Energy Strategy Reviews 42 (2022) 100892

Contents lists available at ScienceDirect

Energy Strategy Reviews

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

# Using robust optimization to inform US deep decarbonization planning

Neha Patankar [a,*], Hadi Eshraghi [b], Anderson Rodrigo de Queiroz [b,c], Joseph F. DeCarolis [b]

[a] Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ, 08544, USA
[b] Department of Civil, Construction, Environmental Engineering, NC State University, Raleigh, NC, 27695, USA
[c] Department of Decision Sciences, School of Business at NC Central University, Durham, NC, 27695, USA

ARTICLE INFO

*Keywords:*
Robust optimization
Energy system planning
Parametric uncertainty
Energy modeling
Monte Carlo analysis

ABSTRACT

US energy system development consistent with the Paris Agreement will depend in part on future fuel prices and technology costs, which are highly uncertain. Energy system optimization models (ESOMs) represent a critical tool to examine clean energy futures under different assumptions. While many approaches exist to examine future sensitivity and uncertainty in such models, most assume that uncertainty is resolved prior to the model run. Policy makers, however, must take action before uncertainty is resolved. Robust optimization represents a method that explicitly considers future uncertainty within a single model run, yielding a near-term hedging strategy that is robust to uncertainty. This work focuses on extending and applying robust optimization methods to Temoa, an open source ESOM, to derive insights about low carbon pathways in the United States. A robust strategy that explicitly considers future uncertainty has expected savings in total system cost of 12% and an 8% reduction in the standard deviation of expected costs relative to a strategy that ignores uncertainty. The robust technology deployment strategy also entails more diversified technology mixes across the energy sectors modeled.

## 1. Introduction

While progress on US federal climate policy remains elusive, a carbon neutral energy system represents the ultimate target [1]. Energy technology deployments over the coming decades aimed at achieving carbon neutrality are subject to considerable uncertainty that should be analyzed when considering specific policy actions. Energy system optimization models (ESOMs) provide a self-consistent framework for such evaluation, allowing analysts to examine future technology and fuel pathways under specific policy objectives. ESOMs employ mathematical programming techniques to minimize the present cost of energy supply or maximize welfare by optimizing the installation and utilization of energy technologies over a user-defined time horizon. These models link energy technologies together in a broad-based network through the flow of energy commodities. Inputs to ESOMs include existing technology capacity, capital costs, fixed and variable operational costs, conversion efficiencies, and emissions coefficients; outputs include installed technology capacity, energy consumed and produced by each technology, energy prices, and emissions. The insights obtained from ESOMs are often used to inform policy decisions, and thus energy models should account for uncertainty that can affect policy outcomes [2–6]. While a variety of methods can be used to quantify model sensitivity and the effects of future uncertainty [7], most of these methods require the selection of uncertain values prior to the model run, and thus represent a learn-then-act approach [8]. However, policy makers are confronted with decisions that must be made before future uncertainty is resolved, which represents an act-then-learn approach [8]. When conducting model-based analysis to inform policy strategy, ESOMs should explicitly account for future uncertainty within the model formulation.

Developing a hedging strategy that provides one single best course of action while accounting for future uncertainty can generally be accomplished by two different methods: stochastic linear programming (SLP) and robust optimization (RO) [9]. SLP requires the modeler to develop a scenario tree, assign subjective probabilities and outcomes to each branch in the tree, and then optimize the model over the whole tree. A key limitation of SLP is the curse of dimensionality [10], whereby the number of decision variables grow exponentially with the number of uncertain parameters and model time periods. Given the large parametric uncertainties in future energy systems, the curse of dimensionality has been a key limiting factor in previous applications of SLP to ESOMs (e.g., [4,11–14]).

A second approach, which does not suffer from the curse of dimensionality, is RO. It combines the features of sensitivity analysis (SA), multi-objective optimization, and stochastic linear programming (SLP) to generate a series of solutions that are progressively less sensitive to the realization of uncertainty associated with the inputs, thereby

* Corresponding author.
*E-mail address:* patankar@princeton.edu (N. Patankar).

https://doi.org/10.1016/j.esr.2022.100892
Received 5 September 2021; Received in revised form 14 April 2022; Accepted 28 June 2022
Available online 15 July 2022

**Nomenclature of parameters and variables**

***Sets***

| | |
|---|---|
| $i \in I$ | Set of all technologies with investment cost |
| $f \in F$ | Set of all fuels |
| $v, \tilde{v} \in V_i$ | Vintages associated with technology $i$ |
| $t, \tilde{t} \in T_i / T_f$ | Time period associated with technology $i$ or fuel $f$ |
| $T$ | All model time periods |
| $\eta$ | Feasible domain for the maximization problem given in constraint (5) |

***Parameters***

| | |
|---|---|
| $IC_{iv}$ | Nominal investment cost of technology $i$ with vintage $v$ [$ Million/GW] |
| $IC^*_{iv}$ | Realization of investment cost of technology $i$ with vintage $v$ [$ Million/GW] |
| $\Delta IC_{iv}$ | Maximum deviation in investment cost of technology $i$ with vintage $v$ [$ Million/GW] |
| $OC_{ivt}$ | Nominal operational cost of technology $i$ with vintage $v$ at time period $t$ [$ Million/GWh] |
| $FC_{ft}$ | Nominal cost of fuel $f$ in time period $t$ [$ Million/GWh] |
| $FC^*_{ft}$ | Realization of cost of fuel $f$ in time period $t$ [$ Million/GWh] |
| $\Delta FC_{ft}$ | Maximum deviation in cost of fuel $f$ in time period $t$ [$ Million/GWh] |
| $\gamma_{ivt}$ | Conversion factor from capacity to activity [h] |
| $D_t$ | Demand at time period $t$ [GWh] |
| $B$ | Matrix of constraint coefficients |
| $b$ | Right-hand-side of constraints |
| $\Gamma$ | Number or percentage of uncertain parameters assuming worst-case values; also referred to as the 'budget of uncertainty' |
| $\epsilon$ | Probability of more than $n$ uncertain parameters assuming their worst-case value |
| $g^{IC}_{\tilde{v}iv}$ | Maximum possible impact of $IC_{iv}$ on $IC_{i\tilde{v}}$ |
| $g^{FC}_{\tilde{t}ft}$ | Maximum possible impact of $FC_{ft}$ on $FC_{f\tilde{t}}$ |
| $X^{max}_t$ | worst-case value of the random variable $\mathbf{X}_t$ |
| $\sigma_t$ | Standard deviation of the random variable $\mathbf{X}_t$ |
| $\rho$ | Temporal correlation coefficient of a technology or a fuel |
| $\mu_t$ | Mean of a random variable $\mathbf{X}_t$ |
| $\mu_{\mathbf{X}_{t+\Delta\|t}}$ | Mean of a random variable $\mathbf{X}_{t+\Delta\|t}$ |
| $\Delta$ | Distance between adjacent time periods or vintages |

***Variables***

| | |
|---|---|
| $\mathbf{CAP}_{iv}$ | Capacity of technology $i$ with vintage $v$ in [GW] |
| $\mathbf{ACT}_{ivt}$ | Activity associated with technology $i$ with vintage $v$ at time period $t$ [GWh] |
| $\mathbf{CON}_{ft}$ | Consumption of fuel $f$ in time period $t$ [GWh] |
| $\mathbf{x}$ | Vector of decision variables in constraint (1c) |
| $\mathbf{W}$ | Continuous variable used to convert objective function into a constraint |
| $\boldsymbol{\eta}^{IC}_{\tilde{v}iv}, \boldsymbol{\eta}^{FC}_{\tilde{t}ft}$ | Independent and symmetrically distributed random variable |
| $\mathbf{p}, \mathbf{q}^{IC}_{i\tilde{v}}, \mathbf{q}^{FC}_{f\tilde{t}}, \mathbf{y}^{IC}_{i\tilde{v}}, \mathbf{y}^{FC}_{f\tilde{t}}$ | Dual variables associated with the feasible space defined in (5) |
| $\mathbf{X}_t$ | Random variable used to represent the uncertain parameters |
| $\mathbf{X}_{t+\Delta\|t} = \mathbf{X}_{t+\Delta}\|\mathbf{X}_t$ | A random variable denoting a value of $\mathbf{X}_{t+\Delta}$ given the value of $\mathbf{X}_t$ |
| $\mathbf{s}^{IC}_{iv}$ | Scaled deviation of uncertain investment cost $IC^*_{iv}$ |
| $\mathbf{s}^{FC}_{ft}$ | Scaled deviation of uncertain fuel cost $FC^*_{ft}$ |
| $\mathbf{p}, \mathbf{q}^{IC}_{iv}, \mathbf{q}^{FC}_{ft}, \mathbf{r}^{IC}_{iv}, \mathbf{r}^{FC}_{ft}$ | Dual variables associated with the feasible space defined in (A.6) |

providing solutions that are robust to the modeled uncertainties [15]. This method requires the modeler to choose a set of uncertain parameters, assign both baseline and worst-case values to those parameters, and then iteratively perform optimization runs, where each iteration includes a given number of model input parameters assuming their worst-case value. In contrast to SLP, RO is a more suitable approach for developing a hedging strategy when considering a large number of uncertain parameters.

The first implementation of RO by Soyster [16] provides a feasible solution under any realization of uncertain parameters. However, this early formulation has two drawbacks. First, the formulation is nonlinear, which results in a high computational burden. Second, the resulting solution is very conservative since the model remains feasible even when the worst-case value is assumed for all uncertain parameters. To improve his approach, subsequent developments were made to explore the trade-off between the conservativeness of the solution and optimality [17–21]. In particular, the approach proposed by Bertsimas and Sim [19] is most relevant in an energy systems planning context, as it lets the decision maker choose their risk tolerance by specifying a parameter called the "budget of uncertainty". The budget of uncertainty represents the number (or percentage) of uncertain model input parameters that take on their worst-case value. When a model assumes the highest budget of uncertainty such that all uncertain parameters take on their worst-case values, the formulation proposed by Soyster [16] becomes a special case of the formulation proposed by Bertsimas and Sim [19]. One limitation of RO over SLP is that it only considers the worst-case value of each parameter, rather than a probability distribution. While [22–24] have developed methodologies to create robust strategies, direct application of the RO methodology to energy system optimization models is limited to [25–30]. We build on this existing literature by introducing the RO formulation for correlated uncertain parameters.

The purpose of this paper is to apply RO to develop insights about robust future technology pathways that achieve an 80% reduction in $CO_2$ emissions below 2005 levels by 2050. The RO formulation is implemented in an open source ESOM called Tools for Energy Model Optimization and Analysis (Temoa), in conjunction with a US input dataset to explore robust technology development pathways that result in deep decarbonization. This paper also makes several methodological contributions. First, we introduce a systematic methodology to form the RO uncertainty set. Second, we provide a methodology to account for the auto-correlation in Temoa's input cost parameters. For example, the range in solar photovoltaic (PV) capital cost in a given time period will be correlated with the cost in the previous time period. This correlated RO methodology (CR-ESOM) is applied to an ESOM for the first time. Third, this paper is the first application of the RO methodology to a large scale representation of the US energy system. The RO formulation presented here is generalized for ESOMs, and can be applied to similar models and datasets to examine a wide range of scenarios.

The remainder of the paper is organized as follows. Section 2 describes the specific RO methodology used in this paper, including the formulation of the RO counterpart for ESOMs with temporal correlation between uncertain data. A description of the US energy system database is given in Section 3. Section 4 presents the methodology

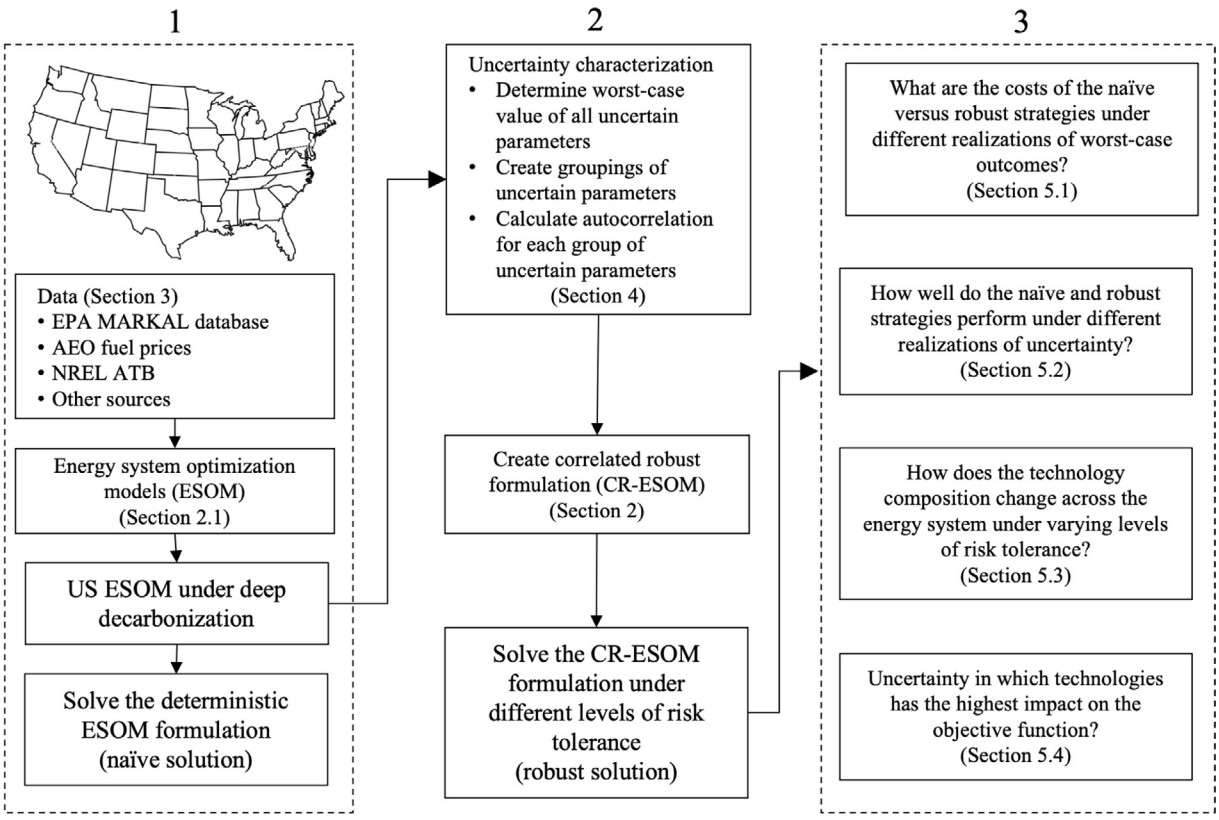


**Fig. 1.** Methodology for obtaining robust alternate policies for GHG emission mitigation under technology and fuel cost uncertainty.

to build the uncertainty set while Section 5 presents the results and provides discussion. Section 6 presents conclusions from the analysis. Appendix provides the RO formulation without temporal correlation for completeness.

