Electrical Power and Energy Systems 68 (2015) 103–114

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Linear programming techniques for developing an optimal electrical system including high-voltage direct-current transmission and storage

C.T.M. Clack [a,b,*], Y. Xie [b], A.E. MacDonald [b]

[a] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO 80305, United States
[b] Earth Systems Research Laboratory (ESRL), National Oceanic and Atmospheric Administration (NOAA), Boulder, CO 80305, United States

ARTICLE INFO

Article history:
Received 8 October 2013
Received in revised form 10 December 2014
Accepted 10 December 2014

Keywords:
Mathematical optimization
Electrical power systems
Methods and techniques
Renewable energy
Transmission networks
Planning

ABSTRACT

The planning and design of an electric power system, including high-voltage direct-current transmission, is a complex optimization problem. The optimization must integrate and model the engineering requirements and limitations of the generation, while simultaneously balancing the system electric load at all times. The problem is made more difficult with the introduction of variable generators, such as wind and solar photovoltaics. In the present paper, we introduce two comprehensive linear programming techniques to solve these problems. Linear programming is intentionally chosen to keep the problems tractable in terms of time and computational resources. The first is an optimization that minimizes the deviation from the electric load requirements. The procedure includes variable generators, conventional generators, transmission, and storage, along with their most salient engineering requirements. In addition, the optimization includes some basic electric power system requirements. The second optimization is one that minimizes the overall system costs per annum while taking into consideration all the aspects of the first optimization. We discuss the benefits and disadvantages of the proposed approaches. We show that the cost optimization, although computationally more expensive, is superior in terms of optimizing a real-world electric power system. The present paper shows that linear programming techniques can represent an electrical power system from a high-level without undue complication brought on by moving to mixed integer or nonlinear programming. In addition, the optimizations can be implemented in the future in planning tools.

1. Introduction

An electric power system is a complex web of power generators, transmission and distribution lines, a small amount of storage, and power consumers, which must be kept in dynamic equilibrium. The electric power generated on the system at any one instance must be consumed somewhere at the same instant. The historical design of electric power systems is an ad hoc method of addition as needed. A description of electric power systems can be found in, e.g., [1,2]. The ad hoc nature of electric power system growth and regeneration can lead to system weaknesses, which can hamper further growth of new generation and transmission. The electric power system design is an ideal candidate for mathematical optimization. There already exists research into different optimization schemes for different aspects of the power system. The research ongoing has wide ranging interests from the optimization of asset scheduling to power flow optimization across a network, for a selection of related optimizations and overviews see, e.g., [3–6].

The optimization of electric power systems becomes even more difficult with the addition of renewable generators, such as wind turbines and solar photovoltaic (PV) cells. The optimization must take into consideration the variable nature of these relatively new forms of power. In recent years, the optimization of wind, solar, and conventional generator systems has attracted strong research. Much of the attention in the research has been to consider high penetration levels of wind and solar PV deployment in the electrical power system, see e.g., [7–11], which is what we consider in the present paper. The variable nature of wind and solar resources implies that for an optimization to be an effective planning tool for electric power systems it must consider large geographic areas, with high-temporal and -spatial resolution discretization.

Since numerous objectives exist in the mathematical *optimization* of an electric power system with wind and solar PV electrical generation plants, one has to choose what is meant by an *optimal*

* Corresponding author at: Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO 80305, United States.
E-mail address: christopher.clack@noaa.gov (C.T.M. Clack).

http://dx.doi.org/10.1016/j.ijepes.2014.12.049
0142-0615/

system. For example, in [7,8] an energy balance optimization was performed whereby the amount of energy produced in a given time is balanced by the energy consumed in the same time interval. Unfortunately from an electric power system perspective, the optimal system that can generate enough power, but not at the correct times is not effective in the real world. The optimization of [7,8] describes only that wind and solar power can help in an electrical power system, but not to what extent. The procedure carried out by [9] is based upon the Ordinary Least Squares (OLS) and Regularized Least Squares (RLS) procedures, see e.g. [12–16], and it finds an optimal solution with respects to matching the electric load at every time step throughout the time interval studied.

The load-matching procedure set out in [9] is novel, but incomplete and incorrect in certain areas, and thus we set out a new formulation in the present paper. The optimization in [9] has the real-world drawback of possibly being extremely inefficient in terms of cost. The methods adopted in [10,11] are both defined as cost-minimization. The former developed a procedure to look at policy implications with growing the variable resource under numerous constraints. The latter performs an iterative approach to the cost-minimization and attempts to search the solution space via discretization. The cost-minimization technique is the most appropriate for real-life decision-making. In the present paper, we develop a unique cost optimization procedure that designs a wind, solar PV (or any other variable generator) and conventional electricity power generation system, while simultaneously designing a HVDC transmission system and deploying storage capabilities.

The purpose of the present paper is to derive two mathematical optimizations that consider the electric power system as a whole, which can be applied to any desired system. The variable and conventional generators, the transmission, the storage, and the electrical demand all need to be modeled in the optimization to create the most realistic system. In reviewing the literature, we did not find a single optimization procedure that performed the modeling of the electric power system holistically. The present paper describes a load-matching procedure and a cost-minimizing procedure. Both optimizations include all of the fundamental characteristics of an electric power system, however, by necessity, they cannot include every single, small, technical detail of a complete system. We develop *both* the load-matching and cost optimization procedures because it is informative to know to what extent wind and solar PV power can contribute to an electric power system, with and without the constraints of cost. We test the codes on a sample system to demonstrate its capabilities. The mathematical optimizations are designed to be used on large geographic-scale electric power systems at a high-temporal and -spatial resolution. The procedures were intentionally designed to be linear programming and not to be nonlinear programming. Choosing LP rather than NLP was done by transforming the problems while retaining all the salient features. In addition, the choice allows for large problems to be solved in a tractable amount of time and computation resources.

The layout of the present paper is as follows: Section 'Electrical load-matching technique' develops a load-matching mathematical linear programming technique for fast and efficient design of large-scale electric power systems, the section includes descriptions of each modeled parameter and its importance. Section 'Cost minimization technique' describes a mathematical cost optimization for large-scale electric power systems and discusses the important features of the procedure. In Section 'Example test case', we show a sample execution of the two mathematical optimizations on a relatively small data set. Finally, in Section 'Discussion and conclusions', we discuss the benefits and disadvantages of the procedures and the most efficient methods to use when deploying the models using GAMS/Cplex software [17].

2. Electrical load-matching technique

The technique of load-matching is to find, from a geometric perspective, the shortest total distance between all of the electricity generation and the electricity demand at each time instance over a specified time interval (usually a day, week, month, or year). To cast this problem effectively in a Linear Programming (LP) framework, we define the objective function as the sum of excess generation, backup generation, electrical losses due to transmission to consumers, and losses from moving electricity between the grid and storage. We then bound the objective function with linear constraints to ensure that the electric demand is met at every time step, the storage is charged and discharged at an appropriate rate, the transmission is constructed as necessary, the backup generation increases and decreases output (ramps) within technical bounds and the (variable) generation plants are not overbuilt.

The minimization problem modeled in the present paper is designed for wind- and solar- dominated systems, however, the methodology can be used, in principle, for optimizing any electricity generation system. It is already well known that all LP problems can be written in the standard (slack) form [18]

$$\text{minimize } f(\mathbf{x}) \triangleq \mathbf{c}^{\mathrm{T}}\mathbf{x}, \tag{1}$$

$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geqslant \mathbf{0}, \tag{2}$$

where $\{c_j\} \in \mathbb{R}^n$, $\{x_j\} \in \mathbb{R}^n$, $\{A_{ij}\} \in \mathbb{R}^{m\times n}$, and $\{b_i\} \in \mathbb{R}^m$. For the sake of brevity, we drop the $\{\}$ notation for the remainder of the present paper. The coefficients in Eq. (1) are known as the costs or weighting factors for each of the primary endogenous variables. The coefficients in Eq. (2) are usually known, or should be calculable from other constraints. For computational efficiency the LP methods set out in the present paper are written in the most appropriate form for compilation by the solvers. We have shown the standard form for reference to check against to make sure the model created does not become infeasible due to conflicting constraints.

