# Optimization via Statistical Emulation and Uncertainty Quantification: Hosting Capacity Analysis of Distribution Networks

**Hailiang Du[1,2], Wei Sun[3], (Member, IEEE), Michael Goldstein[1] and Gareth Harrison[3], (Senior Member, IEEE)**

[1]Department of Mathematical Sciences, Durham University, Durham, UK
[2]Centre for the Analysis of Time Series, London School of Economics, London, UK
[3]School of Engineering, University of Edinburgh, Edinburgh, UK

Corresponding author: Hailiang Du, Wei Sun (hailiang.du@durham.ac.uk, w.sun@ed.ac.uk).

**ABSTRACT** In power systems modelling, optimization methods based on certain objective function(s) are widely used to provide solutions for decision makers. For complex high-dimensional problems, such as network hosting capacity evaluation of intermittent renewables, simplifications are often used which can lead to the 'optimal' solution being suboptimal or nonoptimal. Even where the optimization problem is resolved, it would still be valuable to introduce some diversity to the solution for long-term planning purposes. This paper introduces a general framework for solving optimization for power systems that treats an optimization problem as a history match problem which is resolved via statistical emulation and uncertainty quantification. Emulation constructs fast statistical approximations to the complex computer simulation model in order to identify appropriate choices of candidate solutions for given objective function(s). Uncertainty quantification is adopted to capture multiple sources of uncertainty attached to each candidate solution and is conducted via Bayes linear analysis. It is demonstrated through a hosting capacity case study involving variable wind generation and active network management. The proposed method effectively identified not only the maximum connectable capacities but also a diverse set of near-optimal solutions, and so provided flexible guides for using the existing network to maximize the benefits of renewable generation.

**INDEX TERMS** Hosting capacity; distribution network; distributed generation; wind curtailment; history matching; optimization; statistical emulation; uncertainty quantification.

## I. INTRODUCTION

Effective decision-making for power systems is often addressed using optimization which, for example, has been applied to planning national-level energy policy [1], energy network expansion [2], and down to building level [3] or household systems [4]. Much of the interest is in constructing computer simulation models (physical models that approximate the real systems) to define the objective function based on historic system data or forecasts in order to identify an optimal solution for decision support. The computer simulation model takes input, describing the properties of the system, including the parameters of interest for optimization, and returns the corresponding value of the objective function.

The conventional optimization method is to find the global minimum/maximum solution of an objective function, via solving the zero solution to the derivative of the objective function. In power systems, the optimization task is commonly based on complex high-dimensional systems, for example, unit commitment (UC) and economic dispatch (ED) problems at transmission level, which feature thousands of different generator units at national scales to be optimized simultaneously; and hosting capacity evaluation (HC) of variable renewables at the distribution level, which needs to consider the uncertainty and variance of renewables over a very long horizon at high temporal resolution. Solving these mathematical problems becomes non-trivial and computationally expensive. Therefore, there is a trade-off to

consider between computational tractability and modelling accuracy. Since linear programming (LP) and mixed integer linear programming (MIP) are easier to solve than nonlinear programming problem (NLP), modelling of the physical components in the energy system often uses linearized simplifications [5], [6]. This approach can be found in many large-scale energy system studies. For example, the popular technology-rich TIMES (an acronym for The Integrated MARKAL-EFOM System) model family, used for developing energy policy, fully adopts linear programming and translates each technology via linear representation [7], [8]. However, an LP model could face the challenges of being unable to provide accurate or sufficiently detailed representations of real energy systems. When nonlinear optimization is considered necessary and used to improve the accuracy of models, it may suffer from local minima/maxima even if the problem is tractable. Some sophisticated approaches, for example, Second-Order Cone Programming [9] and Semidefinite programming [10], may be able to help solve problems with more computational cost. Even if the optimal solution is obtained, end users may still find it practically challenging to act precisely according to the optimal solution, given the uncertainties of real-work transition in policy, technology and costs.

It is important to note that the operational environment of the energy system is not static and varies from year to year with uncertainties. The optimal solution found in the certain deterministic case may have less satisfactory performance across future scenarios. Therefore, it is valuable that the modelling method can efficiently explore the full input space by allowing deviations from the optimality, helping to address structural uncertainties that arise when the real world deviates from the fixed consumption adopted by the modelling process. Providing a range of suboptimal solutions could be advantageous since it can increase the number of quality candidate plans for the planner to take into consideration against uncertainty. Taking electricity network expansion as an example, in practice the 'optimal' route for the new line may be found to be impossible due to land permissions or terrain considerations, and alternative suboptimal solutions are valuable. Therefore, a variety of quality solutions with objective function values close to that of the optimal solution offers decision makers flexibility against different scenarios.

When computer simulation models are adopted to make inferences about the behaviour of the real world, uncertainties arise. Some are due to observational error and others to model discrepancy between the computer simulation model and the real system. These uncertainties are rarely considered in conventional optimization as only the optimal solution is delivered to the decision maker.

To overcome the shortcomings of the conventional optimization approach, a general methodology based on statistical emulation and uncertainty quantification is proposed in this paper. The essence of the approach is to treat the optimization problem as a history matching problem [11], [12] where the goal is to locate a set of inputs for which the corresponding outputs meet certain criteria (e.g. the objective function is larger/smaller than some threshold). While Monte Carlo methods can sample the input space to locate such sets, the computational cost can be enormous (if achievable) for complex systems. Therefore, statistical emulations are adopted to explore the input space efficiently by approximating the computer simulation model with statistical models. The corresponding uncertainty is assessed via Bayes linear analysis [13]. Our approach provides not only a set of candidate 'optimal' solutions (due to various sources of uncertainty, one cannot identify the true unique optimal solution just a set of 'optimal' solutions that are indistinguishable from each other) but also quantifies the uncertainty attached to each solution.

A hosting capacity problem is used in this paper to demonstrate the proposed framework. Analysis of 'hosting capacity' for effective utilization of the existing network to improve renewable integration is now a popular research field [14]–[17] . 'Hosting capacity' is usually modelled as an optimization problem, with the objective function of maximum DG power capacity that can be connected, and subject to network constraints including line capacity, bus voltage, and power quality limits [18]–[21]. The solving methods for hosting capacity problems are mainly categorized as analytical methods and intelligent ones. Analytical methods are often used to convert non-convex and nonlinear power flow constraints into linear constraints, and then solve the relaxed model by using commercial software such as CPLEX and GUROBI. Recent developments, for example, include linear programming [22], polygonal inner-approximation [23], second-order cone relaxation [24], and the quadratically constrained method [25]. However, the adopted relaxations lose accuracy and are difficult to apply when the network and control model are complex. Popular intelligent optimization algorithms include Genetic Algorithms (GA) [26], [27] and Particle Swarm Algorithm (PSA) [15], [28], [29]. They are able to solve hosting capacity while maintaining intact network and control models, but require many iterations to find the optimum and do not scale well with complexity.

Given the issues with the above methods, the efficiency of hosting capacity solving techniques still needs to be improved. The proposed framework can tackle the challenges associated with conventional nonlinear nonconvex optimization and also consider uncertainty that is not addressed by deterministic approaches. The proposed method effectively identified the maximum connectable capacities as well as a diverse set of near-optimal alterative solutions, and therefore provided flexible guides for using the existing network to maximize the benefits of renewable generation. In this case study, candidate 'optimal' solutions identified from the proposed approach are found comparable to the

deterministic 'optimal' solution obtained from conventional optimization.

This paper provides four key contributions to energy system optimization. Firstly, a general framework of using history matching to address the optimization problem is proposed to overcome the shortcomings of the conventional optimization approach. Secondly, the proposed framework is demonstrated using a popular problem in renewable energy integration, namely wind farm hosting capacity. Thirdly, statistical emulation is demonstrated, in a complex energy system, to be able to explore the input space in a much more efficient way. Lastly, various sources of uncertainties have been assessed to aid decision support. The essential difference between the proposed optimization approach and the traditional (and intelligent) approaches is that instead of searching for an 'optimal' solution directly/interactively, the proposed approach adapts history matching to remove all the distinguishable 'non-optimal' solutions from the input (search) space, which leaves the reduced space that contains candidate 'optimal' solutions. Furthermore, our aim is to find the class of good decisions not just to carry out an optimization to find the "best" decision.

