uncertainty, and potentially the discovery of energy and reserve prices eliciting efficient flexible demand response.

The rest of this paper evolves as follows. Section II sketches a system reliability constraint, which motivates the uncertainty set construction presented in Section III. Section IV discusses the numerical results, and Section V concludes the paper.

II. Uncertainty Set Motivation

While current practice uses the expected value of variable renewable generation as its available capacity for scheduling it in the DAM, an uncertainty aware DAM model should be able to model the fact that the anticipated average capacity of generator g scheduled for energy and reserves at the DAM, $q_g^E(\tau)$, and $q_g^R(\tau)$ for $\tau = 1, 2, 3, \ldots, 24$ is likely to take a different value in real time. A system reliability constraint is introduced to model renewable generation available capacity at a future hour as a random variable. We use tilde to denote future uncertainty, i.e., $\tilde{q}_g^E(\tau)$, and $\tilde{q}_g^R(\tau)$, and add in the uncertainty aware DAM model a system reliability constraint that should hold for each hour τ for any uncertain realization of the tilded random variables. We define the uncertainty set, $\mathcal{U}(\tau)$, as the hour specific set that contains all realizations $\tilde{q}_g^E(\tau) + \tilde{q}_g^R(\tau)$ during this hour at a reasonable probabilistic guarantee. Dropping the hour designation for notational simplicity, and assuming without loss of generality that the energy bid of demand d, q_d^E, is deterministic, we write the system reliability constraint as follows:

$$\sum_{g\in\mathcal{V}}(\tilde{q}_g^E + \tilde{q}_g^R) + \sum_{g\in\mathcal{C}}(q_g^E + q_g^R) - \sum_{d\in\mathcal{D}} q_d^E \geq Q^R, \qquad \forall\, \tilde{q}_g^E, \tilde{q}_g^R \in \mathcal{U}, \tag{1}$$

where $\mathcal{V}$ is the set of renewable/random-variable-available-capacity generators, $\mathcal{C}$ the set of conventional firm capacity generators, $\mathcal{D}$ the set of demand resources, and Q^R the contingency reserve requirement.

Constraint (1) is difficult to enforce as it must hold true for all possible random variable realizations, thus introducing computational burden. We convert the uncertainty aware market clearing problem to an algorithmically deterministic problem by considering the worst-case value (or values if more than one) over the uncertainty set and solving the problem:

$$\min_{\tilde{q}_g^E,\tilde{q}_g^R\in\mathcal{U}} \sum_{g\in\mathcal{V}}(\tilde{q}_g^E + \tilde{q}_g^R). \tag{2}$$

Denoting the preferably finite solutions of problem (2) by the superscript "wc" for worst case, the system reliability constraint can be written as:

$$\sum_{g\in\mathcal{V}}(q_g^{E,wc} + q_g^{R,wc}) + \sum_{g\in\mathcal{C}}(q_g^E + q_g^R) - \sum_{d\in\mathcal{D}} q_d^E \geq Q^R \rightarrow \lambda^{SR}, \tag{1$'$}$$

where $q_g^{E,wc}, q_g^{R,wc}$ span the solution set of (2), and λ^{SR} is the dual variable at the respective SCED. A desirable uncertainty set would have a unique solution allowing us to consider a single system reliability constraint in each hour. If that unique solution could be written in terms of the optimal value of the decision variables, we might be able to aim for a deterministic iterative algorithm converging to the optimal uncertainty aware DAM clearing solution. Next, we show that this is possible for an appropriately constructed uncertainty set, adopting the established and generally accepted ellipsoid concentration set [5] derived from the inverse of the hourly wind farm covariance matrix, which we can estimate from wind farm available capacity forecasts.

III. Uncertainty Set Construction

In this section, we describe the wind farm clustering (in Subsection III-A) and the cluster uncertainty set construction (in Subsection III-B).

A. Wind Farm Clustering

Given the incidence of relatively high magnitude pairwise correlations and the usually smaller number of forecasts (about 50 ensembles) than the number of wind farms (hundreds in large Balancing Areas) the covariance matrix estimate is positive semi-definite, hence non-invertible. It is therefore necessary to cluster highly correlated wind farms into a number of clusters that is smaller than the number of forecasts.