The load-matching optimization developed in the present paper can be written as

$$\text{Min } \chi = \sum_{\tau}\left(\sum_{\mu} g_{\mu\tau} + c_{\mu\tau} + \mathcal{L}^{ts}_{\mu}\cdot s^{o}_{\mu\tau} - \mathcal{L}^{fs}_{\mu}\cdot s^{i}_{\mu\tau} + \frac{1}{2}\sum_{\alpha}\sum_{\beta}\mathcal{L}^{tr}_{\alpha\beta}\cdot\mathcal{D}_{\alpha\beta}\cdot\mathcal{T}_{\alpha\beta\tau}\right) \tag{3}$$

subject to:

$$\sum_{\phi}\left(b_{\phi\mu}\cdot\sum_{\kappa}x_{\phi\kappa}\cdot r_{\phi\kappa\tau}\right) + g_{\mu\tau} + t_{\mu\tau} - c_{\mu\tau} + s^{i}_{\mu\tau}\cdot\left(1-\mathcal{L}^{fs}_{\mu}\right) - s^{o}_{\mu\tau}\cdot\left(1+\mathcal{L}^{ts}_{\mu}\right) = L_{\mu\tau}, \quad \forall\mu,\tau; \tag{4}$$

$$t_{\mu\tau} = \sum_{\alpha}\mathcal{T}_{\alpha\beta\tau}\cdot\left(1-\mathcal{L}^{tr}_{\alpha\beta}\cdot\mathcal{D}_{\alpha\beta}\right)\Big|_{\beta=\mu} - \sum_{\beta}\mathcal{T}_{\alpha\beta\tau}\Big|_{\alpha=\mu}, \quad \forall\mu,\tau(\alpha\neq\beta); \tag{5}$$

$$T_{\hat{\alpha}\hat{\beta}} \geqslant \mathcal{T}_{\alpha\beta\tau}\big|_{\alpha,\beta=\hat{\alpha},\hat{\beta}} \geqslant 0, \quad \forall\hat{\alpha},\hat{\beta},\tau(\hat{\alpha}>\hat{\beta}); \tag{6}$$

$$\hat{S}_{\mu\tau} = \left[s^{o}_{\mu\tau} - s^{i}_{\mu\tau}\right] + \hat{S}_{\mu(\tau-1)}\cdot\left(1-\mathcal{L}^{rs}_{\mu}\right), \quad \forall\mu,\tau\geq 1; \tag{7}$$

$$C^{s}_{\mu} \geqslant \left(1+\mathcal{R}^{s}_{\mu}\right)\cdot s^{o}_{\mu\tau} \geqslant 0, \quad \forall\mu,\tau; \tag{8}$$

$$0 \leqslant s^{o}_{\mu\tau} \leqslant \mathcal{S}^{\mathrm{D}}_{\mu}\cdot C^{s}_{\mu}, \quad \forall\mu,\tau; \tag{9}$$

$$0 \leqslant s^{i}_{\mu\tau} \leqslant \mathcal{S}^{\mathrm{C}}_{\mu}\cdot C^{s}_{\mu}, \quad \forall\mu,\tau; \tag{10}$$

$$C^g_\mu \geqslant \left(1 + \mathcal{R}^g_\mu\right) \cdot g_{\mu\tau} \geqslant 0, \quad \forall \mu, \tau; \tag{11}$$

$$0 \leqslant g_{\mu\tau} \leqslant g_{\mu(\tau-1)} + \mathcal{G}^u_\mu \cdot C^g_\mu, \quad \forall \mu, \tau \geqslant 1; \tag{12}$$

$$g_{\mu\tau} \geqslant g_{\mu(\tau-1)} - \mathcal{G}^d_\mu \cdot C^g_\mu \geqslant 0, \quad \forall \mu, \tau \geqslant 1; \tag{13}$$

$$0 \leqslant x_{\phi\kappa} \leqslant X_{\phi\kappa}, \quad \forall \phi, \kappa; \tag{14}$$

$$\sum_\phi \left[b_{\phi\mu} \cdot \sum_\kappa \left(x_{\phi\kappa} \cdot \sum_\tau r_{\phi\kappa\tau} \right) \right] \geqslant \mathcal{P}_\mu \cdot \sum_\tau L_{\mu\tau}, \quad \forall \mu; \tag{15}$$

$$\alpha, \hat{\alpha}, \beta, \hat{\beta}, \mu \in \mathcal{N}, \ \phi \in \mathcal{B}, \ \kappa \in \mathcal{V}, \ \tau \in \mathcal{Q}. \tag{16}$$

Eqs. (3)–(16) are mathematically equivalent to Eqs. (1) and (2), but describe the details of the modeled behavior of an electric power system. These equations should be supplemented with three more equations that pertain to limiting the amount of dispatchable generation, storage and transmission that can be built (equivalent to Eq. (14) for variable generators). If the limits are not placed the system could construct overly large conventional generation, transmission and storage facilities at single sites. The equations are

$$C^g_\mu \leqslant \mathcal{U}^g_\mu; \tag{17}$$

$$C^s_\mu \leqslant \mathcal{U}^s_\mu; \tag{18}$$

$$T_{\hat{\alpha}\hat{\beta}} \leqslant \mathcal{U}^t_{\hat{\alpha}\hat{\beta}}. \tag{19}$$

The upper bounds from Eqs. (17)–(19), are user-defined and could be the existing amounts of conventional generation, storage and transmission in an electric power system. Without Eqs. (18) and (19), the optimization would find a feasible minimum that uses no dispatchable generation, because it can build the storage and transmission lines to any value to accommodate all the excess energy required to compensate for low variable generation times and then ship the power over the domain via the transmission capabilities (depending on the rate of change of output of technologies and electrical losses modeled). The extra constraints are partly caused by the fact that the variable generation does not appear explicitly in the objective function. A summary of all the variables and equations can be found in Appendix A.

It is instructive to go through all of the equations and define their roles along with the meaning of each parameter contained within each equation. In Eqs. (3)–(16), all the variables (exogenous and endogenous) have subscripts, which means that there is, in principle, a different value for each subscript. Eq. (16) shows all the different subscripts and their sets. The set $\mathcal{N}$ contains all the nodes in the optimization (a node is a region within the domain that is only connected to other nodes by HVDC transmission lines calculated by the optimization), the set $\mathcal{B}$ contains all the possible site locations for variable generation plants, the set $\mathcal{V}$ contains all the possible variable generation types (e.g. wind, solar PV) and finally the set $\mathcal{Q}$ contains all the time steps to be optimized over. The endogenous variables in Eqs. (3)–(14) are: the installed capacity of variable generation ($x_{\phi\kappa}$), the dispatchable generation used ($g_{\mu\tau}$), the electrical power extracted from storage ($s^i_{\mu\tau}$), the electrical power injected into storage ($s^o_{mu\tau}$), the electrical power curtailed ($c_{\mu\tau}$), the value of all the heads and tails of arcs in the transmission network ($\mathcal{T}_{\alpha\beta\tau}$), the transmission capacity constructed ($T_{\hat{\alpha}\hat{\beta}}$), the amount of energy stored ($\hat{S}_{\mu\tau}$), the capacity of storage constructed (C^s_μ), and the capacity of dispatchable generation constructed (C^g_μ). All other variables in Eqs. (3)–(15) are exogenous and will be explained in the description of each of the equations. A concise explanation of all equations can be found in Table A.8.

Eq. (3) is the objective function and it is minimizing the sum of the dispatchable generation ($g_{\mu\tau}$), the round trip electrical losses associated with injection and extraction of power from storage ($\mathcal{L}^{ts}_\mu \cdot s^o_{\mu\tau} - \mathcal{L}^{fs}_\mu \cdot s^i_{\mu\tau}$), and the electric power curtailed ($c_{\mu\tau}$) over all the nodes ($\mu \in \mathcal{N}$) and time steps ($\tau \in \mathcal{Q}$) along with the sum of the electrical losses due to transmission ($1/2 \cdot \mathcal{L}^{tr}_{\alpha\beta} \cdot \mathcal{D}_{\alpha\beta} \cdot \mathcal{T}_{\alpha\beta\tau}$) over all possible arcs ($\alpha, \beta \in \mathcal{N}; \alpha \neq \beta$) and time steps. The factor of $1/2$ in front of the electrical losses due to transmission term is to compensate for the matrix including duplication. The cost coefficients in the objective function here are the terms in front of the endogenous variables; namely the loss terms for the storage and transmission terms and unity for the dispatchable generation and curtailment. The cost coefficients could be supplemented with additional terms that would vary the penalization of the associated term. For example, if the electrical losses due to transmission had a multiplicative cost coefficient of $\pi = 10$ the cost of transmission in the objective term would be ten times large resulting in far less transmission being installed. This was not done, because the transmission and storage use is preferred to a mismatch of the electric load. The exogenous variables denoted by $\mathcal{L}$ are electrical losses (defined in percentages per unit) and, in particular $\mathcal{L}^{ts}, \mathcal{L}^{fs}, \mathcal{L}^{rs}$, and $\mathcal{L}^{tr}$ denote losses to storage, losses from storage, loss rates within storage and transmission losses, respectively. We display a brief description of all terms in the objective function in Table A.3.

The most critical constraint is described in Eq. (4); the electrical load (demand) is met in every node at every time step without fail. The variables in Eq. (4) are restricted or controlled by the remaining constraints [Eqs. (5)–(15)]. Eq. (4) states that in every node and at every time step the variable generation $[\sum_\phi \sum_\kappa (b_{\phi\mu} \cdot x_{\phi\kappa} \cdot r_{\phi\kappa\tau})]$ plus the dispatchable generation ($g_{\mu\tau}$) plus the electricity transmission flux ($t_{\mu\tau}$) plus the net electrical energy extraction from storage $[(1 - \mathcal{L}^{fs}_\mu) \cdot s^i_{\mu\tau}]$ minus the electrical energy injection to storage $[(1 + \mathcal{L}^{ts}_\mu) \cdot s^o_{\mu\tau}]$ minus electrical energy curtailment ($c_{\mu\tau}$) must balance the electrical load ($L_{\mu\tau}$). The exogenous variable $r_{\phi\kappa\tau}$ is the realizable electric power from variable generation sources ($\kappa \in \mathcal{V}$) at each viable resource site ($\phi \in \mathcal{B}$) at each time step. The variable $r_{\phi\kappa\tau}$ also includes an approximation of electrical power losses due to transmission from the resource region to the node center, where the majority of the electric load resides. The term $b_{\phi\mu}$ is simply a binary filter to allow the optimization to determine which resource site belongs in which node. Due to the importance of the load following term, we have a table (Table A.4) to describe the terms within it and the linking terms to the transmission network in Appendix A.

