# *Low-dimensional space- and time-coupled power system control policies driven by high-dimensional ensemble weather forecasts*

Article

Accepted Version

Warrington, J., Drew, D. and Lygeros, J. (2018) Low-dimensional space- and time-coupled power system control policies driven by high-dimensional ensemble weather forecasts. IEEE Control Systems Letters, 2 (1). pp. 1-6. ISSN 2475-1456 doi: https://doi.org/10.1109/LCSYS.2017.2720467 Available at https://centaur.reading.ac.uk/72612/



To link to this article DOI: http://dx.doi.org/10.1109/LCSYS.2017.2720467

Publisher: IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2720467, IEEE Control Systems Letters

# Low-dimensional space- and time-coupled power system control policies driven by high-dimensional ensemble weather forecasts

Joseph Warrington⋆, *Member, IEEE,* Daniel Drew†, and John Lygeros⋆, *Fellow, IEEE*

***Abstract*—Many predictive control problems can be solved at lower cost if the practitioner is able to make use of a high-dimensional forecast of exogenous uncertain quantities. For example, power system operators must accommodate significant short-term uncertainty in renewable energy infeeds. These are predicted using sophisticated numerical weather models, which produce an ensemble of scenarios for the evolution of atmospheric conditions. We describe a means of incorporating such forecasts into a multistage optimization framework able to make use of spatial and temporal correlation information. We derive an optimal procedure for reducing the size of the look-ahead problem by generating a low-dimensional representation of the uncertainty, while still retaining as much information as possible from the raw forecast data. We then demonstrate application of this technique to a model of the Great Britain grid in 2030, driven by the raw output of a real-world high-dimensional weather forecast from the UK Met Office. We also discuss applications of the approach beyond power systems.**

***Index Terms*—Power systems, predictive control for linear systems, robust control**

## I. Introduction

MANY systems are operated in the presence of high-dimensional disturbance forecasts. Examples include transportation networks with uncertain usage patterns, air traffic control carried out under complex atmospheric conditions, and power systems with estimated short-term renewable power infeeds. This paper focuses on power system dispatch in the presence of uncertain wind power injections, but we note from the outset that the approach we present is general, and could be applied to other uncertainty sources or other applications.

Uncertainty in power systems is accommodated via a reserve mechanism, in which parts of the capabilities of generators and other devices are set aside to compensate for forecast errors as they arise. Recent work [1], [2] has formalized the notion of a planned *affine* response to the uncertainty realizations, achieved in real time through changes to Automatic Generator Control (AGC) parameters [3]. A key challenge of such formulations is to ensure that the solution respects system constraints under different possible realizations of the forecast error. Recent formulations derive policies which satisfy constraints with high probability in cases where only limited forecast data is available, and the true distribution of forecast errors is unknown [2], [4].

⋆ Automatic Control Laboratory, Swiss Federal Institute of Technology (ETH) Zurich, Physikstrasse 3, 8092 Zurich, Switzerland. Contact: {warrington,lygeros}@control.ee.ethz.ch

† Dept. of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, UK. Contact: d.r.drew@reading.ac.uk

Our previous work [1] is based on multistage recourse policies originally developed for robust predictive control [5]. That theory enabled the development of so-called *reserve policies*, a time-coupled generalization of affine responses which allows device actions to depend not only on the forecast error realization at a future power delivery time, but also on realizations that will have been discovered between now and that future time [1], [6]. Such linear recourse approaches are a restriction with respect to general nonlinear policy functions, but the result may often in practice be near-optimal [7], and extensions to non-linear "liftings" of the underlying uncertainty are also available [8], [9]. We note that due to the high dimension of this decision problem, other well-known approaches such as dynamic programming (DP) or dual DP [10] would not be tractable except for very small problem instances.

In contrast to the min-max formulations of [6], [11], we consider the minimization of an *expected* short-run cost. This is because a power system operator who prioritizes optimality for worst-case scenarios, which in practice rarely arise, would in general bring about economic losses in the long run. To optimize expected cost under uncertainty, it becomes necessary to include statistics of the correlation (in space and time) of forecast errors in the cost function [1].

This paper is concerned with accommodating such statistics in a tractable manner for optimization purposes, even when they arise from an underlying model of high dimension. We describe an optimal means of reducing the dimension of the uncertainty model, such that the resulting control policies can be described by a relatively low number of parameters, while still guaranteeing a supply-demand balance under all modelled uncertainty realizations. This is attractive for both practical and computational reasons, and for power systems allows the reserve policy approach of [1] to accept real-world high-dimensional forecasts as input data.

Section II describes the system model and optimization approach. Section III describes the procedure for optimal uncertainty dimension reduction, and Section IV demonstrates the method for a case study using real outputs from the UK Met Office's MOGREPS weather model [12]. Section V concludes and discusses applications beyond power systems.

## II. System model

This section describes the system model. Although we present a power system application in this paper, the notation is general and could in principle be applied to other systems of similar structure with minor modifications. We consider the optimal control of $N_\mathrm{p}$ generic grid-connected devices or participants, such as thermal generators and energy storage facilities, operating amidst uncertain power injections and/or withdrawals, for example from wind power and demand. We aim to operate the controllable devices at minimum expected cost while remaining robust to uncertainty, whose value is revealed stage-wise over a planning horizon. The grid on which the devices operate is modelled as a standard linearized AC network, in which phase angle differences between nodes are small, line losses negligible, and nodal voltage magnitudes equal. Line power flow constraints may be present.