## 2. Methodology

Our approach involves building an input database for the US energy system, implementing robust optimization within Temoa, characterizing future parametric uncertainty, providing robust strategies for greenhouse gas (GHG) emissions mitigation, and analyzing the quality of the robust model solutions. Fig. 1 outlines the flow of information through the paper. In the first part of the paper, denoted by Column 1 in Fig. 1, we compile data from public sources into a relational database representing the US energy system as a single region (Section 3). We feed the data to Temoa under a constraint on $CO_2$ emissions that achieves an 80% reduction below 2005 levels by 2050. The model solution obtained from this step provides the optimal capacity expansion plan that does not consider future uncertainty. Henceforth, this solution is referred to as the 'naïve' solution.

In the second part of the paper, denoted by Column 2 in Fig. 1, we estimate the worst-case values and auto-correlation for the input fuel costs and capital costs (Section 4). We then formulate the RO model in two steps. The first robust (R-ESOM) formulation does not consider auto-correlation in uncertain parameters. We do not present the results for R-ESOM; however, we provide the methodology in the Appendix for completeness. Auto-correlation is important in capacity expansion models because most input parameters are indexed by model time period, but the RO procedure assumes that each parameter value in each time period can be treated independently. As an example, in this basic R-ESOM formulation, natural gas prices can take on their worst-case value in 2030, but the prices remain unaffected in other time periods. Building on this formulation, we construct a correlated RO (CR-ESOM) formulation (Section 2.2). In this version, a parameter taking on its worst-case value in a given time period will, in turn, affect the parameter value in the other time periods through a correlation parameter. The solution obtained by solving CR-ESOM represents the 'robust' solution.

The third part of the paper, denoted by Column 3 in Fig. 1, utilizes the RO modeling framework developed under the first two parts of the paper to rigorously examine the performance of the robust decision making strategy relative to a naive strategy that ignores future uncertainty. These questions focus on four objectives (1) the incremental cost or savings of the robust planning strategy under different realizations of uncertainty, (2) changes to low carbon technology pathways across the energy system under different levels of risk tolerance, and (3) the identification of the uncertain parameters that have the largest effect on system costs, and therefore the robust strategy.

### 2.1. A simplified ESOM formulation

Since this paper is focused on energy system optimization models (ESOMs), we provide a brief description of ESOMs before reviewing previous applications of robust optimization. ESOMs can be described algebraically as a network of linked processes that convert raw energy commodities (e.g., coal, oil, biomass) into end-use demands (e.g., lighting, transport, water heating) through a series of one or more intermediate energy forms (e.g., electricity, gasoline, ethanol). Each process is defined by a set of engineering, economic, and environmental characteristics (e.g., capital, fixed and variable cost, efficiency, capacity factor) associated with converting an energy commodity from one form to another. Processes are linked together in a network via model constraints representing the allowable flow of energy commodities. The objective of ESOMs is to minimize the net present cost of energy supply by utilizing energy technologies and commodities over a user-specified time horizon to meet a set of exogenously specified end-use demands. ESOMs simultaneously make technology investment and operating decisions while enforcing an energy balance between

primary energy resources, secondary fuels, final energy consumption, and end-use energy services. ESOMs are typically formulated as linear programming models, where installed technology capacity is treated as a continuous variable. An introduction to Temoa is available [31] and an up-to-date mathematical formulation of the model can be found in the online documentation [32].

When end-use demands are specified exogenously, a simplified form of a general ESOM formulation with the objective to minimize total system costs can be written as the following linear program:

$$\min \quad \sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} + \sum_{t\in T_f}\sum_{f\in F} FC_{ft}\mathbf{CON}_{ft} \tag{1a}$$

$$\text{s.t.} \quad \sum_{v\in V_i}\sum_{i\in I}\mathbf{ACT}_{ivt} \geq D_t \qquad \forall t\in T \tag{1b}$$

$$\sum_{V_i\leq T_i}\gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t\in T_i, v\in V_i, i\in I \tag{1c}$$

$$B\mathbf{x}\geq b \tag{1d}$$

In the above model (1a)–(1d), $V$, $I$, $T$ and $F$ are the set of all vintages, technologies model time periods and fuels, respectively. Vintage $V$ is a time period in which a technology is built, while time period $T$ represents the period in which technology or fuel is used. $IC_{iv}$, $OC_{ivt}$ and $FC_{ft}$ are the investment cost, operations and maintenance cost (O&M) and fuel costs, respectively. $\mathbf{CAP}_{iv}$ is the decision variable representing available capacity of technology $i$ of vintage $v$. In the above model formulation, the total commodity production from a process is referred to as "activity". $\mathbf{ACT}_{ivt}$ is the decision variable representing activity (i.e., energy output) of technology $i$ of vintage $v$ in time period $t$. $\mathbf{CON}_{ft}$ is the decision variable representing the consumption of fuel $f$ in time period $t$. Moreover, $\gamma_{ivt}$ is the factor that converts available capacity into activity, taking into account the technology-specific availability factor and conversion from capacity to activity units. $D_t$ is the demand in time period $t$. Furthermore, $\mathbf{x}$ represents all other variables in the ESOM, $B$ represents coefficient matrix of the constraints involving those variables, and $b$ represents the right-hand side of the constraints. Objective function (1a) expresses the total discounted cost to be minimized, (1b) is the set of demand satisfaction constraints, where the right-hand-side $D_t$ are the exogenous demands that need to be satisfied, (1c) represents the relationship between available capacity and activity and (1d) is the set of all other constraints such as flow balance, resource availability, technology share, process balance, lower and upper bounds on decision variables, limit on technology lifetime, reserve margin and storage constraints. DeCarolis and Hunter [32] provide a detailed formulation for the constraints included in (1d).

### 2.2. RO formulation for correlated parameters (CR-ESOM)

The first RO formulation proposed by Soyster [16] constructs a solution that is feasible for any realization of uncertain data. As a result, the solution is too conservative in the sense that we give up too much of the optimality of the 'naive' solution in order to guarantee robustness by focusing on the worst-case possibility that all uncertain input parameters assume their worst-case value. Bertsimas and Sim [19], on the contrary, formulate the RO model based on the stipulation that nature will be restricted in its behavior. In other words, only a subset of the uncertain parameters will realize their worst-case values and adversely affect the system cost. This approach will produce robust solutions that remain feasible under most circumstances. Furthermore, the [19] formulation for uncorrelated uncertain parameters has already been adopted in the energy modeling literature [26–28]. Even though we use the correlated version of the RO formulation, we provide the uncorrelated RO formulation in Appendix for completeness.

Since ESOMs usually deal with capacity expansion over decades, addressing the uncertainty in fuel costs and technology investment costs under different future scenarios is critical to developing robust planning and investment strategies. The existing applications of RO in ESOMs assume that the uncertain ESOM parameters are independently random. In reality, the ESOM parameters are not independently random, particularly for different vintages of the same technology through time. For example, the investment cost of technology $i$ with vintage $v_n$, $IC_{iv_n}$, might depend on $IC_{iv_1}$ and vice versa, where $v_1,\ldots,v_n,\ldots,v_{|V_i|}\in V_i$ and $V_i$ is the set of vintages of technology $i$. To model temporal correlation for uncertain parameters, we assume that the investment cost of a given technology vintage is positively correlated with the other vintages of the same technology. In other words, an increase in $IC_{iv_1}$ leads to an increase in $IC_{iv_n}$, $\forall v_n\in V_i$ and a decrease in $IC_{iv_1}$ leads to a decrease in $IC_{iv_n}$. The realization of uncertain investment and fuel cost parameters, $IC^*_{iv}$, $i\in I$ and $FC^*_{ft}$, $f\in F$, is modeled as:

$$IC^*_{iv} = IC_{iv} + \sum_{\tilde{v}\in V_i}\eta^{IC}_{\tilde{v}iv}g^{IC}_{\tilde{v}iv}IC_{iv} \qquad \forall i\in I, v\in V_i \tag{2}$$

$$FC^*_{ft} = FC_{ft} + \sum_{\tilde{t}\in T_f}\eta^{FC}_{\tilde{t}ft}g^{FC}_{\tilde{t}ft}FC_{ft} \qquad \forall f\in F, t\in T_f \tag{3}$$

where, $\eta^{IC}_{\tilde{v}iv}$ and $\eta^{FC}_{\tilde{t}ft}$ are independent and symmetrically distributed random variables that take on values in the domain $[-1,1]$. $g_{\tilde{v}iv}$ is the effect of $IC_{iv}$ on $IC_{i\tilde{v}}$ for $v,\tilde{v}\in V_i$ and $g_{\tilde{t}ft}$ is the effect of $FC_{ft}$ on $FC_{f\tilde{t}}$ for $t,\tilde{t}\in T_f$. In other words, $\eta^{IC}_{\tilde{v}iv}$ determines the impact of an investment cost change of one vintage on the other and $g_{\tilde{v}iv}$ represents the maximum possible impact of vintage $v$ on vintage $\tilde{v},\tilde{v}\in V_i$. For example, from the approach outline in Section 4, if we determine that a change in $IC_{iv}$ may lead to a 20% increase in $IC_{i\tilde{v}}$, then $g_{\tilde{v}iv}=0.2$. Note that under this model there are $|I|.|V_i|+|F|.|T_f|$ potential sources of uncertainty.

Bertsimas and Sim [19] provide the RO formulation for uncertain constraint parameters. To use this formulation, we introduce a dummy variable $\mathbf{W}$ to write the objective function in the form of a constraint. The resulting CR-ESOM formulation can be given as (4a)–(4e).

$$\min \quad \mathbf{W} \tag{4a}$$

$$\begin{aligned}\text{s.t.} \quad \mathbf{W} - \Bigg(&\sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} \\ &+ \sum_{t\in T_f}\sum_{f\in F} FC_{ft}\mathbf{CON}_{ft}\Bigg) \\ &- \max_{\boldsymbol{\eta}^{IC}_{\tilde{v}iv},\boldsymbol{\eta}^{FC}_{\tilde{t}ft}\in\eta}\Bigg(\sum_{\tilde{v}\in V_i}\sum_{v\in V_i}\sum_{i\in I}\eta^{IC}_{\tilde{v}iv}g^{IC}_{\tilde{v}iv}IC_{iv}CAP_{iv} \\ &+ \sum_{\tilde{t}\in T_f}\sum_{t\in T_f}\sum_{f\in F}\eta^{FC}_{\tilde{t}ft}g^{FC}_{\tilde{t}ft}FC_{ft}CON_{ft}\Bigg)\geq 0\end{aligned} \tag{4b}$$

$$\sum_{v\in V_i}\sum_{i\in I}\mathbf{ACT}_{ivt} \geq D_t \qquad \forall t\in T \tag{4c}$$

$$\sum_{V_i\leq T_i}\gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t\in T_i, v\in V_i, i\in I \tag{4d}$$

$$B\mathbf{x}\geq b \tag{4e}$$

$$\eta = \Bigg\{ |\eta^{IC}_{\tilde{v}iv}|\leq 1, \forall i\in I,\ v,\tilde{v}\in V_i; |\eta^{FC}_{\tilde{t}ft}|\leq 1, \forall f\in F,\ t,\tilde{t}\in T_f, \sum_{\tilde{v}\in V_i}\sum_{v\in V_i}\sum_{i\in I}|\eta^{IC}_{\tilde{v}iv}| + \sum_{\tilde{t}\in T_f}\sum_{t\in T_f}\sum_{f\in F}|\eta^{FC}_{\tilde{t}ft}|\leq \Gamma\Bigg\} \tag{5}$$

In the above model, (5) is a feasible domain for the maximization problem given in (4b). In the case of this work, uncertainties are represented in technology investment costs and fuel costs, which affect the objective function of model (1a), and in the robust formulation are present in the structural constraint (4b). However, for a given $\Gamma$, the maximization problem in constraint (4b) provides a protection against infeasibility by assuming that total system cost includes the worst-case cost of $\Gamma$ uncertain parameters. We note that here we refer to model infeasibility as the inability of the mathematical solver

to produce an optimal solution to the mathematical model, i.e., the mathematical model presents an empty feasible region. For the robust model formulation such infeasibilities may arise due to extreme realizations of uncertainty that may change the feasible region of the model in such a way that no values for the decision variables will be able to satisfy all the model constraints. Eq. (4b) maximizes the protection for a given technology capacity ($CAP_{iv}$) and consumption of fuel ($CON_{ft}$) by varying the decision variables ($\boldsymbol{\eta}^{IC}_{\tilde{v}iv}$ and $\boldsymbol{\eta}^{FC}_{\tilde{t}ft}$) over the feasible domain $\eta$. Note that $\boldsymbol{\eta}^{IC}_{\tilde{v}iv}$ and $\boldsymbol{\eta}^{FC}_{\tilde{t}ft}$ are represented as random variables in Eqs. (2) and (3), however, they act as decision variables in the above formulation. The maximization model is formulated as a linear programming problem, as we keep the $CAP_{iv}$ and $CON_{ft}$ constant as defined by the minimization model. Problem (4a)–(4e) can be solved by iteratively solving minimization and maximization models to minimize **W**. To avoid the iterative method, we can simplify model (4a)–(4e) through the application of strong duality, which says that at the optimum, the objective function value of the dual problem is the same as primal problem [33].