Algorithm 1 details the wind farm clustering and cluster seeding. We decide the number of clusters, K, by evaluating the number of the full covariance matrix eigenvalues and their rate of decline. Once the number of clusters is set, we follow the initialization method based on k-means++ [6]. Starting with the pairwise correlations of all wind farms, we select first the highest pairwise correlated pair and seed the first cluster. We proceed with finding the most distant correlation-wise wind farm and seed the second cluster and continue until all clusters have been seeded. Then, we iteratively assign the wind farms to clusters. Each iteration begins by calculating the Pearson correlation coefficient between each of the wind

Algorithm 1 Clustering Method

Inputs: Wind farms $\mathcal{W}$, clusters K, empty cluster sets $\mathcal{W}_k$, $k = 1, \ldots, K$, threshold θ.

Cluster Seeding:

Assign the pair of wind farms with the highest pairwise correlation, r, to $\mathcal{W}_1$.

for $k = 2, ..., K$ **do**

  Calculate correlation $r_{wk'}$ between wind farm w and cluster k' total output, $\forall w \in \mathcal{W} \setminus \mathcal{W}_{k'}$, $k' < k$.

  $d_w = \min_{k'<k}(1 - |r_{wk'}|), \forall w \in \mathcal{W} \setminus \mathcal{W}_{k'}$, $k' < k$.

  Assign wind farm $w' \in \arg \max_{w \in \mathcal{W} \setminus \mathcal{W}_{k'}, k'<k} d_w$ to $\mathcal{W}_k$.

end for

Iterative Assignment to Clusters:

repeat

  Calculate correlation r_{wk} between wind farm $w \in \mathcal{W}$ and the total output of cluster k, $k = 1, \ldots, K$.

  if $\exists\, w \in \mathcal{W}_k, k = 1, \ldots, K$ such that $r_{wk'} > r_{wk}$, $k' \neq k$, **then** assign wind farm w to $\mathcal{W}_{k'}$,

  else assign wind farm $w \notin \mathcal{W}_k, k = 1, \ldots, K$, with $r_{wk'} > \theta$ and $k' \in \arg \max_{k''=1,\ldots,K} r_{wk''}$, to $\mathcal{W}_{k'}$,

  end if

until all $r_{wk} < \theta$, $w \notin \mathcal{W}_k$, $k = 1, \ldots, K$.

Assign remaining wind farms to Uncorrelated set, $\mathcal{W}_{K+1}$.

Output: Cluster sets $\mathcal{W}_k, k = 1, \ldots, K$, and $\mathcal{W}_{K+1}$.

farms and the clusters. The wind farms that have already been assigned to a cluster are confirmed to have the highest correlation with their assigned cluster. If this is not true for any wind farms, they are reassigned to their most highly correlated cluster. Otherwise, the unassigned wind farm, which is most highly correlated to any cluster, is assigned to that cluster as long as that correlation is above a threshold, θ. Once all the unassigned wind farms have correlations below θ, the loop stops, and the remaining wind farms are assigned to the Uncorrelated Cluster. The output of Algorithm 1 is the K clusters and their constituent wind farms, including the Uncorrelated Cluster.

B. Cluster Uncertainty Set Construction

After cluster population has converged, we construct the cluster ellipsoidal uncertainty set defined as

$$\mathcal{U} = \{\tilde{\mathbf{q}} : (\tilde{\mathbf{q}} - \bar{\mathbf{q}})^T \mathbf{\Sigma}^{-1} (\tilde{\mathbf{q}} - \bar{\mathbf{q}}) \leq \rho^2\}, \tag{3}$$

where $\tilde{\mathbf{q}}$ is the vector of values for the total generation of each cluster, $\bar{\mathbf{q}}$ is the cluster output mean across all forecasts, $\mathbf{\Sigma}^{-1}$ is the inverse of the covariance matrix of cluster output across all forecasts. Note that $\mathbf{\Sigma}^{-1}$ determines the shape of the uncertainty set, whereas the radius of the ellipsoid, ρ, determines its size and quantifies a probabilistic guarantee. The boundary of the ellipsoid is a quadratic form with $\mathbf{\Sigma}^{-1}$ a positive definite symmetric matrix, where the positive definite property is guaranteed by the clustering process reduction of dimensionality. The radius is estimated as follows.