### A. Grid-connected devices

Each device $i$ is modelled in state space form as a linear, time-invariant system with $n_i$ states whose output connects it to the grid. Its dynamics at time step $k$ are given by $x^i_{k+1} = \tilde{A}_i x^i_k + \tilde{B}_i u^i_k$, where $\tilde{A} \in \mathbb{R}^{n_i \times n_i}$ and $\tilde{B}_i \in \mathbb{R}^{n_i \times 1}$ are the state and input transition matrices respectively. Input $u^i_k$ is a scalar. Over a time horizon of length $T$, we represent the device's dynamics in stacked form in the usual manner for predictive control, denoting its current state $x^i_0$:

$$\mathbf{x}^i = A_i x^i_0 + B_i \mathbf{u}^i \,, \tag{1}$$

where $\mathbf{x}^i := [x^{i\prime}_1, \ldots, x^{i\prime}_T]' \in \mathbb{R}^{n_i T}$ and $\mathbf{u}^i := [u^{i\prime}_0, \ldots, u^{i\prime}_{T-1}]' \in \mathbb{R}^T$ are the stacked state and input vectors, and $A_i \in \mathbb{R}^{n_i T \times n_i}$ and $B_i \in \mathbb{R}^{n_i T \times T}$ are built from $\tilde{A}_i$ and $\tilde{B}_i$, as in [1]. We use symbol $'$ to signify a vector or matrix transpose. Vector $x^i_0$ is a given initial state.

Each such device has a cost function applicable over the same look-ahead horizon, consisting of constant, linear, and quadratic terms in $\mathbf{x}^i$ (e.g. to penalize a stored energy state) and $\mathbf{u}^i$ (e.g. to penalize generation costs):

$$J_i(\mathbf{x}^i, \mathbf{u}^i) := f^{x\prime}_i \mathbf{x}^i + \frac{1}{2}\mathbf{x}^{i\prime} H^x_i \mathbf{x}^i + f^{u\prime}_i \mathbf{u}^i + \frac{1}{2}\mathbf{u}^{i\prime} H^u_i \mathbf{u}^i + c_i \tag{2}$$

Note that $\mathbf{x}^i$ depends on $\mathbf{u}^i$, and by substitution from (1) it can be shown that for optimization purposes $J_i$ is convex in $\mathbf{u}^i$ if and only if matrix $(B'_i H^x_i B_i + H^u_i)$ is positive semidefinite.

Alongside controllable dynamics, each device $i$ may be the originator of uncontrollable power injections, represented as $r_i + G_i\delta$, where $r_i \in \mathbb{R}^T$ is a nominal component and the term $G_i\delta$ arises from an exogenous random vector $\delta \in \mathbb{R}^{N_\delta T}$, where $N_\delta$ is the uncertainty dimension. The vector takes the form $\delta := [\delta'_1, \ldots, \delta'_T]'$ and is common to the whole system. $G_i \in \mathbb{R}^{T \times N_\delta T}$ is a device-specific mapping.

The value of $\delta$ is unknown at the start of the time horizon. However, we assume that it will be possible to measure sub-vectors $\delta_k$ as they are revealed, sequentially over the horizon (this is discussed further in Section II-C). The uncertainty $\delta$ belongs to a set $\Delta \subset \mathbb{R}^{N_\delta T}$, and we assume that an estimate of the first and second moments of its probability distribution, $\mathbb{E}[\delta]$ and $\mathbb{E}[\delta\delta']$, is available.

Constraints apply to devices' state and input vectors, and these may also depend linearly on the uncertainty vector $\delta$. This is expressed via a set $\mathcal{Z}_i$, defined for each device $i$ as:

$$\mathcal{Z}_i := \left\{ \begin{bmatrix} \mathbf{x}^i \\ \mathbf{u}^i \\ \delta \end{bmatrix} \middle| \; T_i \mathbf{x}^i + U_i \mathbf{u}^i + V_i \delta \le w_i \right\} , \tag{3}$$

where $T_i$, $U_i$, $V_i$, and $w_i$ define inequalities coupling the state, input, and uncertainty vector. Note that $\mathbf{x}^i$ is itself a function of $\mathbf{u}^i$, although this is not made explicit in (3).

### B. Network constraints

The network sees power flows as a linear mapping from each controllable device $i$'s state, $\tilde{C}_i x^i_k$ at each time step $k$, plus terms arising from the uncontrollable power injections associated with each device. Using the stacked notation of (1), the requirement that power injections and withdrawals balance over the network is represented by the following constraint, which has one row for each of the $T$ time steps:

$$\sum_{i=1}^{N_\mathrm{p}} (r_i + G_i\delta + C_i\mathbf{x}^i) = 0, \quad \forall \delta \in \Delta \,, \tag{4}$$

where $C_i := I_T \otimes \tilde{C}_i$ maps the device's state to power injections or withdrawals seen by the network over the horizon. $I_T$ represents the $T \times T$ identity matrix, and $\otimes$ the Kronecker product. Line flow constraints may also be present, and are represented by linear inequalities,

$$\sum_{i=1}^{N_\mathrm{p}} \Gamma_i (r_i + G_i\delta + C_i\mathbf{x}^i) \le \bar{p}, \quad \forall \delta \in \Delta \,, \tag{5}$$

where coefficient matrices $\Gamma_i$ are built from consideration of a standard linearized line flow model [1].