Similar to the methodology presented in Appendix for R-ESOM, through the application of strong duality, (4a)–(4e) can be written as the linear program represented by (6a)–(6j).

$$\min \quad \mathbf{W} \tag{6a}$$

$$\text{s.t.} \quad \mathbf{W} - \left( \sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} + \sum_{t\in T_f}\sum_{f\in F} FC_{ft}\mathbf{CON}_{ft} \right) - \left( \sum_{\tilde{v}\in V_i}\sum_{i\in I} \mathbf{q}^{IC}_{i\tilde{v}} + \sum_{\tilde{t}\in T_f}\sum_{f\in F} \mathbf{q}^{FC}_{f\tilde{t}} + \Gamma\mathbf{p} \right) \geq 0 \tag{6b}$$

$$\sum_{v\in V_i}\sum_{i\in I} \mathbf{ACT}_{ivt} \geq D_t \qquad \forall t \in T \tag{6c}$$

$$\sum_{V_i\leq T_i} \gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t \in T_i, v \in V_i, i \in I \tag{6d}$$

$$B\mathbf{x} \geq b \tag{6e}$$

$$\mathbf{p} + \mathbf{q}^{IC}_{i\tilde{v}} \geq \mathbf{y}^{IC}_{i\tilde{v}} \qquad \forall v \in V_i, i \in I \tag{6f}$$

$$\mathbf{p} + \mathbf{q}^{FC}_{f\tilde{t}} \geq \mathbf{y}^{FC}_{f\tilde{t}} \qquad \forall t \in T_f, f \in F \tag{6g}$$

$$-\mathbf{y}^{IC}_{i\tilde{v}} \leq \sum_{\tilde{v}\in V_i} g^{IC}_{\tilde{v}iv} IC_{iv}\mathbf{CAP}_{iv} \leq \mathbf{y}^{IC}_{i\tilde{v}} \qquad \forall v \in V_i, i \in I \tag{6h}$$

$$-\mathbf{y}^{FC}_{f\tilde{t}} \leq \sum_{\tilde{t}\in T_f} g^{FC}_{\tilde{t}ft} FC_{ft}\mathbf{CON}_{ft} \leq \mathbf{y}^{FC}_{f\tilde{t}} \qquad \forall t \in T_f, f \in F \tag{6i}$$

$$\mathbf{p}, \mathbf{q}^{IC}_{i\tilde{v}}, \mathbf{q}^{FC}_{f\tilde{t}}, \mathbf{y}^{IC}_{i\tilde{v}}, \mathbf{y}^{FC}_{f\tilde{t}} \geq 0 \qquad \forall \tilde{v} \in V_i, \tilde{t} \in T_f, i \in I, f \in F \tag{6j}$$

In Constraint (6b), the last term represents the objective function of the dual problem of the maximization problem given in Constraint (4b), while Constraints (6f)–(6j) are the corresponding constraints of the dual problem. To aid understanding of the dual formulation, consider a case where the budget of uncertainty ($\Gamma$) is 0. In this case, the optimal value of the dual variable **p** is equal to $max(\mathbf{y}^{IC}_{i\tilde{v}}, \mathbf{y}^{FC}_{f\tilde{t}})$ and the other dual variables $\mathbf{q}^{IC}_{i\tilde{v}}, \mathbf{q}^{FC}_{f\tilde{t}} = 0$, making the third term in (6b) irrelevant and thus reducing the formulation (6a)–(6j) to the least-cost problem (1a)–(1d). For a budget of uncertainty $\Gamma > 0$, the model tries to diversify the selection of fuels and technologies to minimize the combined worst-case outcome by optimizing the dual variables of the maximization problem ($\mathbf{p}, \mathbf{q}^{IC}_{i\tilde{v}}, \mathbf{q}^{FC}_{f\tilde{t}}, \mathbf{y}^{IC}_{i\tilde{v}}$ and $\mathbf{y}^{FC}_{f\tilde{t}}$), which are now represented as decision variables in the model (6a)–(6j).

### 2.3. Evaluating the quality of the RO solution

In addition to the direct application of the CR-ESOM formulation described above, we also examine the quality of the robust solution relative to the naive baseline solution by employing Monte Carlo simulation. This approach has been widely used for analyzing the solution quality in the context of SLPs [34–37], however, we are the first to apply solution quality analysis to an ESOM employing RO. Our goal is to compare the quality of the robust solution with the naïve solution under potential future realizations of uncertain parameters.

To investigate the solution quality of the CR-ESOM, the model analysis proceeds as follows. First, we solve the CR-ESOM formulation twice under two budgets of uncertainty: $\Gamma = 0\%$, which represents the naive scenario in which no uncertain parameters take on their worst case value, and $\Gamma = 7\%$, which represents a robust scenario where 7% of the uncertain parameters take on their worst-case value. We chose this budget of uncertainty because the probability that more than 7% of the uncertain parameters take on their worst-case value is negligible, as shown in Fig. 4.

Second, the technology-specific capacity decisions from both the naive and robust scenarios are passed to a standard, non-RO version of the model, which is used within a Monte Carlo simulation in which the values of the uncertain parameters are varied based on the Latin Hypercube Sampling (LHS). The baseline and worst-case values are used to form a uniform range for each parameter, which is used to make the random draws during the Monte Carlo simulation. By drawing from uniform distributions rather than simply assuming worst-case values, we make the assumption that realized parameter values will likely fall within the defined range instead of taking on only extreme values. For each Monte Carlo iteration, the model must use the fixed capacity values from the previous step and take recourse action by optimizing the activity variables given the realized parameter values.

## 3. Case study

This paper applies the methodology described in Section 2 to a representation of the US energy system. The US database draws data from the US EPA MARKAL (USEPA9r) database [38], NREL [39], and the EIA Annual Energy Outlook [40] and includes representation of the residential, commercial, transportation, industrial, and electric sectors. The model time horizon spans from 2017 to 2050, with 5-year time periods beginning in 2020. To represent seasonal and diurnal variations in energy supply and demand, the model must perform an energy commodity balance across a set of time slices that represent different combinations of seasons and times of day. In the input database used in this analysis, we represent three seasons (summer, winter, intermediate) and four times of day: morning (6AM–12PM), afternoon (12PM–3PM), evening (3PM–9PM), and night (9PM–6AM).

The end-use sectors (residential, commercial, transportation, and industrial) include demand technologies that convert secondary energy carriers (e.g., electricity, natural gas, liquid fuels) into useful energy services (e.g., space heating, space cooling, vehicle miles traveled). These energy service demands are specified exogenously and are drawn from the USEPA9r database. For example, the residential sector includes demands for space heating, space cooling, water heating, freezing, refrigeration, lighting, and miscellaneous electricity for appliances.

Data on existing capacities of technologies in the residential, commercial and transportation sectors as well as their techno-economic parameters is drawn from the USEPA9r [38]. These parameters include overnight investment costs, conversion efficiencies and technology lifetimes. Existing capacity data on electric sector technologies is drawn from [40]. We develop our own simplified representation of the industrial sector, which explicitly represents process heat and combined heat and power and handles the remaining industrial demand through fuel share constraints. Finally, instead of modeling extraction and transport

**Table 1**
Sectoral-level detail in the Temoa input database.

| Sector | Description |
|---|---|
| Fuel Supply | Fuel costs are specified exogenously. Baseline projections are drawn from the [40]. There is no limit on fuel availability except for biofuel use in the transportation sector [42]. In addition, the database includes a representation of power-to-X. Temoa can use electricity to produce hydrogen, which can be used directly or converted to methanol and synthetic natural gas. Cost and performance characteristics are drawn from various sources which are detailed in [43]. |
| Electric | The electric sector includes 34 generating technologies. Air pollution control retrofits for coal include low NOx burners, selective catalytic reduction, selective non-catalytic reduction, and flue gas desulfurization. Costs and performance characteristics are largely drawn from the NREL Annual Technology Baseline (ATB) report [39] and EPA US MARKAL database [38]; existing capacity estimates are drawn from the [40]. |
| Transportation | The transportation sector is divided into four modes: road, rail, air, and water. Road transport is modeled with greater detail by dividing it into three subsectors: light duty transportation, heavy duty transportation, and off-highway transportation. The light duty sector includes 6 size classes and 9 different vehicle technologies. Data is largely drawn from USEPA9r [38]. |
| Industrial | Given the high degree of heterogeneity in the industrial sector, it is modeled with a simplified representation. Process heat and CHP (combined heat and power) are modeled explicitly while the other industrial end uses are modeled as a set of fuel share constraints that are calibrated to the [40]. |
| Commercial | The commercial sector includes the following end-use demands: space heating, space cooling, water heating, refrigeration, lighting, cooking, and ventilation. A total of 83 demand technologies are included to meet these end-use demands. Data is largely drawn from USEPA9r [38]. |
| Residential | The residential sector includes the following end-use demands: space heating, space cooling, water heating, freezing, refrigeration, lighting, cooking, and other appliances. A total of 69 demand technologies are included to meet these end-use demands. Data is largely drawn from USEPA9r [38]. |

of fossil fuels such as natural gas, coal and liquid fuels, fuel costs are specified exogenously and are taken from [40]. A brief sectoral description of the input dataset is provided in Table 1. The database itself is publicly available for testing and verification on Zenodo [41].

We created an 'embarrassingly parallel' implementation of the CR-ESOM framework to minimize the computational time [44]. In an embarrassingly parallel implementation, multiple instances of a problem are solved simultaneously on separate computer cores, and the instances have no co-dependency. We use the "joblib" library in python to parallelize all the iterative runs of the model. We run the model using a workstation containing two Multi-Core Intel Xeon E5-2600 series processors, representing a total of 24 compute cores. The resulting linear model is solved using CPLEX. The computational time to solve the deterministic model is 8–9 min while solving the CR-ESOM formulation takes 20–22 min for a given budget of uncertainty.

## 4. Building the uncertainty set

The CR-ESOM formulation requires bounds on the uncertain input parameters and their auto-correlation. The existing literature applying RO to ESOMs ignores this auto-correlation, which diminishes the effect of fuel cost uncertainty by treating the realized fuel cost in each time period as an independent quantity. Though beyond the scope of this analysis, the same approach can be extended to include the correlation between two different parameters; for example, between natural gas and coal prices through time. For simplicity, we categorize the uncertain parameters into two groups: fuel cost and investment cost. We then determine the uncertainty bounds and the auto-correlation coefficients for each group based on various sources of data. The following sections detail the process for grouping the parameters and determining their uncertainty bounds.

### 4.1. Grouping of the parameters

In total, the model has 6415 uncertain cost parameters, which includes investment costs, fixed O&M, variable O&M, and fuel costs. Eshraghi et al. [5] conducted a Monte Carlo analysis using the same model and database and showed that investment cost and fuel cost parameters have the highest impact on the total system cost. As a result, uncertainty in the O&M costs are not considered in this analysis. After excluding the fixed and variable O&M cost parameters, setting bounds on the remaining 2252 uncertain parameters is still a time-consuming task. As a trade-off between time and accuracy, we divide the parameters into 20 subgroups based on similarities in the underlying technology, as shown in Table 2. For example, rooftop solar PV and centralized solar PV, though distinct technologies, are placed in the same subgroup. Similarly, the uncertainty in the cost of electric vehicles arise from the uncertainty in battery technology. As a result, different types of electric vehicles are grouped together to form one category called 'electric vehicles'.

### 4.2. Uncertainty characterization

Fig. 2 provides a general flowchart to quantify the range and auto-correlation for the uncertain input parameters, and involves the application of six criteria. Each criterion corresponds to a different data source for quantifying the uncertainty range of the parameter. If the uncertain parameter satisfies the given criterion, then the corresponding data is collected for the computation of correlation coefficient and uncertainty range. Then we calculate yearly standard deviation and the correlation coefficient for each category considering a five-year time lag. The highest value associated with each fuel cost and capital cost is considered to be its worst-case value in each model time period. Note that a correlation coefficient is applied to all technologies within a category, however, the worst-case values are specific to each technology represented in the model. For example, the correlation coefficient for all petroleum-based fuels is the same for all fuels in the 'oil' category, such as diesel, gasoline, and LPG. However, worst-case values for each individual fuel type and sector combination are obtained from the 2020 Annual Energy Outlook.

