Development of Expansion Planning Methods and Tools for Handling Uncertainty

Final Project Report

M-37

Power Systems Engineering Research Center

Empowering Minds to Engineer
the Future Electric Energy System

Development of Expansion Planning Methods and Tools for Handling Uncertainty

Final Project Report

Project Team
Lizhi Wang, Project Leader
James McCalley
Iowa State University

Christopher L. Demarco
University of Wisconsin – Madison

Graduate Students
Faezeh Akhavizadegan
Abhinav Venkatraman
Iowa State University

Jonathan Snodgrass
University of Wisconsin – Madison

PSERC Publication 20-03

May 2020

For information about this project, contact

Lizhi Wang
Industrial and Manufacturing Systems Engineering Department
Iowa State University
Ames, IA 50011
Phone: 515-294-1757
Fax: 515-294-3524
Email: lzwang@iastate.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

Acknowledgements

This is the final report for the Power Systems Engineering Research Center (PSERC) research project titled "Development of Expansion Planning Methods and Tools for Handling Uncertainty" (M-37). We express our appreciation for the support provided by PSERC's industry members and by the National Science Foundation under the Industry / University Cooperative Research Center program.

We wish to thank our industry advisors, who have provided strong support and constructive feedback to our work:

- Saman Babaei, NYPA
- Aaron Bloom, NREL
- Keel Brian, SRP
- Jay Caspary, SPP
- Orlando Ciniglio, IDAH Power
- Sara Daubenberger, BPA
- Matt Ellis, MISO
- Lei Fan, GE
- Evangelos Farantatos, EPRI
- Anish Gaikwad, EPRI
- Garry Gu, GE
- Robin Hytowitz, EPRI
- Rebecca Johnson, WAPA
- Xiaochun Luo, ISO NE
- Sakshi Mishra, AEP
- Miguel Ortega-Vazquez, EPRI
- Aditya Jayam Prabhakar, MISO
- Jim Price, CAISO
- Curtis Roe, ATC
- Harvey Scribner, SPP
- Eknath Vittal, EPRI
- Mark Westendorf, MISO

Executive Summary

Reliable transmission expansion planning is critical to power systems development. To make reliable and sustainable transmission expansion plans, numerous sources of uncertainty including demand, generation capacity, and fuel cost must be taken into consideration in both spatial and temporal dimensions. This paper presents a new approach to selecting a small number of high-quality scenarios for transmission expansion. The Kantorovich distance of social welfare distributions was used to assess the quality of the selected scenarios. A case study was conducted on a power system model that represents the US Eastern and Western Interconnections, and ten high-quality scenarios out of a total of one million were selected for two transmission plans. Results suggested that scenarios selected using the proposed algorithm were able to provide a much more accurate estimation of the value of transmission plans than other scenario selection algorithms in the literature.

Project Publications:

[1] Faezeh Akhavizadegan, Lizhi Wang, and James McCalley, "Scenario Selection for Iterative Stochastic Transmission Expansion Planning", *Energies* 2020, 13, 1203.

Part I

Scenario Selection for Iterative Stochastic Transmission Expansion Planning

Faezeh Akhavizadegan
Lizhi Wang

Iowa State University

For information about this project, contact

Lizhi Wang
Industrial and Manufacturing Systems Engineering Department
Iowa State University
Ames, IA 50011
Phone: 515-294-1757
Fax: 515-294-3524
Email: lzwang@iastate.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

Table of Contents

List of Figures

List of Tables

1. Introduction

The integrated electric utility, owning both generation and transmission, traditionally develops generation plans first and then plans the transmission necessary to support those generation plans, an approach consistent with the fact that generation investments are normally far more costly than the needed transmission investments. Today, however, there are organizations that own transmission only. There are also service companies, called regional transmission organizations in the US, that own no infrastructure but are responsible for coordinating the planning function. Both of these organizations must solve the transmission expansion planning (TEP) problem under uncertainty regarding technology, quantity, location, and timing of generation investments [22, 31, 42]. Robust and stochastic optimization approaches are two major modeling techniques for transmission planning under uncertainty, which have their own strengths and limitations.

Robust TEP maximizes the performance of the power system under the worst case scenario, which is dependent on the transmission plan [10, 5, 6, 8, 9]. Most solution techniques for robust TEP consist of two iterative steps: one step identifies the worst case scenario for a proposed transmission plan, whereas the other proposes the most robust transmission plan against a pool of worst case scenarios. A major advantage of robust TEP is its robustness against worst case scenarios, which is especially desirable for enhancing long-term resiliency of power system infrastructures. The iterative algorithm also allows the model to efficiently identify worst case scenarios from an enormous solution space with a large number of (or even infinitely many) possible scenarios. On the other hand, this approach often tends to be over-conservative by focusing on the *possibility* and ignoring the *probability* of worst case scenarios.

Stochastic TEP addresses uncertainty with a very different philosophy. It maximizes the average performance of the power system under all scenarios, weighted by their probabilities of occurrence. As such, the optimal stochastic transmission plan finds a good balance between the (positive or negative) impact and the likelihood of all scenarios [30, 2, 29, 34, 35]. However, this approach also faces a dilemma. On the one hand, due to the numerous sources of uncertainty and long planning horizon, a large number of scenarios is necessary to realistically represent the complexity and uncertainty of the TEP problem. On the other hand, however, the computation time of most solution techniques is very sensitive to the number of scenarios, and many algorithms become intractable for even a few dozen scenarios. To address this dilemma, scenario generation and reduction techniques have been proposed, which attempt to identify a small set of high-quality scenarios that represent the whole set of scenarios [13, 15, 20, 11, 16, 28].

Our proposed approach is an iterative stochastic TEP framework, which attempts to combine the advantages of robust and stochastic TEP approaches and to overcome their limitations. A major limitation of the stochastic TEP approach with scenario generation and reduction is the assumption that the set of selected scenarios will be a good representation of the whole set of scenarios for all transmission plans. This is analogous to identifying one scenario as the "worst case scenario" for all transmission plans in the robust TEP approach. The iterative framework in robust TEP acknowledges the need for re-identifying a worst case scenario for each new transmission plan, until the pool of worst case scenarios is sufficiently inclusive of all worst case scenarios. Similarly, the proposed iterative stochastic TEP framework re-identifies a new set of high-quality scenarios for each new transmission plan until the pool of high-quality scenarios is sufficiently representative of the whole set of scenarios. A formal definition of high-quality scenarios is given in Section 2.1. Fig. 1.1 highlights the conceptual differences among robust TEP, stochastic TEP, and the proposed iterative stochastic TEP. The scope of this paper is the highlighted component for selecting high-quality scenarios.

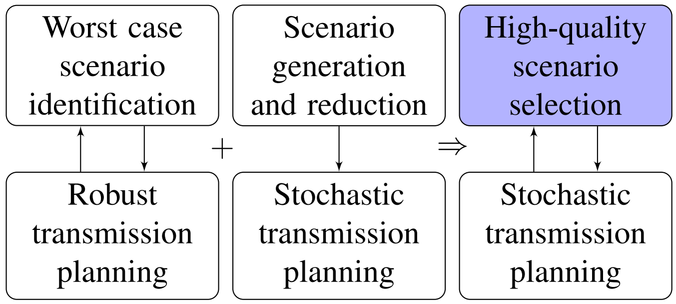


Figure 1.1: Conceptual differences among robust TEP (left), stochastic TEP (middle), and the proposed iterative stochastic TEP (right) frameworks.

Table 1.1: Illustrative example of robust TEP. Numerical values represent social welfare of 10 transmission plans under 10 scenarios. For each transmission plan, the social welfare under the worst case scenario is highlighted.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
p_1	12	46	50	27	33	46	44	28	41	29
p_2	21	10	14	37	46	25	42	30	22	27
p_3	46	36	29	24	23	30	45	33	37	11
p_4	34	49	20	25	21	36	23	14	30	27
p_5	39	48	18	33	15	42	22	28	28	36
p_6	36	13	36	27	30	33	48	40	43	29
p_7	14	45	16	37	28	30	36	24	47	49
p_8	24	11	32	36	26	21	15	40	43	33
p_9	18	22	27	19	40	35	29	23	50	41
p_{10}	38	16	20	46	30	19	31	43	24	40

Consider an illustrative example, in which there are 10 scenarios $\{s_1, ..., s_{10}\}$ and 10 transmission plans $\{p_1, ..., p_{10}\}$, and we try to identify the optimal transmission plan. The number of scenarios and plans are limited to 10 for illustration purposes, but the methods are applicable to situations where there are millions or even infinitely many scenarios and plans. With the values in the table representing social welfare of the transmission plans under different scenarios, Table 1.1 illustrates the solution process of the robust TEP approach, and the optimal transmission plan is determined in Table 1.3 as p_9, which resulted in a higher social welfare than all other transmission plans under their respective worst case scenarios. Table 1.2 illustrates the solution process of the proposed iterative stochastic TEP approach for the same example, in which two high-quality scenarios are selected to represent the whole set of 10 scenarios for each transmission plan. The algorithm proposed in Section 3.3 is used to select these scenarios and assign their probabilities. The optimal transmission plan is determined in Table 1.3 as p_1, which maximizes the weighted average social welfare. The first column of Table 1.3 calculates the average performance of all transmission plans under all scenarios, which shows that the true optimal solution to the stochastic TEP problem is indeed p_1. This example illustrates how the proposed iterative stochastic TEP approach is able to identify the optimal (or close to optimal) solution by selecting a subset of high-quality scenarios chosen specific to each transmission plan.

Table 1.2: Illustrative example of stochastic TEP. Numerical values represent social welfare of 10 transmission plans under 10 scenarios. For each transmission plan, the social welfare under the two high-quality scenarios are highlighted.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
p_1	12	46	50	27	33	46	44	28	41	29
p_2	21	10	14	37	46	25	42	30	22	27
p_3	46	36	29	24	23	30	45	33	37	11
p_4	34	49	20	25	21	36	23	14	30	27
p_5	39	48	18	33	15	42	22	28	28	36
p_6	36	13	36	27	30	33	48	40	43	29
p_7	14	45	16	37	28	30	36	24	47	49
p_8	24	11	32	36	26	21	15	40	43	33
p_9	18	22	27	19	40	35	29	23	50	41
p_{10}	38	16	20	46	30	19	31	43	24	40

Table 1.3: Optimal solutions using robust TEP under the worst case scenario, stochastic TEP with all 10 scenarios, and stochastic TEP under 2 high-quality scenarios.