The structure of the article is as follows: the hosting capacity problem is given in Section II. The general framework of the proposed method is set out in Section III. A case study is presented in Section IV, followed by discussion and conclusions.

## II. HOSTING CAPACITY OPTIMIZATION PROBLEM

### A. HOSTING CAPACITY
This paper uses the problem of distribution network hosting capacity evaluation for wind distributed generation (DG) to demonstrate the proposed framework, which treats an optimization problem as a history matching problem resolved via statistical emulation and uncertainty quantification. Hosting capacity is the maximum capacity of new DGs that are connected at multiple locations across the network without technical (or other) limits being breached.

Active Network Management (ANM) [30], including active generation control, is also considered to help handle the fluctuation of renewables and increase hosting capacity.

### B. CONVENTIONAL OPTIMIZATION APPROACH
The conventional approach for hosting capacity evaluation is to model it as an optimization method. The optimization model adopted is based on AC optimal power flow (ACOPF), as used in a series of established works, so the proposed statistical approach can be readily benchmarked. In this context, it is considered as a reasonable test bed model to demonstrate the methodology. It is acknowledged that there are energy systems, as well as corresponding simulation models, that are far more complex than this.

### 1) OBJECTIVE FUNCTION

In the hosting capacity evaluation, the objective is to maximize the connectable capacity $C$ (rated power, MW) of new wind farms (*WF, G is the set of WFs*) at all potential locations:

$$\max \text{Hosting Capacity} = \max \sum_{g \in G} C_g \qquad (1)$$

The key information required includes the technical details of the network, potential locations for wind farms, demand time series at each bus, and time series of wind production at each potential wind farm location. The time series production from each potential DG is a function of the DG capacity and the level of resource. The hosting capacity is defined across all periods (e.g. hours) within a longer time horizon (e.g. a year).

Additional constraints are modelled to describe the nonlinear network AC power flow, such as the real and reactive nodal power balance, as well as voltage and power flow constraints. The formulation of all these constraints is found in [30].

### 2) POWER FLOW CONSTRAINTS
The optimization model is subject to a range of constraints, where the constraints on the power flow are as follows:

$$V_b^- \leq V_{b,m} \leq V_b^+ \qquad \forall b \in B \qquad (2)$$

$$\left(f_{l,m}^{(1,2),P}\right)^2 + \left(f_{l,m}^{(1,2),Q}\right)^2 \leq (f_l^+)^2 \qquad \forall l \in L \qquad (3)$$

$$
\begin{aligned}
f_{l,m}^{1,P} = g_l \cdot V_{\beta_l^1,m}^2 &- V_{\beta_l^1,m} \cdot V_{\beta_l^2,m} \\
&\cdot \left[g_l \cdot cos\left(\delta_{\beta_l^1,m} - \delta_{\beta_l^2,m}\right) + b_l \right. \\
&\left. \cdot sin\left(\delta_{\beta_l^1,m} - \delta_{\beta_l^2,m}\right)\right]
\end{aligned}
\qquad (4)
$$

$$
\begin{aligned}
f_{l,m}^{1,Q} = -b_l \cdot V_{\beta_l^1,m}^2 &- V_{\beta_l^1,m} \cdot V_{\beta_l^2,m} \\
&\cdot \left[g_l \cdot sin\left(\delta_{\beta_l^1,m} - \delta_{\beta_l^2,m}\right) - b_l \right. \\
&\left. \cdot cos\left(\delta_{\beta_l^1,m} - \delta_{\beta_l^2,m}\right)\right]
\end{aligned}
\qquad (5)
$$

$$
\sum_{l \in L|\beta_l^{1,2}=b}(f_{l,m}^{1,P} + f_{l,m}^{2,P}) + d_{b,m}^P = \sum_{g \in G|\beta_g=b} p_{g,m} + \sum_{\beta_{GSP}=b} p_{GSP,m}
\qquad (6)
$$

$$
\sum_{l \in L|\beta_l^{1,2}=b}(f_{l,m}^{1,Q} + f_{l,m}^{2,Q}) + d_{b,m}^Q = \sum_{g \in G|\beta_g=b} p_{g,m} \cdot tan(\phi_{g,m}) + \sum_{\beta_{GSP}=b} Q_{GSP,m}
\qquad (7)
$$

$$p_{g,m} = C_g \omega_m - p_{g,m}^{curt} \qquad (8)$$

Constraint (2) sets the voltage limits at each bus $b$ (B, set of buses) within maximum/minimum levels. Constraint (3) ensures the flow at each end of lines is not above thermal limits. The two terminal buses for line $l$ are denoted as $\beta_l^1$ and $\beta_l^2$. Equations (4)–(5) ($\forall l \in L$) describe the active and reactive power injections into the end '1' of the lines (the ends of each line are denoted as 1 and 2), which is given in terms of voltage level $V$ and phase $\delta$ by standard Kirchhoff's voltage formula. $g_l$ and $b_l$ are the conductance and susceptance on line $l$, respectively. The active and reactive equations for injection at bus 2 could be obtained by transposing the superscript labels 1 and 2 in (4) and (5). In terms of power flow direction, the direction from the bus injects into its connected line end is defined as positive and the opposite as negative. These constraints are applied for all lines in $l \in L$. When line $l$ contains on-load tap changers (OLTC) or a voltage regulator, the voltage at the start bus of the line would be divided by tap ratio $t_{l,m}$, within $t_l^- \leq t_{l,m} \leq t_l^+$. Equations (6)-(7) ($\forall b \in B$) describe the active and reactive nodal power balance governed by Kirchhoff's current law, where the actual power output of DG ($p_{g,m}$) is modelled as the difference between its potential power output ($C_g \omega_m$) and curtailment ($p_{g,m}^{curt}$) as in (8). $\omega_m$ is the maximum level of output DG can generate relative to its nominal capacity $C_g$, which is determined by the level of variable renewable resources during period $m$. $d_{b,m}^P$ and $d_{b,m}^Q$ are the active or reactive load respectively at bus $b$ during period $m$. $tan(\phi_{g,m})$ is the power factor of DG that may vary between periods.

The main grid can supply power to the distribution network through the grid supply point (GSP). For buses that are connected to the GSP, GSP flow is added to the bus balance in (6)-(7). Constraint (9) sets the import/export limits for the GSP bus. The GSP will be taken as the reference (slack) bus for a power flow model with the voltage angle set at zero, i.e., $\delta_{b_0,m} = 0$.

$$q_{GSP}^- \leq q_{\beta_{GSP},m} \leq q_{GSP}^+ \qquad \forall x \in X \qquad (9)$$

Constraint (10) sets the upper and lower limits on the capacity of new DG g.

$$C_g^- \leq C_g \leq C_g^+ \qquad \forall g \in G \qquad (10)$$

### 3) OPERATION AND CONTROL OF ANM

As implemented, two active network management control schemes are used. The first control is active control of DG output. This acts to curtail the generator output during periods when the level of renewables is high and demand low to avoid voltage or power flow limits being breached. Although this may involve lost revenue for the wind developer, this might be acceptable due to the infrequent occurrence of these periods as the resulting larger capacity

increases the overall energy production and total revenue. Therefore, there is a trade-off between greater generation capacity and the proportion of energy curtailed.

The optimization will determine the variable value of power curtailment ($P_{g,m}^{Curt}$) for each DG in each period. The following constraint applies to limit the curtailment variables within the maximum potential output of $g$:

$$p_{g,m}^{Curt} \leq C_g \omega_m \qquad (11)$$

Here, the Curtailment Ratio (CR) presents the waste level of renewable energy, defined as the ratio of the curtailment against the potential energy that could have otherwise been delivered by these DGs. While curtailment is very effective in managing constraints and enabling larger generator capacity, revenue is reduced. Economic considerations will limit the total amount of curtailed energy that is acceptable to the owner of renewable generators. Such considerations strongly depend on the network connection arrangements, where different "principles of access" govern the scope of curtailment. A simplified approach is taken which places an upper limit on the allowed amount of total curtailment across the network. Mathematically, the total curtailment ratio (TCR) sets a limit on the total curtailed generation (TCG) relative to the total potential generation (TPG) that could otherwise be delivered over the whole period:

$$TCG = \sum_{m \in M} \sum_{g \in G} p_{g,m}^{curt} \qquad (12)$$

$$TPG = \sum_{m \in M} \sum_{g \in G} C_g \omega_m \qquad (13)$$

$$\frac{TCG}{TPG} = TCR \leq \overline{TCR} \qquad (14)$$

The second control is coordinated control of on-load tap changers, where the voltage target of each transformer secondary is treated as variable and chosen for each period within the statutory range.

$$V_{b_{OLTC}}^- \leq V_{b_{OLTC},m} \leq V_{b_{OLTC}}^+ \qquad (15)$$

This control will choose a low target voltage for each transformer secondary when wind production is high and demand low to enable export without breaching the upper voltage limit.