#### 4.2.1. Criteria for investigating uncertainty

The six criteria shown in Fig. 2 are applied to each selected uncertain parameter in the input database. To choose the criterion for an

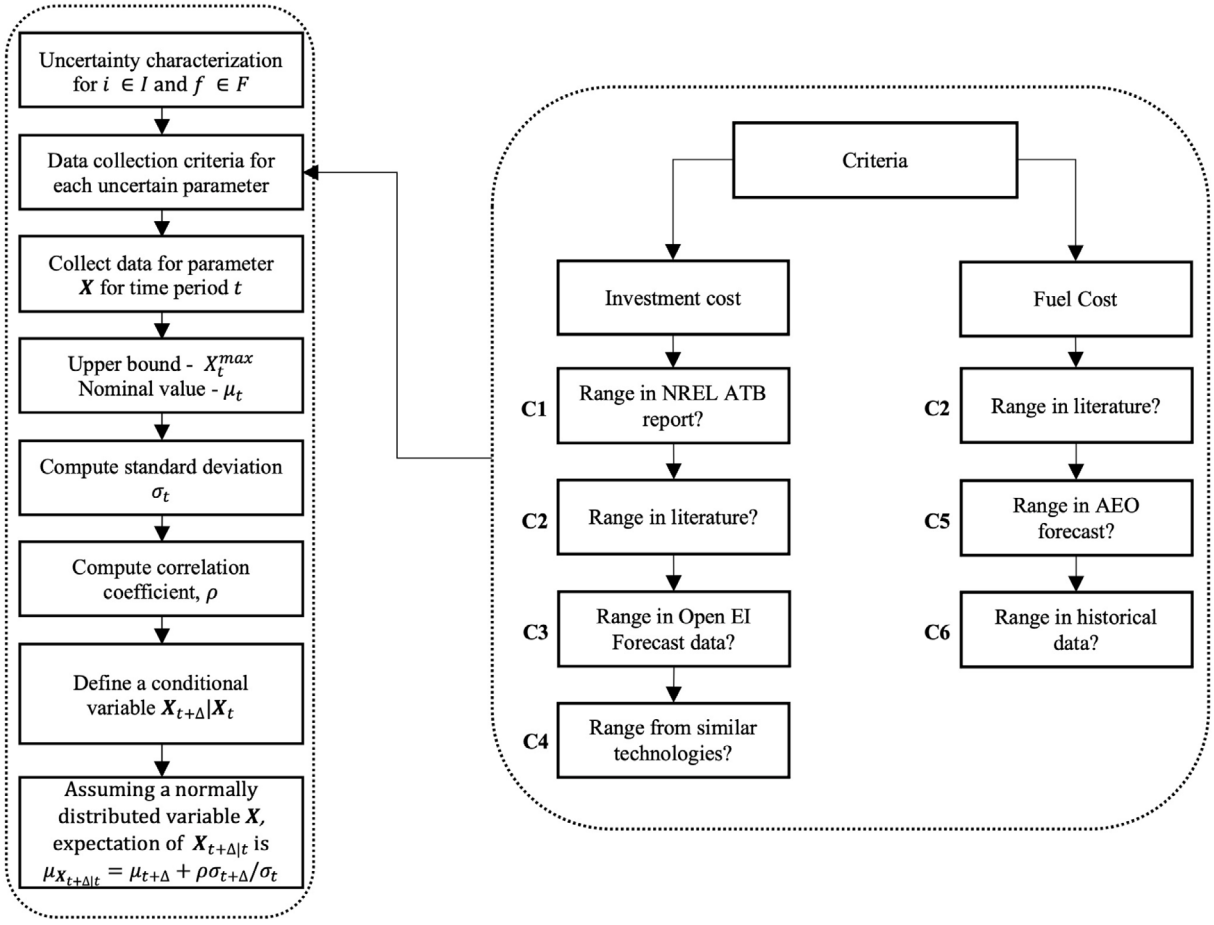


**Fig. 2.** Process to characterize the upper bound (worst-case) values and auto-correlation. The box on the right outlines the six criteria used to select uncertain parameter values in the model. Each criterion corresponds to a different method for quantifying the parameter uncertainty and the correlation coefficient. Note that the baseline values along with the upper bound values define the uniform ranges used in the Monte Carlo simulation.

uncertain parameter, we use an 'if-else' strategy. For example, if the C1 criterion is not satisfied, then we move to criteria C2. To characterize the uncertainty associated with the investment cost of relatively new technologies, we use criteria C1–C4. Criteria C2, C5 and C6 are used to characterize the uncertainty in fuel cost. Details on the six criteria are given below.

C1: Can we obtain the range from NREL Annual Technology Baseline (ATB)? The ATB report provides the investment costs for electricity generators for three scenarios — low, medium and high. We treat the 'high' scenario as the upper bound for the uncertain investment cost parameters that fall in this category.

C2: Is the range proposed in the literature? In some cases, uncertainty ranges for key parameters are provided in the literature. For example, [30] reports the uncertainty bound for biomass supply. If the range in the literature is not for the United States, then we look at the uncertainty ranges defined for other countries.

C3: Can we obtain a range from Open Energy Information (EI) data? OpenEI is a platform created and managed by NREL [45] that collects energy-related data from various national labs, agencies, journal papers, and experts. They are a reliable platform for gathering power plant and transportation sector investment cost data from various sources. We use the data published after 2010 since OpenEI has very few studies that provide data prior to 2010. Not all the studies provide data for all our model time periods, which extend from 2017 to 2050. Hence, we collect the worst-case values, i.e., the highest value of the parameter projection given in the literature for 2020 and 2050. We assume that the 2017 investment and fuel costs are certain since it is a historic year. We linearly interpolate the upper bound for the uncertain investment cost parameter from 2020 and 2050 values for the remaining time periods.

C4: What if the data for the parameter is not readily available? In this case, we divide these parameters into two groups: rapidly advancing technologies and mature technologies. The majority of the residential, commercial and industrial sector technologies have been well developed over several decades. As a result, we assume that all the technologies in these sectors fall under the 'mature' technology category. Technologies related to hydrogen and biofuels are still largely in the research and development phase. Therefore, they are categorized as 'rapidly advancing' technologies. To determine the worst-case value of the investment cost parameter for mature technologies, we follow the methodology proposed in [30] by choosing a representative technology based on data availability. The uncertainty bounds for all the mature technologies are then set based on the uncertainty bound of the representative technology. For the US database, natural gas boiler technology is chosen as a representative technology in the mature category. From the available data for natural gas boilers, we find that the worst-case value is 21.6% higher than its nominal value. Hence, the upper bound for all mature technologies is set at 21.6% above their nominal values. We apply a similar methodology for the 'rapidly advancing' technology category, with solar PV used as the representative technology.

**Table 2**
Application of uncertainty characterization method to the input database. We divide 2252 parameters into 20 categories.

| Category | Criteria | | | | | |
|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | C4 | C5 | C6 |
| | Investment cost | | | | | |
| Coal power plant | ✓ | | | | | |
| Natural gas power plant | ✓ | | | | | |
| Nuclear power plant | ✓ | | | | | |
| Solar photovoltaic | ✓ | | | | | |
| Onshore wind turbine | ✓ | | | | | |
| Li-ion battery | ✓ | | | | | |
| Bio-IGCC | | ✓ | | | | |
| Electric vehicles | ✓ | | | | | |
| Diesel vehicles | | | ✓ | | | |
| Diesel hybrid vehicles | | | ✓ | | | |
| Gasoline vehicles | | | ✓ | | | |
| Gasoline hybrid vehicles | | | ✓ | | | |
| Plug-in hybrid vehicles | | | ✓ | | | |
| Rapidly advancing technologies | | | | ✓ | | |
| Mature technologies | | | | ✓ | | |
| | Fuel cost | | | | | |
| Natural gas | | | | | ✓ | |
| Coal | | | | | ✓ | |
| Oil | | | | | ✓ | |
| Biomass | | ✓ | | | | |
| Nuclear | | | | | | ✓ |

**Table 3**
Auto-correlation for the uncertain fuel cost and technology cost parameters.

| Category | Type | Auto correlation coefficient |
|---|---|---|
| Coal power plant | combined cycle | 0.334 |
| | combined cycle with CCS | 0.547 |
| Natural gas power plant | combined cycle | 0.135 |
| | combined cycle with CCS | 0.644 |
| | combustion turbine | 0.527 |
| Nuclear power plant | – | 0.597 |
| Solar | Solar photovoltaic | 0.534 |
| | Solar thermal | 0.809 |
| Onshore wind turbine | – | 0.309 |
| Li-ion battery | – | 0.996 |
| Bio-IGCC | – | 0.752 |
| Electric vehicles | – | 0.691 |
| Diesel vehicles | – | 0.174 |
| Diesel hybrid vehicles | – | 0.606 |
| Gasoline vehicles | – | 0.597 |
| Gasoline hybrid vehicles | – | 0.108 |
| Plug-in hybrid vehicles | – | 0.233 |
| Rapidly advancing technologies | – | 0.534 |
| Mature technologies | – | 1 |
| Natural gas | – | 0.5798 |
| Coal | – | 0.3857 |
| Oil | – | 0.3015 |
| Biomass | – | 0.5868 |
| Nuclear | – | 0.6701 |

C5: Is the range available in the EIA Annual Energy Outlook (AEO)? We use this category for determining the fuel cost uncertainty. AEO (2020) provides the fuel costs for eight different scenarios. The scenario that gives the highest price for a given fuel across all time periods is used as the worst-case scenario. This scenario provides the upper bound for future fuel costs.

C6: Can we determine the range from the historical data? If the AEO does not provide the upper bound for fuel cost, then the historical upper bound is set as the worst-case value of the fuel. For the US database, only uranium falls in this category. The upper bound for this fuel remains the same through the time horizon.

The results obtained from the above analysis are summarized in Table 2.

#### 4.2.2. Correlation coefficient

After categorizing the uncertain fuel cost and investment cost parameters, we determine the correlation coefficient $\rho$ for each category. First, to compute the correlation coefficient for fuel costs, we use historical fuel cost data from [40] and compute auto-correlation with a 5-year lag to match the length of future model time periods. Next, we consider the uncertainty in investment cost. Similar to the correlation in fuel costs, the investment cost of technology is also temporally correlated. For example, the investment cost of solar PV has been decreasing for the last decade. As a result, it will likely decrease for the next several years although not at the same rate. To compute the auto-correlation coefficient for the technology investment cost, we use historical data collected by OpenEI [45]. However, for some technologies, the historical data collected by OpenEI is not sufficient to compute correlation. Consequently, we also include future projection data through 2050 to compute the auto-correlation coefficients. For simplicity, we assume that the temporal correlation for mature technologies is 100%, such that when a mature technology takes on its worst-case value in a particular time period, it retains that worst-case value for all preceding and subsequent time periods. Since auto-correlation for mature technologies is 100%, the parameter takes on the same value for all other time periods once uncertainty is resolved for one period. Note that as a relatively mature technology, the cost of diesel vehicles fluctuates over time depending on the share of different types and makes of the vehicle technology. Data for the projections of the manufacturing cost of new diesel vehicles demonstrate a lack of a clear upward or downward trend, resulting in low auto-correlation of diesel vehicles. In contrast, lithium-ion batteries have experienced a significant cost decline over the last decade, resulting in very high auto-correlation. Auto-correlation for the uncertain fuel cost and technology cost parameters is given in Table 3. For coal, natural gas and solar categories, the coefficient is reported by technology type.

#### 4.2.3. Determining the effects of correlation on the uncertainty bounds

Using the correlation coefficient and the bounds on uncertain parameters, in this section, we determine the impact of correlation on the uncertainty bounds. We demonstrate the methodology for uncertain fuel costs, and the same methodology is extended for uncertain investment cost parameters. To do so, we assume that the price of fuel $f$ is a normally distributed random variable, $\mathbf{X}$ with mean, $\mu$. We assume that the mean value of fuel $f$ in time period $t$ is equal to the reference case scenario given by the [40]. The AEO scenario with the highest fuel cost is considered the worst-case scenario for each fuel cost parameter, $X_t^{max}$.

Due to the auto-correlation for fuel cost parameters, $\mathbf{X}_t$ is correlated with $\mathbf{X}_{t+\Delta}$, where $\Delta$ is a difference between adjacent time-periods. The correlation coefficient $\rho$ is given in Table 3. Assume that both $\mathbf{X}_t$ and $\mathbf{X}_{t+\Delta}$ are normally distributed random variables. As a result, $\mathbf{X}_{t+\Delta|t} = \mathbf{X}_{t+\Delta}|\mathbf{X}_t$ is also a normally distributed random variable. In this study, we want to quantify the impact of worst-case fuel costs in time period $t$ on fuel costs in time period $t+\Delta$. Therefore, the mean of $\mathbf{X}_{t+\Delta|t}$, $\mu_{\mathbf{X}_{t+\Delta|t}}$, can represent the expected value of $\mathbf{X}_{t+\Delta}$ given that $\mathbf{X}_t$ assumes the worst-case value, $X_t^{max}$. The mean of $\mathbf{X}_{t+\Delta|t}$ is then given by

$$\mu_{\mathbf{X}_{t+\Delta|t}} = \mu_{t+\Delta} + \rho(\mathbf{X}_t - \mu_t)\sigma_{t+\Delta}/\sigma_t \tag{7}$$

where, $\sigma_{t+\Delta}/\sigma_t$ can be written as $(X_{t+\Delta}^{max} - \mu_{t+\Delta})/(X^{max} - \mu_t)$. Similarly, the expected value of $\mathbf{X}_{t+2\Delta}$ given the value of $\mathbf{X}_t$ is given by Eq. (8).

$$\mu_{\mathbf{X}_{t+2\Delta|t}} = \mu_{t+2\Delta} + \rho(\mu_{\mathbf{X}_{t+\Delta|t}} - \mu_{t+\Delta})\sigma_{t+2\Delta}/\sigma_{t+\Delta} \tag{8}$$

Note that $t+\Delta$ corresponds to the next time period and $t+2\Delta$ to the time period after that. The mean of the random variable $\mathbf{X}_{t+\Delta|t}$ can be given as (9).

$$\mu_{\mathbf{X}_{t+\Delta|t}} = \mu_{t+\Delta} + \rho(X_t^{max} - \mu_t)\frac{(X_{t+\Delta}^{max} - \mu_{t+\Delta})}{(X^{max} - \mu_t)} = \mu_{t+\Delta} + \rho(X_{t+\Delta}^{max} - \mu_{t+\Delta}) \tag{9}$$

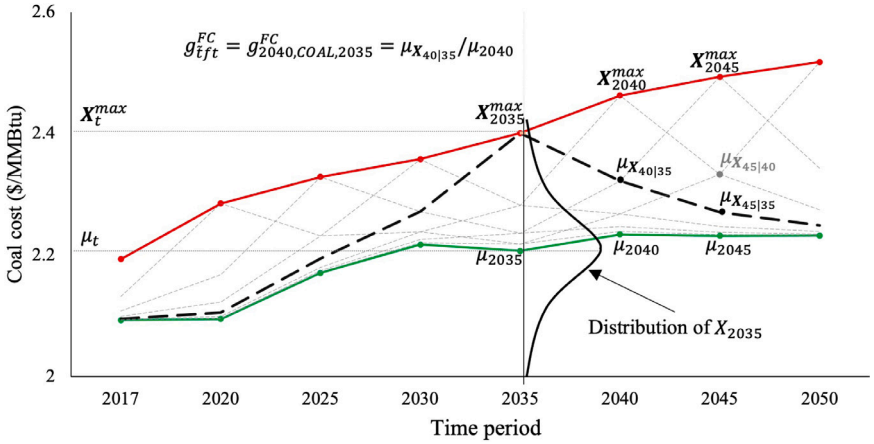


**Fig. 3.** Effect of the correlation in coal costs. The red line shows the maximum worst-case values $\mathbf{X}_t^{max}$ from the [40] while the green line shows the reference case values $\mu_t$. The dark dotted line shows a scenario where coal cost assumes maximum worst-case value in 2035, $\mathbf{X}_{2035}^{max}$. $\mu_{\mathbf{X}_{40|35}}$ and $\mu_{\mathbf{X}_{45|40}}$ denote the impact of worst-case coal cost in 2035 on the year 2040 and 2045, respectively. The lighter dotted lines show the cost trajectory when it hits the maximum worst-case value in other time periods. The equation on the top shows the computation of $g_{tft}^{FC}$ required for the CR-ESOM formulation.