Transmission Plan	Robust (Worst scenario)	Stochastic (All scenarios)	Stochastic (High-quality scenarios)
p_1	12	35.6	46×0.5+28×0.5=37.0
p_2	10	27.4	42×0.3+22×0.7=28.0
p_3	11	31.4	24×0.5+37×0.5=30.5
p_4	14	27.9	34×0.4+21×0.6=26.2
p_5	15	30.9	39×0.5+22×0.5=30.5
p_6	13	33.5	30×0.8+48×0.2=33.6
p_7	14	32.6	45×0.5+24×0.5=34.5
p_8	11	28.1	36×0.5+21×0.5=28.5
p_9	18	30.4	22×0.6+40×0.4=29.2
p_{10}	16	30.7	38×0.6+20×0.4=30.8

Scenarios generation and reduction has been a topic of great interest in the power systems literature. Most existing methods use clustering [11, 24, 23, 37] or sampling methods to reduce the number of scenarios from a randomly generated initial set. In a recent review article, Park et al. [36] compared four methods for scenario reduction using a two-stage stochastic transmission planning model, including random sampling, importance sampling [33], distance based method [13, 12, 21], and stratified scenario sampling [36]. They used these methods to reduce a whole set of 20 scenarios to smaller subsets and compared their pros and cons. Other methods also include extended and improved initial-center-refined and weighted k-means [28], data-driven [44], forward selection [16], moment-based [20, 27, 14], and objective-based [43] approaches.

The proposed scenario selection approach in this paper differs from previous methods for scenario generation and reduction in three major ways. First, we generate scenarios with explicit consideration of temporal and spatial correlations, including generation investment and retirement decisions in response to demand, fuel cost, and transmission capacity. Second, we use the Kantorovich distance of social welfare distributions to assess the quality of the selected scenarios. Third, a different subset of high-quality scenarios is selected for each transmission plan candidate. In comparison, most existing methods in the literature ignore the correlations among the scenarios, select a subset of scenarios based on their similarity rather than their implications on social welfare, and use the same subset of scenarios for all transmission plans.

The remaining sections of this report are the outline of the report. In Section 2., we describe the proposed approach to selecting a set of high-quality scenarios for the transmission planning problem. Section 3. presents a heuristic algorithm for solving the generation expansion problem and selecting high-quality scenarios. To demonstrate the effectiveness of our proposed models and methods, we reported a case study of the US Eastern and Western Interconnection in Section 4.. Finally, we provide summary remarks in Section 5. and restate the key findings and conclusions. We also discuss directions for future work in the same section.

2. Model formulation

In this section, we describe and explain the models that we propose to identify a small set of high-quality scenarios for a given transmission plan. In the following, the variables and parameters used in the model are described.

Sets and Indices

$\mathcal{Y}$	Set of planning years
$\mathcal{T}$	Set of load blocks
$\mathcal{B}$	Set of buses
$\mathcal{L}$	Set of transmission lines
$\mathcal{H}$	Set of generation technologies
$\mathcal{G}^{\text{ER}}(b, y)$	Set of existing renewable generators at bus b in year y
$\mathcal{G}^{\text{EN}}(b, y)$	Set of existing non-renewable generators at bus b in year y
$\mathcal{G}^{\text{E}}(b, y)$	Set of existing generators at bus b in year y: $\mathcal{G}^{\text{E}}(b, y) = \mathcal{G}^{\text{ER}}(b, y) \cup \mathcal{G}^{\text{EN}}(b, y)$
$\mathcal{G}^{\text{CR}}(b, y)$	Set of candidate renewable generators at bus b in year y
$\mathcal{G}^{\text{CN}}(b, y)$	Set of candidate non-renewable generators at bus b in year y
$\mathcal{G}^{\text{C}}(b, y)$	Set of candidate generators at bus b in year y: $\mathcal{G}^{\text{C}}(b, y) = \mathcal{G}^{\text{CR}}(b, y) \cup \mathcal{G}^{\text{CN}}(b, y)$
$\mathcal{G}^{\text{R}}(b, y)$	Set of renewable generators at bus b in year y: $\mathcal{G}^{\text{C}}(b, y) = \mathcal{G}^{\text{ER}}(b, y) \cup \mathcal{G}^{\text{CR}}(b, y)$
$\mathcal{G}(b, y)$	Set of existing and candidate generators at bus b in year y: $\mathcal{G}(b, y) = \mathcal{G}^{\text{E}}(b, y) \cup \mathcal{G}^{\text{C}}(b, y)$
$\mathcal{G}^{\text{E}}_h(b, y)$	Set of existing generators of technology h at bus b in year y

Parameters

r	Discount rate, in %
C^{LS}	Cost of load-shedding, in \$/MWh
T_t	Operating duration of load block t, in hour
H_g	Heat rate for generator g, in Mbtu/MWh
$F^{\text{C}}_{g,y}$	Fuel cost of generator g's technology in year y, in \$/Mbtu
C^{OM}_g	Fixed Operation and Maintenance (O&M) cost for generator g's technology, in \$/MW
C^{I}_g	Amortized annual investment cost for generator g's technology, in \$/MW
C^{V}_g	Variable O&M cost for generator g's technology, in \$/MWh
$C^{\text{V}}_{g,y}$	Variable generation cost for generator g in year y: $C^{\text{V}}_{g,y} = H_g F^{\text{C}}_{g,y} + C^{\text{V}}_g$, in \$/MWh
$D_{b,t,y}$	Demand at bus b for load block t in year y, in MW
$V_{b,t,y}$	Monetary value of energy consumption at bus b for load block t in year y, in \$/MWh
$B_{i,j}$	Susceptance of transmission line (i, j), in siemens
$F^{\max}_{i,j,y}$	Capacity of transmission line (i, j) in year y, in MW
$C_{g,t}$	Capacity factor of generator g for load block t, in %
$P^{\max}_g$	Production capacity of generator g, in MW
α_y	Renewable portfolio requirement in year y, in %

Decision Variables

$p_{g,b,t,y}$	Power production from generator g at bus b for load block t in year y, in MW
$f_{i,j,t,y}$	Power flow through transmission line (i, j) for load block t in year y, in MW
$d^{\mathrm{LS}}_{b,t,y}$	Load shedding at bus b for load block t in year y, in MW
$\theta_{b,t,y}$	Voltage angle at bus b for load block t in year y, in radians
$x_{g,b,y}$	Binary variable indicating whether the generator g at bus b exists in year y ($x_{g,b,y} = 1$) or not ($x_{g,b,y} = 0$). For an existing generator $g \in \mathcal{G}^{\mathrm{E}}(b, y)$, $x_{g,b,y} = 0$ means the generator has retired; and for a candidate generator $g \in \mathcal{G}^{\mathrm{C}}(b, y)$, $x_{g,b,y} = 1$ means the generator has been added to the existing generation capacity.
$p^{\mathrm{LMP}}_{b,t,y}$	Locational marginal price at bus b for load block t in year y, in \$/MWh
Ψ	Social welfare of energy producers and consumers in the power system, in \$.

2.1 Definition of high-quality scenarios

For a given transmission plan, a scenario includes three elements: demand for all buses, fuel cost for all generation technologies, and generation capacity for all generators, throughout the entire planning horizon. Such definition of a scenario captures the temporal and spatial trajectory of the power system, which we view as evolving as a function of a given transmission plan.

High-quality scenarios selected using our proposed approach satisfy two requirements. First, the correlations between generation capacity and other elements (i.e., demand and fuel cost) are reflected. This is because generation investments and retirement decisions are made by decentralized and for-profit generation companies, which have been found to be sensitive to demand and fuel cost [15]. Second, the probabilistic distance between the distribution of social welfare resulting from high-quality scenarios and that from all scenarios is small (the smaller the distance, the higher the quality).

2.2 Estimation of generation investment and retirement

For given demand and fuel cost elements of a scenario, we propose to estimate the generation capacity element of the scenario by solving a simplified generation expansion planning (GEP) problem, which has the following features.

- We cast the GEP problem as a bilevel optimization model, in which the upper level maximizes the net present value of the investment by determining the investment of new generation capacity and the retirement of existing generators, whereas the lower level computes the optimal power flow (used by the upper level to calculate the revenue of power generation) as a result of the GEP decisions from the upper level. Similar models have been used in other studies such as [6, 45].

- It uses a single upper level decision maker for all generation investment and retirement decisions in all buses. This is a simplifying approximation of the power systems, in which investment and retirement decisions are made by multiple generation companies in a decentralized manner. Similar approximations have been used in other studies such as [15, 20].
- The GEP model identifies generation investments based on profit maximization accounting for fixed and variable operational costs, investment costs, and load shedding costs, subject to renewable portfolio standards and generation adequacy requirements, attributes common to other studies [15, 20, 47, 32, 1, 38].

The bilevel optimization model of the GEP problem is formulated as follows, where the upper level decision variable x represents generation investment and retirement decisions $x_{g,b,y}, \forall g, b, y$, and the lower level decision variable p represent both dispatch decisions $p_{g,b,t,y}, \forall g, b, t, y$ and locational marginal prices $p^{\text{LMP}}_{b,t,y}, \forall b, t, y$.

$$\max_{x,p} \quad f^{\text{GEP}}(x,p) \tag{1}$$

$$\text{s.t.} \quad A_1 x \leq b_1 \tag{2}$$

$$x \text{ binary}. \tag{3}$$

$$p \in \operatorname{argmin}\{d_2^\top p : A_2 x + B_2 p \leq b_2, p \geq 0\}. \tag{4}$$

The upper level objective function $f^{\text{GEP}}(x,p)$ is to maximize the profit of generation companies, which can be estimated as

$$\begin{aligned} f^{\text{GEP}}(x,p) = & \sum_{\substack{g\in\mathcal{G}(b,y),b\in\mathcal{B},\\ t\in\mathcal{T},y\in\mathcal{Y}}} \frac{p_{g,b,t,y}\max\{p^{\text{LMP}}_{b,t,y} - C^{\text{V}}_{g,y}, 0\}T_t}{(1+r)^y} x_{g,b,y} \\ & - \sum_{\substack{g\in\mathcal{G}(b,y),\\ b\in\mathcal{B},y\in\mathcal{Y}}} \frac{C^{\text{OM}}_h P^{\max}_g}{(1+r)^y} x_{g,b,y} - \sum_{\substack{g\in\mathcal{G}^{\text{C}}(b,y),\\ b\in\mathcal{B},y\in\mathcal{Y}}} \frac{C^{\text{I}}_h P^{\max}_g}{(1+r)^y} x_{g,b,y}. \end{aligned} \tag{5}$$

The first term in the objective function is the estimated revenue from selling power generation at the locational marginal prices. The second term in the objective function is the O&M cost for existing and new generators, and the third term is the investment cost for new generators. All future cost terms are discounted to the current year to calculate the net present value. All fixed and variable investment costs are amortized so that investments towards the end of planning horizon would not be disincentivized, since only part of the investment cost appropriated for the planning horizon was calculated to offset the benefit of the investment.