The optimization was solved using the nonlinear solver Interior Point OPTimizer (IPOPT). By solving the above optimization, the main outputs of the model are the capacities of each generator. Values of time series variables, including wind production, curtailment, and control actions, are also retained. While the desired results could be the global optimal, there is no guarantee, in common with other well-known nonlinear solvers.

### C. CHALLENGES IN CONVENTIONAL OPTIMIZATION APPROACH

Directly solving the above optimization model for hosting capacity evaluation could be very computationally challenging. Firstly, the model complexity derives from the fact that it is a strictly non-convex nonlinear programming problem (NLP), being based on ACOPF formulations in order to capture the key network operational constraints, such as both magnitudes and phase values of bus voltages [31]. Secondly, it is a dynamic programming that seeks to optimize a set of unique variables (i.e. the DG capacity) over a long period of variable renewable and demand scenarios to capture the full range of operational conditions, i.e. high temporal resolution over a long horizon [16], [32]. This limits the length of time series that can be readily handled within the capability of the solver and computing intractability. In previous work [30], [16], the time series was subject to a scenario reduction process to deliberately reduce the number of periods in order to allow a larger network to be modelled.

Apart from the computational challenges, the fact that only a single deterministic solution is generated without any alternatives also leads to challenges for practical implementation. The exogenous uncertainty that is not considered in the model, such as policy and regulation changes, may play a decisive role. A reasonable scenario could be that the connectable capacity of a wind farm identified by the model would become unrealistic due to construction limits or planning permission; having a range of quality and distinctive results enables the user to choose a suboptimal solution that will enable consequent best use of the network by redeploying available capacity.

## III. OPTIMIZATION VIA HISTORY MATCHING
The proposed framework treats an optimization problem as a history matching problem. History matching (an alternative approach to model calibration [11], [12]) is usually adopted to identify the regions of input space of a model where the corresponding model outputs 'match' the observations with consideration of the associated uncertainties. It starts with the full input space of the model and gradually reduces the volume of the input space by removing the "implausible" inputs "wave by wave" and eventually ends up with a subspace of the input space where inputs can produce model outputs consistent with observations. The task of the optimization problem can be considered as a history matching problem that finds a collection of input choices for which the corresponding "computer simulation model" output meets some criteria (instead of looking for consistency with the observations), for example, that the output is smaller than a threshold.

Fig. 1 presents a flow chart illustrating the methodology. Small samples from the entire input space are initially fed into the computer simulation model (simulator). These model runs are then used to construct and train a statistical emulator. A statistical emulator is a statistical model that approximates the simulation model (see III B for details). As

a statistical emulator is much (computationally) cheaper to run than the simulation model, it helps to explore the input space in a much more efficient way. Bayes linear analysis (see III B for more details) is adopted to quantify the uncertainty due to the use of the emulator. With a statistical emulator, a large set of inputs sampled from the input space can be fed into the emulator. The corresponding outputs can be used to define a threshold, e.g. the $95th$ percentile (one may use a lower or higher percentile depending on how well the emulator approximates the simulation model; lower/higher percentiles may increase/decrease the number of total emulation waves.) of the emulation outputs, so that inputs that lead to outputs above the threshold are discarded from the input space (with the consideration of associated uncertainty). A small set of samples from the reduced input space can then be fed into the computer simulation model again to build a new statistical emulator. And a large set of input samples from the reduced input space can then be fed into the new simulator to define a new threshold and to further reduce the volume of the input space. This process is repeated until the input space is not significantly reduced further, which leaves the reduced input space comprising candidate solutions for the optimization problem. Note that as the emulator is only an approximation of the computer simulation model, there is uncertainty attaching to each candidate solution. When the uncertainty ranges of the outputs from two candidate solutions overlap, it makes the candidate solutions indistinguishable from each other.
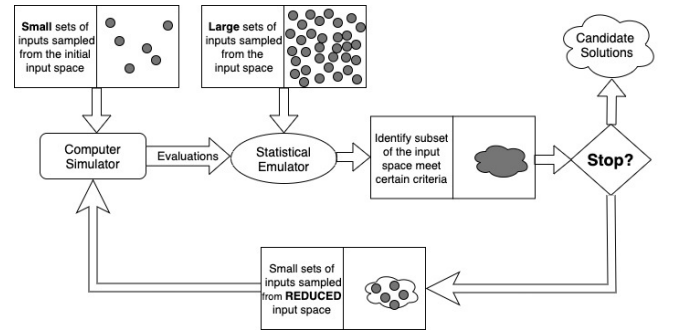


**FIGURE 1.** Flow chart of the proposed methodology that identifies a subset of the input space that meets certain criteria.

### A. REFRAMING THE OPTIMIZATION PROBLEM
The hosting capacity model in Section II needs to be reframed to employ history matching. Instead of the model choosing optimal DG capacities within the optimization, the new model is to simulate the optimal network operation under given specific generator capacities; that is, the capacities of each wind farm are considered as the input of the computer simulation model. The objective function is changed to:

$$min\ curtailment = min \sum_{g \in G} p_{g,m}^{curt} \quad \forall m \in M \quad (16)$$

It is also subjected to constraints (2)-(11) for each individual period, except the dynamic constraints (12)-(14) that sum over all periods. The rest of the input data remains the same: time series of wind speed levels, demand, and potential wind farm locations. In this way, the original dynamic model of hosting capacity optimization is converted to a snapshot operational model that can be solved separately in each period. This reframe presents a significant reduction in terms of the optimization size, for example, 8,760 times less in the number of variables than the original hosting capacity model if a one-year horizon is considered. Such size reduction is essential for an NLP model, which is otherwise likely to be intractable. While the reframed model still needs to be run many times in time sequence, the total solving time is generally much faster than solving a large dynamic NLP at once.

The initial outputs of the reframed simulation model are the power generation and curtailment from each wind farm plus the control variable values (OLTC target voltages and import/export power at GSP) which minimize power curtailment in each period. Based on these outputs, the final performance metrics measure how the capacity input scenario relates to the total potential power generation and total power curtailment ratio.

For a given input of wind speed, the TPG is a linear combination of the capacities $\{C_1 - C_g\}$ of the wind farms:

$$TPG = b_1 C_1 + \ldots + b_g C_g + \ldots + b_G C_G \qquad (17)$$

where each term reflects the power generation by each wind farm. Wind farms experiencing the same wind speed profiles will have the same $b_n$ coefficients.

Rather than being a dynamic constraint on the original optimization, the total power curtailment ratio $TCR$ is the output of the computer simulation model $f^*$ acting to minimize the power curtailment of each wind farm in each period across the horizon. This is a function of the capacities and controls, as well as of wind and demand:

$$TCR = f^*(C_1, C_g, \ldots, C_G, Ctrls, Wind, Demand, Network) \qquad (18)$$

The conventional optimization approach naturally respects the constraints imposed by the network and other limits. In this computer simulation model $f^*$, the network limits (voltage and power flow) are respected by the control system operation (mainly by controlling generator outputs) at minute by minute. Potential violations of network security constraints during high wind periods due to large installed capacity are avoided by reducing the generator outputs, potentially down to zero. This has the effect of pushing up the curtailment ratio which the emulation process considers. It is, of course, possible for network limits to be violated should demand be too high, but this would not occur in a properly designed network as the primary aim of the network operator is to supply the demand.