The CR-ESOM formulation given in (6a)–(6j) requires the computation of $g_{tft}^{FC}$, which represents the ratio of the worst case value and nominal value of an uncertain fuel parameter for a given time period. As per the above computations, $g_{(t+\Delta)ft}^{FC}$ can be given by

$$g_{(t+\Delta)ft}^{FC} = \mu_{\mathbf{X}_{t+\Delta|t}}/\mu_{t+\Delta} \tag{10}$$

Fig. 3 shows an example of coal costs over the time horizon given that the worst case occurs in every time period. Since the difference in adjacent model time periods $\Delta = 5$, $\mu_{\mathbf{X}_{40|35}}$ represents the coal cost in 2040 given that coal cost in 2035 assumes worst-case value. Similarly, $\mu_{\mathbf{X}_{45|35}}$ represents the coal cost in 2045 given that coal cost in 2035 assumes worst-case value. If coal cost assumes worst-case value in 2035 and 2040, then the coal cost in 2045 will be maximum of $\mu_{\mathbf{X}_{45|35}}$ and $\mu_{\mathbf{X}_{45|40}}$.

## 5. Robust optimization results

We use the CR-ESOM formulation to address the basic questions outlined in Fig. 1. In Section 5.1, we develop three scenarios based on running the CR-ESOM formulation under different budgets of uncertainty. We then fix the capacity decisions from each scenario and allow the model to take recourse action by optimizing the associated activity variables under budgets of uncertainty ranging from 0% to 100%. The goal of the exercise is to observe how the system costs change under each scenario as the budget of uncertainty varies. As a reminder, the budget of uncertainty represents the percentage of uncertain parameters taking on their worst-case value. It can also be interpreted as the level of risk tolerance. When $\Gamma = 0\%$, all parameters take on their nominal values and thus, the model ignores uncertainty, which implies a high-risk decision strategy. When $\Gamma = 100\%$, all uncertain parameters take on their worst-case value, and the model returns a solution that is maximally conservative, implying a highly risk-averse decision strategy.

In Section 5.2, we perform the solution quality analysis described in Section 2.3. The Naive ($\Gamma = 0\%$) and Robust ($\Gamma = 7\%$) scenarios are subject to a Monte Carlo simulation in which the capacity variables are fixed, and the model takes recourse action under different realizations of the uncertain parameters. The goal of this analysis is to rigorously test and compare the performance of the robust and naive strategies under a wide array of conditions.

In Section 5.3, we vary the budget of uncertainty ($\Gamma$) from 0% to 100%. Unlike in the previous sections, the capacity and activity variables are solved simultaneously. The goal is to observe how the technology pathways in each sector change as the risk tolerance is adjusted.

In Section 5.4, we identify the uncertain parameters that have the largest effect on system cost under different budgets of uncertainty. Since ESOMs are computationally intensive, the goal of this exercise is to identify the uncertain parameters with the highest impact, which can be prioritized for further refinement in future modeling efforts.

### 5.1. Performance of the robust optimization strategy

Fig. 4 presents the total system cost versus the budget of uncertainty ($\Gamma$) under three different scenarios: 'Naive', 'Robust', and 'Robust Nominal'. For each scenario, the optimization is performed in two steps. In the first step, capacity deployment decisions are made under a particular scenario-specific planning strategy using the CR-ESOM formulation. In the second step, the model utilizes the fixed capacity decisions from the first step to making optimal technology utilization decisions under different budgets of uncertainty.

In the 'Naive' scenario, the model optimizes technology capacity assuming no future uncertainty ($\Gamma = 0\%$). In the 'Robust' scenario, the model optimizes technology capacity assuming $\Gamma = 7\%$. We chose this budget of uncertainty because the probability that more than 7% of the uncertain parameters take on their worst-case value becomes negligible, as shown in Fig. 4. As a result, decision-makers will have a little economic incentive to choose a solution with higher robustness.

In the second step, under both the 'Naive' and 'Robust' scenarios, the capacity decisions remain fixed and ($\Gamma$) ranges from 0% to 100%. The total system cost calculated in this second step accounts for the model's recourse actions under different realizations of uncertainty, under which different numbers and combinations of parameters take on their worst-case values. More significant recourse actions in the second step will result in higher system costs. In general, the 'Robust' scenario accounts for the possibility of worst-case parameter values in the capacity planning stage, which help limit the required recourse actions in the second step under higher budgets of uncertainty.

We also test a 'Robust Nominal' scenario where in the first step $\Gamma$ varies from 0%–100%. Then in the second step, for each $\Gamma$ value, the capacity decisions are fixed, and the model performs the optimization under nominal conditions with $\Gamma = 0\%$. Thus, results from the 'Robust Nominal' scenario provide a useful point of comparison by quantifying the added cost of hedging capacity decisions only to find that none of the parameters take on their worst-case values in the future.

As shown in Fig. 4, the naïve solution is more expensive than the robust solution with one exception: if all uncertain parameters assume

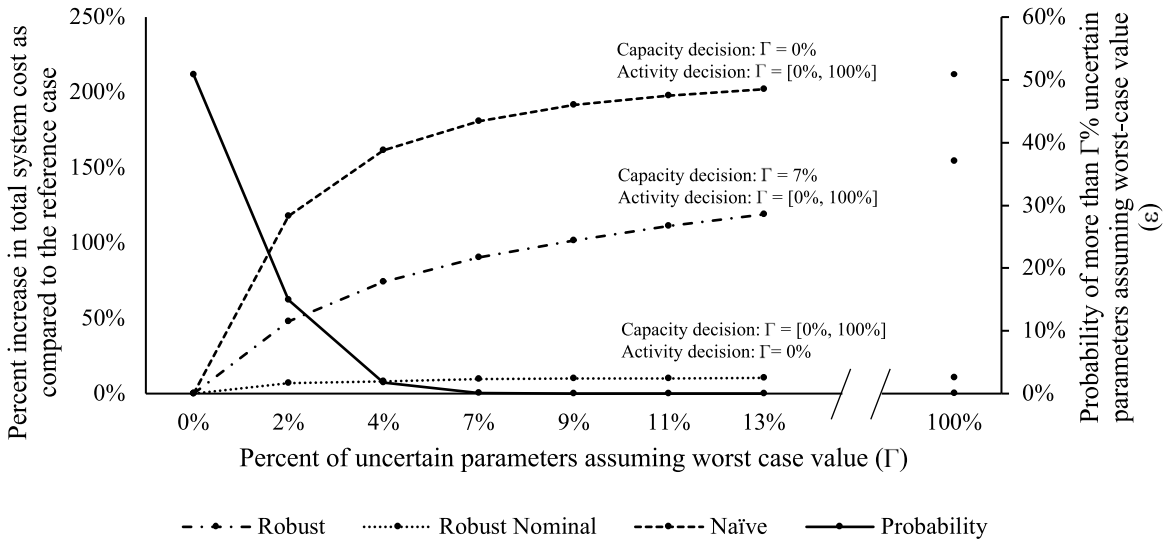


**Fig. 4.** Percent change in the total system cost as a function of the budget of uncertainty. 'Naïve' represents the cost of the naïve solution as the budget of uncertainty ($\Gamma$) increases; 'Robust' represents the same for the robust solution; 'Robust Nominal' represents the cost of the robust solution if all uncertain parameters assume their nominal value. The solid line labeled 'Probability' indicates the probability that the number of parameters taking on their worst-case value will exceed a given budget of uncertainty. The probability that more than 7% of uncertain parameters assuming their worst-case value is negligible.

their nominal value in the future, the extra expense associated with hedging under the robust strategy is not required. Since the naive solution ignores uncertainty while making capacity decisions, it must take more expensive recourse actions when some subset of uncertain parameters assume their worst-case values. Fig. 4 indicates that the risk of the robust strategy is lower than the risk of the naïve one: at a 7% budget of uncertainty, the robust solution will only cost 6% more than the naïve solution if no parameters assume their worst-case value. However, the naïve solution may cost 91% more than the robust solution if 7% of the uncertain parameters assume their worst-case value.

### 5.2. Assessing the quality of the robust solution

In reality, most uncertain parameters will not assume their worst-case value simultaneously. Instead, the realization of uncertain parameters will more likely be within the uncertainty range, defined on the lower end by the nominal value and on the upper end by the worst-case value. To account for this behavior, we analyze the quality of the robust solution following the procedure outlined in Section 2.3. We then compare the quality of the RO solution with the naïve solution under potential future realizations of uncertain parameters. To do so, we fix the capacity decisions based on the RO solution with $\Gamma$ = 7% and solve the least cost formulation (1a)–(1d) for 2000 different realizations of uncertain parameters. We follow the same procedure for the naïve solution. The spread of total system cost for the RO and naïve solution under varying realizations of uncertain parameters is shown in Fig. 5. This approach allows us to estimate the economic savings associated with following a robust versus naïve strategy. Fig. 5 suggests that the expected cost of the RO solution, $\mu_{rob}$, is 12% lower than the expected cost of the naïve solution, $\mu_{det}$, while the standard deviation of the naïve solution is 8% higher than the robust solution. Thus, the results suggest that making investment decisions that explicitly consider future uncertainty yields significant economic benefits, with a modest decrease in the potential range of system costs.

### 5.3. Robust emission mitigation pathways considering future uncertainty

This section investigates technology pathways consistent with US commitments under the Paris Agreement under different robustness assumptions. Fig. 6 shows the change in energy system activity in three representative time periods (2020, 2030, and 2050) under a gradually increasing budget of uncertainty. Each stacked bar plot for each time period contains three elements, from left to right: a single bar indicating the naive solution ($\Gamma = 0\%$), a series of 12 bars representing a range in the budget of uncertainty ($\Gamma = [2\%, 24\%]$), and a single bar indicating the worst-case scenario ($\Gamma = 100\%$). The strategy with $\Gamma = 100\%$ represents the [16] hedging approach whereby all uncertain parameters assume their worst-case value. However, varying the budget of uncertainty, as proposed by Bertsimas and Sim [19], allows us to identify the solutions along a continuum. For $\Gamma$ varying between 0% to 24%, the RO approach encourages diversification in fuels and technology to hedge against uncertainty. These intermediate solutions do not emerge in the more traditional [16] approach. For a budget of uncertainty greater than 24%, the RO solution remains relatively constant, though there are exceptions. A similar diversification in optimal results is observed for the numerical simulations presented by Bertsimas and Sim [19]. In Fig. 6, we vary the budget of uncertainty to observe the technology deployment options available to decision-makers based on their risk tolerance. Note that the industrial sector is excluded from Fig. 6 as we only represent process heating and CHP in detail while assuming a fixed fuel share for the rest of the sector. Moreover, only the variation space heating and space cooling is shown in detail for the residential and commercial sector as the rest of the sector-specific technology composition does not change significantly.