### C. Finite horizon optimization

We wish to operate the power system devices for minimum expected cost by choosing control policies $\mathbf{u}_i = \pi_i(\delta)$ for each device $i$. Each policy specifies in advance how device $i$ should react to $\delta$ as its value unfolds. Applying the state dynamics (1) to the objectives and constraints described above yields the following optimization problem:

$$\min_{\text{Causal } \pi_i(\cdot)} \quad \sum_{i=1}^{N_\mathrm{p}} \mathbb{E}[J_i(A_i x^i_0 + B_i\pi_i(\delta), \pi_i(\delta))] \tag{6a}$$

$$\text{s.t.} \sum_{i=1}^{N_\mathrm{p}} r_i + G_i\delta + C_i(A_i x^i_0 + B_i\pi_i(\delta)) = 0 \,, \forall \delta \in \Delta \tag{6b}$$

$$\sum_{i=1}^{N_\mathrm{p}} \Gamma_i (r_i + G_i\delta + C_i(A_i x^i_0 + B_i\pi_i(\delta))) \le \bar{p}, \forall \delta \in \Delta \tag{6c}$$

$$\begin{bmatrix} A_i x^i_0 + B_i\pi_i(\delta) \\ \pi_i(\delta) \\ \delta \end{bmatrix} \in \mathcal{Z}_i \,, \forall \delta \in \Delta \tag{6d}$$

"Causal $\pi_i(\cdot)$" means that the input determining the state at time $k$ can depend only on elements known up to time $k$:

$$u^i_{k-1} = [\pi_i(\delta)]_{k-1} = \phi^i_k(\delta_1, \ldots, \delta_k) \,, \tag{7}$$

for some function $\phi^i_k : \times^k_{j=1} \mathbb{R}^{N_\delta} \to \mathbb{R}$. We allow $u^i_{k-1}$ to depend on $\delta_k$ on the assumption that controller gains on each device can be adjusted to respond to time step $k$'s disturbance as it is revealed in continuous time [1]. For power systems, this is the manner in which equality constraint (6b) can be satisfied in real time for all realizations of $\delta$ [2].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2720467, IEEE Control Systems Letters

For tractability, we limit $\pi_i(\cdot)$ to *affine* policies with a nominal term $e_i \in \mathbb{R}^T$ and a matrix $D_i \in \mathbb{R}^{T \times N_\delta T}$ of linear responses to $\delta$:

$$\mathbf{u}^i = \pi_i(\delta) = D_i\delta + e_i\,. \tag{8}$$

Matrix $D_i$ is composed of blocks of dimension $1 \times N_\delta$, where the $(l, m)$ block $[D_i]_{l,m}$ determines the response at time $l$ to the uncertainty realization discovered at time $m$. Causality requirement (7) is encoded in a constraint that $D_i$ be block-lower-triangular, i.e., $[D_i]_{l,m} = 0$ for all $m > l$.

We also define the expected cost $\tilde{J}_i(x_0^i, D_i, e_i) := \mathbb{E}[J_i(\mathbf{x}^i, \mathbf{u}^i)]$, which contains the statistics $\mathbb{E}[\delta]$ and $\mathbb{E}[\delta\delta']$ as well as the terms in (2). Problem (6) can then be written

$$\min_{(D_i,e_i)\in\mathcal{F}_i(x_0^i)} \quad \sum_{i=1}^{N_\mathrm{p}} \tilde{J}_i(x_0^i, D_i, e_i) \tag{9a}$$

$$\text{s.t.} \sum_{i=1}^{N_\mathrm{p}} r_i + G_i\delta + C_i(A_i x_0^i + B_i(D_i\delta + e_i)) = 0\,, \quad \forall \delta \in \Delta \tag{9b}$$

$$\sum_{i=1}^{N_\mathrm{p}} \Gamma_i(r_i + G_i\delta + C_i(A_i x_0^i + B_i(D_i\delta + e_i))) \le \overline{p}\,, \quad \forall \delta \in \Delta\,, \tag{9c}$$

where $\mathcal{F}_i(x_0^i)$ is the set of constraint-admissible, causal policies for the current state of device $i$:

$$\mathcal{F}_i(x_0^i) := \left\{ (D_i, e_i) \;\middle|\; \begin{array}{ll} [D_i]_{l,m} = 0, & \forall m > l \\ \begin{bmatrix} A x_0^i + B_i(D_i\delta + e_i) \\ D_i\delta + e_i \\ \delta \end{bmatrix} \in \mathcal{Z}_i, \forall \delta \in \Delta \end{array} \right\}$$

### D. Solution method

The difficulty of solving problem (9) depends to a great extent on how the uncertainty is modelled. In this paper we are concerned with data arising from a numerical prediction model producing a finite ensemble of $N_\mathrm{ens}$ forecast scenarios,

$$\Delta := \{\delta^{(n)},\, n = 1, \ldots, N_\mathrm{ens}\}\,.$$

In this case the inequality constraint (9c), as well as those present in the definitions of each device's constraint set $\mathcal{F}_i(x_0^i)$, can be replaced by $N_\mathrm{ens}$ sets of scenario constraints.[1]