## B. STATISTICAL EMULATION

For an optimization problem, if one could sample and simulate the entire input space, the optimal solution would be located. The conventional Monte Carlo approach can be adopted to uniformly draw samples from the input space and then run the computer simulation model for every input drawn. This approach is, however, computationally unfavourable for high-dimensional complex systems (such as hosting capacity, where the operational power system contains controls, power supplies, demands and wind speed profiles) as it is impossible to evaluate the computer simulation model with enough inputs in order to account for uncertainty over the entire input space. Therefore, statistical emulation is adopted to address this challenge, where a statistical model (emulator) is fitted to a small number of simulation model evaluations. For high-dimensional complex systems, one single evaluation of a computer simulation model for a specific set of inputs might take hours, days or even weeks (for example, the galaxy formation model [33]), while it usually takes ~ms to run a statistical emulator. A statistical emulator approximates the simulation model for any inputs in the input space, thus, even for inputs that have not been evaluated, the emulator can provide an approximation of the simulation model output (with uncertainty attached due to the approximation) at that input. Statistical emulation has been used to explore the input space and assess uncertainty due to unknown inputs of the simulation model for various complex systems, e.g. cosmology [33], oil reservoirs [34], meteorology [35], and energy [36]–[38].

A common choice for the emulator [12] is adopted in the paper, where an individual component of the output is modelled by a statistical function $f(\mathbf{x})$, where $\mathbf{x}$ is treated as a vector of random variables. For example for the optimization problem defined in III.A, the output TCR is modelled by a statistical function $f(C_1, C_g, \ldots, C_G)$ where the inputs $C_1, C_g, \ldots, C_G$ are treated as random variables. The function $f(\mathbf{x})$ contains stochastic terms, given by:

$$f(\mathbf{x}) = \sum_i \beta_i \, g_i(\mathbf{x}) + u(\mathbf{x}) \qquad (19)$$

where $B = \{\beta_i\}$ are unknown scalars and $g_i(\mathbf{x})$ are known deterministic functions of the input vector $\mathbf{x}$ (e.g., polynomials). $Bg(\mathbf{x})$ expresses global variation in $f(\mathbf{x})$, where the functional forms $g_i$ (for example, $g_i$ is some power function of x) are chosen and the specification for the elements of $B$ are fitted based on an analysis of a set of simulation model evaluations. In practice, a common way to construct $Bg(\mathbf{x})$ is simply conducting a linear regression fit. $u(\mathbf{x})$ expresses local variation and is represented as a second-order stationary stochastic process, with a correlation function which expresses the notion that the correlation between the value of $u$ for any two values $\mathbf{x}$, $\mathbf{x}'$ is a decreasing function of the distance between the input values. Various choices for the form of the correlation function have

been proposed in the literature (see, for example, [35]) with different meritorious aspects. Following [12], the correlation function used in this paper is of the form:

$$Corr(u(\boldsymbol{x}), u(\boldsymbol{x}')) = exp\left(-\left(\frac{\| \boldsymbol{x} - \boldsymbol{x}' \|}{\theta}\right)^2\right) \qquad (20)$$

where $\theta$ is a tuning parameter specified through the analysis.

To construct the emulator there is a rich literature that covers statistical modelling of complex functions (see [39] for a good introduction and [35] for an example). Following [12], we adopted a simple approach to construct the emulator, choosing the functional forms $g_i$ using least squares fitting and then fitting the correlation function for $u(\boldsymbol{x})$ to the residuals by trial and error based on cross-validation.

### C. UNCERTAINTY QUANTIFICATION VIA BAYES LINEAR ANALYSIS

There are three major sources of uncertainty that need to be accounted for when solving an optimization problem. Firstly, the computer simulation model is an imperfect analogue of the system. Even with the "best" choice for the model input, the output will almost certainly not reflect the system behaviour (reality) precisely. Secondly, if there are data involved, there is the uncertainty induced by measurement error. Lastly, as the statistical emulator acts as an approximation to the computer simulation, additional uncertainty is introduced. One of the major drawbacks of conventional optimization is that it only provides a solo "optimal" solution without taking account of uncertainties. A natural way to conduct uncertainty quantification is to adopt a Bayesian approach [40].

Full Bayes analysis can be very informative when both the prior specification and the analysis are carefully conducted. Bayes linear analysis [13] is partial but easier, faster, and more robust [12]. The application of a full Bayesian analysis requires enormous computation cost for complex systems. The more tractable Bayes linear approach has had some success (see, for example, [11], [41]) for complex systems, and avoids much of the computational burden of the full Bayesian approach. In this paper, Bayes linear analysis is adopted to conduct uncertainty quantification.

Instead of specifying the full distribution of prior and posterior in full Bayes analysis, the Bayes linear approach is based only on the mean, variance, and covariance specification. Similarly, instead of updating the entire distribution in full Bayes analysis, the Bayes linear approach only updates the expectation and variance through Bayes linear adjustment (see [13] for the details of the mathematical formula). In general, Bayes linear analysis may be considered as a fast approximation to a full Bayes analysis. Note that Bayes linear analysis does not require the model to be linear; "linear" refers to the linearity properties of expectation. If one treats expectation as primitive, Bayes linear analysis is simply giving the appropriate analysis under a direct partial specification of means, variances, and covariances (see [13] for more details).

## IV. CASE STUDY

A real UK distribution network is used for the case study, as shown in Fig. 2. It is a 61 buses 33/11kV weakly meshed network and data can be found in [42]. The rating of the substation is 60 MVA and the maximum demand of the network is 38.2 MW. The 15 MVA interconnector in the network is modelled as a PV bus with target voltage of 1pu. Voltage limits are taken to be +6/-6% of nominal. A whole year of hourly demand and potential wind power profiles are used. Five potential locations for wind generation are considered. Two different wind resource profiles are employed, with wind farms 1-3 (at buses 1108, 1106 and 1105) considered sufficiently close geographically to use one profile and wind farms 4-5 (at buses 1114 and 1115) close enough to use the other. The one-year wind (wind area 1) and demand are shown in Figure 3. (The respective $b_n$ coefficients for TPG in the simulation model (17) are 390,915 for wind farms 1-3 and 341,880 for wind farms 4-5.)

### A. CONVENTIONAL OPTIMIZATION

The optimization was conducted using the full year's hourly time series of local demands and wind speed at different wind farm locations. For illustration, the maximum amount of curtailment (TCR) allowed across the wind farms is 10%. The solution is further constrained by requiring that the curtailment ratio for each individual wind farm is also no more than 10%.