Eqs. (5) and (6) bound the behavior of the transmission network. The term on the left-hand side of Eq. (5) also appears in constraint Eq. (4), and is known as the transmission flux ($t_{\mu\tau}$). The transmission flux term is a free endogenous variable and, as such, can take positive and negative values. For the purposes of modeling the transmission network to be constructed, we introduce the concept of arcs (transmission lines) with heads and tails, for a more detailed explanation of network optimization theory, we refer the reader to, e.g., [19–21]. A head represents the power flow out from an arc and a tail describes the power flow into an arc. The heads and tails are represented by the columns and the rows of $\mathcal{T}_{\alpha\beta\tau}$, respectively. The indices $\alpha, \beta \in \mathcal{N}$ are dummies that are evaluated at each of the nodes μ. Eq. (5) states that the transmission flux is equal to the sum of all the heads entering node μ multiplied by one minus the transmission losses encountered $[\sum_\alpha \mathcal{T}_{\alpha\beta\tau} \cdot (1 - \mathcal{L}^T_{\alpha\beta} \cdot \mathcal{D}_{\alpha\beta})|_{\beta=\mu}]$ minus the sum of all the tails leaving node μ $(\sum_\beta \mathcal{T}_{\alpha\beta\tau}|_{\alpha=\mu})$. The exogenous variable $\mathcal{D}_{\alpha\beta}$ is the distance matrix containing the length of each arc between α and β. When the variable $t_{\mu\tau}$ is positive, the node is considered (in terms of network optimization) a *supply*

node, when its negative, it is considered a *sink node* and when it is zero it is considered a *transshipment node*. The second transmission constraint endogenously finds the arc capacities ($T_{\hat{\alpha}\hat{\beta}}$) by being greater than or equal to the heads and tails of each arc at all time steps ($\mathcal{T}_{\alpha\beta\tau}|_{\alpha,\beta=\hat{\alpha},\hat{\beta}}$). The arc capacity matrix is a lower triangular matrix to avoid duplication of arc capacities. The optimization procedure considers the transmission to be a direct power flow balance. There is no attempt to mimic the voltage phase shift which is highly nonlinear. However, the power flow balance approximation is a reasonable representation for high-voltage direct-current (HVDC) transmission network at a high level [1]. The use of an HVDC transmission instead of high-voltage alternating-current (HVAC) is due to the nonlinear nature of HVAC, which significantly complicates the optimization. However, the HVDC transmission can be through of as an approximation of HVAC in terms of power flow because it includes electrical losses and it describes the transmission at a high level.

The group of Eqs. (7)–(10) describe the functionality of the electric storage contained within the optimization. Eq. (7) calculates the electrical energy stored in any given node at any given time ($\hat{S}_{\mu\tau}$). The right-hand side of Eq. (7) is the difference between the net electric power injected into storage minus the power removed from storage [$s^{o}_{\mu\tau} - s^{i}_{\mu\tau}$] plus the remaining stored electrical energy from the previous time step [$(1 - \mathcal{L}^{rs}_{\mu}) \cdot \hat{S}_{\mu(\tau-1)}$]. The process is iterative within the optimization. The optimization assumes that $\hat{S}_{\mu 0}$ is zero, but it can be prescribed before the start of the optimization. The capacity of the storage contained within each node (C^{s}_{μ}) is determined by Eq. (8). It states that the capacity must be greater than or equal to the electric power injected to storage plus a reserve margin ($\mathcal{R}^{s}_{\mu}$) for all nodes and time steps [$(1 + \mathcal{R}^{s}_{\mu}) \cdot \hat{S}_{\mu\tau}$]. The amount of electric power allowed to be injected into storage and removed from storage at each time step is modeled by Eqs. (9) and (10), respectively. The exogenous variables $\mathcal{S}^{C}_{\mu}$ and $\mathcal{S}^{D}_{\mu}$ are the charge and discharge rates for storage in each node, which is a percentage of capacity (C^{s}_{μ}). Eqs. (7)–(10) all interact directly with each other to model the electric storage properties. These properties dictate how electric storage influences the objective function and main constraint Eq. (4). The storage is modeled assuming it is pumped hydroelectric, where the cost of the generator is largest component of the cost compared with the portion of the cost associated with the energy stored. If battery storage was to be modeled, the left hand side term in (8) would be the capacity in terms of energy and the right-hand side term would be $\hat{S}_{\mu\tau}$, thus representing the energy stored being the dominate factor. Storage has an important impact on the optimization; it allows the system to reduce curtailment at times of over-generation in large areas, and reduce dispatchable generation in times of under production all without the use of transmission. This changes the need for generation and consumption of power at all times to be synchronized, because the storage can hold electrical energy for a later time. This is different to transmission that only allows the shifting of power spatial rather than temporally.

Eqs. (11)–(13) model the attributes of the variable generation in the optimization procedure. The capacity of the dispatchable generation is dictated by Eq. (11), which behaves in exactly the same way as Eq. (8). The capacity (C^{g}_{μ}) in each node is greater or equal to the dispatchable generation plus some reserve requirement ($\mathcal{R}^{g}_{\mu}$). In Eqs. (12) and (13), the dispatchable generation ramp rates (electric power output variability) are taken into consideration. Eq. (12) states that the electric power output from the dispatchable generation in each node at each time step ($g_{\mu\tau}$) must be less than or equal to the dispatchable output in the node at the previous time step ($g_{\mu(\tau-1)}$) *plus* an allowed up-ramp ($\mathcal{G}^{u}_{\mu} \cdot C^{g}_{\mu}$). In Eq. (13), the down-ramp is controlled by stating that the generation in each node at each time step must be greater than or equal to the previous time step *minus* an allowed down-ramp ($\mathcal{G}^{d}_{\mu} \cdot C^{g}_{\mu}$). The exogenous variables $\mathcal{G}^{u}_{\mu}$ and $\mathcal{G}^{u}_{\mu}$ determine the percentage of up-ramp and down-ramp allowed per time step, respectively. The terms found in Eqs. (7)–(13) are collected in Table A.5 for ease of reference.

Eq. (14) sets the upper bound amount of variable generation at each site. The bounds would have to be determined from available land areas, topography of the landscape, socio-political infringements, and multiple other constraints.

The final constraint in Eq. (15) is to enforce a lower bound on the amount of variable generation to be used in each node ($\sum_{\phi}\sum_{\kappa}\sum_{\tau} b_{\phi\mu} \cdot x_{\phi\kappa} \cdot r_{\phi\kappa\tau}$). The right-hand side term determines the percentage ($\mathcal{P}_{\mu}$) of the total electric load ($\sum_{\tau} L_{\mu\tau}$) to be met by variable generation, which can be different for every node μ. The remaining terms for the optimization are collated in Tables A.6 and A.7 in Appendix A for reference.

The procedure set out in Eqs. (3)–(16) is a high-level attempt at designing an electric power system that includes variable generation, conventional generation, HVDC transmission, and electric storage by optimizing in terms of electric power wasted (load-matching). The optimization relies on many inputs, some of which are engineering constraints; $\mathcal{L}^{ts}, \mathcal{L}^{fs}, \mathcal{L}^{rs}, \mathcal{L}^{tr}, \mathcal{D}_{\alpha\beta}, \mathcal{G}^{u}_{\mu}$ and $\mathcal{G}^{d}_{\mu}$, some are policy- or regulatory- imposed; $\mathcal{P}_{\mu}, \mathcal{R}^{s}_{\mu}$ and $\mathcal{R}^{g}_{\mu}$, others are socio-economically derived; $L_{\mu\tau}$ and $X_{\phi\kappa}$, and one is meteorologically prescribed; $r_{\phi\kappa\tau}$. The success or failure of such a routine will be based on the accuracy of all of the exogenous parameters described, which is particularly true when discussing new resources that are asynchronous and variable in nature (e.g. wind and solar PV). The procedure will find the best blend of generators, transmission, and storage that will result in a system with the least amount of electric energy wasted, which is of paramount importance when studying variable generators being connected to an electric power system where its variability can be both a positive and negative effect.

The drawback to the optimization represented in Eqs. (3)–(16) is that while it is optimal with respect to the dispatchable generation used and electrical variable generated energy curtailed, when transferring to the real-world, the costs of such a system may be much higher than necessary, causing the system to not be implemented. Indeed, one can imagine that the construction of such a system may *use* more energy in deploying the generators than is *saved* by matching the electricity load optimally. However, the load-matching optimization was developed for two important reasons. First, it is scientifically instructive to determine the upper bounds for variable generation to match the electricity load without the constraint of costs. Secondly, the mathematics of the cost optimization and load-matching optimization are similar for certain aspects. We show a test case in Section 'Example test case'.

3. Cost minimization technique

In Section 'Electrical load-matching technique', the load-matching optimization routine was developed and explained in detail. In the present section, we will develop a cost optimization routine. The cost optimization is superior to the load-matching optimization for real world applications for the development of a free-market solution to the incorporation of variable generation into an electric power system.