The composition of the energy system changes dramatically when we account for uncertainty. In Fig. 6(a), the naïve solution with an emission target achieves 40% electricity generation by nuclear power plants in 2050. However, the robust solution suggests diversifying the sources of electricity generation. Carbon capture and sequestration (CCS), underground hydrogen storage, and solar thermal play a more significant role even though these technologies are not a cost-effective choice in the naïve case. In other words, if we assume a rational but naive cost-minimizing decision maker, then the optimal technology mix would not include these technologies. The relative capacity of the technologies, however, depends on the decision maker's risk tolerance. Interestingly, with an increased budget of uncertainty, the investment in new CCS decreases while investment in new nuclear increases. Bertsimas and Sim [19] observe similar results whereby the diversity of the solution decreases under the highest budgets of uncertainty. The rationale behind this observation is that when $\Gamma = 100\%$, solving the RO model is equivalent to solving a deterministic model where all cost

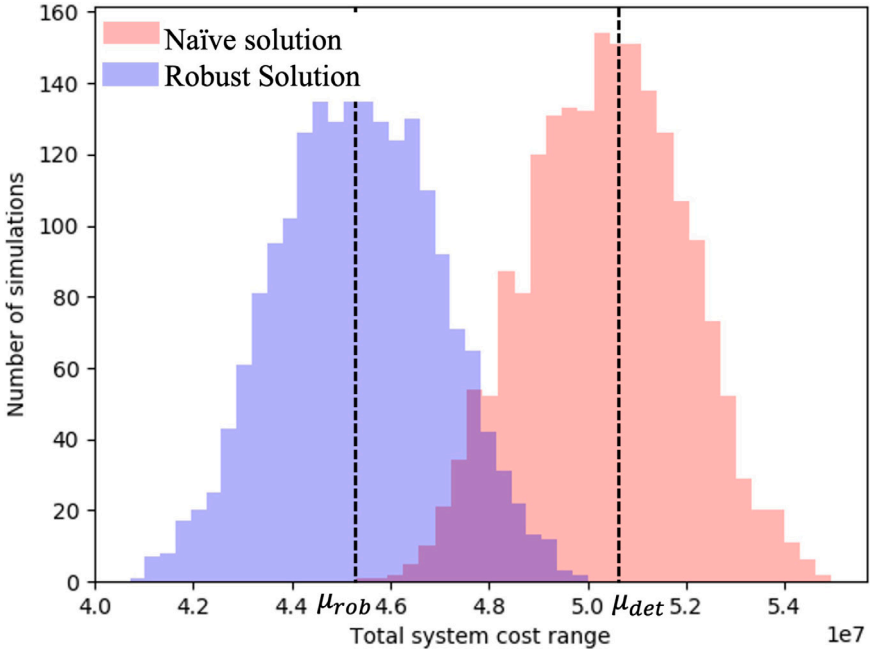


**Fig. 5.** Distribution of system cost for the naïve ($\Gamma = 0\%$) and robust solutions ($\Gamma = 7\%$) under 2000 uncertain parameter realizations produced via Monte Carlo simulation.

parameters are at their worst-case value. As a result, the model chooses the least-cost option instead of diversifying the resources. Overall electricity generation in 2050 increases by approximately 20% in a robust solution as compared to the naïve solution regardless of the budget of uncertainty. This result indicates that increasing electrification in the end-use sector enhances diversification. The RO solution at $\Gamma < 10\%$ also suggests delaying some investments in wind and solar generation until after 2030 because of high uncertainty in their investment cost. Instead, the required emission reduction is achieved using nuclear power plants and CCS. Note that this insight is driven by our selection of worst-case values. As a result, future work is needed to refine the bounds and correlation coefficients to increase the confidence in the insights. With the increase in CCS, the robust solution suggests delaying the retirement of coal power plants even though natural gas combined cycle power plants are cheaper to operate. The main reason behind this shift is that the relative variation in coal prices is lower than the variation in natural gas prices. Thus the retirement of coal is delayed until 2030. After 2030, with the reduced uncertainty in the cost of renewable generation, coal is replaced by solar PV and onshore wind turbines.

Figs. 6(b) and 6(c) show the changes in space heating and space cooling for commercial and residential sectors, respectively. The commercial sector results remain consistent across all $\Gamma$ values, with some trade-offs between natural gas boilers and furnaces. In the residential sector, there is a switch from geothermal to natural gas heat pumps as $\Gamma$ increases from 2% to 16%, and then a large scale shift to electric heat pumps in the worst-case scenario ($\Gamma = 100\%$). Overall, heat pumps appear to be a robust technology option across the full range of uncertainty. De Villiers and Matibe [46] show a similar solution and suggest mitigating emissions from the residential sector by increasing electrification.

The transportation sector, plotted in Fig. 6(d), undergoes more drastic changes when the uncertainty is considered compared to the residential and commercial sectors. The first significant shift from the naïve solution is reducing the dependency on electric cars alone and replacing them with PHEVs and hybrid vehicles. High uncertainty in electricity prices caused by highly uncertain natural gas prices affects the transportation sector's electric vehicle share. The model results indicate that the share of PHEVs and hybrid vehicles depends on the preferred level of risk tolerance.

As mentioned above, the diversity of the resources in the robust solution tends to decrease at the highest budgets of uncertainty. Consequently, the share of technologies within each sector depends on the budget of uncertainty selected. The naïve solution suggests electrification of vehicle transport; however, more robust strategies prioritize the electrification of process heating in the industrial sector over investment in electric vehicles. Since the process heating technologies in the industrial sector have conservative worst-case values, the model tends to choose them over electric cars, which have a wider potential cost range. This observation illustrates a broader point: the results are contingent on how the worst-case values are calibrated.

### 5.4. Importance of a parameter in achieving a robust solution

For a given budget of uncertainty, the CR-ESOM formulation determines the uncertain parameters, among all 2252 uncertain parameters considered for this analysis, that produce the highest impact on the total system cost. Identifying the most impactful parameters helps to prioritize data development associated with technologies and parameters that play a key role in the model. To do so, we sort the shadow prices corresponding to the feasible domain given in Eq. (5), in decreasing order. Specifically, we concatenate optimal values of the variables $\mathbf{y}^{IC}_{i\tilde{v}}$ and $\mathbf{y}^{FC}_{f\tilde{t}}$ and put them in descending order, since a higher value of $\mathbf{y}$ leads to a higher impact of the uncertain parameter on the cost function. We plot the parameters corresponding to the highest $\mathbf{y}^{IC}_{i\tilde{v}}$ and $\mathbf{y}^{FC}_{f\tilde{t}}$ in Fig. 7. As we move from top to bottom, the importance of the parameter decreases. As we move from left to right, the robustness of the solution increases. For example, for $\Gamma = 1$, an investment cost of light-duty vehicles influences the objective function the most (top-left corner). For $\Gamma = 50$, solar PV investment cost becomes the most influential parameter (top-right corner). As we move down a column in the triangle, corresponding to a given budget of uncertainty, the importance of the parameter in determining the robust solution decreases. For example, at $\Gamma = 50$, the investment cost uncertainty

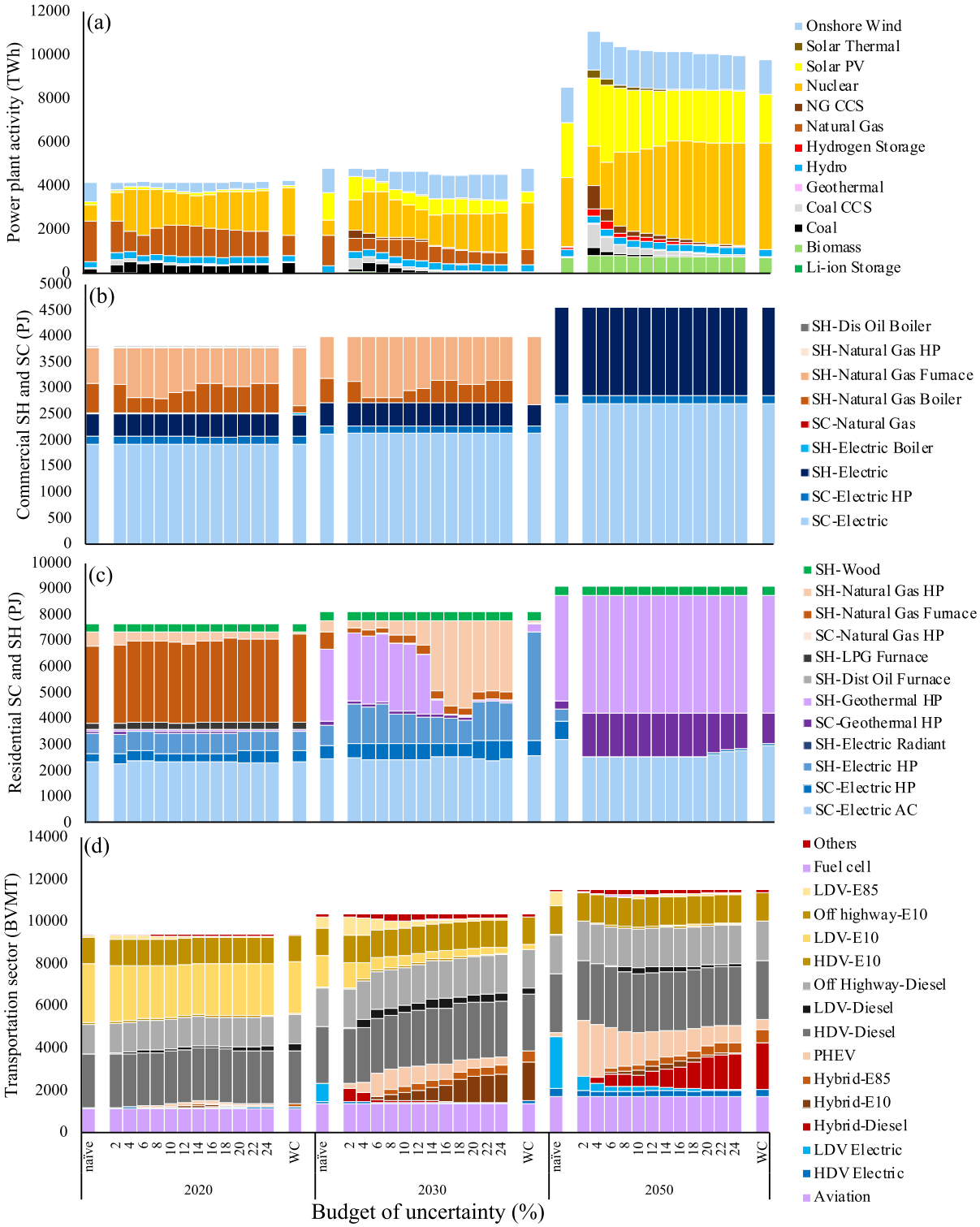


**Fig. 6.** Activity of technologies in different sectors as a function of the budget of uncertainty and three different time periods (2020, 2030, 2050): (a) electricity generation, (b) commercial space cooling ('SC') and heating ('SH'), (c) residential space cooling and space heating, and (d) transportation. The 'WC' stacked bars represent the technology activity in a given sector and time period under the worst-case outcome ($\Gamma = 100\%$), and 'naïve' represents the reference scenario where uncertainty is ignored ($\Gamma = 0\%$).

related to solar water heating in the commercial sector (far right-bottom corner) is less influential in determining the robust solution than the investment cost of solar PV (top-far right). It is important to note that the RO methodology tries to minimize the product of the worst-case value and optimal capacity of a technology and thus the ordering of the most influential parameters can change. When $\Gamma$ is lower than 30, this product for solar PV and bioenergy with CCS (BECCS) is not large enough, as their optimal capacity is small. As we increase $\Gamma$, the optimal capacity of these technologies increases and hence, they show up as more important in Fig. 7.

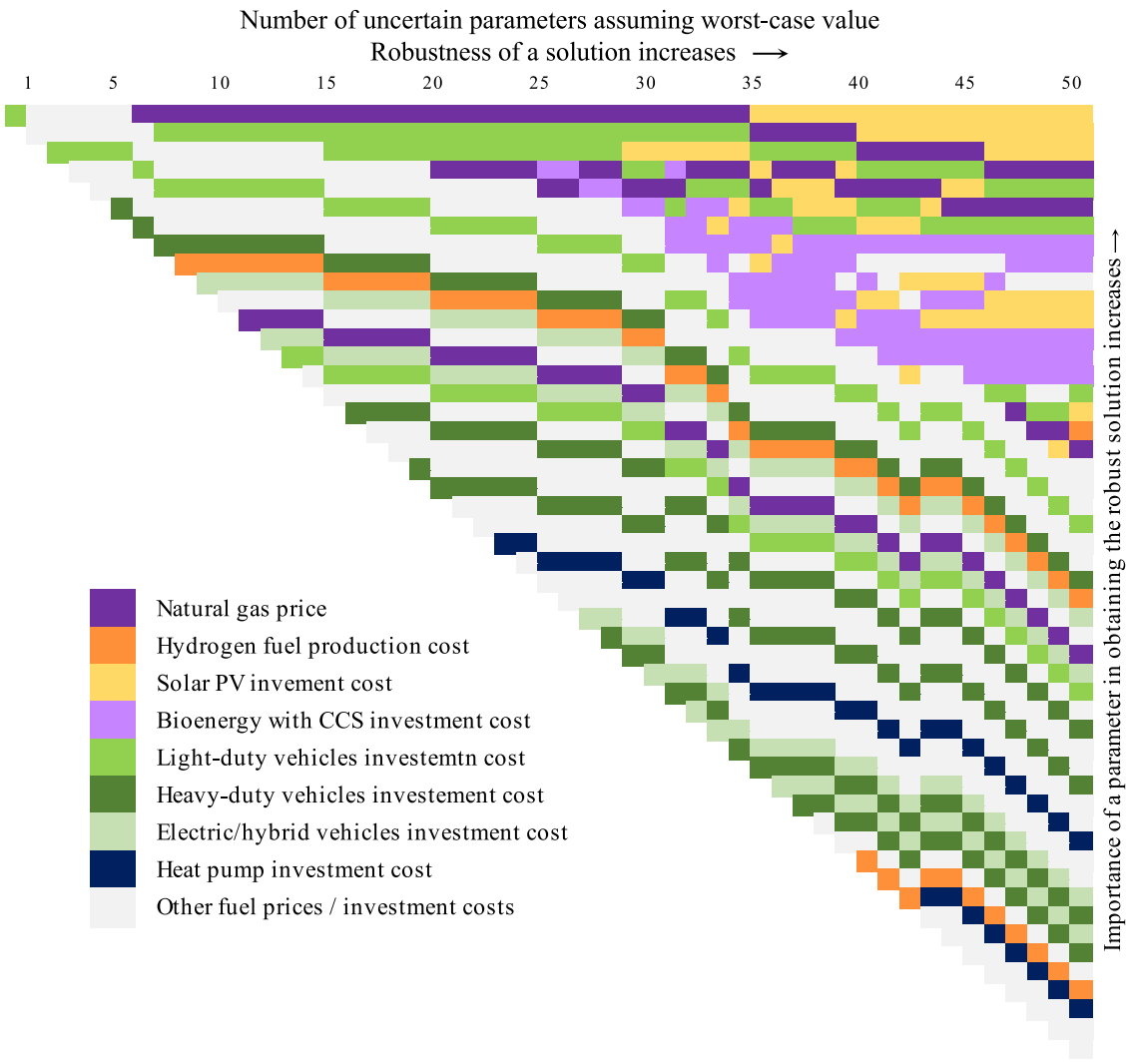


**Fig. 7.** Identification of the most sensitive parameters as a function of the budget of uncertainty. The importance of a parameter increases as we move from bottom to top and the robustness of a solution increases as we move from left to right. The top right corner shows the most impactful parameter, investment cost of solar PV, when budget of uncertainty is 50, i.e., 50 uncertain parameters assume their worst-case value.