Directly using the full year of hourly data, the optimization takes around 3.8 hours to execute on a typical PC (16 GB RAM and 3 GHz processor). For reference, the scenario reduction approach presented in [30] allows the optimization to find a single approximate solution within 55 minutes, albeit with some reduction in accuracy. Table I shows the solution from this conventional optimization approach. The total hosting capacity reaches 52.7 MW, with the highest capacity available at wind farms 2 and 5, and the lowest capacity at wind farm 1.
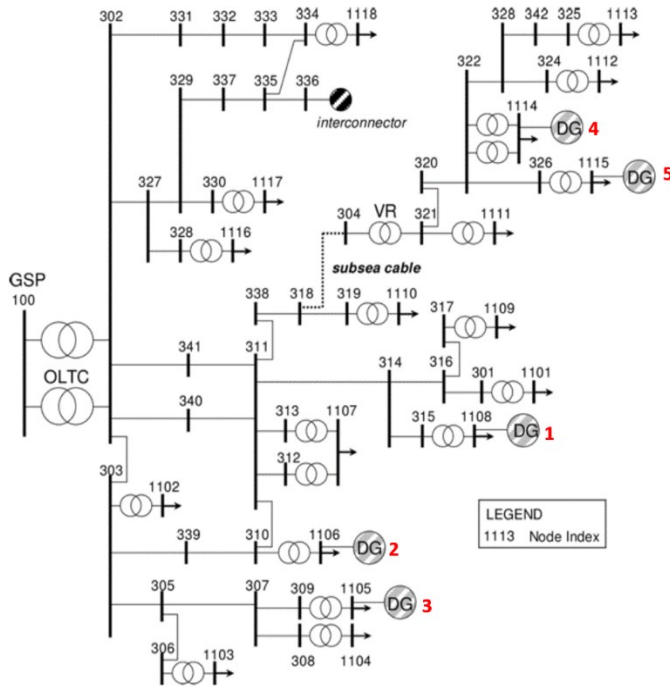
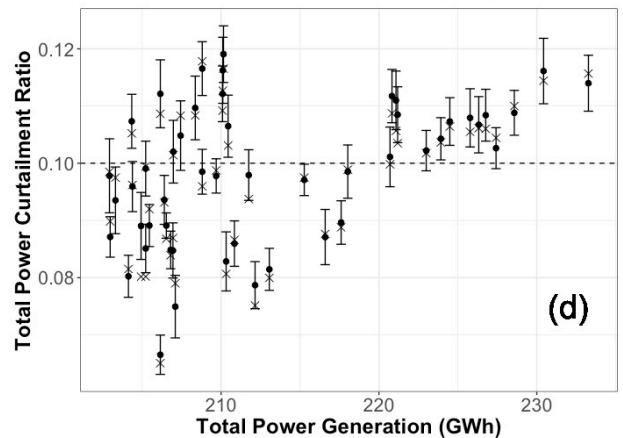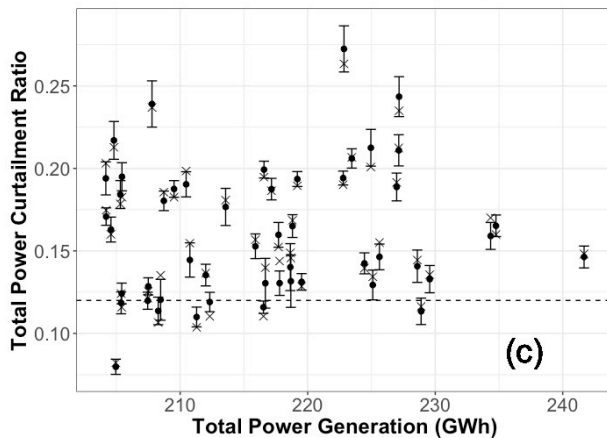FIGURE 2. A 38-kV five-bus radial distribution network diagram



FIGURE 3. One-year wind and demand data

TABLE I
"OPTIMAL" SOLUTION FROM CONVENTIONAL APPROACH.

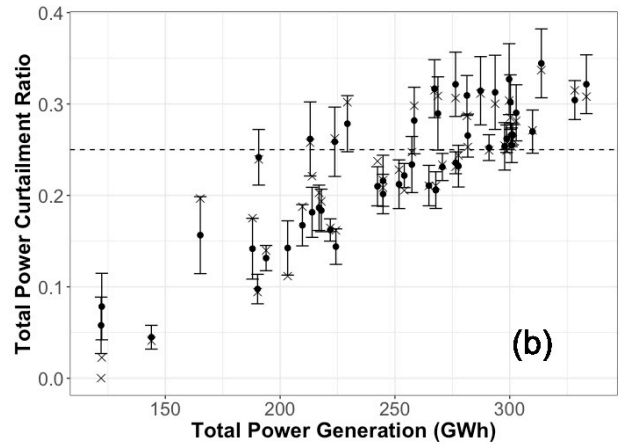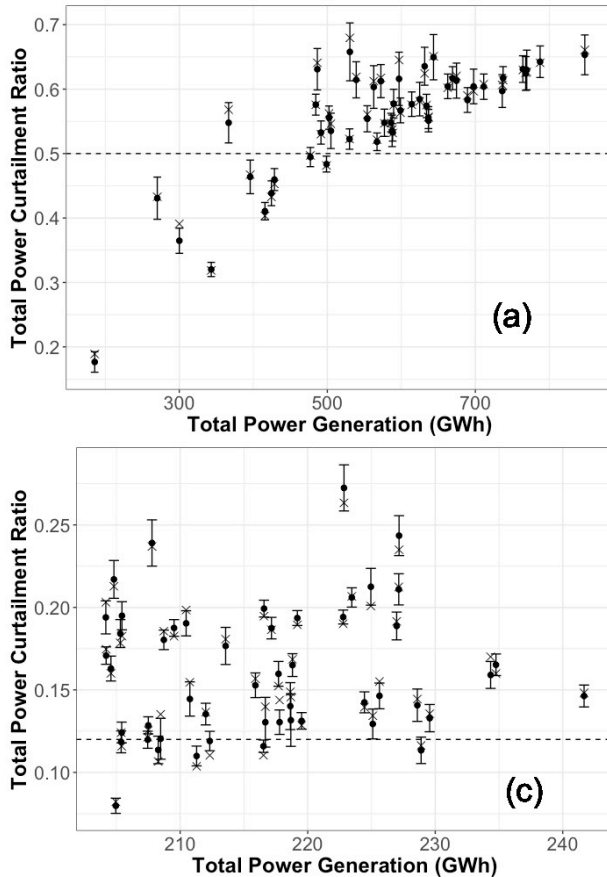| Wind farm ID | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|
| Optimal capacity (MW) | 5.0 | 14.8 | 9.5 | 8.6 | 14.9 | 52.7 |
| Curtailment ratio | 10% | 10% | 10% | 10% | 10% | 10% |
| Full generation (GWh) | 17.9 | 53.1 | 34.1 | 35.9 | 62.0 | 203.0 |



FIGURE 4. First (a), third (b), fourth (c) and fifth (d) wave emulation validations. The mean values of the emulation approximation for inputs drawn from an independent test set are marked as black dots, error bars are two standard deviations around this mean. Note each set of inputs (five capacities) determines the Total Power Generation (X-axis). Cross is actual output from the computer simulation model. Dashed line is the threshold used for history matching.
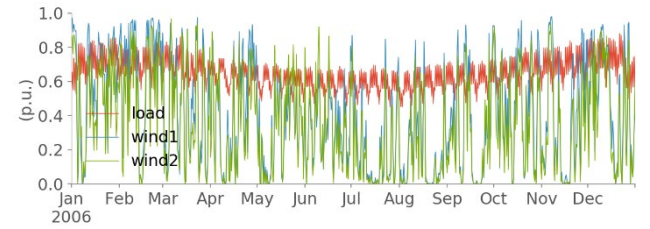
## B. OPTIMIZATION VIA HISTORY MATCHING

Following the proposed framework, this optimization problem is treated as a history matching problem. The simulation model is defined in (19) and (20). The inputs we want to explore are the capacities of each wind farm, and the outputs are TPG and TCR. As the wind farm capacities fully determine TPG given the wind speed data, we only need to build a statistical emulator for TCR. Following Fig 1., an initial design for the history matching samples input parameter values (wind farm capacities) using a Latin hypercube [43] with a coarse constraint that each farm capacity ranges between 0 and 50MW. Such a Latin hypercube design aims to select the values of inputs that are space-filling in order to avoid sample bias. A set of 50 simulation model evaluations of TCR were generated using 50 Latin hypercube input samples, with the same hourly demand and wind speed data. There is a trade-off between the number of scenarios and convergence speed; if a smaller number of scenarios are considered at each wave of emulation, the emulator might not approximate the simulation model well which would lead to more emulation waves to achieve similar results as using larger number of scenarios. Using more scenario runs, however, are more computationally costly. One could adjust the number of scenarios runs at each different wave of emulation. In this manuscript, for illustrative purpose, we used 50 scenarios at all waves of emulation. It takes about 30 seconds computing time to evaluate the simulation model for a given capacity scenario. Moreover, the evolution of multiple scenarios could be parallel and so the total time would further reduce significantly.

The first emulator $f^{(1)}$ was fitted using these model evaluations and the performance of the emulator in terms of predicting TCR was tested with an independent test set of 50 model evaluations. The best global fit $(Bg(\mathbf{x}))$ is found with an adjusted $R^2$ of 0.87. For the correlation function $u(\mathbf{x})$, the parameter $\theta$ is chosen so that the emulator successfully predicts more than 95% of the evaluations to within 3 standard deviations. (i.e. the Three Sigma Rule [44]).