1) Estimation of Radius: The radius of the ellipsoid is determined so as to enforce a reasonable probabilistic guarantee that a realization of a total hourly cluster output is contained in that hour's ellipsoid. This is estimated by ensuring that the ellipsoid includes all forecasts after excluding some outliers. The radius magnitude needed associated with hour τ for the uncertainty set to contain forecast i is:

$$\rho_i = \sqrt{(\tilde{\mathbf{q}}_i - \bar{\mathbf{q}})^T \mathbf{\Sigma}^{-1} (\tilde{\mathbf{q}}_i - \bar{\mathbf{q}})}, \tag{4}$$

where $\bar{\mathbf{q}}_i$ is the vector of total cluster outputs associated with the ith forecast and hour τ. Algorithm 2 details the estimation of the maximum radius, $\rho_{\max}$. The ρ_i values are

Algorithm 2 Maximum Radius Estimation

Inputs: forecasts I, cluster output forecasts $\tilde{\mathbf{q}}_i$, mean cluster output $\bar{\mathbf{q}}$, inverse covariance matrix $\mathbf{\Sigma}^{-1}$, parameters a and t, hour τ.

Calculate ρ_i using (4), $i = 1, \ldots, I$.

Exclude the largest $m(\tau) = a + \lfloor \tau/t \rfloor$ values (outliers) of $\rho_i, i = 1, \ldots, I$.

Identify the largest remaining value of ρ_i as $\rho_{\max}$.

Outputs: $\rho_{\max}$.

ordered in increasing magnitude, the $m(\tau)$ largest magnitudes are excluded and the remaining maximal ρ value is selected. We use $m(\tau) = a + \lfloor \tau/t \rfloor$, where a and t are adjustable parameters. The a and the t are parameters that control the number of outliers excluded every hour, and the additional are excluded due to increasing uncertainty as the day progresses, respectively. The rationale of estimating the radius magnitude in this manner is that we exclude outliers in increasing numbers as the hour is further removed from the time the DAM closed and the forecasts were estimated. Indeed, the later in the day ahead horizon, the larger the forecast error — this is clearly evidenced in our data set.

2) Calculation of Cluster Output Worst Case: For the Uncorrelated Cluster set, $\mathcal{W}_{K+1}$, we use the minimum forecast to calculate the wind farm worst-case value. For the first K Clusters, the worst case is the solution to the linear objective function convex quadratic constraint problem:

$$\min_{\tilde{q}_g \in \mathcal{U}} \sum_g \tilde{q}_g. \tag{5}$$

Proposition 1. *For a positive definite cluster output covariance matrix* $\mathbf{\Sigma}$*, the unique solution to problem* (5) *is*

$$\mathbf{q}^{wc} = -\rho_{\max} \frac{\mathbf{\Sigma 1}}{\sqrt{\mathbf{1}^T \mathbf{\Sigma 1}}} + \bar{\mathbf{q}}, \tag{6}$$

where $\mathbf{q}^{wc}$ *is the vector of generation values for each cluster at* $\rho_{\max}$ *estimated by Algorithm* 2*, and* $\mathbf{1}$ *is an appropriate dimension vector of ones.*

Proof. Given a closed ellipsoid defined by (3) and seeking to minimize (5), a closed-form solution can be obtained by noting that the extreme point must lie on the boundary of the ellipsoid and be where the cost vector, $\mathbf{c}$, is normal to the ellipsoid boundary. Writing the ellipsoid boundary as $f(\tilde{\mathbf{q}}) = (\tilde{\mathbf{q}} - \bar{\mathbf{q}})^T \boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{q}} - \bar{\mathbf{q}}) \leq \rho^2$, the gradient is then $\nabla f(\tilde{\mathbf{q}}) = 2\boldsymbol{\Sigma}^{-\mathbf{1}}(\tilde{\mathbf{q}} - \bar{\mathbf{q}})$. Let this be a multiple of the cost vector, $\mathbf{c}$, $2\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{q}} - \bar{\mathbf{q}}) = \gamma \mathbf{c}$, and solve for the point $\tilde{\mathbf{q}} = \frac{\gamma}{2}\boldsymbol{\Sigma}\mathbf{c} + \bar{\mathbf{q}}$. To find the value of the coefficient, γ, we solve for the boundary of the ellipsoid, $(\frac{\gamma}{2}\boldsymbol{\Sigma}\mathbf{c} + \bar{\mathbf{q}} - \bar{\mathbf{q}})^T \boldsymbol{\Sigma}^{-1}(\frac{\gamma}{2}\boldsymbol{\Sigma}\mathbf{c} + \bar{\mathbf{q}} - \bar{\mathbf{q}}) = \rho^2$, which yields $\left(\frac{\gamma}{2}\right)^2 (\boldsymbol{\Sigma}\mathbf{c})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\mathbf{c}) = \rho^2$. Because $\boldsymbol{\Sigma}$ is a covariance matrix, it is symmetric, and hence $\left(\frac{\gamma}{2}\right)^2 \mathbf{c}^T \boldsymbol{\Sigma}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\mathbf{c}) = \rho^2$, and solving for γ we have $\gamma = \pm 2\rho/\sqrt{\mathbf{c}^T \boldsymbol{\Sigma}\mathbf{c}}$. To find the minimum, we take the negative value of γ, i.e., $\tilde{\mathbf{q}}^{wc} = -\rho\boldsymbol{\Sigma}\mathbf{c}/\sqrt{\mathbf{c}^T\boldsymbol{\Sigma}\mathbf{c}} + \bar{\mathbf{q}}$. Replacing $\mathbf{c}$ with the all ones vector and ρ with $\rho_{\max}$, we get (6). □