The cost optimization for an electric power system formulated in the present paper is written as

$$\text{Min } \psi = \sum_{\phi}\sum_{\kappa} \mathcal{C}^{v}_{\phi\kappa} \cdot x_{\phi\kappa} + \sum_{\mu} \mathcal{C}^{g}_{\mu} \cdot C^{g}_{\mu} + \sum_{\mu} \mathcal{C}^{s}_{\mu} \cdot C^{s}_{\mu} + \sum_{\hat{\alpha}}\sum_{\hat{\beta}} \mathcal{C}^{tr}_{\hat{\alpha}\hat{\beta}} \cdot \mathcal{D}_{\hat{\alpha}\hat{\beta}} \cdot T_{\hat{\alpha}\hat{\beta}} + \sum_{\mu}\left[\left(\mathcal{C}^{f}_{\mu} + \mathcal{C}^{c}_{\mu} \cdot \mathcal{H}_{\mu} \cdot \mathcal{F}\right) \cdot \sum_{\tau} g_{\mu\tau}\right] \tag{20}$$

subject to Eqs. (4)–(16). Here the notation $\mathcal{C}$ denotes cost coefficients. Specifically, $\mathcal{C}^v_{\phi\kappa}$ is the capital cost (per unit generator constructed) of the variable generator of type κ at location ϕ, $\mathcal{C}^g_\mu$ and $\mathcal{C}^s_\mu$ are the capital costs of the dispatchable generators and storage facilities in node μ, $\mathcal{C}^{tr}_{\hat{\alpha}\hat{\beta}}$ is the cost of each HVDC transmission line per unit power per unit length, $\mathcal{C}^f_\mu$ is the cost of fuel for the dispatchable generators per unit energy, and $\mathcal{C}^c_\mu$ is the additional carbon tax cost per ton of CO_2. The capital costs are considered to be the amortized cost over the optimization time period. The exogenous variables $\mathcal{H}_\mu$ and $\mathcal{F}$ represent the *heat rate* of the dispatchable generation fleet (the amount of heat energy needed to produce a unit of electricity) in node μ and the carbon content of the fuel type (per unit of thermal energy). The product of the three terms $\mathcal{C}^c_\mu \cdot \mathcal{H}_\mu \cdot \mathcal{F}$ result in a cost-per-unit energy of the dispatchable generation in each node μ. The other terms in the objective function have been introduced in the previous section. A summary of all these variables can be found in Appendix A.

The objective function ψ is the total generation, transmission, and storage costs for an electric power system for a selected time frame. Each term is weighted by their respective costs, which means two things; first, to be an effective planning tool for an optimal system the costs need to be accurate considering, among other things, the discount rates, the incentives and the marginal costs of the technology, and secondly, the costs can be tuned to discover the cost at which a technology will become competitive in the generation fleet. In the optimization Eqs. (20) and (4)–(16), it is assumed that there is only one type of dispatchable generator (natural gas combined cycle) as we are considering a high penetration-level, variable-generation system, however it is a trivial addition to expand the optimization to include a variety of dispatchable generators each with their own costs, ramp rates, reserve requirements, and carbon costs using the same process as shown here. It is also possible to *blend* all the different dispatchable generators into a single average one and use that in the exact optimization shown.

Both optimizations, in Eqs. (3)–(20) with Eqs. (4)–(16) assume no knowledge of the existing electric power system, unless specified by the user. The optimizations can take the generation into account by removing it from the electric load in Eq. (4). The transmission can be added in as well, by aggregating the existing line capacities and then changing the bounds on the size of the transmission capacity in the optimization; however, it is assumed that the HVDC transmission envisioned by the optimization would be an overlay to the existing network, which is taken into account by losses considered within each node μ. The power of the optimization set out in (20) with (4)–(16) resides with the fact that the system is always the cheapest or most economical. If you change one parameter, it will find the most economical electric power system that is possible and this allows for a detailed examination of what is realistically possible from an economical and technical standpoint. In Section 'Example test case', we show a test case of the cost optimization procedure in action. We compare the load-matching optimization against the cost optimization to show how they differ.

One final idea for an optimization considered in the present paper is a hybrid optimization,

$$\text{minimize } \xi = \psi + \lambda \cdot \chi, \tag{21}$$

where we minimize the sum of the load-matching optimization and cost optimization regularized by λ. Of course, Eq. (21) is subject to Eqs. (4)–(16). We do not perform the actual optimization here. However, in Section 'Discussion and conclusions' we briefly discuss some analysis of the role of λ in determining an optimal system. We mention it because it is an interesting future field of study, more specifically, when forecast data is incorporated to generate a hybrid weather forecast load-matching and cost optimization.

4. Example test case

We conduct an example test case of the two optimizations set out in Sections 2 and 3. To keep the analysis simple, we only take three forms of generation into account: solar PV, onshore wind, and natural gas. For the wind and solar PV power output estimates, we utilize weather data from seven-day forecasts of the Flow-following Icosahedral Model (FIM) [22] over a domain roughly equivalent to the contiguous USA. The data is for the first 750 h of 2008 (winter) and hours 4000–4750 of 2008 (summer). The natural gas plants are assumed to be back-up generation for when the wind and solar PV cannot meet the electrical demand. We only constrain how the plant operates in terms of the equations within the optimization, we assume other constraints to be negligible in terms of the current methodology. The price data for cost optimization is taken from the transparent cost database median projections for 2030 for the different technologies [23], which equates to \$1,425.70/kW for onshore wind, \$2,652.40/kW for utility-scale PV, \$1,033.10/kW for natural gas combined cycle. We amortized the cost at a discount rate of 5% over 30 years. The cost of transmission is set at \$1.25/kW-mile, while the cost of storage is held at \$1,500.00/kW. The price of natural gas is assumed to be \$6.60/MM Btu; the average price for natural gas to electricity producers in the USA between 2000 and 2010 [24]. The unit size of the generators, storage, and transmission are assumed to be 2.52 MW for wind turbines, 20 MW for solar PV plants, multiples of MW for both storage and natural gas plants, and MW-miles for the transmission. The dispatchable generators are taken to be only natural gas combined cycle plants.

The weather data is transformed into normalized electric power ($r_{\phi\kappa\tau} \in [0,1]$) that designates the percentage power output of a FIM cell for wind and solar PV. The transformation from weather data to electric power is performed by creating power modeling algorithms. The process of converting weather data to power in this manner is described in [25,26]. The basic approach is to take the salient variables (wind speed, solar irradiance, etc.) from a numerical weather prediction model and process them through computer code that mimics the behavior of a wind turbine and solar PV panel. The output will take into account the engineering constraints of the technologies as well as the weather components. To get a normalized value the output from the power model is divided by the capacity of the technology modeled.

We make the simplifying assumption that all variables are not a function of space or node (e.g. $\mathcal{C}^v_{\phi\kappa} = \mathcal{C}^v, \mathcal{R}^s_\mu = \mathcal{R}^s, \mathcal{L}^{tr}_{\alpha\beta} = \mathcal{L}^{tr}$, and $\mathcal{P}_\mu = \mathcal{P}$). The upper bounds for the amount of variable generators ($X_{\phi\kappa}$) to be positioned is calculated such that only half of each FIM cell can be utilized. The electrical losses from and to storage are considered the same and take the value of 5% ($\mathcal{L}^{ts} = \mathcal{L}^{fs} = 5 \cdot 10^{-2}$). The losses for energy in storage is taken to be 0.01% per time step ($\mathcal{L}^{rs} = 10^{-4}$). The electrical losses in the HVDC transmission is assumed to be 0.5% per 100 miles ($\mathcal{L}^{tr} = 5 \cdot 10^{-5}$). The operating reserve requirements for storage and natural gas are the same at 5% ($\mathcal{R}^s = \mathcal{R}^g = 5 \cdot 10^{-2}$). The natural gas and storage ramp rates and charge rates, respectively, are considered the same at 20% ($\mathcal{G}^u = \mathcal{G}^d = \mathcal{S}^C = \mathcal{S}^D = 2 \cdot 10^{-1}$). The amount of variable generation has to be larger than 20% ($\mathcal{P} = 2 \cdot 10^{-1}$) and the carbon tax is set at \$20 per ton of carbon ($\mathcal{C}^c = 2 \cdot 10^1$). The heat rate of the natural gas fleet is set at 6.43 MM Btu/MWh ($\mathcal{H} = 6.43$) and the carbon content of natural gas is 0.0532 tons/MM Btu ($\mathcal{F} = 5.32 \cdot 10^{-2}$) [27]. The transmission is based on the US being divided into sixteen equal regions and the node center being the largest city in each region, which forms the basis of matrix $\mathcal{D}_{\alpha\beta}$. The transmission is further limited to only allow transmission between adjacent nodes, such that every node has access to the power that crosses its borders. The electric load was obtained by means of historical data for 2008 from Federal

Regulatory Commission (FERC) form 714 [28]. We also assume that there is no underlying electric power system, that is the optimizations are trying to find a new electric power system given all the information mentioned above. For the load-matching optimization, the transmission upper bounds is set to 50 GW per arc corridor (or line), while the storage upper bound is set to 2 GW per node.