It is trivial to show that the following two constraints together imply equality constraint (9b):

$$\sum_{i=1}^{N_\mathrm{p}} (r_i + C_i A_i x_0^i + C_i B_i e_i) = 0 \tag{10a}$$

$$\sum_{i=1}^{N_\mathrm{p}} (G_i + C_i B_i D_i) = 0 \tag{10b}$$

Constraint (10a) ensures the nominal power infeeds (i.e., those when $\delta = 0$) sum to zero, and constraint (10b) ensures that all forecast errors $\delta$ are matched by the compensating actions of devices, described by policy matrices $D_i$. Moreover, equations (10a) and (10b) are equivalent to (9b) if $\Delta$ has been constructed to include the origin in the interior of its convex hull [1].[2]

## III. Dimensionality reduction

If $N_\delta$, the dimension of the uncertainty set, is large, then choosing a reaction to each of the elements of $\delta$ requires heavy computation. Moreover, the high complexity of the controller may be unacceptable to the practitioner.

### A. Response basis vectors

One way of reducing this burden is to choose *a priori*, based on the characteristics of the forecast data, a limited number of vectors $\rho_k^p \in \mathbb{R}^{N_\delta}$, $p = 1, \ldots, P$, that form a basis for the system's response to uncertainty during any given time step $k$. The matrix $D_i$ in equation (8) would then take the form

$$\mathbf{u}^i = D_i\delta + e_i = \sum_{p=1}^{P} \bar{D}_i^p R^p \delta + e_i \tag{11}$$

where $\bar{D}_i^p \in \mathbb{R}^{T\times T}$ is a lower-triangular matrix of scalar response coefficients for component $p$ of the basis,

$$\bar{D}_i^p = \begin{bmatrix} d_{i,0,0}^p & 0 & \cdots & 0 \\ d_{i,1,0}^p & d_{i,1,1}^p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_{i,T-1,0}^p & d_{i,T-1,1}^p & \cdots & d_{i,T-1,T-1}^p \end{bmatrix}, \tag{12}$$

where the lower-triangular structure ensures that the system response does not rely on parts of the vector $\delta$ that have not yet been revealed. Each matrix $R^p \in \mathbb{R}^{T\times N_\delta T}$ is a block-diagonal matrix common to all participants $i$, containing row vectors weighting the elements of vectors $\delta_1, \ldots, \delta_T$:

$$R^p = \begin{bmatrix} \rho_1^{p\prime} & 0 & \ldots & 0 \\ 0 & \rho_2^{p\prime} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \rho_T^{p\prime} \end{bmatrix}. \tag{13}$$

Following a straightforward reformulation of problem (9) using substitution (11), there will now only be $\mathcal{O}(N_\mathrm{p} P T^2)$ optimization variables to choose in relation to the $D_i$ matrices, as opposed to order $\mathcal{O}(N_\mathrm{p} N_\delta T^2)$ in the original problem.[3] The basis vectors and choice of $P$ affect how much of the (spatial and temporal) correlation information encoded in $\mathbb{E}[\delta\delta']$ enters into the optimization through the cost function (9a) via terms of the form $R^p\mathbb{E}[\delta\delta']R^{p\prime}$.

[1] If set $\Delta$ does not have a finite number of elements, then this constraint enumeration approach cannot be used directly. Alternatively, if it consists of a finite but very large number of scenarios, then the added constraints may become too numerous for the problem to be tractable. In both these cases, an outer approximation such as a polytope may have to be found, following which a tractable reformulation employing auxiliary matrix variables is available [1]. However, in the case of ensemble weather forecasts, it is currently typical for $N_\mathrm{ens}$ to extend only up to a few dozen scenarios, such that it remains preferable to enumerate the constraints by scenario.

[2] In the case where $N_\mathrm{ens} < N_\delta T + 1$, set $\Delta$ will consist of coplanar points and its convex hull will therefore have an empty interior. However the practitioner is still likely to prefer to implement the sufficient constraints (10a) and (10b), on the assumption that the $N_\mathrm{ens}$ scenarios were sampled from a distribution with full-dimensional support, and subsequent samples will not in general be coplanar with those already modelled.

[3] Note that with minor adjustments to the formulation, one could in principle use a different value of $P$ for each time step $k$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2720467, IEEE Control Systems Letters

The question arises how to choose the $P$ basis vectors $\rho_k^p$ for each time step $k$, with $P \ll N_\delta$, such that

1) Constraint (10b) is satisfied, meaning the $D_i$-matrices still track any disturbance $\delta \in \Delta$ after imposing the structural constraint (11);
2) The resulting basis is in some sense "optimal", in that it allows the system to respond in directions that matter most in the space of possible uncertainty realizations.