Proposition 2. *The worst-case sum*

$$\mathbf{1}^T \mathbf{q}^{wc} = -\rho_{\max}\sqrt{\mathbf{1^T \Sigma 1}} + \mathbf{1^T}\bar{\mathbf{q}}, \tag{7}$$

changes only through adjusting $\rho_{\max}$, as $\sqrt{\mathbf{1^T\Sigma 1}}$ and $\mathbf{1^T}\bar{\mathbf{q}}$ are invariant to clustering and outlier selection.

Proof. We show that the quantities $\sqrt{\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}}$ and $\mathbf{1}^T\bar{\mathbf{q}}$ in (7) are invariant to clustering. The latter is easy to see as expectation is a linear operator, so clustering does not change the sum of means: $\sum_i E[\tilde{\mathbf{q}}_i] = E[\sum_i \tilde{\mathbf{q}}_i]$. To prove the invariance of the first quantity, we introduce some notation: $\mathbf{M}$ is the $W \times I$ zero mean data matrix, where W is the number of wind farms and I is the number of forecasts; $\mathbf{C}$ is the clustering matrix, where $\mathbf{C}_{kw}$ is 1 if wind farm w is in cluster k, otherwise 0. $\mathbf{C}$ is $K \times W$ where K is the number of clusters. In this notation, the covariance matrix is $\boldsymbol{\Sigma} = \frac{1}{I-1}\mathbf{M}\mathbf{M}^T$. Therefore, the clustered covariance matrix is $\boldsymbol{\Sigma}_C = \frac{1}{I-1}(\mathbf{CM})(\mathbf{CM})^T$. The clustered covariance matrix is $K \times K$, so we pre- and post- multiply by the ones vector of length K: $\mathbf{1}_K^T \boldsymbol{\Sigma}_C \mathbf{1}_K = \frac{1}{I-1}\mathbf{1}_K^T \mathbf{CMM^T C^T 1}_K$. Because each wind farm is only assigned to one cluster, the columns of $\mathbf{C}$ have only one nonzero entry. Therefore, $\mathbf{1}_K^T\mathbf{C} = \mathbf{1}_W^T$. Using this identity, we have: $\mathbf{1}_K^T \boldsymbol{\Sigma}_C \mathbf{1}_K = \frac{1}{I-1}\mathbf{1}_W^T \mathbf{M}\mathbf{M}^T \mathbf{1}_W$, i.e., $\mathbf{1}_K^T \boldsymbol{\Sigma}_C \mathbf{1}_K = \mathbf{1}_W^T \boldsymbol{\Sigma}\mathbf{1}_W$. The sum of all elements of the covariance matrix remains constant regardless of clustering. □

3) Worst-Case Assignment to Individual Farms and Estimation of Risk Coefficients: Risk coefficients can be estimated from cluster worst cases and inherited by individual wind farms. More importantly, the risk coefficients can be used to write the worst-case values in the system reliability constraint as a linear multiple of the SCED decision variable q_g, namely substituting into (1′) $q_g^{wc} = \alpha_g q_g$ we get:

$$\sum_{g\in\mathcal{V}} \alpha_g(q_g^E + q_g^R) + \sum_{g\in\mathcal{C}}(q_g^E + q_g^R) - \sum_{d\in\mathcal{D}} q_d^E \geq Q^R \rightarrow \lambda^{SR}, \tag{1''}$$