Fig. 1 shows the variable generation locations from the load-matching optimization procedure for the two time periods investigated. The top panel shows the solution for the 750 h at the beginning of the year (winter) and the bottom panel shows the solution for the 750 h in the middle of the year (summer). We investigated two time periods to try to understand the performance of the optimization. The region for which variable generation can be installed is a rectangle and, therefore, does not exactly fit all possible sites within the contiguous US. The blue squares are FIM cells (approximate area of 900 km^2) which have wind turbines installed and the darker the shade the more installed (as denoted by the legend in Figs. 1 and 2). The yellow/red squares are the FIM cells that have solar PV installed in them. The gray lines denote the HVDC transmission lines constructed by the optimization, with the thickness of the line representing the relative capacity. The images only show the capacity of the wind, solar PV, and transmission infrastructure. The power flow, or instantaneous generation, is not shown. The size of the lines are thickest at 50 GW and all other lines are normalized to this value, to see the relative difference. The same normalization is applied to Fig. 2. For the load-matching optimization, the HVDC transmission lines are all constructed to maximum capacity (50 GW). It should be noted that we are not suggesting that a single line of a specific capacity be constructed, rather that the total capacity be built between the two nodes. The total installed transmission throughout the system is 4200 GW. The location of the natural gas plants and storage are not shown by the optimization as they are constructed on a nodal basis (that is 2 GW of storage in a specific node).

Both the winter and summer load-matching optimizations had the same configuration; namely 146,345 rows, 271,011 columns and 4,360,544 non-zeros. The optimizations each took 315 s to complete (≈110,000 iterations) and output the results. The memory required was 2.5 GB and was run on a single processor in deterministic mode. The optimal objective function for the two load-matching optimization were $\mathcal{O}(10^7)$. For the winter, the load-matching procedure was optimal at 94.75% of the electrical demand met by variable sources (and extraction from storage). The summer month is better at 96.99%, which is likely due to extra solar PV installations (255.09 GW in summer compared with 116.28 GW in winter) and the better solar resource in the summer, which is correlated to the demand. It can be seen in Fig. 1 that there is much more wind power installed in the summer than in the winter optimization (2006.89 GW compared with 1033.11 GW), a consequence of the poorer summer wind power availability. The natural gas installations reduce from winter to summer (from 1547.97 GW to 1216.11 GW) that can be attributed to the decrease in variability of wind and increase in use of solar PV for the daytime peak load. The storage and transmission

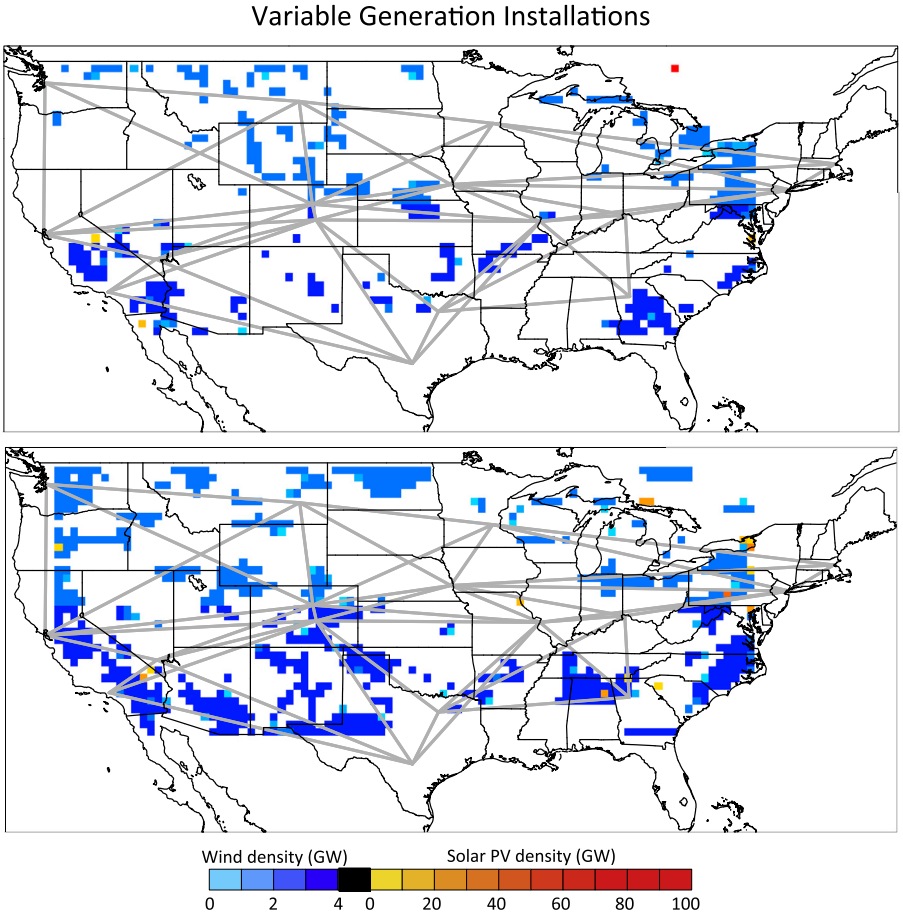

Fig. 1. The location of wind (blue) and solar (yellow) installed plants (MW/km^2) for the load-matching optimization. Top panel is for winter and the bottom is for summer. The gray lines are the HVDC transmission constructed by the optimization, where the thickness represents the capacity (here all equal to the upper bound of 50 GW). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

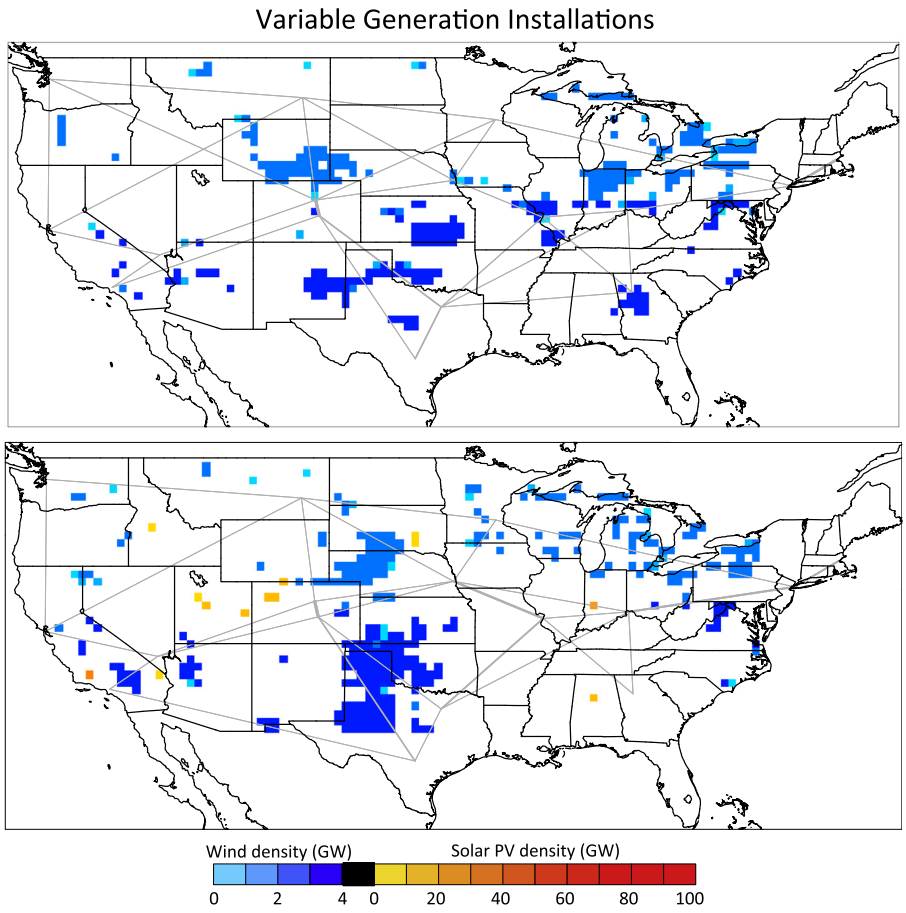

Fig. 2. The location of wind (blue) and solar (yellow) installed plants (MW/km^2) for the cost optimization. Top panel is for winter and the bottom is for summer. The gray lines are the HVDC transmission constructed by the optimization, where the thickness represents the scale of the capacity. The difference between these images and those in Fig. 1 is striking, but they are all *optimal* electric power systems, providing power every time step of the optimization without fail. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

constructed in both scenarios are the same, as the optimization maxed out the use of both (2100 GW of transmission capacity and 32 GW of storage capacity). The computed yearly amortized cost of the winter load-matching optimization is $357.91 billion, while the summer is $449.26 billion. The estimated yearly carbon dioxide emissions from the winter and summer optimizations are 91.44 and 55.80 million metric tons, respectively.