The upper level constraints (2) and (3) are defined as follows:

$$
\begin{aligned}
& x_{g,b,y} \geq x_{g,b,y+1}, && \forall g \in \mathcal{G}^{\mathrm{E}}(b,y), b \in \mathcal{B}, y \in \mathcal{Y} && (6)\\
& x_{g,b,y} \leq x_{g,b,y+1}, && \forall g \in \mathcal{G}^{\mathrm{C}}(b,y), b \in \mathcal{B}, y \in \mathcal{Y} && (7)\\
& \sum_{\substack{g\in\mathcal{G}^{\mathrm{R}}(b,y),\\ b\in\mathcal{B}}} P_g^{\max} x_{g,b,y} \geq \alpha_y \sum_{\substack{g\in\mathcal{G}(b,y),\\ b\in\mathcal{B}}} P_g^{\max} x_{g,b,y}, && \forall y \in \mathcal{Y} && (8)\\
& \sum_{\substack{g\in\mathcal{G}(b,y),\\ b\in\mathcal{B}}} P_g^{\max} x_{g,b,y} \geq 1.2 \sum_{b\in\mathcal{B}} \max_{t\in\mathcal{T}} D_{b,t,y}, && \forall y \in \mathcal{Y} && (9)\\
& x_{g,b,y} \in \{0,1\}, && \forall g \in \mathcal{G}(b,y), b \in \mathcal{B}, y \in \mathcal{Y}. && (10)
\end{aligned}
$$

Here, constraints (6) and (7) allow the retirement of existing generators and investment of new generators; Constraint (8) imposes a renewable portfolio standard type of requirement on the minimal percentage of generation capacity being renewable; Constraint (9) imposes a generation adequacy requirement, where the total generation capacity for each year must exceed 120% of predicted peak demand; and Constraint (10) is the definition of binary decision variables.

Constraint (4) defines the lower level problem, which takes generation investment and retirement decisions from the upper level as input, and solves the optimal power flow problem to determine power dispatches and locational marginal prices throughout the planning horizon. This lower level problem can be defined as follows, which needs to be solved for all $t \in \mathcal{T}$ and $y \in \mathcal{Y}$.

$$
\begin{aligned}
\min \quad & \sum_{\substack{g\in\mathcal{G}^{\mathrm{E}}(b,y)\\ b\in\mathcal{B}}} C^{\mathrm{V}}_{g,y} p_{g,b,t,y} + \sum_{b\in\mathcal{B}} C^{\mathrm{LS}} d^{\mathrm{LS}}_{b,t,y} && && (11)\\
\text{s.t.} \quad & \textstyle\sum_{g\in\mathcal{G}^{\mathrm{E}}(b,y)} p_{g,b,t,y} + \sum_{i\in\mathcal{B}} f_{i,b,t,y} && && \\
& - \textstyle\sum_{j\in\mathcal{B}} f_{b,j,t,y} = D_{b,t,y} - d^{\mathrm{LS}}_{b,t,y}, && \forall b \in \mathcal{B} && (12)\\
& 0 \leq d^{\mathrm{LS}}_{b,t,y} \leq D_{b,t,y}, && \forall b \in \mathcal{B} && (13)\\
& f_{i,j,t,y} = B_{i,j}(\theta_{i,t,y} - \theta_{j,t,y}), && \forall (i,j) \in \mathcal{L} && (14)\\
& -\pi \leq \theta_{b,t,y} \leq \pi, && \forall b \in \mathcal{B} && (15)\\
& -F^{\max}_{i,j,y} \leq f_{i,j,t,y} \leq F^{\max}_{i,j,y}, && \forall (i,j) \in \mathcal{L} && (16)\\
& 0 \leq p_{g,b,t,y} \leq C_{g,t} P_g^{\max} x_{g,b,y}, && \forall g \in \mathcal{G}^{\mathrm{E}}(b,y), b \in \mathcal{B}. && (17)
\end{aligned}
$$

The lower level objective function (11) is to minimize the production cost and load shedding cost. Constraint (12) enforces Kirchhoff's current law that requires nodal balance of power generation, in flow, out flow, demand, and load shedding; Constraint (13) defines the lower and upper bounds of load shedding; Constraint (14) enforces Kirchhoff's voltage law, which computes power flows through transmission lines based on voltage angles and susceptance of the power network; Constraint (15) defines the bounds of voltage angles; Constraint (16) limits the power flow within transmission capacity; and Constraint (17) limits the power production within existing generators' capacity, which depends on generation investment and retirement decisions.

2.3 Definition of social welfare

Social welfare is a commonly used objective for transmission planning [7, 41, 19, 40, 39, 25]. We measure the quality of selected scenarios based on the similarity between social welfare distributions resulting from selected and whole set of scenarios, rather than the similarity of the elements of the scenarios. For a given transmission plan under a given scenario, the social welfare (Ψ) is defined as the summation of producers' surplus and consumers' surplus:

$$\begin{aligned}\Psi = \quad & f^{\mathrm{GEP}}(x,p) + \sum_{\substack{b\in\mathcal{B},t\in\mathcal{T},\\ y\in\mathcal{Y}}} \frac{(V_{b,t,y} - p^{\mathrm{LMP}}_{b,t,y})D_{b,t,y}T_t}{(1+r)^y} \\ & - \sum_{\substack{b\in\mathcal{B},t\in\mathcal{T},\\ y\in\mathcal{Y}}} \frac{(C^{\mathrm{LS}} + V_{b,t,y} - p^{\mathrm{LMP}}_{b,t,y})d^{\mathrm{LS}}_{b,t,y}T_t}{(1+r)^y}.\end{aligned} \tag{18}$$

Here, the first term is the objective function of the GEP model, which represents producers' surplus; the second and third terms are consumers' surplus, which includes monetary valuation of energy consumption less prices of energy and economic cost of load shedding. All cost and benefit terms in the social welfare definition are discounted to year 0, so that it reflects the net present value of all transactions throughout the planning horizon.

2.4 The Kantorovich distance

We use the Kantorovich distance [26] to measure the probabilistic distance between the two distributions of social welfare resulting from the selected high-quality scenarios and those from the whole set of scenarios:

$$\mathcal{D}^{\mathrm{Kan}}(\mathcal{S}^{\mathrm{H}}, \mathcal{S}^{\mathrm{W}}) := \sum_{s\in\mathcal{S}^{\mathrm{W}}} p_s \min_{s'\in\mathcal{S}^{\mathrm{H}}} |\Psi_s - \Psi_{s'}|. \tag{19}$$

Here, the notations are defined as follows.

- $\mathcal{S}^{\mathrm{W}}$: the whole set of scenarios
- $\mathcal{S}^{\mathrm{H}}$: the high-quality set of scenarios
- p_s: probability of scenario s
- Ψ_s: social welfare under scenario s.

3. Solution techniques

3.1 Algorithm for GEP calculation

For a given transmission plan and two elements (i.e., demand and fuel cost) of a given scenario, we propose the following Algorithm 1 to calculate the generation capacity element by solving the GEP probelm (1)-(4). In this algorithm, we use $\mathcal{L}^{\text{OPF}}(\hat{x})$ to denote the parametric lower level optimal dispatch problem with a given generation investment decision $\hat{x}$:

$$p \in \operatorname{argmin}\{d_2^\top p : A_2\hat{x} + B_2 p \leq b_2, p \geq 0\},$$

and we use $\mathcal{U}^{\text{GEP}}(\hat{p})$ to denote the upper level GEP problem with a given dispatch decision $\hat{p}$:

$$\max_{x,\hat{p}}\{f^{\text{GEP}}(x,\hat{p}) : A_1 x \leq b_1, x \text{ binary}\}.$$

Algorithm 1 Algorithm for the GEP probelm (1)-(4)

```
Choose an arbitrary x^1. Define k = 1 and m = 1. Set the hyper-parameter M as a positive integer.
while m ≤ M do
    solve L^OPF(x^k) and let p^k denote an optimal solution.
    solve U^GEP(p^k) and let x^{k+1} denote an optimal solution.
    if |f^GEP(x^k, p^k) − f^GEP(x^{k−1}, p^{k−1})| < ϵ then
        update m ← m + 1;
    else
        update m ← 0.
    end if
    update k ← k + 1.
end while
(x*, p*) ∈ argmax{f^GEP(x^k, p^k) : ∀k}.
```

The termination criterion of this algorithm is that the improvement in the GEP objective function has been lower than the threshold ϵ for M consecutive iterations. To accelerate convergence of the algorithm, dispatch of existing generators from the previous iteration is used for the next iteration, and the dispatch rate (ratio of dispatch over full generation capacity) of new generators is set to be the average of existing ones at the same bus in the same year:

$$\hat{p}_{g,b,t,y} = \frac{\sum_{\hat{g}\in\{\mathcal{G}(b,y):x_{g,b,y}=1\}} p_{\hat{g},b,t,y}}{\sum_{\hat{g}\in\{\mathcal{G}(b,y):x_{g,b,y}=1\}} P_{\hat{g}}^{\max}} P_g^{\max}, \quad \forall g \in \{\mathcal{G}(b,y) : x_{g,b,y} = 0\}, b \in \mathcal{B}, t \in \mathcal{T}, y \in \mathcal{Y}.$$

There exist numerous algorithms for solving bilevel optimization models such as (1)-(4), e.g., branch-and-bound [3] and KKT reformulation with big-M parameters [18]. After experimenting with multiple algorithms, we found the proposed heuristic to yield a good balance between computation time and solution quality. The termination criterion was based on the maximum number of iterations.

3.2 Algorithm for social welfare estimation

We use a linear regression model to provide a computationally efficient estimation of social welfare. Conceptually, high-quality scenarios could be selected by first calculating the social welfare for all scenarios and then selecting a subset to minimize the probabilistic distance between the distributions of social welfare resulting from the high-quality and whole set of scenarios. However, this approach requires solving the GEP problem millions of times using the time consuming Algorithm 1. Alternatively, our strategy is to train a regression model to estimate social welfare and select the high-quality scenarios based on the estimated rather than actual social welfare values. If trained efficiently, the regression model requires only a small set of training data to provide reasonably accurate estimation, thus we only need to use Algorithm 1 to calculate the actual social welfare for a small number of scenarios to produce the training data.

The multiple linear regression model uses the average load for each year, load block, and bus and the average fuel cost for each technology and each year as explanatory variables and social welfare as the response variable:

$$\hat{\Psi} = \beta_0 + \sum_{b\in\mathcal{B}} \beta_b^D \bar{D}_b + \sum_{y\in\mathcal{Y}} \beta_y^D \bar{D}_y + \sum_{t\in\mathcal{T}} \beta_t^D \bar{D}_t + \sum_{h\in\mathcal{H}} \beta_h^F \bar{F}_h^C + \sum_{y\in\mathcal{Y}} \beta_y^F \bar{F}_y^C, \tag{20}$$

where

- $\bar{D}_b = \frac{1}{|\mathcal{T}|\cdot|\mathcal{Y}|} \sum\limits_{t\in\mathcal{T}} \sum\limits_{y\in\mathcal{Y}} D_{b,t,y}$,
- $\bar{D}_y = \frac{1}{|\mathcal{T}|\cdot|\mathcal{B}|} \sum\limits_{t\in\mathcal{T}} \sum\limits_{b\in\mathcal{B}} D_{b,t,y}$,
- $\bar{D}_t = \frac{1}{|\mathcal{B}|\cdot|\mathcal{Y}|} \sum\limits_{b\in\mathcal{B}} \sum\limits_{y\in\mathcal{Y}} D_{b,t,y}$,
- $\bar{F}_h^C = \frac{1}{\sum\limits_{b\in\mathcal{B}} \sum\limits_{y\in\mathcal{Y}} |\mathcal{G}_h^{\mathrm{E}}(b,y)|} \sum\limits_{b\in\mathcal{B}} \sum\limits_{y\in\mathcal{Y}} \sum\limits_{g\in\mathcal{G}_h^{\mathrm{E}}(b,y)} F_{g,y}^{\mathrm{C}}$,
- $\bar{F}_y^C = \frac{1}{\sum\limits_{b\in\mathcal{B}} |\mathcal{G}^{\mathrm{E}}(b,y)|} \sum\limits_{b\in\mathcal{B}} \sum\limits_{g\in\mathcal{G}^{\mathrm{E}}(b,y)} F_{g,y}^{\mathrm{C}}$, and
- $\beta_0, \beta_b^{\mathrm{D}}, \beta_y^{\mathrm{D}}, \beta_t^{\mathrm{D}}, \beta_h^{\mathrm{F}}$, and β_y^{F} are regression coefficients that need to be estimated from training data.