As the emulator is an approximation of the simulation model, a poor approximation would lead to unreliable candidate solutions in the end. Therefore, before moving onto the next wave emulation, the performance of the emulator needs to be properly assessed. Fig. 4a shows the performance of the first emulator in the first wave emulation (wave 1), where TCR is plotted against TPG. Almost all the simulation model outputs lie within the predicted interval, which indicates a reasonably good fit of the emulator to the computer simulation model. Fig. 4a also shows that the TCR of all the design points are much higher than 10%, which is due to the fact that the input capacity range is too wide. This first wave emulation, however, provides information to reduce the input (search) space. A threshold for output TCR

was defined subjectively to be $T_x^{(1)} = 0.5$ (alternatively, the threshold can be defined based on a percentile of the output). Given this first emulator $f^{(1)}$, 1,000,000 point Latin Hypercube samples of five input capacities are fed into the first emulator. Remove all the input $\mathbf{x} \equiv (C_1, C_2, C_3, C_4, C_5)$ for which $E(f^{(1)}(\mathbf{x})) - 3std(f^{(1)}(\mathbf{x})) > T_x^{(1)}$, where the expected value $E(f^{(1)}(\mathbf{x}))$ and the standard deviation $std(f^{(1)}(\mathbf{x}))$ of the emulation outcome are obtained through Bayes linear updates. It is almost certain that the removed inputs are not 'optimal' solutions. About 18% of the input sets remain, which means the first wave emulation reduces the volume of the input space by 82%.

The computer simulation model's behaviour over the reduced input space can then be studied. Fifty samples from the reduced input space are fed into the simulation model (wave 2) and a new emulator $f^{(2)}$ is built based on the evaluations. The best global fit $(Bg(\mathbf{x}))$ is found giving an adjusted $R^2$ of 0.91. Defining a threshold of the output TCR to be $T_x^{(2)} = 0.3$, a large set of inputs in the reduced input space from wave 1 emulation was fed into $f^{(2)}$. After removing all the input $\mathbf{x}$ for which $E(f^{(2)}(\mathbf{x})) - 3std(f^{(2)}(\mathbf{x})) > T_x^{(2)}$, about 11% of the input sets remain, which means the second wave emulation reduces the search space further by 89%. In the third wave emulation, emulator $f^{(3)}$ (with adjusted $R^2$ of 0.88) is built and fits the simulation model well over the reduced input space (see Fig. 4b). The exceptions are the points corresponding to low TPG, which seem to overestimate TCR. Under the threshold $T_x^{(3)} = 0.25$, the input space is further reduced by 80%.

At the fourth wave emulation, emulator $f^{(4)}$ is built with adjusted $R^2$ of 0.87. Figure 4c shows the results on an independent test set of evaluations. There is one input resulting in TCR less than 10%. Note the objective is not only to achieve TCR $\leq$ 10%, but also to generate as much power as possible. If zero capacity is assigned to all the wind farms, their curtailment ratio would be zero as well as the TPG. Therefore, an additional constraint based on TPG is imposed to reduce the input space. Defining the threshold to be $T_x^{(4)} = 0.12$ and the TPG of the conventional "optimal" solution to be $TPG^\star$, all values of $\mathbf{x}$ for which $E(f^{(4)}(\mathbf{x})) - 3std(f^{(4)}(\mathbf{x})) > T_x^{(4)}$ and corresponding to TPG $\geq 0.9TPG^\star$ were removed; this further reduces the input space by 90%.

At the fifth wave emulation, emulator $f^{(5)}$ is built with adjusted $R^2$ of 0.91. Fig. 4d shows the results on an independent test set of evaluations. Note there are already several inputs that result in TPG and TCR comparable to the conventional "optimal" solution. Under the threshold $T_x^{(5)} = 0.10$, we remove all values of $\mathbf{x}$ for which $E(f^{(5)}(\mathbf{x})) - 3std(f^{(5)}(\mathbf{x})) > T_x^{(5)}$ and the corresponding TPG $\geq 0.95TPG^\star$. This further reduces the input space by 75%.

After five waves of emulations, we now have only 18% × 11% × 20% × 10% × 25% = 0.0099% of the original input parameter volume. This is a space of

acceptable candidate solutions which can be accessed through emulation runs. Any candidate solutions can still be evaluated through the simulation model.

Table II presents 5 candidate solutions (S1-S5). For each solution, the capacity (Cap in MW) of wind farms (WF) and its curtailment ratio (CR) are listed as well as the TPG (in brackets) and TCR.

TABLE II
FIVE CANDIDATE SOLUTIONS SAMPLED FROM FINAL REDUCED SPACE.

|    |     | WF1   | WF2   | WF3   | WF4   | WF5   | Total          |
|----|-----|-------|-------|-------|-------|-------|----------------|
| S1 | Cap | 5.9   | 14.7  | 10.5  | 8.4   | 13.7  | 53.2 (203.4GWh) |
|    | CR  | 17.9% | 9.8%  | 14.6% | 8.7%  | 6.0%  | 10.31%         |
| S2 | Cap | 3.3   | 14.2  | 11.0  | 9.1   | 14.3  | 51.9 (199.6GWh) |
|    | CR  | 0%    | 8.2%  | 16.8% | 12.8% | 8.0%  | 10.26%         |
| S3 | Cap | 5.5   | 15    | 10.1  | 8.5   | 13.8  | 52.9 (203.2GWh) |
|    | CR  | 14.7% | 10.7% | 12.8% | 9.3%  | 6.3%  | 9.98%          |
| S4 | Cap | 5.0   | 14.3  | 10.1  | 8.3   | 14.9  | 52.6 (202.8GWh) |
|    | CR  | 10.3% | 8.5%  | 12.9% | 8.2%  | 10.1% | 9.89%          |
| S5 | Cap | 4.6   | 13.7  | 8.9   | 9.6   | 15.1  | 51.9 (201.3GWh) |
|    | CR  | 6..7% | 6.7%  | 7.2%  | 15.5% | 10.7% | 9.81%          |

All solutions in Table II produce TPG and TCR comparable to the conventional ``Optimal" solution (Table I). In fact, candidate solution 3 produces larger TPG and smaller TCR than the conventional ``Optimal" solution. This is due to the conventional ``Optimal" solution not only requiring TCR $\leq 10\%$ but also each wind farm's CR $\leq 10\%$, while the emulation relaxes the second constraint. The candidate solutions from our approach explore the diversity of solutions that might be of interest. Moreover, if any specific requirement (e.g. curtailment ratio for wind farm 5 no more than 6%, as in candidate S1) or additional limit emerges when implementing the plan (e.g. high additional costs for wind farm 3 to use the land, so S5 become more preferable), extra constraints can be added when selecting candidate solutions.

While a set of quite different candidate solutions are found, by curtailing at critical periods (e.g high wind and low demand), the network is able to operate reliably. DG tends to raise the voltage at its connected bus given its power injection into the network. When there is a potential voltage rise issue during high DG output, a certain amount of curtailment is performed by the optimization to maintain the voltage at or below the allowed upper value. To clearly compare the impact of different candidate capacity solutions on voltage profiles over a year, using WF5 as an example, Fig. 5 shows the total hours when the voltage at bus 1115 (where WF5 is connected) reached its upper limit and curtailment occurred. As can be seen, the larger size DG (e.g the largest WF5 at S5 vs smallest WF5 at S1) at the same location does raise the voltage to its limit more frequently, also causing more curtailments.
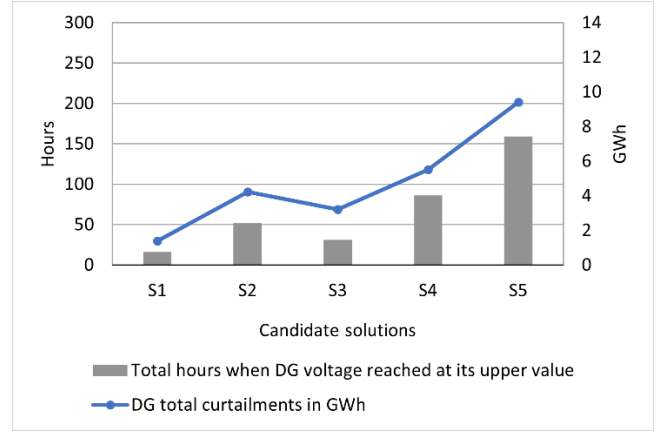


FIGURE 5. Impact of DG size on voltage and curtailment

## C. MODEL DISCREPANCY

No matter how carefully a physical model is designed, it is an imperfect approximation of the real system. The impact of model discrepancy plays a key role in linking the model and the real world, yet it is often neglected in energy system optimization problems. Following [12], there are two types of model discrepancy: internal and external. Internal discrepancy can be assessed by experiments on the computer simulation model and it provides a lower bound on the model discrepancy. External discrepancy refers to the aspects of model discrepancy that have not been addressed by the assessment of internal discrepancy; it could be estimated by expert judgement.