### B. Conditions for a feasible response basis

We now develop conditions under which a reduced-dimension response to uncertainty can still balance the system, i.e., satisfy constraint (10b). We make the following assumptions, which can straightforwardly be made to hold for reasonable forecast models, and for devices connected to a single node of the network:

**Assumption 1.** *Each matrix $G_i \in \mathbb{R}^{T\times N_\delta T}$, which models how uncertainty $\delta$ contributes to participant $i$'s uncontrollable power injections, is block-diagonal with the form $G_i = \mathrm{blkdiag}(g_{i,1}', \dots, g_{i,T}')$, for some vectors $g_{i,k} \in \mathbb{R}^{N_\delta}$.*

**Assumption 2.** *As described in the appendix of [1], each controllable device $i$ has been modelled such that its input appears with unity scaling factor as a net power injection on the network, leading to the simplification $C_iB_i = I_T$.*

Under Assumptions 1 and 2, consider the case $P = 1$, i.e. only one basis component is allowed. Then from examination of the $(k, k)$ block of constraint (10b), for each $k$ the only permissible choices of basis vector $\rho_k^1$ are those satisfying

$$\sum_{i=1}^{N_\mathrm{p}} g_{i,k} + \left(\sum_{i=1}^{N_\mathrm{p}} d_{i,k,k}^1\right)\rho_k^1 = 0\,. \tag{14}$$

Equation (14) restricts $\rho_k^1$ to be a scalar multiple of $\sum_{i=1}^{N_\mathrm{p}} g_{i,k}$. In such a case, the control policy can only be driven by a single parameter, namely the scalar $\sum_{i=1}^{N_\mathrm{p}} g_{i,k}'\delta_k$.

If we allow more basis components, i.e., $P > 1$, more nuanced responses to the uncertainty become possible. In particular, the extra freedom allows the optimization to produce responses governed by different weightings of the elements of vector $\delta_k$. Constraint (14) becomes

$$\sum_{i=1}^{N_\mathrm{p}} g_{i,k} + \sum_{p=1}^{P}\left(\sum_{i=1}^{N_\mathrm{p}} d_{i,k,k}^p\right)\rho_k^p = 0\,. \tag{15}$$

We define $\mathcal{R}_k(\rho_k^1, \dots, \rho_k^P)$ as an affine subspace of $\mathbb{R}^{N_\delta}$, parameterized by basis vectors $\rho_k^1, \dots \rho_k^P$, of summed, or aggregated, system responses at time $k$ that are feasible for constraint (10b),

$$\mathcal{R}_k(\rho_k^1, \dots, \rho_k^P) := \left\{ \alpha_1\rho_k^1 + \ldots + \alpha_P\rho_k^P \;\middle|\; \sum_{i=1}^{N_\mathrm{p}} g_{i,k} + \sum_{p=1}^{P}\alpha_p\rho_k^p = 0 \right\}. \tag{16}$$

Then $(\rho_k^1, \dots, \rho_k^P)$ is a permissible basis as long as there exist scalars $(\alpha_1, \dots, \alpha_P)$ that satisfy $\sum_{i=1}^{N_\mathrm{p}} g_{i,k} + \sum_{p=1}^{P}\alpha_p\rho_k^p = 0$ for given vectors $g_{i,k}$.

### C. Optimal response basis

For each time step $k$, the basis vectors should be chosen from permissible parameterizations of $\mathcal{R}_k$ such that as much as possible of the information in the original set of data points is preserved for the purpose of system optimization.[4] Formally, we wish to maximize the mutual information between points $\delta_k^{(n)}$ and their projections onto the space spanned by $\{\rho_k^p\}_{p=1}^P$. The following proposition describes how to achieve this.

**Proposition 1.** *Define matrix $M_{g,k} := I - g_k(g_k'g_k)^{-1}g_k'$, where $g_k = \sum_{i=1}^{N_p} g_{i,k}$, and the matrix*

$$W_k := M_{g,k}'\left[\delta_k^{(1)}, \cdots, \delta_k^{(n)}\right]\begin{bmatrix}\delta_k^{(1)\prime}\\ \vdots \\ \delta_k^{(n)\prime}\end{bmatrix} M_{g,k}\,. \tag{17}$$

*Then the matrix $\mathbf{R}_k := [\lambda_{1,k}, \lambda_{2,k}, \cdots, \lambda_{P-1,k}, g_k] \in \mathbb{R}^{N_\delta\times P}$ (where $\lambda_{1,k}, \dots, \lambda_{P-1,k} \in \mathbb{R}^{N_\delta}$ are the first $P-1$ eigenvectors of matrix $W_k$ sorted in descending order of the magnitudes of their corresponding eigenvalues), maximizes the mutual information between scenario data points $\delta_k^{(n)}$ and their projections $\mathbf{R}_k(\mathbf{R}_k'\mathbf{R}_k)^{-1}\mathbf{R}_k'\delta_k^{(n)} =: \Pi_k\delta_k^{(n)}$. This corresponds to maximizing the determinant of the covariance matrix of the projected points,*

$$\det\left(\frac{1}{N_\mathrm{ens}}\sum_{n=1}^{N_\mathrm{ens}}\left(\Pi_k\delta^{(n)} - \Pi_k\mathbb{E}[\delta]\right)\left(\Pi_k\delta^{(n)} - \Pi_k\mathbb{E}[\delta]\right)'\right).$$

*To achieve this maximization, for each time step $k$, matrices $R^p$ defined in equation (13) should therefore be populated according to $\rho_k^p = \lambda_{p,k}$ for $p = 1, \dots, P-1$, and $\rho_k^P = g_k$ for $p = P$.*