3.3 Algorithm for high-quality scenario selection

We present the following algorithm for selecting a small number of high-quality scenarios.

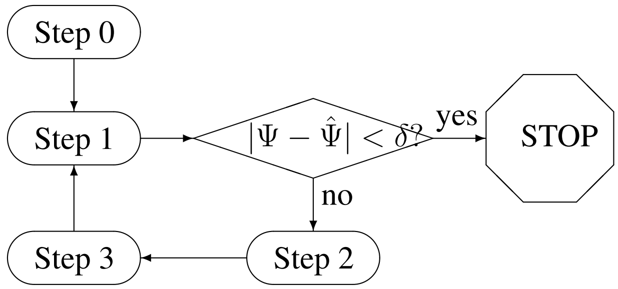


Figure 3.2: Diagram of algorithm for scenario selection.

Step 0: Initialization. Randomly select $\mathcal{S}^{\mathrm{H}}$ and set the estimated social welfare as $\hat{\Psi}_s = 0, \forall s \in \mathcal{S}^{\mathrm{H}}$. Initialize training dataset as empty.
Step 1: Social welfare calculation. Solve the GEP problem using Algorithm 1 and use its optimal solution (x^*, p^*) and equation (18) to calculate the actual social welfare for the set of high-quality scenarios $\mathcal{S}^{\mathrm{H}}$.
Checkpoint: The algorithm finishes if the error between actual and estimated social welfare is small enough and proceeds to Step 2 otherwise.
Step 2: Social welfare estimation. Add new results from Step 1 to the training dataset, obtain updated regression parameters for model (4.5), and then use the updated model to estimate social welfare values for the whole set of scenarios $\mathcal{S}^{\mathrm{W}}$.
Step 3: Scenario selection. Update the set of high-quality scenarios $\mathcal{S}^{\mathrm{H}}$ by minimizing its Kantorovich distance from $\mathcal{S}^{\mathrm{W}}$, $\mathcal{D}^{\mathrm{Kan}}(\mathcal{S}^{\mathrm{H}}, \mathcal{S}^{\mathrm{W}})$, which was defined in equation (19). Various heuristic algorithms, such as the golden section search method [4], can be used in this step. As proved by Dupacova et al. [13], the probabilities of high-quality scenarios are given by $p_s = \sum_{i \in I(s)} p_i, \forall s \in \mathcal{S}^{\mathrm{H}}$, where $I(s) = \{i \in \mathcal{S}^{\mathrm{W}} : |\hat{\Psi}_s - \hat{\Psi}_i| \leq |\hat{\Psi}_s - \hat{\Psi}_j|, \forall j \in \mathcal{S}^{\mathrm{W}}\}$.

4. Case study

We demonstrate the effectiveness of the proposed method in a case study, in which two transmission expansion plans are being evaluated for the US Eastern and Western Interconnections. A total of one million scenarios were used to represent the uncertainty of the power system over the next 15 years, and the proposed algorithm is deployed to select ten high-quality scenarios for the two transmission plans.

4.1 Data and computational settings

We use the same dataset for the US Eastern and Western Interconnections as used in [17] with some modifications. The dataset contains 169 buses, 730 transmission lines, 1,640 existing generators and 1,568 candidate generators, representing the transmission infrastructures of the North American power grid. The locations of the 169 buses are shown in Fig. 4.3. Demand for each year is divided into 19 load blocks. There are 60 generation technologies and fuel types, including coal, gas, oil, nuclear, hydro, geothermal, biomass, wind, and solar. Approximately 30% of existing generation capacity is renewable, and this ratio is required to increase by 1% each year, so that it will reach 45% by the end of year 15. One million demand and fuel cost elements of the scenarios were randomly generated with an average 1% annual growth rate for both. All algorithms were implemented in Matlab, and CPLEX 12 was used as the mixed integer linear programming solver.

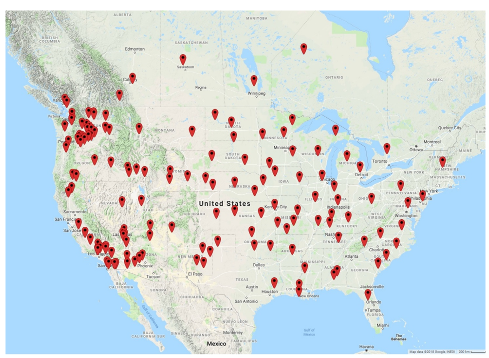


Figure 4.3: Locations of the 169 buses.

4.2 Visualization of power system status

We designed a circular figure to visualize the status of the power system, as shown in Fig. 4.4, which shows the status of the system in year 0. All buses are represented by black dots and arranged in a circle according to their bus numbers from # 1 at the 3 o'clock position to # 169 in a clockwise direction. The blue lines inside this circle represent transmission lines that connect buses. Due to the circular arrangement of the dots, the lengths of the line segments in the figure are not proportional to the actual lengths of the transmission lines. The first layer outside the buses represents demand, with the lengths of the purple bars being proportional to the loads. Dark purple was used to indicate load shedding. The second layer represents generation capacity, with the lengths of the green bars being proportional to the generation capacity. Renewable and nonrenewable generation capacities are differentiated with light green and dark green colors, respectively. The third layer represents fuel costs, with the lengths of the orange bars being proportional to the average fuel cost for the generators at the associated buses. At the price of distorted geographical locations of the buses, this figure presents all three elements (demand, generation, and fuel cost) of scenarios considered in this study, allowing the decision maker to have an overview of the system status at once.

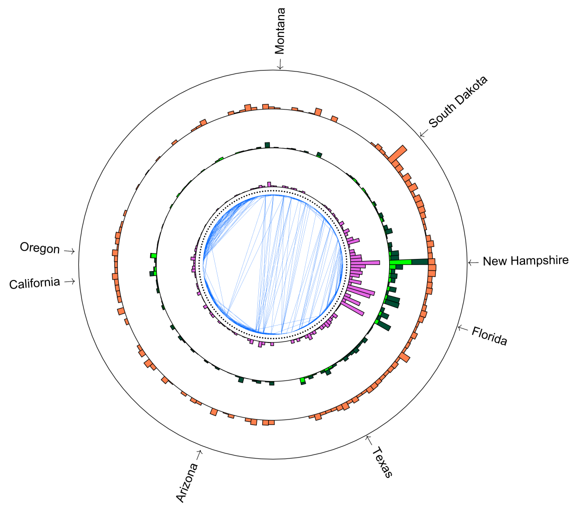


Figure 4.4: Status of power system in year 0. The inner circle represents the transmission network, and the outer layers with purple, green, and orange bars represent demand, generation capacity, and fuel cost, respectively.

4.3 Validation of algorithm for social welfare estimation

The algorithm for social welfare estimation from Section 3.2 was used in Step 2 of the algorithm for high-quality scenario selection. It took approximately 10 minutes to calculate the actual social welfare for one scenario. Using a termination criterion as $\max_s \frac{|\Psi_s - \hat{\Psi}_s|}{\Psi_s} < 1\%$ for the training scenarios, the algorithm finished after calculating the actual social welfare for 50 scenarios (in approximately 8 hours), which would have taken 19 years for the whole set of 1 million scenarios. To validate the quality of social welfare estimation, we calculated the actual social welfare for 100 random scenarios that were not used in the training set using Algorithm 1. Fig. 4.5 compared the estimated and actual social welfare of these 100 scenarios, and the minimum, mean, and maximum of $\frac{|\Psi_s - \hat{\Psi}_s|}{\Psi_s}$ were 0.00%, 0.35%, and 0.95%, respectively, for these 100 scenarios.

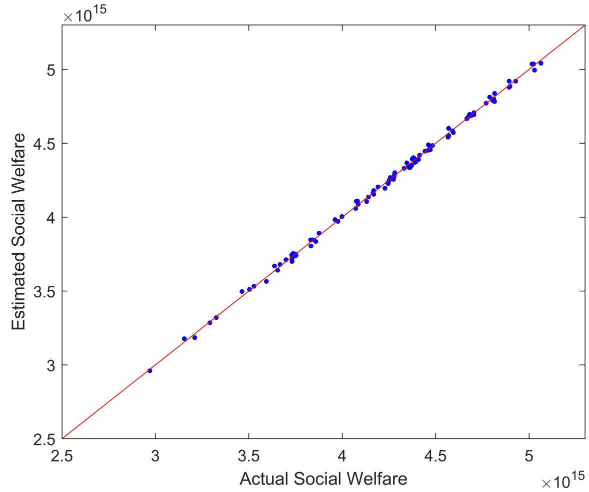


Figure 4.5: The performance of the social welfare estimation model on 100 validation scenarios.

4.4 Scenario selection for two transmission plans

We demonstrate the effectiveness of the scenario selection algorithm in Section 3.3 using two transmission plans.

- **Transmission plan 1:** existing transmission capacity will stay constant for the next 15 years without any expansion.
- **Transmission plan 2:** the capacities of 26 congested lines will be doubled in year 5, and another 27 congested lines will be doubled in year 10.

Ten high-quality scenarios were selected for each transmission plan, and four of them are shown in Figure 4.6. These figures illustrate how the power system would evolve in both the temporal and spatial dimensions with different transmission capacities under different scenarios. Although the selected scenarios demonstrated temporal and spatial correlations, it is not straightforward to interpret their social welfare implications, which demonstrated the need for the proposed algorithms for selecting high-quality scenarios.

As a comparison with the proposed algorithm, we also used three other approaches to select ten scenarios for the two transmission plans:

- **Random selection:** ten scenarios were randomly selected from the whole set of one million scenarios.
- **K-means:** ten scenarios were selected using the K-means method [28] based on demand and fuel cost information of the whole set of one million scenarios.