In the previous section, uncertainty in the five input capacity values as well as the uncertainty due to emulation approximation have been accounted for, but there is no consideration of model discrepancy. To interpret the simulation model solutions for future planning, the internal discrepancy due to weather and demand profiles can be assessed. Due to the complexity of the weather system and consumer behaviours, it is impossible to produce detailed hourly wind speed and demand forecasts years into the future. Instead, we can assess the mean effects by running the simulation model with fixed wind farm capacities but varying the weather and demand profiles in the historical data. Linear perturbations are applied to the wind speed and demand data. Note if the variation (that reflects the uncertainty due to structural discrepancy) strongly depends on the 5 wind capacities, we can emulate this variation as well, which would provide extra information for decision makers to select a reliable solution. After analysing the experimental results, it turns out that the changes in TCR respond to the changes in weather profile or demand profile linearly, but for a different set of capacity values the response rates are different. Fig. 6 shows how the five candidate solutions listed in Table II respond to the changes in demand and wind speed, with the slope indicating sensitivity. Increasing the wind speed or decreasing the demand would

increase the TCR, and wind speed clearly has more impact than demand. Quantifying the response to the changes in demand and weather profiles provides valuable information for decision support. This paper is focused on addressing the optimization problem, where the solutions that the simulation model suggests are explored for their reliability. If this involved working directly with decision makers, these discrepancies could be introduced at the optimization stage.
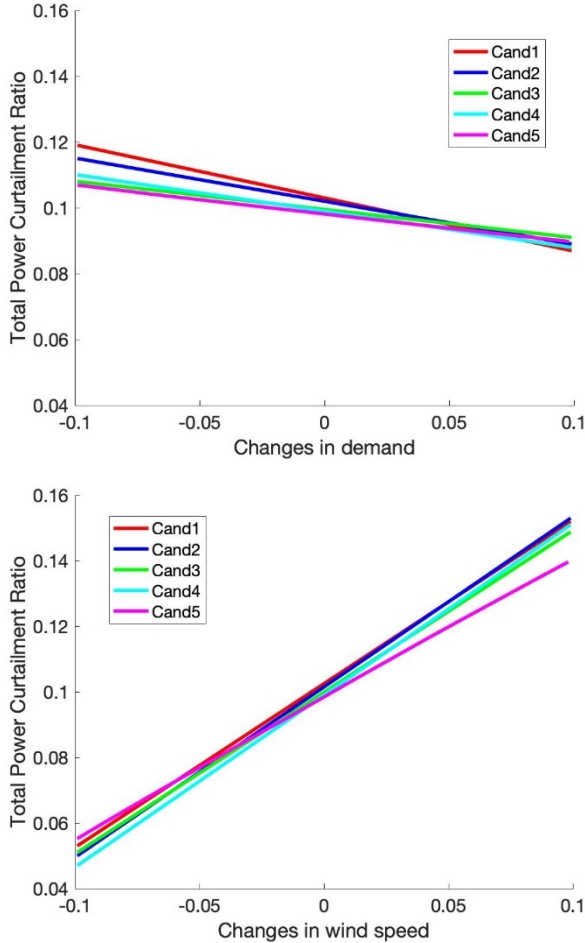


**FIGURE 6**. **Total Curtailment Ratio responses to the changes in demands (above) and wind profiles (below). The lines correspond to each candidate solution in Table II.**

## V. DISCUSSION AND CONCLUSION

While the hosting capacity optimization is solved as an exemplary problem, the proposed emulation-based framework presents an attractive general way to address optimization problems in power systems. Statistical emulation is effective in carrying out this approach and Bayes linear analysis provides proper uncertainty quantification. Careful structural discrepancy assessment and multi-level emulation are essential parts of this methodology, which overcomes the shortcomings of conventional optimization. The proposed methodology is able to access the space of acceptable candidate solutions and provides proper uncertainty quantifications attached to the solutions.

For the studied hosting capacity problem, as there is a trade-off between the total capacity and curtailment ratio, a constraint upon the curtailment ratio (no more than 10%) is needed to conduct the optimization. Such optimization problems can be extended to estimate a Pareto boundary based on power generation and the curtailment ratio. From the perspective of optimization, it is well known that a bi/multi-objective optimization problem usually admits an infinite number of non-inferior solutions (theoretical limits), which form the outermost boundary of achievable performance, the Pareto boundary [45]. A noninferior solution on the Pareto boundary is considered to be Pareto-optimal in the sense that no other solution can improve the performance of some objectives without reducing other objective(s). Solutions on the Pareto boundary can be identified using the proposed method, and therefore allow relaxation of the 10% constraint on the curtailment ratio, which would provide additional valuable information for decision support.

In future study, the statistical emulation-based method offers scope to enhance the precision of the control system by reducing the time interval, improving representation of weather conditions by extending the time series of wind and demand, and also by enabling representation of technologies, such as storage, which adds inter-temporal constraints.

## REFERENCE

[1]     M. Zeyringer, J. Price, B. Fais, P. H. Li, and E. Sharp, "Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather," *Nature Energy*, vol. 3, no. 5, pp. 395–403, Apr. 2018.

[2]     M. Chaudry, N. Jenkins, M. Qadrdan, and J. Wu, "Combined gas and electricity network expansion planning," *Applied Energy*, vol. 113, pp. 1171–1187, Jan. 2014.

[3]     P. H. Shaikh, N. B. M. Nor, P. Nallagownden, I. Elamvazuthi, and T. Ibrahim, "A review on optimized control systems for building energy and comfort management of smart sustainable buildings," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 409–429, Jun. 2014.

[4]     M. A. A. Pedrasa, T. D. Spooner, and I. F. MacGill, "Coordinated scheduling of residential distributed energy resources to optimize smart home energy services," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 134–143, Sep. 2010.

[5]     C. Coffrin and P. Van Hentenryck, "A linear-programming approximation of AC power flows," *INFORMS Journal on Computing*, vol. 26, no. 4, pp. 718–734, Sep. 2014.

[6]     H. Zhang, V. Vittal, G. T. Heydt, and J. Quintero, "A mixed-integer linear programming approach for multi-stage security-constrained transmission expansion planning," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1125–1133, May 2012.

[7]     Y. Li, B. Feng, G. Li, J. Qi, D. Zhao, and Y. Mu, "Optimal distributed generation planning in active distribution networks considering integration of energy storage," *Applied Energy*, vol. 210, pp. 1073–1081, Jan. 2018.

[8]     H. Zhang, W. Chen, and W. Huang, "TIMES modelling of transport sector in China and USA: Comparisons from a decarbonization perspective," *Applied Energy*, vol. 162, pp. 1505–1514, Jan. 2016.

[9]     C. Huang, F. Li, T. Ding, Z. Jin, and X. Ma, "Second-order cone programming-based optimal control strategy for wind energy

conversion systems over complete operating regions," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 263–271, Jan. 2015.

[10] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *International Journal of Electrical Power and Energy Systems*, vol. 30, no. 6–7, pp. 383–392, Jul. 2008.

[11] P. S. Craig, M. Goldstein, A. H. Seheult, and J. A. Smith, "Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments," in *Case Studies in Bayesian Statistics*, 1997, pp. 37–93.

[12] M. Goldstein and N. Huntley, "Bayes linear emulation, history matching, and forecasting for complex computer simulators," in *Handbook of Uncertainty Quantification*, Springer International Publishing, 2017, pp. 9–32.

[13] M. Goldstein and D. Wooff, *Bayes linear statistics: Theory and methods*. John Wiley & Sons, Ltd, 2007.

[14] L. F. Ochoa, C. J. Dent, and G. P. Harrison, "Distribution network capacity assessment: Variable DG and active networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 87–95, Feb. 2010.

[15] Y. Xiang *et al.*, "Coordinated DG-Tie planning in distribution networks based on temporal scenarios," *Energy*, vol. 159, pp. 774–785, 2018.