In Fig. 2, we show the variable generation locations for the electric power systems produced by the two cost optimizations. The difference between the two methods is now striking. The locations of the wind and solar PV installations is completely different, and yet both techniques produce an *optimal* network. The optimal network from the cost minimization is more compact and less different between summer and winter. The optimal-cost solution provided 89.49% of the electricity from variable generation in the winter and 82.94% in the summer. The wind installed is 812.98 GW in the winter optimization and 897.40 GW in the summer optimization, while the solar PV constructed is 0 GW in winter and 171.67 GW in summer. The solar is zero in the winter as it is not cost competitive over that time period. The natural gas installations are 229.80 GW in winter and 284.25 GW in summer. There is no storage installed in either run, as it is not cost competitive. The transmission line construction is significantly reduced compared with the load-matching optimization at 407.66 GW for winter and 572.23 GW for the summer. The estimated cost for the winter optimization is $127.39 billion, while the summer optimization comes to $193.27 billion. The difference in cost has several factors, but a major one is the difference in the amount of electricity required between the two periods considered. The estimated yearly carbon dioxide emissions from the two optimization runs are 183.24 and 316.44 million metric tons for the winter and summer, respectively. The cost optimizations have the same number of variables and constraints; 146,073 rows, 153,839 columns and 4,150,626 non-zeros, which is not overly different to the load-matching optimization, however, the computational difficult has increased. The cost optimizations take 1820 s to complete (≈269,000 iterations), but still only require 2.5 GB of memory and a single processor. The optimal objective function for both of the cost optimizations was $\mathcal{O}(10^{10})$.

The load-matching and cost optimization example results are summarized in Tables 1 and 2. The tables show that the cost optimization results in lower capacity installations and reduced yearly costs, which is of paramount importance in a realistic simulation of the electric grid. The load matching optimization has a higher utilization of variable generation, and thus a lower carbon emission footprint, illustrating that it is finding the upper bound with regards to possible variable generation meeting the electric load. Even though the load-matching optimization results in lower carbon emissions, and higher variable generation, it comes at a high price (≈2.8 times for winter and ≈2.3 times for summer). The carbon emissions in the calculations only come from the burning of fossil fuels, however, in reality, the construction of the variable generation will come with emissions of their own, and so the benefits may be outweighed by the added emissions due to

Table 1
Comparison between the winter version of the load matching and cost optimizations.

Result – winter	Load matching	Cost optimization
Wind Capacity (GW)	1033.11	**812.98**
Solar Capacity (GW)	116.28	**0.00**
Natural Gas Capacity (GW)	1547.97	**229.80**
Storage Capacity (GW)	32.00	**0.00**
Transmission Capacity (GW)	4200.00	**407.66**
Average Transmission Utilization (%)	20.3	**30.0**
Yearly Cost (USD$)	357.91 billion	**127.39 billion**
Variable Generation (%)	**94.75**	89.49
Carbon Emissions (million metric tons)	**91.44**	183.24
Computational Expense (core seconds)	**315**	1820
Iterations	≈**110,000**	≈269,000

The bold font signifies that the optimization is superior in terms of that metric for the system.

Table 2
Comparison between the summer version of the load matching and cost optimizations.

Result – summer	Load matching	Cost optimization
Wind Capacity (GW)	2006.89	**897.40**
Solar Capacity (GW)	255.09	**171.67**
Natural Gas Capacity (GW)	1216.11	**284.25**
Storage Capacity (GW)	32.00	**0.00**
Transmission Capacity (GW)	4200.00	**572.23**
Average Transmission Utilization (%)	19.6	**27.9**
Yearly Cost (USD$)	449.26 billion	**193.27 billion**
Variable Generation (%)	**96.99**	82.94
Carbon Emissions (million metric tons)	**55.80**	316.44
Computational Expense (core seconds)	**315**	820
Iterations	≈**110,000**	≈269,000

The bold font signifies that the optimization is superior in terms of that metric for the system.

construction of extra capacity. Indeed, the load-matching optimization constructs far more natural gas in these scenarios as well, of the order of 4–7 times the capacity of cost optimization. The additional capacity is not restricted to generation, the load-matching over builds transmission and storage as well. The trade-off for a more realistic optimization via cost compared with the load-matching is computational expense. The cost optimization is 6 times the computational time and 2.5 times the iterations of the solver. The additional time per iteration is caused by the more complex objective function.

The optimization schemes developed in the present paper have been tested on numerous examples. The results, with regards to performance, are very similar between runs, thus we do not show them here. In general, the cost optimization takes longer (on the order of six times as long) to complete than the load-matching optimization.

5. Discussion and conclusions

In the present section, we will discuss the differences between the two formulations for optimizing an electric power system and the difficulties associated with each. The load-matching optimization is computationally faster than the cost optimization, which is due to the less complicated relationship between the constraints and the objective function. The load-matching does produce a system that has 10% more variable generation than the cost optimization, but the costs are three times higher. Indeed, there is more wind, solar PV, natural gas plants, storage, and transmission installed. The added capacity increases the societal and environmental impacts because the installation of variable generation is not without carbon emissions. The true power of the cost optimization is that one can interpret what a perfect free market would develop at certain cost levels. The load-matching optimization can show how closely a variable generation-dominated electric grid can meet the load, without any regard to cost. In other words, the load-matching can be considered an upper bound to meeting the load without dispatchable generation and curtailment. The cost optimization will find the most economical system, which will be more realistic. In the authors' opinion, the cost optimization has the most real-world use because in reality the generators must pay for themselves to be viable economically, otherwise they will not be built in the first place.

Both optimizations show different solutions for the two time periods. The difference can be associated with the fact that they are optimal to only that time period and are, therefore, likely to be *suboptimal* for any other time period. The optimizations will become more and more accurate the longer the time period over which they are optimized on. Of course, this problem can never be fully resolved, but the variability can be reduced to tolerable levels. Moreover, for variable generation-dominated electric grids, the higher the spatial and temporal discretization and longer time series, the better the optimization result will be. The difficulty in achieving this will be computational, the higher the resolution and the longer time series will result in a much more intensive problem. For example, to conduct a one-year optimization, with the same resolution as the current optimizations would require 30–50 GB of memory and several tens of hours of computing time.

The weather and load data do vary on inter-annual timescales and so the problem variability will always persist. However, the weather and load data can be described in a statistical manner. One technique that the present technique can handle is a quasi-stochastic optimization. To carry this out long time series of weather-driven power data and electric data are required, which can be distilled down to a mean value and a standard deviation. The optimization can then pre-process the power and load data to provide stochastic datasets to optimize upon. The stochastic version of the optimization would have to be performed numerous times to converge on a most appropriate solution; or to give a probability solution. Stochastic optimization will lead to a possible solution that is sub-optimal for each individual time horizon, but hedges for different weather patterns over a longer time horizon, resulting in a more appropriate solution for the longer term.

The biggest challenge of the optimizations in the present paper is the modeling of the HVDC transmission system. The number of transmission endogenous variables rises by the number of the nodal locations to the power of the new time steps. For the optimizations here, an additional time step would add sixteen new transmission variables. Testing of the optimizations has shown that sixteen nodes over the continental US is optimal in terms of the trade-off between the computing time (and resources) required and the resulting electric power system. The cost optimization, although superior to the load-matching in reality, has the disadvantage of being controlled by the prices of the technologies. To get a worthy result from the cost optimization realistic prices must be input to the procedure. For both optimization, for variable generator-dominated systems, accurate weather data is also essential for a credible result. In fact, for variable generation the weather data is the most critical input.

The present paper has derived two optimizations that can efficiently simulate a large-scale electric grid that incorporates variable generation and transmission simultaneously. The procedure is intended to be used on large geographic domains with high spatial- and temporal- resolution weather, topological, and electrical data. The optimizations included are expandable to include any generator type, any transmission network configuration, and any regulations required to simulate an effective electric grid over a pre-defined domain. Other optimizations of electric grids have

been introduced into the literature, but each has a significant deficiency, which the present methodology corrects and improves upon. For example, in [9] their try to do a load-matching optimization, but find that they have a numerical instability problem when using a natural gas constraint. The reason for the apparent issue, is that the ramp rate was only dependent on the previous time step output. In the present methodology, this is avoided, because the ramp rate is controlled by the previous time step and the capacity of the dispatchable fleet. In addition, [9] does not take into account the transmission component of the system, which is included in the present paper. It is noted in [9] that the addition of transmission is a difficult challenge due to its degeneracy and difficulty in representation, even as an approximation. Other load matching techniques in the literature have similar limitations with regards to the addition of transmission within the optimization routine.

The cost optimization introduced in [11], is not an LP method, but rather a trial-and-error method that tests billions of combinations of generators and storage to find the least cost solution. It also does not take into account transmission. It is also over a very limited geographic domain, so expansion of that methodology will become limited, due to the number of iterations it takes to find a minimum (~28 billion combinations). It finds the least-cost option to power the electric grid to 99.99%, which is not the true cost-optimal solution for such a system. The present paper, introduces the cost optimization that will find the theoretical minimum for a system of any geographic size. The cost optimization considers any generators, transmission, and storage. The only true cost optimization available in the literature is found in [10]. In the authors' opinion, the economic evaluation of the generators is very accurate, however, the treatment of the weather discretization is unfavorable for variable generation. In [10], they take the peak and median day each month only for their optimization period, this caused poor optimization performance. This is illustrated in the fact that the summer and winter months in the current example optimizations are different, which is caused by the different weather regimes. The only way to perform optimization including variable generation is to include all available weather information (be that hourly, 15-min, etc.). The other major drawback for the optimization in [10] is that it is over a small domain, and the transmission only exists where transmission currently exists. It does allow expansion, but only over the same corridors.