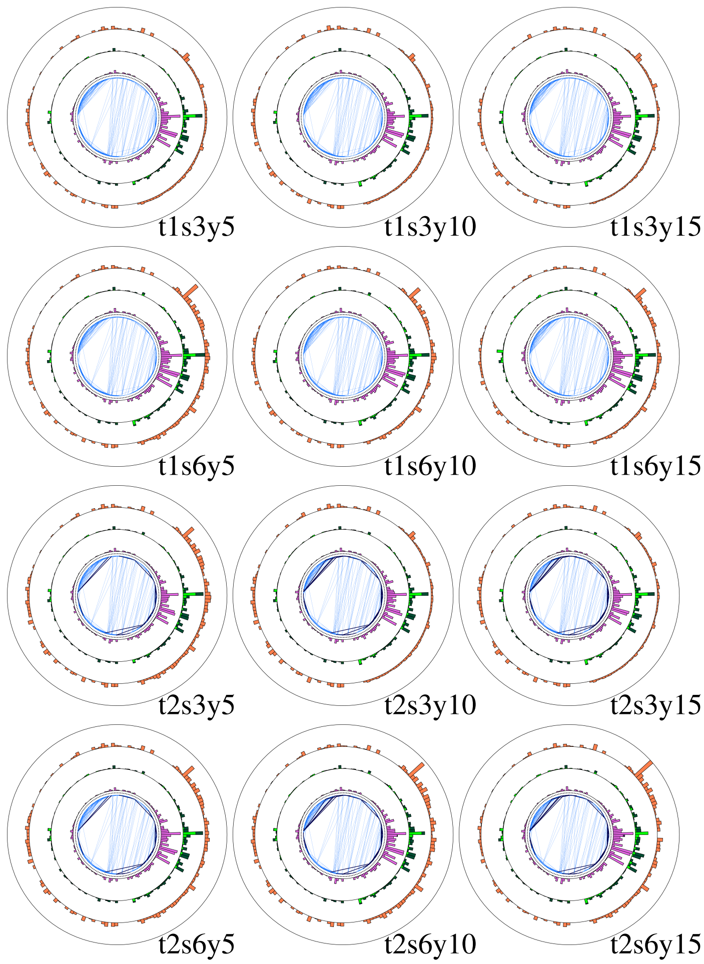


Figure 4.6: High-quality scenarios for transmission plans 1 (t1) and 2 (t2), under scenarios 3 (s3) and 6 (s6), and for years 5 (y5), 10 (y10), and 15 (y15).

- **K-medoids:** ten scenarios were selected using the K-medoids method [46] based on demand and fuel cost information of the whole set of one million scenarios.

Table 4.4: Social welfare (Ψ, in \$$10^{15}$) and probability ($p^{\mathrm{H}}$) of 10 selected scenarios and the Kantorovich distance (D^{Kan} in \$$10^{19}$) for transmission plan 1.

Scenario	Our approach		Random		K-means		K-medoids	
	Ψ	p^{H}	Ψ	p^{H}	Ψ	p^{H}	Ψ	p^{H}
1	3.35	0.04	3.68	0.16	3.44	0.06	3.56	0.09
2	3.68	0.08	4.03	0.21	3.72	0.13	3.82	0.13
3	3.88	0.10	4.31	0.15	4.02	0.14	4.09	0.15
4	4.07	0.13	4.37	0.04	4.23	0.16	4.26	0.14
5	4.23	0.13	4.38	0.01	4.35	0.03	4.26	0.04
6	4.36	0.12	4.38	0.11	4.39	0.18	4.44	0.18
7	4.47	0.12	4.58	0.14	4.55	0.11	4.56	0.09
8	4.58	0.11	4.66	0.05	4.58	0.06	4.68	0.10
9	4.72	0.11	4.71	0.04	4.68	0.11	4.81	0.06
10	4.93	0.05	4.79	0.09	4.92	0.04	4.93	0.04
D^{Kan}	4.29		7.11		5.22		5.31	

Table 4.5: Social welfare (Ψ, in \$$10^{15}$) and probability ($p^{\mathrm{H}}$) of 10 selected scenarios and the Kantorovich distance (D^{Kan} in \$$10^{19}$) for transmission plan 2.

Scenario	Our approach		Random		K-means		K-medoids	
	Ψ	p^{H}	Ψ	p^{H}	Ψ	p^{H}	Ψ	p^{H}
1	3.35	0.04	3.68	0.16	3.44	0.06	3.57	0.09
2	3.68	0.08	4.03	0.21	3.73	0.13	3.82	0.13
3	3.88	0.10	4.31	0.14	3.99	0.14	4.06	0.15
4	4.07	0.13	4.37	0.04	4.23	0.16	4.27	0.14
5	4.24	0.13	4.38	0.01	4.36	0.03	4.27	0.04
6	4.37	0.12	4.38	0.11	4.40	0.18	4.45	0.18
7	4.47	0.12	4.58	0.14	4.58	0.06	4.58	0.09
8	4.59	0.11	4.66	0.05	4.59	0.11	4.66	0.10
9	4.72	0.11	4.71	0.05	4.69	0.11	4.82	0.06
10	4.93	0.05	4.79	0.09	4.94	0.04	4.92	0.04
D^{Kan}	4.22		7.09		5.40		5.32	

Results from these methods are summarized in Tables 4.4 and 4.5, in which the ten selected scenarios from each method are sorted in ascending order of social welfare and listed in the same ten rows (e.g., the row for scenario 1 shows four different worst case scenarios selected using the four methods). Computationally, the proposed approach took approximately 8 hours, whereas the other methods only took seconds.

4.5 Cost benefit analysis of transmission expansion

The analysis from Section 4.4 enables cost benefit analysis of transmission plan 2. As shown in Table 4.6, the difference between average social welfare across the whole set of one million scenarios for the two transmission plans is \$5.23 trillion, which reveals the social benefit of transmission plan 2 throughout the 15-year horizon. The proposed approach estimated this difference as \$4.99 trillion using the 10 high-quality scenarios. In comparison, the K-means approach over-estimated this value as \$7.21 trillion, K-medoids under-estimated it as \$2.16 trillion, and the random selection approach estimated this value to be negligible. These results suggested that the proposed approach was able to estimate the social benefit of transmission plans reasonably accurately, which also allows it to be potentially integrated into the iterative stochastic TEP framework proposed in Fig. 1.1.

Table 4.6: Average social welfare (in $\$10^{15}$) for the two transmission plans estimated using different methods.

	Plan 1	Plan 2	Difference
All 10^6 scenarios	4.26234	4.26757	0.00523
10 scenarios, our approach	4.22598	4.23097	0.00499
10 random scenarios	4.38840	4.38840	0.00000
10 scenarios, K-means	4.28801	4.29522	0.00721
10 scenarios, K-medoids	4.33987	4.34203	0.00216

5. Concluding remarks

In this section we discuss in detail the analytical basis for Scenario Selection for Iterative Stochastic TEP, analyze the case study results and provide an comparsion with the sata of the art methods.

The main contributions of this paper include the iterative stochastic TEP framework and a scenario selection method for transmission planning. The new framework aimed to combine the strengths of both robust TEP and stochastic TEP. The new scenario selection method had three salient features: (1) correlations between generation capacity and demand, fuel cost, and transmission capacity are explicitly captured, (2) high-quality scenarios are selected to minimize the Kantorovich distance of social welfare distributions between selected and the whole set of scenarios, and (3) the set of high-quality scenarios is specific for each transmission plan, which enables this approach to interact iteratively with a stochastic TEP model in the proposed framework.

A case study was conducted to demonstrate the proposed approach, in which a 169-bus network was used to represent the US Eastern and Western Interconnections. When compared with three other methods for scenario selection, our algorithm was found to provide a more accurate estimation of the economic value of transmission plans.

The proposed approach is not without its limitations and caveats. For example, the generation expansion planning model was a simplified estimation of the actual decision-making process, which involves far more realistic and complex constraints such as multiple decision makers and risk hedging constraints. We also acknowledge that there are numerous other sources of uncertainty in realistic transmission planning projects, which are not explicitly taken into consideration in the proposed approach, such as investment cost, growth rate of distributed energy resources, renewable energy production, energy policies, etc. Besides Kantorovich distance, other probabilistic distances could also be used as the selection criterion for high-quality scenarios. Moreover, the social welfare evaluation could be extended to include environmental and reliability benefits of the power system. Co-optimization of GEP and TEP is another potential topic for future research.

References

[1] Jamshid Aghaei, Mohammad Amin Akbari, Alireza Roosta, and Amir Baharvandi. Multiobjective generation expansion planning considering power system adequacy. *Electric Power Systems Research*, 102:8–19, 2013.

[2] Ali Bagheri, Jianhui Wang, and Chaoyue Zhao. Data-driven stochastic transmission expansion planning. *IEEE Transactions on Power Systems*, 32(5):3461–3470, 2017.

[3] Jonathan F Bard and James T Moore. A branch and bound algorithm for the bilevel programming problem. *SIAM Journal on Scientific and Statistical Computing*, 11(2):281–292, 1990.

[4] Yen-Ching Chang. N-dimension golden section search: Its variants and limitations. *In 2009 2nd International Conference on Biomedical Engineering and Informatics*, pages 1–6. IEEE, 2009.

[5] Bokan Chen, Jianhui Wang, Lizhi Wang, Yanyi He, and Zhaoyu Wang. Robust optimization for transmission expansion planning: Minimax cost vs. minimax regret. *IEEE Transactions on Power Systems*, 29(6):3069–3077, 2014.

[6] Bokan Chen and Lizhi Wang. Robust transmission planning under uncertain generation investment and retirement. *IEEE Transactions on Power Systems*, 31(6):5144–5152, 2016.

[7] Sebasti´an De La Torre, Antonio J Conejo, and Javier Contreras. Transmission expansion planning in electricity markets. *IEEE transactions on power systems*, 23(1):238–248, 2008.

[8] Shahab Dehghan and Nima Amjady. Robust transmission and energy storage expansion planning in wind farm-integrated power systems considering transmission switching. *IEEE Transactions on Sustainable Energy*, 7(2):765–774, 2016.

[9] Shahab Dehghan, Nima Amjady, and Antonio J Conejo. Adaptive robust transmission expansion planning using linear decision rules. *IEEE Transactions on Power Systems*, 32(5):4024– 4034, 2017.

[10] Shahab Dehghan, Ahad Kazemi, and Nima Amjady. Multi-objective robust transmission expansion planning using information-gap decision theory and augmented constraint method. *IET Generation, Transmission & Distribution*, 8(5):828–840, 2014.

[11] Ershun Du, Ning Zhang, Chongqing Kang, Jianhua Bai, Lu Cheng, and Yi Ding. Impact of wind power scenario reduction techniques on stochastic unit commitment. *In Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO), 2016 Second International Symposium on*, pages 202–210. IEEE, 2016.

[12] Jitka Dupaˇcov´a, Giorgio Consigli, and Stein W Wallace. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100(1-4):25–53, 2000.

[13] Jitka Dupaˇcov´a, Nicole Gr¨owe-Kuska, andWerner R¨omisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3):493–511, 2003.

[14] Ali Ehsan, Qiang Yang, and Ming Cheng. A scenario-based robust investment planning model for multi-type distributed generation under uncertainties. *IET Generation, Transmission & Distribution*, 12(20):4426–4434, 2018.