[16] W. Sun and G. P. Harrison, "Wind-solar complementarity and effective use of distribution network capacity," *Applied Energy*, vol. 247, pp. 89–101, Aug. 2019.

[17] R. Torquato, D. Salles, C. Oriente Pereira, P. C. M. Meira, and W. Freitas, "A Comprehensive Assessment of PV Hosting Capacity on Low-Voltage Distribution Systems," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 1002–1012, 2018.

[18] W. Sun, G. P. Harrison, and S. Z. Djokic, "Distribution network capacity assessment: Incorporating harmonic distortion limits," in *IEEE Power and Energy Society General Meeting*, 2012, pp. 1–7.

[19] C. J. Dent, L. F. Ochoa, G. P. Harrison, and J. W. Bialek, "Efficient Secure AC OPF for Network Generation Capacity Assessment," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 575–583, 2010.

[20] A. Dubey and S. Santoso, "On Estimation and Sensitivity Analysis of Distribution Circuit's Photovoltaic Hosting Capacity," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2779–2789, 2017.

[21] J. F. Franco, A. Procopiou, J. Quiros-Tortos, and L. Ochoa, "Advanced Control of OLTC-Enabled LV Networks with PV Systems and Electric Vehicles," *IET Generation, Transmission & Distribution*, vol. 13, no. 14, pp. 2967–2975, 2019.

[22] A. Borghetti, "Using mixed integer programming for the volt/var optimization in distribution feeders," *Electric Power Systems Research*, vol. 98, pp. 39–50, 2013.

[23] S. Wang, S. Chen, L. Ge, and L. Wu, "Distributed Generation Hosting Capacity Evaluation for Distribution Systems Considering the Robust Optimal Operation of OLTC and SVC," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1111–1123, 2016.

[24] Z. Tian, W. Wu, B. Zhang, and A. Bose, "Mixed-integer second-order cone programing model for VAR optimisation and network reconfiguration in active distribution networks," *IET Generation, Transmission and Distribution*, vol. 10, no. 8, pp. 1938–1946, May 2016.

[25] J. F. Franco, L. F. Ochoa, and R. Romero, "AC OPF for smart distribution networks: An efficient and robust quadratic approach," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4613–4623, 2018.

[26] S. Ganguly and D. Samajpati, "Distributed Generation Allocation on Radial Distribution Networks Under Uncertainties of Load and Generation Using Genetic Algorithm," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 688–697, 2015.

[27] B. R. Pereira, G. R. M. Martins da Costa, J. Contreras, and J. R.

S. Mantovani, "Optimal Distributed Generation and Reactive Power Allocation in Electrical Distribution Systems," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 975–984, 2016.

[28] A. M. El-Zonkoly, "Optimal placement of multi-distributed generation units including different load models using particle swarm optimisation," *IET Generation, Transmission & Distribution*, vol. 5, no. 7, p. 760, 2011.

[29] A. Ameli, S. Bahrami, F. Khazaeli, and M.-R. Haghifam, "A Multiobjective Particle Swarm Optimization for Sizing and Placement of DGs from DG Owner's and Distribution Company's Viewpoints," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1831–1840, 2014.

[30] L. F. Ochoa, C. J. Dent, and G. P. Harrison, "Distribution Network Capacity Assessment: Variable DG and Active Networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 87–95, 2010.

[31] A. Keane *et al.*, "State-of-the-art techniques and challenges ahead for distributed generation planning and optimization," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1493–1502, May 2013.

[32] S. Pfenninger, "Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability," *Applied Energy*, vol. 197, pp. 1–13, Jul. 2017.

[33] I. Vernon, M. Goldsteiny, and R. G. Bowerz, "Galaxy formation: A Bayesian uncertainty analysis," *Bayesian Analysis*, vol. 5, no. 4, pp. 619–670, Dec. 2010.

[34] J. Cumming and M. Goldstein, *Bayes Linear Uncertainty Analysis for Oil Reservoirs Based on Multiscale Computer Experiments*. Oxford University Press, 2009.

[35] D. Williamson *et al.*, "History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble," *Climate Dynamics*, vol. 41, no. 7–8, pp. 1703–1729, Aug. 2013.

[36] M. Xu, A. Wilson, and C. J. Dent, "Calibration and sensitivity analysis of long-term generation investment models using Bayesian emulation," *Sustainable Energy, Grids and Networks*, vol. 5, pp. 58–69, Mar. 2016.

[37] 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.

[38] A. L. Wilson, C. J. Dent, and M. Goldstein, "Quantifying uncertainty in wholesale electricity price projections using Bayesian emulation of a generation investment model," *Sustainable Energy, Grids and Networks*, vol. 13, pp. 42–55, Mar. 2018.

[39] A. O'Hagan, "Bayesian analysis of computer code outputs: A tutorial," *Reliability Engineering and System Safety*, vol. 91, no. 10–11, pp. 1290–1300, Oct. 2006.

[40] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, Aug. 2001.

[41] P. S. Craig, M. Goldstein, J. C. Rougier, and A. H. Seheult, "Bayesian forecasting for complex systems using computer simulators," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 717–729, Jun. 2001.

[42] A. Keane, L. F. Ochoa, E. Vittal, C. J. Dent, and G. P. Harrison, "Enhanced utilization of voltage control resources with distributed generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 252–260, Feb. 2011.

[43] M. D. McKay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, vol. 21, no. 2, p. 239, May 1979.

[44] F. Pukelsheim, "The Three Sigma Rule," *The American Statistician*, vol. 48, no. 2, p. 88, May 1994.

[45] L. A. Zadeh, "Optimality and Non-Scalar-Valued Performance Criteria," *IEEE Transactions on Automatic Control*, vol. 8, no. 1, pp. 59–60, Jan. 1963.

**Hailiang Du** is an Assistant Professor in the Department for Mathematical Sciences of Durham University, UK. His principal research interest is in uncertainty quantification, forecast interpretation and evaluation for decision support.

His current energy-related work includes uncertainty quantification for complex computer simulators representing energy economics and engineering systems, forecasting energy consumption and assessing retrofit options for UK high-rise buildings, and assessing climate change impact on future integrated energy systems.

**Wei Sun** (S'11, M'15) received the Ph.D. degree in power system engineering from the University of Edinburgh, U.K., where he is currently a Research Associate on multi-vector energy systems.

Dr. Sun's research interest is in the development and application of optimization methods to the planning, design and operation of smart energy systems. He is currently working on building a multi-scale energy system integration architecture for the National Centre for Energy Systems Integration (CESI) and previously was a lead researcher on Hydrogen's Value in the Energy system (HYVE). He has also contributed to EPSRC-funded projects including Realising Energy Storage Technologies in Low-carbon Energy Systems (RESTLESS) and Adaptation and Resilience In Energy Systems (ARIES).
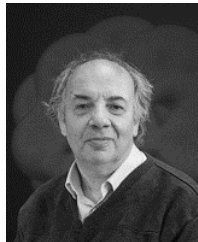
**Michael Goldstein** is Professor of Statistics in the Department for Mathematical Sciences of Durham University, UK. His principal research interest is in Bayesian statistics and he has published extensively in this area. Michael's current energy-related work includes the uncertainty analysis of hierarchical energy systems models, studying the relationship between mathematical and computer models of energy systems, and the real energy systems they attempt to describe, with a view to enabling better model-based decisions in industry and government.

**Gareth P. Harrison** (M'02, SM'14) is Bert Whittington Chair of Electrical Power Engineering and Deputy Head of the School of Engineering at the University of Edinburgh. He holds a Bachelor's degree and a Doctorate from the same institution.

Professor Harrison leads research activity across a wide area including integration of renewable energy within multi-vector energy systems, renewable resource assessment, climate change impacts on energy systems; and carbon footprints of energy systems. He is Associate Director of the National Centre for Energy Systems Integration (CESI), was previously Principal Investigator of the Adaptation and Resilience in Energy Systems project and is currently a Co-investigator on a range of EPSRC and EU projects covering energy storage, hydrogen, conventional generation and offshore renewable energy.

Professor Harrison is a Chartered Engineer, a Fellow of the IET, a Senior Member of the IEEE and an Affiliate of the ACCA.