*Proof.* The proof follows the argument made by Rolle [13, §3.1] for constrained Principal Component Analysis. The sum of squared distances of a group of points from their projections onto a given $P$-dimensional subspace, constrained to include vector $g_k$, is minimized when the first $P-1$ eigenvectors of $W_k$ are used as a basis for the remaining $P-1$ dimensions. This also maximizes the determinant of the covariance matrix of the projected points. □

**Corollary 1.** *The optimal response to a high dimensional uncertainty when $P = 1$ is constrained in the manner described in equation (14), i.e. $\rho_k^1 = \beta g_k$ for any scalar $\beta \neq 0$. When $P > 1$ is permitted, it is optimal to use the same vector plus the first $P-1$ eigenvectors of $W_k$.*

Note that since $M_{g,k}\delta$ is the residual vector of the projection of $\delta$ onto the ray in direction $g_k$, $g_k$ is itself in the nullspace of $M_{g,k}$ (and also in the nullspace of $W_k$). Therefore all eigenvectors of $W_k$ are by construction orthogonal to $g_k$.

[4] In principle, optimal basis vectors $\rho_k^p$ could be chosen as part of problem (9). However, alongside other numerical issues this would introduce awkward bilinearities, since matrices $R^p$ multiply other optimization variables. Therefore we restrict ourselves to an *a priori* choice before solving (9).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2720467, IEEE Control Systems Letters

## IV. Numerical example

We now demonstrate the approach on the output of a real-world weather forecast model, applied to a lumped model of the Great Britain (GB) power system based on real data. Subsection IV-A explains how the uncertainty scenarios $\delta^{(n)}$ are extracted from the weather model, and Subsection IV-B explains how these feed into the power system model.

### *A. Weather forecast model*

We consider the output of a numerical weather prediction model in which atmospheric dynamics are solved over a large grid containing $N_{\mathrm{grid}}$ cells. It consists of an *ensemble* of forecast scenarios, $q^{(n)} = [q_1^{(n)\prime}, q_2^{(n)\prime}, \ldots, q_T^{(n)\prime}]'$ for $n = 1, \ldots, N_{\mathrm{ens}}$. Each sub-vector $q_k^{(n)} \in \mathbb{R}_+^{N_{\mathrm{grid}}}$ contains the value of the weather variable modelled at each of $N_{\mathrm{grid}}$ points, at time step $k$ in scenario $n$.

In this simulation $q_k^{(n)}$ represents modelled wind speeds. To obtain wind power forecasts, $q_k^{(n)}$ must be mapped to normalized power availability via a wind turbine characteristic function, $y_k^{(n)} = f(q_k^{(n)})$, where $f : \mathbb{R}_+^{N_{\mathrm{grid}}} \to [0,1]^{N_{\mathrm{grid}}}$.

Only some forecast locations will be relevant to the power system. Assuming for ease of presentation that wind farm locations coincide with grid points, the wind power output at time $k$ under scenario $n$ can be expressed in the vector

$$y_k^{\mathrm{wf}(n)} := \Phi^{\mathrm{wf}} y_k^{(n)}, \tag{18}$$

where the rows of matrix $\Phi^{\mathrm{wf}} \in \mathbb{R}^{N_\delta \times N_{\mathrm{grid}}}$ contain a non-zero element equal to the wind farm rating for the appropriate grid point. $N_\delta$ is the number of wind farms, and therefore the dimension of forecast errors arising for wind farms, as described below. We define $y^{\mathrm{wf}(n)} := [y_1^{\mathrm{wf}(n)\prime}, \ldots, y_T^{\mathrm{wf}(n)\prime}]'$ and $\mathbb{E}[y^{\mathrm{wf}}] = \frac{1}{N_{\mathrm{ens}}} \sum_{n=1}^{N_{\mathrm{ens}}} y^{\mathrm{wf}(n)}$.

Relating this to the model described in Section II-A, each device $i$'s vector of uncontrolled power outputs over the forecast horizon, $r_i \in \mathbb{R}^T$, is then given by

$$r_i := r_i^{\mathrm{other}} + \Psi_i \mathbb{E}[y^{\mathrm{wf}}] \tag{19}$$

where $r_i^{\mathrm{other}}$ denotes uncontrolled flows that are unrelated to the wind model, such as power demand. Each block-diagonal matrix $\Psi_i \in \mathbb{R}^{T \times N_\delta T}$ contains "ones" selecting the appropriate elements of $\mathbb{E}[y^{\mathrm{wf}}]$, or contains only zeroes if participant $i$ does not include a wind farm.

We define $\delta := \hat{y}^{\mathrm{wf}} - \mathbb{E}[y^{\mathrm{wf}}]$ as the difference between forecast and realised quantities. Then the matrices $\Psi_i$ also describe how forecast errors feed into the grid as described in Section II-A, i.e. $G_i = \Psi_i$. This respects Assumption 1.