[15] Yonghan Feng and Sarah M Ryan. Scenario construction and reduction applied to stochastic power generation expansion planning. *Computers & Operations Research*, 40(1):9–23, 2013.

[16] Yonghan Feng and SarahMRyan. Solution sensitivity-based scenario reduction for stochastic unit commitment. *Computational Management Science*, 13(1):29–62, 2016.

[17] Armando Luis Figueroa-Acevedo. Opportunities and benefits for increasing transmission capacity between the US Eastern and Western interconnections. PhD thesis, Iowa State University, 2017.

[18] José Fortuny-Amat and Bruce McCarl. A representation and economic interpretation of a two-level programming problem. *The Journal of the operational Research Society*, pages 783–792, 1981.

[19] Lina P Garcés, Antonio J Conejo, Raquel García-Bertrand, and Rubén Romero. A bilevel approach to transmission expansion planning within a market environment. *IEEE Transactions on Power Systems*, 24(3):1513–1522, 2009.

[20] Esteban Gil, Ignacio Aravena, and Raúl Cárdenas. Generation capacity expansion planning under hydro uncertainty using stochastic mixed integer programming and scenario reduction. *IEEE Transactions on Power Systems*, 30(4):1838–1847, 2015.

[21] Nicole Growe-Kuska, Holger Heitsch, and Werner Romisch. Scenario reduction and scenario tree construction for power management problems. *In 2003 IEEE Bologna Power Tech Conference Proceedings*, volume 3, pages 7–pp. IEEE, 2003.

[22] Reza Hemmati, Rahmat-Allah Hooshmand, and Amin Khodabakhshian. State-of-the-art of transmission expansion planning: Comprehensive review. *Renewable and Sustainable Energy Reviews*, 23:312–319, 2013.

[23] VH Hinojosa and J Velásquez. Stochastic security-constrained generation expansion planning based on linear distribution factors. *Electric Power Systems Research*, 140:139–146, 2016.

[24] Victor Hinojosa, Esteban Gil, and Ignacio Calle. A stochastic generation capacity expansion planning methodology using linear distribution factors and hourly load modeling. *In 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6. IEEE, 2018.

[25] Masoud Jenabi, Seyyed Mohammad Taghi Fatemi Ghomi, and Yves Smeers. Bi-level game approaches for coordination of generation and transmission expansion planning within a market environment. *IEEE Transactions on Power systems*, 28(3):2639–2650, 2013.

[26] LV Kantorovich. On the transfer of masses, dok 1. Akad. Nauk SSSR, 37:227–229, 1942.

[27] Jinghua Li, Liu Ye, Yan Zeng, and Hua Wei. A scenario-based robust transmission network expansion planning method for consideration of wind power uncertainties. *CSEE Journal of Power and Energy Systems*, 2(1):11–18, 2016.

[28] Chaofan Lin, Chengzhi Fang, Yonglin Chen, Shiyu Liu, and Zhaohong Bie. Scenario generation and reduction methods for power flow examination of transmission expansion planning. *In Power and Energy Systems (ICPES), 2017 IEEE 7th International Conference on*, pages 90–95. IEEE, 2017.

[29] Yixian Liu, Ramteen Sioshansi, and Antonio J Conejo. Multistage stochastic investment planning with multiscale representation of uncertainties and decisions. *IEEE Transactions on Power Systems*, 33(1):781–791, 2018.

[30] Juan Alvarez López, Kumaraswamy Ponnambalam, and Víctor H Quintana. Generation and transmission expansion under risk using stochastic programming. *IEEE Transactions on Power Systems*, 22(3):1369–1378, 2007.

[31] Sara Lumbreras and Andrés Ramos. The new challenges to transmission expansion planning. Survey of recent practice and literature review. *Electric Power Systems Research*, 134:19–29, 2016.

[32] Bryan Palmintier and Mort Webster. Impact of unit commitment constraints on generation

expansion planning with renewables. *In Power and Energy Society General Meeting, 2011 IEEE*, pages 1–7. IEEE, 2011.
[33] Anthony Papavasiliou and Shmuel S Oren. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, 61(3):578–592, 2013.
[34] Heejung Park and Ross Baldick. Transmission planning under uncertainties of wind and load: Sequential approximation approach. *IEEE Transactions on Power Systems*, 28(3):2395– 2402, 2013.
[35] Heejung Park, Ross Baldick, and David P Morton. A stochastic transmission planning model with dependent load and wind forecasts. *IEEE Transactions on Power Systems*, 30(6):3003–3011, 2015.
[36] SangWoo Park, Qingyu Xu, and Benjamin Hobbs. Comparing scenario reduction methods for stochastic transmission planning. *IET Generation, Transmission & Distribution*, 2019.
[37] Salvador Pineda and Juan M Morales. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, 33(6):7162– 7170, 2018.
[38] Seyyed A Rashidaee, Turaj Amraee, and Mahmud Fotuhi-Firuzabad. A linear model for dynamic generation expansion planning considering loss of load probability. *IEEE Transactions on Power Systems*, 33(6):6924–6934, 2018.
[39] Enzo E Sauma and Shmuel S Oren. Proactive planning and valuation of transmission investments in restructured electricity markets. *Journal of Regulatory Economics*, 30(3):358–387, 2006.
[40] Enzo E Sauma and Shmuel S Oren. Economic criteria for planning transmission investment in restructured electricity markets. *IEEE Transactions on Power Systems*, 22(4):1394–1405, 2007.
[41] GB Shrestha and PAJ Fonseka. Congestion-driven transmission expansion in competitive power markets. *IEEE Transactions on Power Systems*, 19(3):1658–1665, 2004.
[42] I de J Silva, Marcos J Rider, Rub´en Romero, and Carlos AF Murari. Transmission network expansion planning considering uncertainty in demand. *IEEE Transactions on Power Systems*, 21(4):1565–1573, 2006.
[43] M Sun, F Teng, I Konstantelos, and G Strbac. An objective-based scenario selection method for transmission network expansion planning with multivariate stochasticity in load and renewable energy sources. *Energy*, 145:871–885, 2018.
[44] Mingyang Sun, Jochen Cremer, and Goran Strbac. A novel data-driven scenario generation framework for transmission expansion planning with high renewable energy penetration. *Applied Energy*, 228:546–555, 2018.
[45] Jianhui Wang, Mohammad Shahidehpour, Zuyi Li, and Audun Botterud. Strategic generation capacity expansion planning with incomplete information. *IEEE Transactions on Power Systems*, 24(2):1002–1010, 2009.
[46] Qun Wang, Wenlue Dong, and Li Yang. A wind power/photovoltaic typical scenario set generation algorithm based on Wasserstein distance metric and revised k-medoids cluster. *In Proceedings of the CSEE*, volume 35, pages 2654–2661, 2015.
[47] Ying Zhou, Lizhi Wang, and James D McCalley. Designing effective and efficient incentive policies for renewable energy in generation expansion planning. *Applied Energy*, 88(6):2201– 2209, 2011.

Part II

Improving Computational Efficiency in Adaptive Expansion Planning Problems for Handling Uncertainty

James McCalley
Abhinav Venkatraman, Graduate Student
Patrick Maloney, Graduate Student

Iowa State University

For information about this project, contact:

James McCalley
Department of Electrical and Computer Engineering
Iowa State University
Ames, Iowa 50011
Phone: 515-294-4944
Email: jdm@iastate.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

Table of Contents

List of Figures

List of Tables

1. Introduction

This report describes the work performed by J.McCalley, A.Venkatraman, and P.Maloney in the project "M37 – Development of expansion planning methods and tools for handling uncertainty." The objective of this part of the work is to identify ways to increase computational efficiency in a particular form of expansion planning problems under uncertainty. This particular form has been developed at Iowa State University and is called Adaptive Expansion Planning (AEP).

In this introductory chapter, we describe three different forms of deterministic expansion planning problems in Section 1.1. Section 1.2 provides a characterization of uncertainty together with two different ways of handling it within expansion planning problems. Section 1.3 describes the system used in this project to test the developed approaches. Section 1.4 motivates the objective associated with the work described herein. The contents of the rest of the report are described in Section 1.5.

1.1 Deterministic expansion planning problems

The power system community began developing expansion planning applications in the early 1970's. The initial efforts concentrated on solving the generation expansion planning (GEP) problem. A simplified expression of the GEP problem is given in (1)

min

$$\text{NPW}\left\{\sum_{t}^{N}\text{InvCosts}(\boldsymbol{x}_G(t))+\sum_{t}^{N}\text{OpCosts}(\boldsymbol{x}_G(t))\right\} \qquad \textbf{(1)}$$

subject to

$$\text{Constraints}(\boldsymbol{x}_G(t), \boldsymbol{p}(t)), \quad t=1,\dots N$$

where the objective function expresses net present worth (NPW) over N years of investment and operational costs; the vector $\boldsymbol{x}_G(t)$ represents the possible generation investments at each year $t=1,\dots,N$; and "Constraints" include constraints on reliability (reserve requirements), capacity updates from year to year, forced retirements or builds, and operations (DC power flow equations, and generation and line flow limits). These constraints depend on the decision variables (in this case, $\boldsymbol{x}_G(t)$) and on the parameter set $\boldsymbol{p}(t)$, both of which vary through the decision horizon as indicated by their dependence on the year t. An important modeling feature is how the operational cost term in the objective function is computed. A rigorous approach would perform an 8760 hour production cost evaluation for every year, but doing so is computationally intense. Therefore, the operational costs are approximated by evaluating them over a limited number of operating conditions. A comprehensive reference was developed in 1984 on the GEP problem [1].

The simplest version of the GEP problem does not represent transmission at all, implying all generation is represented at a single busbar. An improved version represents existing transmission, and requires auxiliary angle variables, $\boldsymbol{\theta}(t)$, but it gives no option for expanding the transmission system. Initial steps in formulating and solving problems involving transmission investment were taken in the 1980's. The resulting problem was referred to as the transmission expansion planning (TEP) problem. A simplified version of the TEP problem may be stated as in (2):

$$
\begin{aligned}
&\min \\
&\quad \mathrm{NPW}\left\{\sum_{t}^{N}\mathrm{InvCosts}(\boldsymbol{x}_T(t))+\sum_{t}^{N}\mathrm{OpCosts}(\boldsymbol{x}_T(t))\right\} \qquad \textbf{(2)}\\
&\text{subject to}\\
&\quad \mathrm{Constraints}(\boldsymbol{x}_T(t),\boldsymbol{p}(t),\boldsymbol{\theta}(t)),\quad t=1,\ldots N
\end{aligned}
$$

where the main difference between the formulations for the GEP and TEP problems lies in the identity of the decision variables. In the GEP problem, the decision variables, $\boldsymbol{x}_G$, are the capacities associated with the generation expansion options. In the TEP problem, the decision variables, $\boldsymbol{x}_T$, may represent capacities associated with the transmission expansion options but only if a transportation model (which looks like a capacitated pipe and thus ignores the effects of impedance, similar to a DC line) is used for transmission investments. Capturing the effects of impedance in a single optimization problem requires[1] that the decision variables $\boldsymbol{x}_T$ be binary (representing whether each invested line is built or not), but in this case, the problem becomes a mixed integer nonlinear program (MINLP). Nonlinearities arise because of product terms between the phasor angles (of buses terminating the invested circuit) and the binary investment decision variable of the circuit. This nonlinearity can be avoided without loss of modeling fidelity via use of the so-called disjunctive model (also called the "Big M" model), in which case the problem is a mixed integer linear program (MILP). Although the MILP is considerably more tractable than the MINLP, it is still computationally daunting for even modest-size network representations (e.g., having ~500-1000 buses), much less those used in normal industry planning (e.g., having 20000-70000 buses).