This result also suggests that resolving the uncertainty in solar PV investment cost would assist in achieving a robust strategy. Furthermore, if the uncertainty is irreducible through improved data analysis, then the parameter range and auto-correlation should be well-calibrated. Other critical costs are the natural gas fuel cost, investment cost of light duty vehicle technologies, especially PHEV, investment cost of coal with CCS, BECCS, nuclear, hydrogen electrolysis, and synthetic natural gas production. Since the goal of the analysis is to achieve greenhouse gas emissions reductions, the low carbon technologies in the electric sector tend to have a higher impact on the objective function than the technologies with GHG emissions.

## 6. Conclusions

Highly uncertain input data is a reality of using ESOMs. Since the insights obtained from such models are often used to inform policy, the uncertainties should be explicitly considered to build robust strategies. The main goal of this paper is to implement robust optimization in an ESOM in order to explore deep decarbonization pathways that are robust to uncertainties in future fuel costs and investment costs.

Here we introduce a robust optimization (CR-ESOM) framework that is the first to consider the auto-correlation in uncertain parameters. Because the technology and fuel specific input cost parameters are indexed by time period, it is important to consider how the assumption of a worst-case value achieved in a particular time period affects the value of that parameter in other time periods. Ignoring auto-correlation can diminish the effect of any single worst-case value realized in a single time period. On the other hand, auto-correlation does add additional data requirements and complexity to the formulation. Future applications should weight the costs and benefits of including auto-correlation, taking into account the nature of the problem being addressed.

The results indicate that pursuing a robust strategy has significant monetary benefits relative to pursuing a naive strategy that ignores future uncertainty. The degree of monetary benefits from the robust strategy depends on how uncertainty is resolved in the real world. Our analysis indicates that more than 7% of the input parameters assuming their worst-case value has a negligible probability. At such a budget of uncertainty, the robust pathway entails diversification of both fuel supply and technologies deployed. US climate policy should focus on ways to diversify fuel and technology pathways across the energy system, in order to make it more robust to future uncertainty.

Several open issues deserve further investigation. First, the process of estimating auto-correlation associated with uncertain parameters needs improvement. For some parameters, data for computing auto-correlation is limited. For example, to compute the correlation coefficient for industrial coal prices, we use the weighted average of the other types of coal. For future work, sensitivity analysis focused on the correlation coefficients is needed to ensure the robustness of the strategy.

Second, the formulation in this paper only considers auto-correlation for uncertain parameters but does not consider the correlation between different uncertain parameters, which exist in the real world. Such a formulation can be written as a natural extension of model (6a)–(6j). Third, we assume the uncertain parameters are a normally distributed random variable when estimating auto-correlation across time periods. However, the CR-ESOM formulation itself ignores information about the parameter distributions and merely works with the nominal value and parameter deviation. Fourth, the data and methodology used to determine the ranges for uncertain parameters needs to be refined over time. Many technologies lack a detailed cost range, and under a conservative assumption about their worst-case value, can be over emphasized by the RO model. A critical focus of this work needs to be an extended effort to develop an improved representation of uncertain input parameters. Section 5.4 outlines a method to target the most sensitive input parameters in the model. Fig. 7 can serve as a guide to prioritize future data collection efforts. Given the data intensive nature of ESOMs generally and this RO method specifically, it would be beneficial to make this data gathering a community effort, with contributions and critical reviews provided by a broad array of modelers and analysts.

Despite the daunting task of quantifying the uncertainty for solving the robust optimization model, it can provide valuable insights to policymakers who must act before uncertainty is resolved. While further research is needed to improve the construction of the uncertainty set and the underlying data, the utility and computational tractability of RO can be an essential tool for addressing parametric uncertainty in ESOMs. While the application of robust optimization yields a hedging strategy, this analysis should nonetheless be viewed as an exercise to explore the decision space when the parametric uncertainty is explicitly considered. In addition, we emphasize that models alone cannot provide a solution in such complex decision landscapes but can yield insight that informs decision making.

## CRediT authorship contribution statement

**Neha Patankar:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Original draft, Writing – review & editing. **Hadi Eshraghi:** Data curation, Methodology, Visualization. **Anderson Rodrigo de Queiroz:** Conceptualization, Methodology, Writing – review & editing. **Joseph F. DeCarolis:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Software, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix. Uncorrelated robust optimization formulation

In order to implement the RO formulation proposed by Bertsimas and Tsitsiklis [33], we need to rewrite the objective function of ESOM in the form of a constraint as given below in Eqs. (A.1a)–(A.1e).

$$\min \quad \mathbf{W} \tag{A.1a}$$

$$\text{s.t.} \quad \mathbf{W} - \Bigg( \sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} + \sum_{t\in T_f}\sum_{f\in F} FC_{ft}\mathbf{CON}_{ft} \Bigg) \geq 0 \tag{A.1b}$$

$$\sum_{v\in V_i}\sum_{i\in I} \mathbf{ACT}_{ivt} \geq D_t \qquad \forall t \in T \tag{A.1c}$$

$$\sum_{V_i \leq T_i} \gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t \in T_i, v \in V_i, i \in I \tag{A.1d}$$

$$B\mathbf{x} \geq b \tag{A.1e}$$

where, $\mathbf{W}$ is a dummy variable introduced to write the model's objective function in the form of a constraint. Since we assume uncertainty in fuel cost and technology investment cost, only constraint (A.1b) contains uncertain parameters when attempting to solve an extension of model (A.1a) under uncertainty.

In the R-ESOM formulation, we assume that the uncertain ESOM parameters are independently random. We provide a formulation for the uncertain investment cost and fuel costs; however, the formulation can be extended to consider the uncertainty in fixed operations and maintenance cost. Let $I$ be the set of technologies with uncertain investment cost and $F$ be the set of fuels with uncertain prices. The elements of the vector, $IC_{iv}$ i.e., investment cost of a technology $i$ with vintage $v$, for $i \in I$ and $FC_{ft}$ i.e., cost of fuel $f$ in time period $t$ for $f \in F$, are assumed to be subjected to uncertainty. Let the uncertain realization associated with investment cost and fuel cost be denoted as $IC^*_{iv}$ and $FC^*_{ft}$, respectively. $IC^*_{iv}$ belongs to a symmetrical interval $[IC_{iv} - \Delta IC_{iv}, IC_{iv} + \Delta IC_{iv}]$ defined by the modeler, where $IC_{iv}$ and $\Delta IC_{iv}$ are used to represent nominal values and deviation magnitudes, respectively. This interval is centered at the point forecast $IC_{iv}$, while $\Delta IC_{iv}$ measures the range of the estimate. Similarly, $FC^*_{ft}$ belongs to a symmetrical interval $[FC_{ft} - \Delta FC_{ft}, FC_{ft} + \Delta FC_{ft}]$ defined by the modeler. The scaled deviation $\mathbf{s}^{IC}_{iv}$ of uncertain investment cost $IC^*_{iv}$ and $\mathbf{s}^{FC}_{ft}$ of uncertain fuel cost $FC^*_{ft}$ from its nominal value can then be defined as (A.3) and (A.3), respectively.

$$\mathbf{s}^{IC}_{iv} = \frac{IC^*_{iv} - IC_{iv}}{\Delta IC_{iv}} \qquad \forall i \in I, v \in V_i \tag{A.2}$$

$$\mathbf{s}^{FC}_{ft} = \frac{FC^*_{ft} - FC_{ft}}{\Delta FC_{ft}} \qquad \forall f \in F, t \in T_f \tag{A.3}$$

The scaled deviation $\mathbf{s}^{IC}_{iv}$ belongs to $[-1, 1]$ as $(IC^*_{iv} - IC_{iv})$ varies within $[-\Delta IC_{iv}, \Delta IC_{iv}]$. Similarly, $\mathbf{s}^{FC}_{ft}$ belongs to $[-1, 1]$. The aggregate scaled deviation for the constraint with uncertain parameters, $\sum_{v\in V_i}\sum_{i\in I} |\mathbf{s}^{IC}_{iv}| + \sum_{t\in T_f}\sum_{f\in F} |\mathbf{s}^{FC}_{ft}|$ can take any value between 0 and $N$, where $N$ is the total number of uncertain parameters. However, it is unlikely that all of the coefficients take their worst cases simultaneously. Consequently, the true value of $\sum_{v\in V_i}\sum_{i\in I} |\mathbf{s}^{IC}_{iv}| + \sum_{t\in T_f}\sum_{f\in F} |\mathbf{s}^{FC}_{ft}|$ can be assumed to be in a narrower range as given in (A.4), i.e.,

$$\sum_{v\in V_i}\sum_{i\in I} |\mathbf{s}^{IC}_{iv}| + \sum_{t\in T_f}\sum_{f\in F} |\mathbf{s}^{FC}_{ft}| \leq \Gamma \tag{A.4}$$

where $\Gamma \in [0, N]$, referred to as the budget of uncertainty of the constraint containing uncertain parameters, is used to adjust the level of conservatism of the solution. Its value reflects the decision maker's tolerance for risk. Thus, $\Gamma = 0$ returns the "naive" solution with no "protection" against uncertainty and $\Gamma = N$ yields a very conservative solution since it represents all uncertain parameters taking their worst-case values at the same time. For any values between 0 and $N$, the decision maker makes a trade-off between the robustness of the solution and the related cost. Based on [19], the general ESOM model (A.1a)–(A.1e), and on Eqs. (A.2)–(A.4), the R-ESOM formulation can be written as (A.5a)–(A.6).

$$\min \quad \mathbf{W} \tag{A.5a}$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{W} - \Bigg( \sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} + \sum_{t\in T_f}\sum_{f\in F} FC_{ft}\mathbf{CON}_{ft} \Bigg) \\ & - \max_{\mathbf{s}^{IC}_{iv}, \mathbf{s}^{FC}_{ft} \in S} \Bigg( \sum_{v\in V_i}\sum_{i\in I} \Delta IC_{iv} CAP_{iv}\mathbf{s}^{IC}_{iv} + \sum_{t\in T_f}\sum_{f\in F} \Delta FC_{ft} CON_{ft}\mathbf{s}^{FC}_{ft} \Bigg) \geq 0 \end{aligned} \tag{A.5b}$$

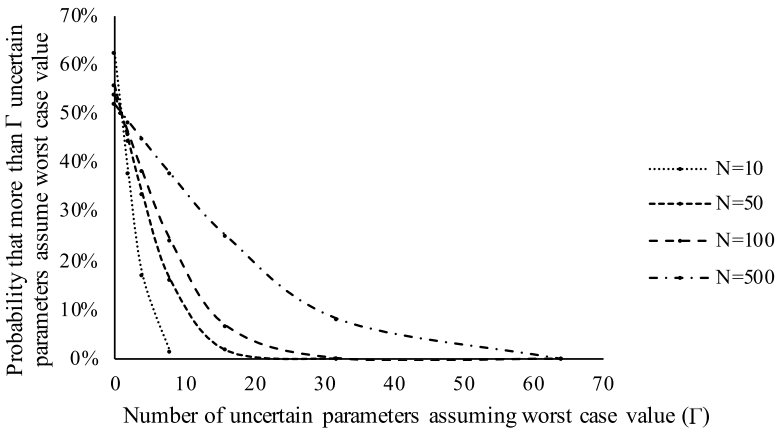


**Fig. A.1.** Probability that the number of uncertain variables assuming their worst-case value exceeds $\Gamma$. $N$ represents the total number of uncertain parameters in a constraint. For example, for an optimization problem with 500 uncertain parameters in a constraint, if we set budget of uncertainty to $\Gamma = 30$, then the probability that more than 30 uncertain variables assume their worst value is less than 10%.

$$\sum_{v\in V_i}\sum_{i\in I}\mathbf{ACT}_{ivt} \geq D_t \qquad \forall t \in T \tag{A.5c}$$

$$\sum_{V_i\leq T_i}\gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t \in T_i, v \in V_i, i \in I \tag{A.5d}$$

$$B\mathbf{x} \geq b \tag{A.5e}$$

$$S = \left\{ |\mathbf{s}_{iv}^{IC}| \leq 1, \forall i \in I, v \in V_i; |\mathbf{s}_{ft}^{FC}| \leq 1, \forall f \in F, t \in T_f; \sum_{v\in V_i}\sum_{i\in I}|\mathbf{s}_{iv}^{IC}| + \sum_{t\in T_f}\sum_{f\in F}|\mathbf{s}_{ft}^{FC}| \leq \Gamma \right\} \tag{A.6}$$

In the above model, (A.6) is a feasible domain for the maximization problem given in (A.5b). For a given $\Gamma$, the maximization problem in constraint (A.5b) provides a protection against infeasibility by assuming that total system cost includes the worst-case cost of $\Gamma$ uncertain parameters. It maximizes the protection for a given $CAP_{iv}$ and $CON_{ft}$ by varying $\mathbf{s}_{iv}^{IC}$ and $\mathbf{s}_{ft}^{FC}$ over $S$. Note that the maximization model is formulated as a linear programming problem, as we keep the $CAP_{iv}$ and $CON_{ft}$ constant as defined by the minimization model. Problem (A.5a)–(A.5e) can be solved by iteratively solving minimization and maximization models to minimize $\mathbf{W}$. To avoid the iterative method, we can simplify model (A.5a)–(A.5e) through the application of strong duality, which says that at the optimum, objective function value of the dual problem is the same as primal problem [33]. The dual of the maximization problem is given below in (A.7a)–(A.7f)

$$\min \quad \sum_{v\in V_i}\sum_{i\in I}\mathbf{q}_{iv}^{IC} + \sum_{t\in T_f}\sum_{f\in F}\mathbf{q}_{ft}^{FC} + \Gamma\mathbf{p} \tag{A.7a}$$

$$\text{s.t.} \quad \mathbf{p} + \mathbf{q}_{iv}^{IC} \geq \Delta IC_{iv}\mathbf{r}_{iv}^{IC} \qquad \forall v \in V_i, i \in I \tag{A.7b}$$

$$\mathbf{p} + \mathbf{q}_{ft}^{FC} \geq \Delta FC_{ft}\mathbf{r}_{ft}^{FC} \qquad \forall t \in T_f, f \in F \tag{A.7c}$$

$$-\mathbf{r}_{iv}^{IC} \leq \mathbf{CAP}_{iv} \leq \mathbf{r}_{iv}^{IC} \qquad \forall v \in V_i, i \in I \tag{A.7d}$$

$$-\mathbf{r}_{ft}^{FC} \leq \mathbf{CON}_{ft} \leq \mathbf{r}_{ft}^{FC} \qquad \forall t \in T_f, f \in F \tag{A.7e}$$

$$\mathbf{p}, \mathbf{q}_{iv}^{IC}, \mathbf{q}_{ft}^{FC}, \mathbf{r}_{iv}^{IC}, \mathbf{r}_{ft}^{FC} \geq 0 \qquad \forall v \in V_i, i \in I, t \in T_f, f \in F \tag{A.7f}$$