### *B. Power system model*

A lumped network model[5] of the island of Great Britain (GB) in 2030 was created, in order to simulate a look-ahead optimization over 12 one-hour steps. The electricity supply capacities and gas generation costs were based on the National Grid (NG) Future Energy Scenarios (FES) 2016 [14]; the transmission zones were based on the 7 simplified regions presented in [15], with transmission capacities between these zones as estimated for 2030 in NG's public ELSI model.[6]

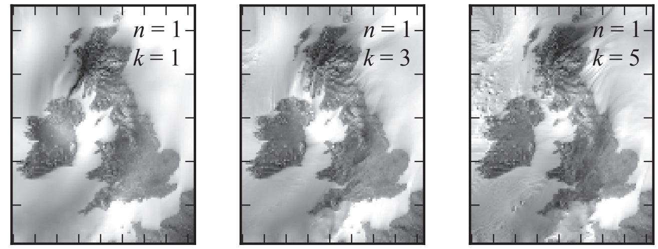


Fig. 1. Slices from the ensemble forecast output of the MOGREPS model, showing wind speed for scenario $n = 1$. Black represents a wind speed of zero, and white represents speeds of at least 15 m/s. The forecast was carried out at 03:00 on 2nd Nov 2014, and $k$ is measured from this time.

Non-wind generation was modelled as 49 controllable devices, in part according to the layout of existing generation and in part under reasonable assumptions given the long-term locational incentives offered by GB's TNUoS charging regime.[7] Each zone contained generation classed as Biomass/CHP, Gas, Hydro, Interconnectors, Nuclear, Other Renewables, or Storage.

The simulation was fed by the Met Office MOGREPS model of the region, as illustrated in Fig. 1. This contains a control forecast plus 11 perturbed variations; for demonstration purposes these were considered as $N_{\mathrm{ens}} = 12$ equally-weighted scenarios. Each scenario contains $N_{\mathrm{grid}} = 230{,}287$ spatial points evenly covering the geographical area in a $2 \times 2$ km grid, and hourly time steps out to 36 hours ahead, of which the first 12 were used. 337 wind farm locations were derived from real-world data, with capacities consistent with projections for 2030. As a result, the dimension of the uncertainty introduced to the model is also $N_\delta = 337$. The normalized wind power mapping $f(\cdot)$ took value 0 for argument below 4 m/s, increased from this point to a maximum of 1 at 16 m/s, and kept this value up until a cut-out speed of 24 m/s, above which point it took value 0.

The computer had an Intel Core i7 CPU at 2.3 GHz (4 cores with 2 threads each), and 16 GB RAM. Implementation was in Python 2.7 and Gurobi 6.5.0.

### *C. Summary of approach*

The following steps were used to construct the reduced-order response to an $(N_{\mathrm{ens}} \times N_{\mathrm{grid}} \times T)$-dimensional raw forecast dataset for given approximation order $P$:

(a) Convert raw wind speed scenarios $q^{(n)}$ to wind power availability $y^{(n)}$;
(b) Map power availability at the $N_{\mathrm{grid}}$ raw forecast points to wind farm power availability at the $N_\delta$ wind farm locations, $y^{\mathrm{wf}(n)} = \Phi^{\mathrm{wf}} y^{(n)}$;
(c) For each time step $k$, form matrix $W_k$ according to (17), with $\delta_k^{(n)} = y_k^{\mathrm{wf}(n)} - \mathbb{E}[y_k^{\mathrm{wf}}]$;
(d) Set $\rho_k^P = g$, and if $P > 1$ set $\rho_k^p = \lambda_{p,k}$ as described in Proposition 1 for $p = 1, \ldots, P-1$, for all steps $k$;
(e) Form matrices $R^p$ and insert into problem (9) using relationship (11).

[5] http://control.ee.ethz.ch/index.cgi?page=publications&action=details&id=5656.
[6] http://www2.nationalgrid.com/WorkArea/DownloadAsset.aspx?id=39022
[7] http://www2.nationalgrid.com/UK/Industry-information/System-charges/Electricity-transmission/Transmission-network-use-of-system-charges/

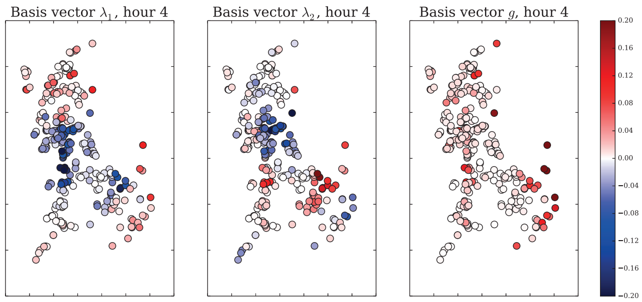


Fig. 2. Basis vectors for $P = 3$, for responding to correlated wind power forecast errors during hour $k = 4$. The figures show visualizations of vectors $\lambda_1$, $\lambda_2$, and $g_k$, as derived from the method in Proposition 1. Clear structures arise from the underlying forecast data; within each component, forecast errors at blue points are negatively correlated with those at red points. Note all elements of $g_k$ (third plot) are positive, since these relate higher-than-expected wind power availability to grid power injections.

### D. Results

The following configurations were simulated:

(1) *Perfect foresight:* A deterministic optimization was carried out with full knowledge of the realization of $\delta$, (a) for the nominal case $\delta = 0$, and (b) for each $\delta$ scenario and averaging the resulting costs.

(2) *Full time-coupled (lower-triangular) policy matrices $D_i$:* The uncertainty model reduction procedure of Proposition 1 was performed.

(3) *Diagonal policy matrices $D_i$:* As with (2) above, but with the recourse matrices $D_i$ restricted to block-diagonal form. The effect of this is to disallow time coupling in the control policy. This resembles current secondary reserve mechanisms, as discussed in [1].