The co-optimized expansion planning (CEP) problem combines the GEP and the TEP to identify both generation and transmission expansion. A simple approach to doing this is to iterate on sequential GEP and TEP solutions, usually solving the GEP first because it is most often the most costly; such an approach is illustrated in Figure 1.1. Indeed, Figure 1.1 not only indicates a software application of GEP and TEP to achieve a CEP, but it also indicates what has generally been done in the industry in previous years, via human processes.

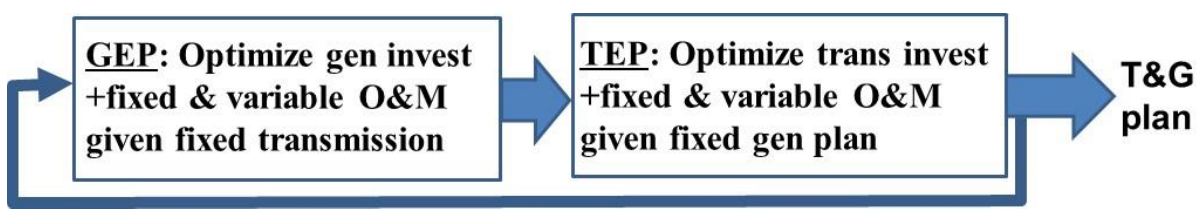


Figure 1.1: Iterative CEP

Although the iterative CEP generally approaches the optimal solution, the number of iterations to get close can be computationally expensive [2, 3]. A better approach is to solve both problems within a single optimization, as indicated in (3):

---

[1] One may also avoid solving the nonlinear TEP problem by using an algorithm that iterates on a linearized TEP problem, updating the linearization in each iteration.

$$\min \; \text{NPW}\left\{\sum_{t}^{N}\text{InvCosts}(\boldsymbol{x}_G(t), \boldsymbol{x}_T(t)) + \sum_{t}^{N}\text{OpCosts}(\boldsymbol{x}_G(t), \boldsymbol{x}_T(t))\right\} \tag{3}$$

subject to

$$\text{Constraints}(\boldsymbol{x}_G(t), \boldsymbol{x}_T(t), \boldsymbol{p}(t), \boldsymbol{\theta}(t)), \quad t = 1, \ldots N$$

or, by representing both generation and transmission decision variables as

$$\boldsymbol{x}_{GT}(t) = \begin{bmatrix} \boldsymbol{x}_G(t) \\ \boldsymbol{x}_T(t) \end{bmatrix}$$

we may express (3) more compactly as

$$\min \; \text{NPW}\left\{\sum_{t}^{N}\text{InvCosts}(\boldsymbol{x}_{GT}(t)) + \sum_{t}^{N}\text{OpCosts}(\boldsymbol{x}_{GT}(t))\right\} \tag{4}$$

subject to

$$\text{Constraints}(\boldsymbol{x}_{GT}(t), \boldsymbol{p}(t), \boldsymbol{\theta}(t)), \quad t = 1, \ldots N$$

The CEP problem may also be extended to include expansion in distribution systems and distributed energy resources (DER) [4], natural gas systems [5], and transportation systems [6]. Water system operation and expansion may also be cooptimized with electric system expansion, and though there is little published work to this end, references [7, 8, 9] motivate the need.

Co-optimization is the simultaneous identification of two or more related classes of investment decisions within one optimization strategy. Since the decision to build generation at a certain location affects the decision to build or expand transmission at that location, a co-optimization of these two related decisions must be as good as, or better than if they were identified through sequential (or iterative) solution of GEP and TEP. Thus, solving CEP is consistent with the goals of a vertically integrated organization that invests in, owns, and operates generation, transmission, and distribution systems, or any two of these. In addition, it is useful for regional transmission organization (RTO) responsible for coordinating electric infrastructure plans across a region, as it identifies a "best possible" expansion plan, and in so doing provides a benchmark against which all other expansion plans may be compared. Finally, CEP is useful for organizations intending to invest in just one of generation or transmission. The reason for this is that it provides a basis to anticipate the behavior of other network investors, enabling the planner to anticipate, for example, how generation investment may respond to transmission expansion under the assumption that investors will be attracted to make investments that minimize the long-term cost of investing and operating the network [10, 11]. But only under this assumption are CEP solutions predictive. In general, CEP is better thought of as a way to explore the future, resulting in what we may call guided exploration where we as the planner impose certain constraints on the decision variables (generation, transmission investments) and/or on the parameters (e.g., demand growth, fuel prices, carbon emissions, technology cost), that characterize aspects of how we think the future may occur, observe the solution, and then we metaphorically turn the telescope, i.e., we change the constraints, and look out into another direction of interest. In considering this metaphor, we immediately recognize that, although we may have reason to believe some directions are more likely representations of the future than other directions, we must ultimately embrace the fact that our infrastructure investment decision must be made under uncertainty. We present associated problem formulations in Section 1.2.

1.2 Expansion planning under uncertainty

Expansion planning is necessarily addressing future conditions, and so uncertainty in modeling that future is unavoidable. There are various ways to account for uncertainty, the simplest of which is to model deterministically, obtain a solution, and then perform sensitivity studies to observe the level of change in the solution for an expected amount of change in a decision variable or parameter. A second approach, called robust optimization, assumes that the uncertainty space of data is constrained to an uncertainty space (or "set") and finds the best solution that is feasible for all the realizations of uncertainties that lie in the uncertainty space under consideration.

A third approach, or really a class of approaches, identifies certain possible futures, usually referred to as scenarios, and then determines investment plans that perform "well" in some sense, when exposed to one or more of the scenarios. The variation among these approaches depends on how the planner judges the performance. Many RTOs today use such scenario-based decision approaches to make transmission planning recommendations. For example, they might choose the plan that minimizes the probability-weighted summation of costs (investment and operational) across all scenarios, or they might evaluate "a range of plausible scenarios made up of different generation portfolios, and identifies the transmission reinforcements found to be necessary in a reasonable number of those scenarios," as indicated in a California ISO (CAISO) planning document [12] and as similarly expressed by the Mid-Continent ISO in a report describing its multi-value projects (MVPs) [13]. An alternative approach would choose the plan that minimizes the maximum regret across the scenarios, where regret is, for each plan and for each scenario, the difference between the cost of the plan and the cost of the least-cost plan in that scenario. Two additional scenario-based approaches are traditional stochastic programming (TSP) and adaptive expansion planning (AEP) which are further described in Sections 1.2.3 and 1.2.4; these descriptions depend on understanding the view taken for "uncertainties," a view we provide in Section 1.2.1. In Section 1.2.4, we compare and contrast TSP and AEP.

1.2.1 Characterization of uncertainty

We view that uncertainties may be classified as global and local. Global uncertainties are uncertainties for which different realizations within the range of likely values produce significantly different results. Typical examples for expansion planning problems include emissions policies, large demand shifts, coal or nuclear retirements, extremes in fuel prices, extended drought, and dramatic change in technology investment costs. In contrast, local uncertainties are uncertainties characterized by a range of values a parameter may take under a global realization; for example, under a "low" load growth or fuel price scenario, the annual load growth may vary ± 0.5 % and the annual fuel price change may vary ± 1%. Local uncertainties may also be referred to as parametric uncertainties. The difference between these two classes of uncertainties is illustrated in Figure 1.2.

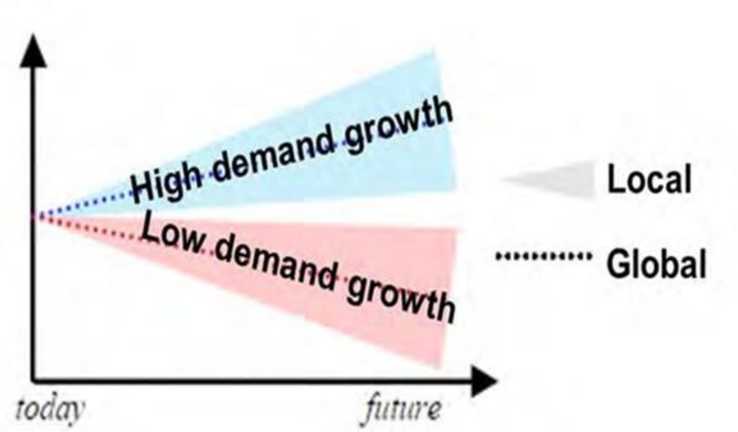


Figure 1.2: Local and global uncertainties, using demand growth as an exemplar

1.2.2 Traditional stochastic programming (TSP)

A formulation of TSP for co-optimizing both generation and transmission is given in (5):

$$\min_{x_{GT1}} \text{NPW}\left[\left\{\text{InvCosts}(\boldsymbol{x}_{GT1}) + \sum_{t=1}^{N_1}\text{OpCosts}(\boldsymbol{x}_{GT1})\right\} + \sum_{s=1}^{N_S}\Pr_s \times \min_{x_{GT2s}}\left\{\text{InvCosts}(\boldsymbol{x}_{GT2s} \mid \boldsymbol{x}_{GT1}) + \sum_{t=N_1+1}^{N}\text{OpCosts}(\boldsymbol{x}_{GT2s} \mid \boldsymbol{x}_{GT1})\right\}\right]$$

subject to

$$\left\{\text{Constraints1}(\boldsymbol{x}_{GT1}, \boldsymbol{p}_1(t), \boldsymbol{\theta}_1(t)),\ \ t=1,\ldots N_1\right\}$$
$$\left\{\text{Constraints2s}(\boldsymbol{x}_{GT2s} \mid \boldsymbol{x}_{GT1}), \boldsymbol{p}_{2s}(t), \boldsymbol{\theta}_{2s}(t)),\ \ t=N_1+1,\ldots N\right\},\ \ s=1\ldots N_s$$

(5)

The two-stage TSP is characterized by ***two decision points*** separated by the occurrence of ***uncertainty***. The two decision points are represented by two sets of decision variables:

- Here and now variables $\boldsymbol{x}_{GT1}$: These variables represent investment decisions made in the first decision period, $t=1,\ldots,N_1$. They are subject to the constraints associated with the first decision period as denoted by Constraints1. Decisions in this period are made in ignorance of what will occur in the future.
- Wait and see variables $\boldsymbol{x}_{GT2s}$: These variables represent investment decisions made in the second decision period, $t=N_1+1,\ldots,N$. We consider that the second decision period is subject to uncertainty, as characterized by scenarios $s=1,\ldots,N_s$. We also consider that these decisions are conditioned on the decisions made in the first decision period, which gives rise to the notation $\boldsymbol{x}_{GT2s} \mid \boldsymbol{x}_{GT1}$. These decisions are made after the future is revealed, i.e., after the scenario is known, and so the $\boldsymbol{x}_{GT2s}$ are often referred to as recourse[2] variables in that they

[2] In normal conversational usage, the word "recourse" may be loosely understood to be an action that taken by a person once the person has made some decision to bring trouble to themself.

enable the decision maker to adjust for the fact that the first state decisions had to be made without knowing which scenario would occur. These decisions are subject to constraints associated with the second decision period, given the knowledge of the scenario that occurred (as indicated by the parameter *s*), as denoted by Constraints2s, and given the decisions that were made in the first decision period.