The cost optimization of the present paper, allows for a completely new infrastructure of transmission to be built on top of any existing network supplied within the optimization. The mathematical derivation of the transmission flux term allows efficient calculation of the transmission network simultaneously to the generation and storage network. It allows solvers to breakdown the problem in a logical manner. In the literature there does not exist any optimizations that take into consideration the generators, transmission, and storage simultaneously over large geographic domains for all time steps within the time period investigated.

The hybrid optimization, introduced in Eq. (21), can take the advantages of both optimization routines and blend them together to give a result that is both economically favorable and matches the load effectively. The hybrid optimization is not a free market solution and as such is not as economical as the cost optimization, but it could provide sensitivity analysis between the two optimizations. From the results of the optimizations performed in the present paper, we can find approximate values for the regularization factor λ. From the cost optimizations we have

$$\psi \approx 2 \cdot 10^{10}, \quad \chi_{\psi} \approx 8 \cdot 10^{7} \Rightarrow \lambda = \mathcal{O}(10^{2}) \tag{22}$$

and from the load-matching optimizations we find

$$\psi_{\chi} \approx 4 \cdot 10^{10}, \quad \chi \approx 2 \cdot 10^{7} \Rightarrow \lambda = \mathcal{O}(10^{3}). \tag{23}$$

In other words, if $\lambda \gg 10^3$ the hybrid optimization will act as the load-matching optimization, if $\lambda \ll 10^2$ the hybrid optimization will perform like the cost optimization, and if $\mathcal{O}(10^2) \leqslant \lambda \leqslant \mathcal{O}(10^3)$ the hybrid optimization blends the cost and load-matching objective functions effectively. Although we do not perform the hybrid optimization, this analysis has important consequences. On inspection of Eq. (21), it can be seen that the dimensions of λ is \$/MWh. Hence, the *cost* of dispatchable or curtailment must be $\mathcal{O}(10^2) = \$100$/MWh before the load-matching technique becomes important, which explains why the cost optimization is so much more economical.

In the present paper, we have developed two robust formulations for optimizing an electric power system that includes conventional and variable generators, HVDC transmission, and storage capacity. Both optimizations include salient features of an electric system from a high-level perspective, and they can both be expanded to increase the accuracy of the modeling of the engineering issues. The optimizations derived in the present paper are the first to design the generation, transmission, and storage networks simultaneously for an electric grid over large spatial domains at a high level while considering each time step of the time period in question. The optimizations are fully expandable to incorporate more generator, transmission, and storage types, as well as regulatory and policy constraints. The work contained in the present paper will be incorporated into higher-resolution optimizations over larger times series in future work. The cost optimization will be deployed to investigate the impact of variable generation within the US electric power system. The implementation of such a large scale problem in a novel area is due to the optimization being formulated in LP rather than NLP to keep the it tractable in terms of time and computer resources.

Acknowledgements

The authors would like to thank A. Alexander, A. Dunbar, J. Wilczak, and C. Sosa for their helpful discussion and recommendations for the paper. The work contained within the present paper was funded by the Office of Oceanic and Atmospheric Research at the National Oceanic and Atmospheric Administration.

Appendix A. Summary of symbols, variables, and equations

The present appendix contains a summary of all the symbols, variables, and equations employed by the optimization contained in the paper. The appendix is designed to help the reader understand the mathematical description. In the optimizations presented in the paper, there are two types of variables; the first is exogenous, or computed by the optimizations, and they are denoted in the tables with a *V*; the second is endogenous, or prescribed by the user (or defined by set rules), and are denoted by *P* in tables. We display the purpose of each equation utilized in the present paper in tabular form for ease of reference.

In Table A.3, the symbols for the two objective functions, Eqs. (3) and (20), are described. The table clearly displays which of the symbols exists in the different objective functions. It is also shown which variables are used in the constraints (and therefore will not be repeated in other tables).

The contents of Table A.4 display the symbols included in Eqs. (4)–(6). The symbols which are both in Eqs. (4)–(6) and contained within Table A.3 are not repeated. The important variable combinations that are shown and explained for clarity. The purpose of these equations is to ensure the electrical load is met at every instant within the optimization at every node, even when considering transmitting the power to different locations, which is a fundamental requirement of any electrical power system.

Table A.3
Symbols utilized within the objective functions contained in Eqs. (3) and (20).

Symbol	Load matching	Cost optimization	Included in constraints	Description
$x_{\phi\kappa}$	–	V	V	The installed capacity of each variable generation (in MW) at each resource site
C^g_μ	–	V	V	The installed capacity of dispatchable generation (in MW) within each node
C^s_μ	–	V	V	The installed capacity of storage (in MW) within each node
$T_{\hat{\alpha}\hat{\beta}}$	–	V	V	The installed capacity of transmission (in MW) between each transmission node
$g_{\mu\tau}$	V	V	V	The amount of dispatchable power required (in MW) each time step within each node
$\mathcal{T}_{\alpha\beta\tau}$	V	–	V	The power flow within the transmission network (in MW) between each transmission node at each time step
$s^o_{\mu\tau}, s^i_{\mu\tau}$	V	–	V	The power transmitted to (and from) storage within each node at each time step
$c_{\mu\tau}$	V	–	V	The amount of generated power that is curtailed (in MW) within each node at each time step
$\mathcal{L}^T_{\alpha\beta}$	P	–	P	The fractional electrical power losses (% per mile) when transmission is performed between each transmission node
$\mathcal{D}_{\alpha\beta}, \mathcal{D}_{\hat{\alpha}\hat{\beta}}$	P	P	P	The geodesic distance between each transmission node center (in miles)
$\mathcal{L}^{ts}_\mu, \mathcal{L}^{fs}_\mu$	P	–	P	The fractional electrical losses (%) due to sending power to (and from) storage within each node
$\mathcal{C}^v_{\phi\tau}, \mathcal{C}^g_\mu, \mathcal{C}^s_\mu$	–	P	–	The amortized capital costs (currently in USD$) of the generators (variable, conventional, and storage) per MW installed, including fixed operation and maintenance (O/M)
$\mathcal{C}^T_{\hat{\alpha}\hat{\beta}}$	–	P	–	The amortized capital costs (currently in USD$) of the transmission lines per MW-mile, including fixed O/M
$\mathcal{C}^f_\mu$	–	P	–	The cost (currently in USD$) of the fuel for dispatchable generation per MWh
$\mathcal{C}^c_\mu, \mathcal{H}_\mu, \mathcal{F}$	–	P	–	The cost (currently in USD$) of carbon per ton, the heat rate of the dispatchable fleet, and the carbon content of the dispatchable fuel within each node

Table A.4
Summary of symbols for Eqs. (3)–(5) along with explanation of major term combinations.

Symbol	Description
$x_{\phi\kappa}, g_{\mu\tau}, \mathcal{L}^{fs}_\mu, s^o_{\mu\tau}, s^i_{\mu\tau}, c_{\mu\tau}, \mathcal{T}_{\alpha\beta\tau}, \mathcal{L}^T_{\alpha\beta}, \mathcal{D}_{\alpha\beta}, T_{\hat{\alpha}\hat{\beta}}$	These symbols were described in Table A.3
$t_{\mu\tau}$	The power transmission flux for each of the nodes (in MW). The term acts as a critical link between the transmission network capacity optimization, Eqs. (5) and (6), and the load fulfillment constraint, Eq. (4). It is this term that allows efficient operation of the simultaneous optimization of generators and transmission [V]
$b_{\phi\mu}$	The binary mask term that designates which resource sites belong to which node [P]
$r_{\phi\kappa\tau}$	The realizable electrical power (%) for each variable generator at each resource site at each time step. The *most important* term in the optimization when considering weather driven renewable energy. All the information about the weather is contained in this term, any reduction in spatial or temporal resolution degrades the optimizations' realism [P]
$L_{\mu\tau}$	The electrical load (in MW) to be fulfilled in each node at each time step, which is dependent on human activity. The competing signals of the weather and the electrical load are what cause the optimal design of an electrical system with renewable energy so difficult [P]
$\sum_\phi \sum_\kappa (b_{\phi\mu} \cdot x_{\phi\kappa} \cdot r_{\phi\kappa\tau}) + g_{\mu\tau}$	The total generation (in MW) at each time step within each node
$s^i_{\mu\tau} \cdot \left(1 - \mathcal{L}^{fs}_\mu\right) - s^o_{\mu\tau}$	The net contribution (in MW) to generation due to storage in each node at each time step
$\sum_\alpha \mathcal{T}_{\alpha\beta\tau} \cdot \left(1 - \mathcal{L}^T_{\alpha\beta} \cdot \mathcal{D}_{\alpha\beta}\right)\Big\vert_{\beta=\mu}$	The net sum of all inbound power transmission (in MW) to each node at each time step
$\sum_\beta \mathcal{T}_{\alpha\beta\tau}\Big\vert_{\alpha=\mu}$	The outbound power transmission (in MW)

Table A.5
Summary of symbols for Eqs. (7)–(13).