The resulting costs and computation times are given in Table I, and an example of the reduced forecast model is shown in Fig. 2 for $P = 3$. Costs decreased rapidly for even modest choices of $P$ above 1, and costs also reduced when time-coupled responses were permitted.[8] Solver time increased steeply beyond $P = 3$ in case (2), which may in part be related to the numerical conditioning of the problem. In case (3), time-decoupled policies, the diagonal form of matrices $\overline{D}_i^p$ allowed the solver to eliminate the off-diagonal variables and solution times are lower. The average reserve cost under perfect foresight, case (1b), represents a lower bound (0.439% in this case) on the cost that one could aim to achieve for any causal policy with a robustness requirement; configuration (2) with $P$ as low as 3 was almost able to match this.

## V. Conclusions

We have demonstrated a means of solving stochastic look-ahead optimization problems made difficult by high-dimensional forecast uncertainty, with particular application to power system control. The method allows the raw output of a real-world numerical forecast model to be incorporated into balanced system dispatch decisions, and preserves the main spatial and temporal correlations while substantially reducing the number of decision variables required.

Although this paper has focused on a power system application, our approach to uncertainty model reduction is generic and can be extended to other systems with similar structure to (9), in which expected costs for multiple devices or agents should be minimized in the presence of local constraints and coupling (equality and/or inequality) constraints.

[8]On the test hardware it was not possible to solve the unreduced model ($P = N_\delta$), which has over 2.4 million decision variables. We also note that since $N_{\text{ens}} = 12$ in this case, the full original forecast information can in fact be reconstructed with $P = 12$, although this was also too large to solve on the test hardware.

TABLE I
NUMERICAL RESULTS

| Configuration | Vars[1] | Cost ($\times 10^6$) | Reserve cost | Solver time[2] (s) |
|---|---|---|---|---|
| (1a), $\delta = 0$ | 588 | 7.39775 | — | 0.020 |
| (1b), average | 588 | 7.43023 | 0.439% | 0.020 |
| (2), $P = 1$ | 4410 | 7.47856 | 1.092% | 4.36 |
| (2), $P = 3$ | 12054 | 7.43194 | 0.462% | 8.57 |
| (2), $P = 5$ | 19698 | 7.43087 | 0.448% | 1086 |
| (3), $P = 1$ | 1176 | 7.66483 | 3.610% | 2.53 |
| (3), $P = 3$ | 2352 | 7.57503 | 2.397% | 3.76 |
| (3), $P = 5$ | 3528 | 7.52922 | 1.777% | 6.25 |

[1] Excludes variables trivially constrained to zero.
[2] Plus basis computation time in cases (2) and (3) of 2.77 s.

## References

[1] J. Warrington, P. Goulart, S. Mariethoz, and M. Morari, "Policy-Based Reserves for Power Systems," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4427–4437, 2013.
[2] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-Constrained Optimal Power Flow: Risk-Aware Network Control under Uncertainty," *SIAM Review*, vol. 56, no. 3, pp. 461–495, Jan. 2014.
[3] J. Warrington, S. Mariethoz, and M. Morari, "Time-sequence reserve products for electricity markets," in *International Conference on the European Energy Market*, Stockholm, Sweden, 2013.
[4] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on conditional value at risk and distributional robustness," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 116–125, Nov. 2015.
[5] P. Goulart, E. Kerrigan, and J. Maciejowski, "Optimization over state feedback policies for robust control with constraints," *Automatica*, vol. 42, no. 4, pp. 523–533, 2006.
[6] Á. Lorca, X. A. Sun, E. Litvinov, and T. Zheng, "Multistage Adaptive Robust Optimization for the Unit Commitment Problem," *Operations Research*, vol. 64, no. 1, pp. 32–51, Jan. 2016.
[7] D. Bertsimas, D. A. Iancu, and P. A. Parrilo, "Optimality of Affine Policies in Multistage Robust Optimization," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 363–394, May 2010.
[8] C. Wang, C.-J. Ong, and M. Sim, "Model Predictive Control Using Segregated Disturbance Feedback," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 831–840, 2010.
[9] P. N. Beuchat, J. Warrington, T. H. Summers, and M. Morari, "Performance Bounds for Look-Ahead Power System Dispatch Using Generalized Multistage Policies," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 474–484, Jan. 2016.
[10] M. V. F. Pereira and L. M. V. G. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Mathematical Programming*, vol. 52, no. 1-3, pp. 359–375, May 1991.
[11] D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng, "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem," *IEEE Trans. Power Sys.*, vol. 28, no. 1, pp. 52–63, 2013.
[12] Met Office, "The Met Office ensemble system (MOGREPS)." [Online]. Available: http://www.metoffice.gov.uk/research/weather/ensemble-forecasting/mogreps
[13] J. Rolle, "Optimal subspaces and constrained principal component analysis," *Elec. J. of Lin. Algebra*, vol. 10, no. 1, Jan. 2003.
[14] National Grid, "Future Energy Scenarios 2016," Tech. Rep., Jul. 2016. [Online]. Available: http://fes.nationalgrid.com
[15] A. Lawson, M. Goldstein, and C. J. Dent, "Bayesian framework for power network planning under uncertainty," *Sustainable Energy, Grids and Networks*, vol. 7, pp. 47–57, Sep. 2016.