where α_g is obtained from $\alpha_g = q_g^{wc}/\hat{q}_g^{prev}$, with $\hat{q}_g^{prev}$ being a tentative dispatch decision, say from a previous iteration. The Uncertainty Aware DAM is cleared with the system reliability constraint and the first iteration of coefficients. The dispatch is used to update risk coefficients, and the process continues until SCED decisions converge. We have observed that $\hat{q}_g^{prev}$ converges reasonably fast to the final optimal decision and with it converges the risk coefficient as well, which quantifies the price that a wind farm ought to obtain from the dual of the reliability constraint, i.e., λ^{SR} derated by α_g just like a binding transmission constraint dual is derated by the shift factor. Note that the risk coefficients must be in $[0, 1]$; otherwise, e.g., when the worst case is bigger than the dispatch because of a binding transmission capacity constraint resulting in $\alpha_g = q_g^{wc}/\hat{q}_g^{prev} > 1$, they are clipped to 1.

IV. Numerical Results

We demonstrate the proposed uncertainty set construction using sanitized data from a large U.S. Balancing Area. We consider two days, one in December 2021 (Day 1), and one in July 2022 (Day 2), with 192 and 206 available wind farms, respectively, and 54 hourly forecasts. For both days, we selected $K = 4$ clusters, threshold value $\theta = 0.15$, and outlier parameters $a = 5$ and $t = 3$. Algorithms 1 and 2 were implemented in Python 3.8.13.

In Fig. 1, we present the 25 largest eigenvalues of the covariance matrix for both Days, Hour 1. Notably, the largest four eigenvalues capture 65%-95% of the explained variance, justifying the selection of four clusters! The large first eigenvalue means that one of the clusters is much larger than the others. Similar results are observed across all hours.

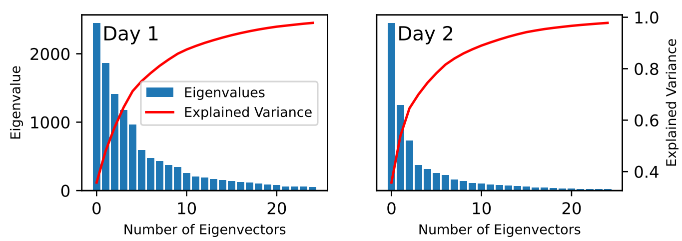

Fig. 1. Largest 25 eigenvalues for Day 1 and 2, Hour 1, covariance matrix, and cumulative explained variance of the largest eigenvectors.

In Fig. 2, we illustrate the application of the clustering method (Algorithm 1) for Day 1, Hour 1. The left heat map shows the pairwise correlation matrix of the wind farms before assignment to a cluster. The right heat map shows the output of Algorithm 1, where wind farms are clustered in four definite boxes with high intra-cluster correlations (e.g., in the largest

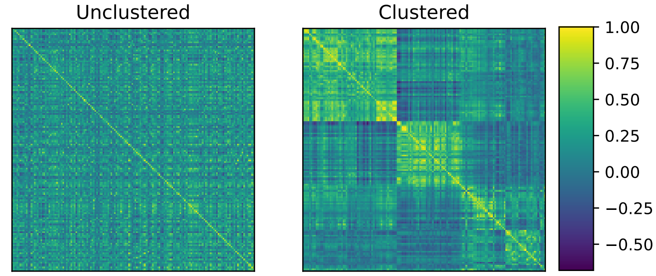

Fig. 2. Heat maps of wind farm pairwise-correlations for Day 1, Hour 1; unclustered (left, random order), and clustered (right, output of Algorithm 1).

cluster, the median correlation coefficient is 0.7) and low inter-cluster correlations.

In Fig. 3, we illustrate the maximum radius estimation (Algorithm 2) for Day 1, Hours 1, 12, and 24. The ρ values are sorted, and the excluded outliers are marked in red. As the hours grow later, more outliers are excluded according to Algorithm 2, which is the desired outcome due to growing forecasting uncertainty.

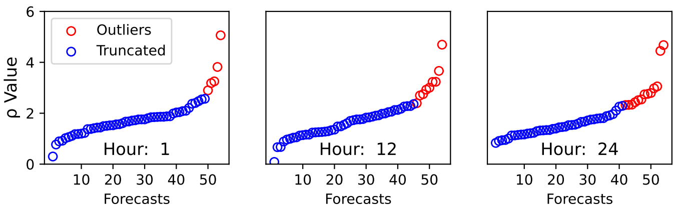

Fig. 3. Illustration of Algorithm 2 for Day 1, Hours 1, 12, and 24.