Where, $\mathbf{p}, \mathbf{q}_{iv}^{IC}, \mathbf{q}_{ft}^{FC}, \mathbf{r}_{iv}^{IC}, \mathbf{r}_{ft}^{FC}$ are the dual variables associated with the feasible domain given in (A.6). The simplified version of (A.5a)–(A.5e) is given in (A.8a)–(A.8j) where we replace the maximization problem with the objective function of the dual problem. Constraints of (A.7a)–(A.7f) can be added as they are without changing the solution to the problem (A.5a)–(A.5e).

$$\min \quad \mathbf{W} \tag{A.8a}$$

$$\text{s.t.} \quad \mathbf{W} - \left( \sum_{v\in V_i}\sum_{i\in I} IC_{iv}\mathbf{CAP}_{iv} + \sum_{t\in T_i}\sum_{v\in V_i}\sum_{i\in I} OC_{ivt}\mathbf{ACT}_{ivt} + \sum_{t\in T_f}\sum_{i\in I} FC_{ft}\mathbf{CON}_{ft} \right) - \left( \sum_{v\in V_i}\sum_{i\in I}\mathbf{q}_{iv}^{IC} + \sum_{t\in T_f}\sum_{f\in F}\mathbf{q}_{ft}^{FC} + \Gamma\mathbf{p} \right) \geq 0 \tag{A.8b}$$

$$\sum_{v\in V_i}\sum_{i\in I}\mathbf{ACT}_{ivt} \geq D_t \qquad \forall t \in T \tag{A.8c}$$

$$\sum_{V_i\leq T_i}\gamma_{ivt}\mathbf{CAP}_{iv} \geq \mathbf{ACT}_{ivt} \qquad \forall t \in T_i, v \in V_i, i \in I \tag{A.8d}$$

$$B\mathbf{x} \geq b \tag{A.8e}$$

$$\mathbf{p} + \mathbf{q}_{iv}^{IC} \geq \Delta IC_{iv}\mathbf{r}_{iv}^{IC} \qquad \forall v \in V_i, i \in I \tag{A.8f}$$

$$\mathbf{p} + \mathbf{q}_{ft}^{FC} \geq \Delta FC_{ft}\mathbf{r}_{ft}^{FC} \qquad \forall t \in T_f, f \in F \tag{A.8g}$$

$$-\mathbf{r}_{iv}^{IC} \leq CAP_{iv} \leq \mathbf{r}_{iv}^{IC} \qquad \forall v \in V_i, i \in I \tag{A.8h}$$

$$-\mathbf{r}_{ft}^{FC} \leq CON_{ft} \leq \mathbf{r}_{ft}^{FC} \qquad \forall t \in T_f, f \in F \tag{A.8i}$$

$$\mathbf{p}, \mathbf{q}_{iv}^{IC}, \mathbf{q}_{ft}^{FC}, \mathbf{r}_{iv}^{IC}, \mathbf{r}_{ft}^{FC} \geq 0 \qquad \forall v \in V_i, t \in T_f, i \in I, f \in F \tag{A.8j}$$

Bertsimas and Sim [19] show that irrespective of the distribution of uncertain parameter values, we can compute the probability, $\epsilon$, that the number of uncertain parameters assuming their worst-case value exceeds the specified budget of uncertainty $\Gamma \in [0, N]$, where $N$ is the total number of uncertain parameters. In other words, if the decision maker chooses a budget of uncertainty $\Gamma$, then the probability of the system cost going beyond the optimal objective function value of CR-ESOM is less than $\epsilon$. The probability, $\epsilon$, can be calculated as $1 - \Phi((\Gamma - 1)/\sqrt{N})$, where $\Phi(.)$ represents the cumulative density function (CDF) of a normal distribution. Note that in the results section, we represent $\Gamma$ in terms of percent of total uncertain parameters assuming their worst-case value, however, we define $\Gamma$ as an integer in $[0, N]$ for the CR-ESOM formulation. Fig. A.1 plots $\epsilon$ assuming a different total number of uncertain parameters ($N$).

## References


[1] O. Hoegh-Guldberg, D. Jacob, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K. Ebi, F. Engelbrecht, J. Guiot, et al., Impacts of 1.5 C global warming on natural and human systems, in: Global Warming of 1.5° C. An IPCC Special Report, IPCC Secretariat, 2018.

[2] H. Turton, L. Barreto, Long-term security of energy supply and climate change, Energy Policy 34 (15) (2006) 2232–2250.

[3] R. Kannan, N. Strachan, Modelling the UK residential energy sector under long-term decarbonisation scenarios: Comparison between energy systems and sectoral modelling approaches, Appl. Energy 86 (4) (2009) 416–428.

[4] N. Patankar, A.R. de Queiroz, J.F. DeCarolis, M.D. Bazilian, D. Chattopadhyay, Building conflict uncertainty into electricity planning: A South Sudan case study, Energy Sustain. Dev. 49 (2019) 53–64.

[5] H. Eshraghi, A.R. de Queiroz, J.F. DeCarolis, US energy-related greenhouse gas emissions in the absence of federal climate policy, Environ. Sci. Technol. 52 (17) (2018) 9595–9604.

[6] N. Patankar, H.G. Fell, A.R. de Queiroz, J. Curtis, J.F. DeCarolis, Improving the representation of energy efficiency in an energy system optimization model, Applied Energy 306 (2022) 118083.

[7] X. Yue, S. Pye, J. DeCarolis, F.G. Li, F. Rogan, B.Ó. Gallachóir, A review of approaches to uncertainty assessment in energy system optimization models, Energy Strategy Rev. 21 (2018) 204–217.

[8] A. Kann, J.P. Weyant, Approaches for performing uncertainty analysis in large-scale energy/economic policy models, Environ. Model. Assess. 5 (1) (2000) 29–46.

[9] M.-C. Hu, B.F. Hobbs, Analysis of multi-pollutant policies for the US power sector under technology and policy uncertainty using MARKAL, Energy 35 (12) (2010) 5430–5442.

[10] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory, SIAM, 2014.

[11] A. Kanudia, R. Loulou, Robust responses to climate change via stochastic MARKAL: The case of Québec, European J. Oper. Res. 106 (1) (1998) 15–30.

[12] S. Messner, A. Golodnikov, A. Gritsevskii, A stochastic version of the dynamic linear programming model MESSAGE III, Energy 21 (9) (1996) 775–784.

[13] R. Loulou, A. Lehtila, Stochastic programming and tradeoff analysis in TIMES, 2012, TIMES Version 3.3 User Note.

[14] J.A. Bennett, C.N. Trevisan, J.F. DeCarolis, C. Ortiz-García, M. Pérez-Lugo, B.T. Etienne, A.F. Clarens, Extending energy system modelling to include extreme weather risks and application to hurricane events in Puerto Rico, Nat. Energy 6 (3) (2021) 240–249.

[15] V. Mulvey, Zenios, Robust optimization of large-scale systems, Oper. Res. 43 (2) (1995) 264–281.

[16] A.L. Soyster, Convex programming with set-inclusive constraints and applications to inexact linear programming, Oper. Res. 21 (5) (1973) 1154–1157.

[17] L. El Ghaoui, F. Oustry, H. Lebret, Robust solutions to uncertain semidefinite programs, SIAM J. Optim. 9 (1) (1998) 33–52.

[18] A. Ben-Tal, A. Nemirovski, Robust optimization–methodology and applications, Math. Program. 92 (3) (2002) 453–480.

[19] D. Bertsimas, M. Sim, The price of robustness, Oper. Res. 52 (1) (2004) 35–53.

[20] D. Bertsimas, A. Thiele, Robust and data-driven optimization: modern decision making under uncertainty, in: Models, Methods, and Applications for Innovative Decision Making, INFORMS, 2006, pp. 95–122.

[21] D. Bertsimas, D.B. Brown, C. Caramanis, Theory and applications of robust optimization, SIAM Rev. 53 (3) (2011) 464–501.

[22] P. Nahmmacher, E. Schmid, M. Pahle, B. Knopf, Strategies against shocks in power systems–An analysis for the case of Europe, Energy Econ. 59 (2016) 455–465.

[23] T. Niet, B. Lyseng, J. English, V. Keller, K. Palmer-Wilson, I. Moazzen, B. Robertson, P. Wild, A. Rowe, Hedging the risk of increased emissions in long term energy planning, Energy Strategy Rev. 16 (2017) 1–12.

[24] V. Krey, K. Riahi, Risk hedging strategies under energy system and climate policy uncertainties, in: Handbook of Risk Management in Energy Production and Trading, Springer, 2013, pp. 435–474.

[25] F. Babonneau, J.-P. Vial, R. Apparigliato, Robust optimization for environmental and energy planning, in: Uncertainty and Environmental Decision Making, Springer, 2009, pp. 79–126.

[26] F. Babonneau, A. Kanudia, M. Labriet, R. Loulou, J.-P. Vial, Energy security: a robust optimization approach to design a robust European energy supply via TIAM-WORLD, Environ. Model. Assess. 17 (1–2) (2012) 19–37.

[27] D. Lorne, S. Tchung-Ming, The French biofuels mandates under cost uncertainty–an assessment based on robust optimization, 2012, Les Cahiers de L'Economie.

[28] M. Labriet, C. Nicolas, S. Tchung-Ming, A. Kanudia, R. Loulou, Energy decisions in an uncertain climate and technology outlook: How stochastic and robust methodologies can assist policy-makers, in: Informing Energy and Climate Policies using Energy Systems Models, Springer, 2015, pp. 69–91.

[29] S. Moret, M. Bierlaire, F. Maréchal, Robust optimization for strategic energy planning, Informatica 27 (3) (2016) 625–648.

[30] S. Moret, F. Babonneau, M. Bierlaire, F. Maréchal, Decision support for strategic energy planning: A robust optimization framework, European J. Oper. Res. 280 (2) (2020) 539–554.

[31] K. Hunter, S. Sreepathi, J.F. DeCarolis, Modeling for insight using tools for energy model optimization and analysis (Temoa), Energy Econ. 40 (2013) 339–349.

[32] J. DeCarolis, K. Hunter, Tools for energy model optimization and analysis (Temoa), 2020.

[33] D. Bertsimas, J.N. Tsitsiklis, Introduction to Linear Optimization, Vol. 6, Athena Scientific Belmont, MA, 1997.

[34] G. Bayraksan, D.P. Morton, Assessing solution quality in stochastic programs, Math. Program. 108 (2–3) (2006) 495–514.

[35] G. Bayraksan, D.P. Morton, Assessing solution quality in stochastic programs via sampling, in: Decision Technologies and Applications, Informs, 2009, pp. 102–122.

[36] W.-K. Mak, D.P. Morton, R.K. Wood, Monte Carlo bounding techniques for determining solution quality in stochastic programs, Oper. Res. Lett. 24 (1–2) (1999) 47–56.

[37] A.R. de Queiroz, A Sampling-Based Decomposition Algorithm with Application to Hydrothermal Scheduling: Cut Formation and Solution Quality, Citeseer, 2011.

[38] C. Lenox, R. Dodder, C. Gage, D. Loughlin, O. Kaplan, W. Yelverton, EPA US nine-region MARKAL database: database documentation, 2013, US Environmental.

[39] L.J. Vimmerstedt, S. Akar, C.R. Augustine, P.C. Beiter, W.J. Cole, D.J. Feldman, P. Kurup, E.J. Lantz, R.M. Margolis, T.J. Stehly, et al., 2019 Annual Technology Baseline, Technical Report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[40] AEO, Annual energy outlook, 2020, https://www.eia.gov/outlooks/aeo/pdf/AEO2020.pdf. (Accessed 25 Aguest 2021).

[41] N. Patankar, H. Eshraghi, A.R. de Queiroz, J.F. DeCarolis, TemoaProject: Using Robust Optimization Techniques to Inform US Climate Policy, Zenodo, 2021, http://dx.doi.org/10.5281/zenodo.5450870, For details on how to run the model, please visit the project website: http://temoacloud.com.

[42] S. Babaee, A.S. Nagpure, J.F. DeCarolis, How much do electric drive vehicles matter to future US emissions? Environ. Sci. Technol. 48 (3) (2014) 1382–1390.

[43] N.S. Patankar, Addressing Uncertainty in Energy System Optimization Models over a Long Planning Horizon, North Carolina State University, 2019.

[44] P. Wilkinson, Parallel Programming: Techniques and Applications using Networked Workstations and Parallel Computers, 2/E, Pearson Education India, 2006.

[45] NREL, Open EI, 2019, https://openei.org/apps/TCDB/. (Accessed 25 June 2020).

[46] M. De Villiers, K. Matibe, Greenhouse Gas Baseline and Mitigation Options for the Residential Sector, Vol. 11, Energy & Development Research Centre, Cape Town, 2000, Citeseer.