Symbol	Description
$s^o_{\mu\tau}, \mathcal{L}^{ts}_\mu, s^i_{\mu\tau}, C^s_\mu, C^g_\mu, g_{\mu\tau}$	These symbols were described in Table A.3
$\hat{S}_{\mu\tau}$	The amount of energy (in MWh) available in the storage reservoir within each node [V]
$\mathcal{L}^{rs}_\mu$	The fractional electrical losses (%) from the storage reservoir within each node per time step [P]
$\mathcal{R}^s_\mu, \mathcal{R}^g_\mu$	The reserve margin for storage and gas (%) [P]
$\mathcal{S}^D_\mu, \mathcal{S}^C_\mu$	The (dis) charge rate for storage within nodes (% of the capacity) [P]
$\mathcal{G}^d_\mu, \mathcal{G}^u_\mu$	Dispatchable down- and up- ramp rates within each node (% of the capacity) [P]

Table A.6
Summary of symbols for Eqs. (14)–(19).

Symbol	Description
$x_{\phi\kappa}, b_{\phi\tau}, r_{\phi\kappa\tau}, L_{\mu\tau}, C^s_\mu, T_{\hat{\alpha}\hat{\beta}}$	These symbols were described in Table A.3 & A.4
$X_{\phi\kappa}$	The upper bounds for the variable generators. It specifies the maximum MW of capacity allowed in each resource site [P]
$\mathcal{P}_\mu$	The fraction (%) of the electrical load that must be met by variable generation [P]
$\mathcal{U}^s_\mu, \mathcal{U}^t_\mu$	The upper bounds for the storage and transmission (in MW) for each node. These terms are only prescribed for the load matching optimization [P]

Table A.7
Summary of indices and sets used throughout the present paper.

Symbol	Description
$\mathcal{N}$	The set of all nodes
$\mathcal{B}$	The set of all renewable energy resource sites
$\mathcal{V}$	The set of all variable generator types
$\mathcal{Q}$	The set of all time steps within the optimization
α, β	The indices for the head and tail of the transmission power flow arcs
$\hat{\alpha}, \hat{\beta}$	The indices denoting the head and tail of each transmission node
μ	Indices for the power producing/ consuming nodes
ϕ	Index of the renewable energy resource sites
κ	Defines each of the variable generator types
τ	The index for the time steps

Table A.8
Brief description of equations used in the optimizations.

Equation	Description and purpose
3.	**The objective function of the load matching optimization**. It minimizes the total of the dispatchable generation consumed, the electrical power curtailed, the roundtrip electrical losses for storage, and the electrical losses in transmission
19.	**The objective function of the cost optimization**. It minimizes the total cost of the installation of generation capacity, the installation of storage, the installation of transmission infrastructure, and the fuel consumed by the dispatchable generators (with the possible inclusion of a carbon tax)
4.	**The load following constraint**. The most fundamental constraint in the optimization. It combines the variable generation, the dispatchable generation, the net storage flux, the transmission flux, and the curtailment of variable generation such that the electrical load is exactly met at every time step in every node throughout the optimization
5.	**The transmission flux link equation**. The equation that allows the simultaneous optimization of the transmission and generation network. It calculates the net flux for each node, when taking electrical losses into account for each arc. It is a Direct Current (DC) approximation
6.	**The transmission capacity constraint**. It determines the capacity of each transmission line by always being greater than the power flow within the lines
7.	**The storage reservoir**. The equation that keeps track of the energy stored within each node. It computes the net flux for storage with the addition of the previous time step reservoir value while taking into account the loss of energy while in storage
8.	**The storage capacity constraint**. It determines the capacity of each node's storage by always being greater than the power in storage multiplied by a reserve margin
9. & 10.	**The storage charge and discharge constraints**. It restricts the charge and discharge rate for the storage facilities in each node at each time step. The rate is tied to the capacity of the storage
11.	**The dispatchable capacity constraint**. It determines the capacity of each node's dispatchable generators by always being greater than the dispatchable power required at any time step multiplied by a reserve margin
12. & 13.	**The dispatchable ramp rate constraints**. It restricts the up and down ramp rate for the dispatchable facilities in each node at each time step. The rate is tied to the capacity of the storage and the previous dispatchable power generated
14.	**The variable generator upper bounds**. Constrains the variable generator size (in MW) at each resource site to be below a set limit, usually defined by limitations on space
15.	**The RPS constraint**. Ensures the optimization provides a pre-designated percentage of the electrical load from variable generation
16.	**Elements and Sets**. Describes the elements and sets used in the other equations
17. & 18.	**Storage and transmission upper bounds**. Only utilized in the load matching optimization. It limits the storage and transmission that can be constructed
20.	**The objective function of a hybrid optimization**. It minimizes a blend of the cost and load matching optimizations, regularized by λ

Table A.5 show the symbols from the dispatchable and storage constraints, Eqs. (7)–(13). These equations deal with the installed capacity of the dispactables and storage plants, as well as the ramping and (dis) charge rates. The purpose of the equations is to constrain the optimization with respect to the engineering capabilities of the conventional generation and generic storage facilities. Again, symbols explained previously will not be repeated.

We show the final symbols in Table A.6 which are defined in the remaining Eqs. (14)–(19). This table completes the summary of all the endogenous and exogenous variables employed in the present paper. Of course, again, we do not repeat earlier defined variables. Additionally, for extra clarity, we display the indices and sets characterized in the present paper in Table A.7.

In the final table of the present appendix, we explain the purpose of each equation utilized in the paper. The brief description of each equation is not meant to replace the explanation in the text, but rather it is designed to support it and be used as a quick reference guide. The content is in Table A.8.

References

[1] von Meier A. Electric power systems: a conceptual introduction. Hoboken (NJ): Wiley-IEEE Press; 2006.
[2] Blume SW. Electric power system basics for the nonelectrical professional. Hoboken (NJ): Wiley-IEEE Press; 2007.
[3] Joel Theis, Electric power system asset optimization. National Energy Technology Laboratory, Report No: DOE/NETL-430/061110; 2011.
[4] Castronuovo ED. Optimization advances in electric power systems. Nova Publishers; 2008.
[5] Bai X, Wei H. Semi-definite programming-based method for security-constrained unit commitment with operational and optimal power flow constraints. Gener Transm Distrib IET 2009;3:182–97.
[6] Farhat IA, El-Hawary ME. Optimization methods applied for solving the short-term hydrothermal coordination problem. Electr Power Syst Res 2009;79:1308–20.
[7] Archer CL, Jacobson MZ. Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements. JGR 2003;108:4289.
[8] Archer CL, Jacobson MZ. Evaluation of global wind power. JGR 2005;110:12110.
[9] Short W, Diakov V. Matching Western US electricity consumption with wind and solar resources. Wind Energy 2012;1099.
[10] Nelson J, Johnston J, Mileva A, Fripp M, Hoffman I, Petros-Good A, et al. High-resolution modeling of the western North American power system demonstrates low-cost and low-carbon futures. Energy Policy 2012;43:436–47.
[11] Budischak C, Sewell D, Thomson H, Mach L, Veron DE. Cost-minimized combinations of wind power, solar power and electrochemical storage, powering the grid up to 99.9% of the time. J Power Sources 2013;225:60–74.
[12] Golub GH, van Loan CF. An analysis of the total least squares problem. SIAM J Matrix Anal Appl 1980;21:185.
[13] Golub GH, Hansen PC, O'Leary DP. Tikhonov regularization and total least squares. SIAM J Matrix Anal Appl 1990;21:185.
[14] Markovsky I. Bibliography on total least squares and related methods. Stat Interface 2010;3:329.
[15] Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. Soviet Math Dokl 1963;4:1035–8.

[16] Tikhonov AN, Arsenin VY. Solution of Ill-posed problems. Washington: Winston and Sons; 1977.
[17] http://www.gams.com/dd/docs/solvers/cplex.pdf.
[18] Fletcher R. Practical methods of optimization. Hoboken (NJ): John Wiley & Sons Inc.; 1987. p.150.
[19] Du DZ, Pardalos PM. Network optimization problems: algorithms, applications and complexity. River Edge (NJ): World Scientific Publishing Co.; 1993.
[20] Ahuja RK, Magnanti TL, Orlin JB. Network flows: theory, algorithms, and applications. Upper Saddle River (NJ): Prentice Hall; 1993.
[21] Bazaraa MS, Jarvis JJ, Sherali HD. Linear programming and network flows. Hoboken (NJ): John Wiley & Sons; 2010.
[22] Earth Systems Research Laboratory, Flow-following Icosahedral model. <http://fim.noaa.gov/fimdocu_rb.pdf>.
[23] National Renewable Energy Laboratory, Transparent cost database. <http://en.openei.org/wiki/Transparent_Cost_Database>.
[24] Energy Information Administration, Natural gas prices. <http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_m.htm>.
[25] Clack CTM, Alexander A, Choukulkar A, MacDonald AE. Improvements in wind resource assessment incorporating wind shear and consideration of resource variability. Wind Energy 2015. Submitted for publication.
[26] Clack CTM, Alexander A, MacDonald AE. Modeling solar irradiance and solar PV power output to create a resource assessment using linear multiple multivariate regression. Solar Energy 2015. Submitted for publication.
[27] Energy Information Administration, Office of Integrated Analysis and Forecasting, Voluntary Reporting of Greenhouse Gases Program, Table of Fuel and Energy Source: Codes and Emission Coefficients. <http://www.eia.gov/oiaf/1605/coefficients.html>.
[28] Federal Energy Regulatory Commission, Form 714. <http://www.ferc.gov/docs-filing/forms/form-714/overview.asp>.