In Fig. 4, we compare our clustering method to a PCA method, which reduces the dimensionality of the forecasts according to the eigenvectors of the 4 largest eigenvalues of the data covariance matrix and is similar to [2]. For comparison purposes, instead of polytope uncertainty sets used in [2], we use an ellipsoid as described in Eq. (3) and Algorithm 2. Interestingly, both methods give similar total hourly worst cases, although our clustering method distributes them to individual wind farms in a manner that is more likely to be acceptable to renewable generation market participants. In Day 1, our clustering method tracks the minimum forecast closely in the morning hours and is less than the minimum forecast in the later hours of the day. The profile for Day 2 shows a greater differentiation between the minimum forecast and the clustering worst case, because the wider range of forecasts leads to higher ρ values.

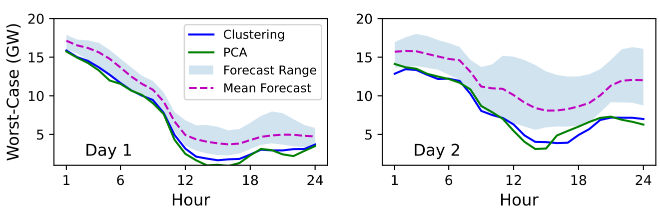

Fig. 4. Comparison of worst-case wind generation output calculated with the Clustering and PCA methods.

In Fig. 5, we present the risk coefficients for Day 1, calculated assuming the wind plants are scheduled at their mean forecast (left) and their max forecast (right). The size of the circles corresponds to the relative percentage of wind generation available in each cluster. Note, the clusters are reseeded every hour and so the membership of the clusters changes. We also mention that after a couple of iterations, SCED decisions remained within 1% tolerance. Throughout the day, the risk coefficients of the mean forecast are generally higher than those of the max forecast, which schedules more uncertain capacity. This is also reflected in the higher additional reserve

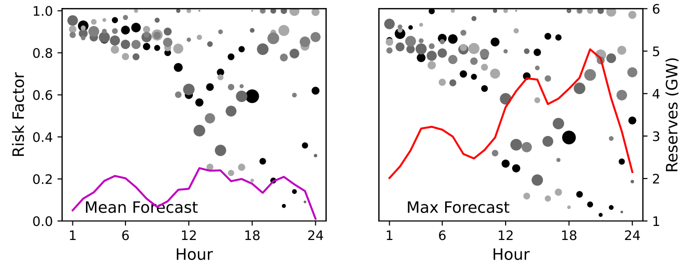

Fig. 5. Risk coefficients and additional reserve requirements calculated using the Clustering Method for Day 1.

requirements. In the morning, risk coefficients hover around 0.9 indicating that the worst case is relatively close to the forecast, but around noon, risk coefficients decrease sharply corresponding to the separation of the worst case from the forecast observed in Fig. 4. The decrease in the max forecast scenario is more drastic thus requires a greater increase in reserves in the afternoon.

V. Conclusion

In order to incorporate uncertainty awareness in the DAM MILP problem, a new system reliability constraint has been proposed, which relies on estimated uncertainty sets. We propose and explore an ellipsoidal uncertainty set construction applicable to systems with hundreds of wind farms. This yields a worst-case performance of individual wind farms and in a large system application compares favorably with other approaches such as PCA and quantifies risk coefficients impacting dynamic reserve requirements and pricing. Future work on the assignment of system risk to individual risky assets and consumption may be useful. The paper improves on current non-risk aware dispatch by incorporating risk measures in the DAM clearing process and provides the basis for future uncertainty related DAM payments. Work is in progress to quantify the benefits from the proposed DAM clearing model, which includes the new system reliability constraint, compared with the current practice, as well as evaluate the performance of alternative approaches to constructing uncertainty sets.

References

[1] D. Bertsimas, E. Litvinov, X.A. Sun, J. Zhao, and T. Zheng, “Adaptive robust optimization for the security constrained unit commitment problem,” *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, 2013.

[2] C. Ning and F. You, “Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods,” *Comput. Chem. Engineering*, vol. 112, pp. 190–210, 2018.

[3] P. Li, X. Guan, J. Wu and X. Zhou, “Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1594–1605, 2015.

[4] C. Li, J. Zhao, T. Zheng and E. Litvinov, “Data-driven uncertainty sets: Robust optimization with temporally and spatially correlated data,” in 2016 IEEE Power and Energy Soc. Gen. Meeting (PESGM), 2016.

[5] H. Cramer, *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1974, pp. 284–85.

[6] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” Stanford University Infolab, Stanford, 2006.