The uncertainty is represented by N_S possible scenarios $s=1,\ldots,N_S$, each of which is associated with a probability Pr_S. Intuitive insight into TSP is gained by visualizing it via a tree structure, as shown in Figure 1.3. Here, the boxes represent decision periods, with each branch emanating from a box corresponding to the specific choice indicated in the oval. The ovals represent chance nodes, with each branch emanating from an oval corresponding to a specific scenario (i.e., a set of realizations of what is considered uncertain in the problem). In this particular TSP tree, there are three possible choices in the first decision period, the uncertainty is captured by two scenarios, and there are two possible (recourse) choices in the second decision period.

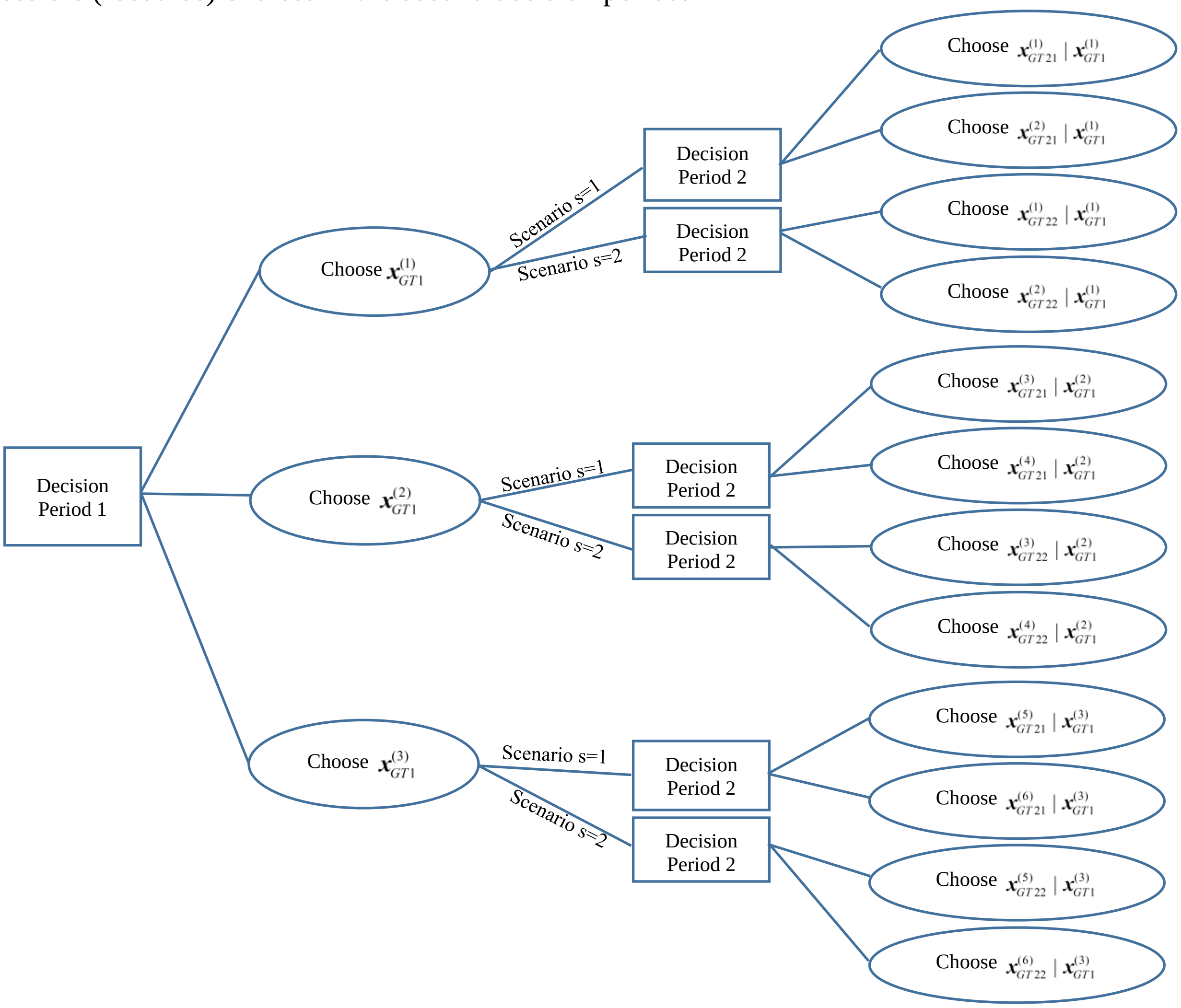


Figure 1.3: Tree structure of a TSP

1.2.3 Adaptive Expansion Planning (AEP)

A formulation of AEP for co-optimizing both generation and transmission is given in (6):

$$\min_{\boldsymbol{x}_{GT}(t),\Delta\boldsymbol{x}_{GTs}(t)} \mathrm{NPW}\left[\underbrace{\sum_{t=1}^{N}\mathrm{CoreCosts}(\boldsymbol{x}_{GT}(t))}_{\text{Cost of core investments}}+\beta\times\underbrace{\sum_{t=1}^{N}\sum_{s=1}^{N_s}\mathrm{Pr}_s\times\mathrm{AdaptationCosts}(\Delta\boldsymbol{x}_{GTs}(t))}_{\text{Expected cost of adaptation investments}}+\underbrace{\sum_{t=1}^{N_1}\mathrm{Pr}_s\times\mathrm{OpCosts}(\boldsymbol{x}_{GT}(t)+\Delta\boldsymbol{x}_{GTs}(t))}_{\text{Expected operating cost}}\right]$$

subject to

$$\left\{\mathrm{Constraints1}(\boldsymbol{x}_{GT}(t)),\quad t=1,\ldots N\right\}$$

$$\left\{\mathrm{Constraints2s}(\boldsymbol{x}_{GT}(t)+\Delta\boldsymbol{x}_{GTs}(t),\boldsymbol{p}_s(t),\boldsymbol{\theta}_s(t)),\quad t=1,\ldots N\right\},\ s=1\ldots N_s$$

(6)

The objective in adaptive expansion planning (AEP) is to identify a core set of investments, as represented by $\boldsymbol{x}_{GT}(t),\ t=1,...,N$, through the entire decision horizon that are "good" relative to the set of specific scenarios considered. There is, of course, a cost to making the core set of investments. The measure of how "good" the core is relative to a single scenario is quantified by the cost of adapting the core to the conditions of that scenario via a set of adaptation investments $\Delta\boldsymbol{x}_{GTs}(t)$. Thus, the objective in AEP is to minimize the cost of the core investments, the expected cost of the adaptations, and the expected operational cost.

We associate to the analytical description three illustrations. Consider Figure 1.4, where $\boldsymbol{x}_{GT}$ is a chosen plan in the "investment space." It is a vector such that each element of the vector consists of a capacity investment of the generation resource or transmission circuit corresponding to that element. Assume that we identify the plan $\boldsymbol{x}_{GT}$ deterministically, under a single scenario that we think accurately characterizes the future. We think of $\boldsymbol{x}_{GT}$ as the core investment, i.e., it is what we actually build, and there is a cost to build it, denoted by $\mathrm{CoreCosts}(\boldsymbol{x}_{GT}(t))$. After we build it, we discover that the future scenario we expected does not happen, i.e., we were wrong, and the real future differs from the scenario we used in identifying $\boldsymbol{x}_{GT}$. We refer to the scenario that actually occurs as scenario *s*. Thus, we need to adapt our investment plan $\boldsymbol{x}_{GT}$ so that it is feasible[3] under the conditions of scenario *s*, and this requires that we make investments denoted by $\Delta\boldsymbol{x}_{GTs}$. Then the "point" in the feasible region of the "investment space" can be designated as $\boldsymbol{x}_{GT}+\Delta\boldsymbol{x}_{GTs}$, as illustrated in Figure 1.4.

[3] We may also require that the adaptation results in our adapted plan be optimal under the conditions of scenario *s*,

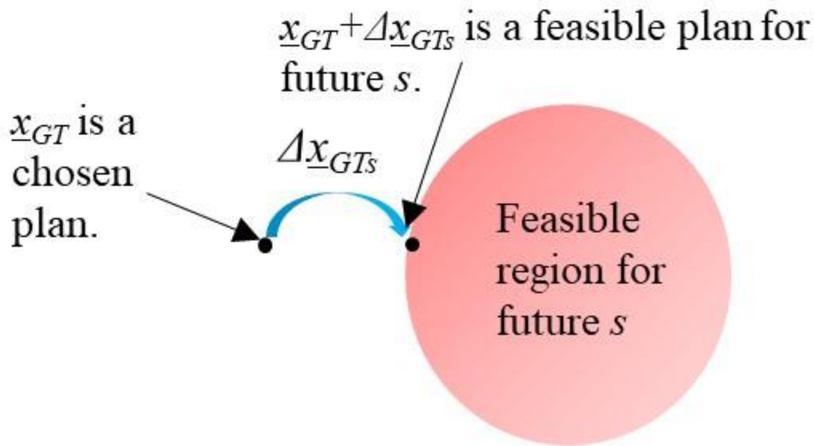


Figure 1.4: Adapting a plan

We emphasize four important terms that have been used in the previous paragraph:

- "Core" is the core investments, designated by $\boldsymbol{x}_{GT}$.
- "Core Cost" is the cost of the core investment, designated by $\text{CoreCosts}(\boldsymbol{x}_{GT}(t))$
- "Adaptation" is the additional investment necessary for plan $\underline{x}$ to acceptably perform under scenario *s*. It is designated by $\Delta\boldsymbol{x}_{GTs}$.
- "Adaptation Cost" of adapting a plan $\boldsymbol{x}_{GT}$ to a scenario *s* is the minimum cost to move $\boldsymbol{x}_{GT}$ to a feasible design in scenario *s*. It measures the additional cost to plan $\boldsymbol{x}_{GT}$ if scenario *s* happens. It is designated by $\text{AdaptationCosts}(\Delta\boldsymbol{x}_{GTs}(t))$. We observe that $\Delta\boldsymbol{x}_{GTs} = 0$ if plan $\boldsymbol{x}_{GT}$ is designed under scenario *s*.

We now take an additional step by considering that we are uncertain about the future in that we believe one of several scenarios could happen: any of scenarios $s = 1$, $s = 2$, ..., or $s = N_s$ could happen. Assuming we can obtain (or assign) probabilities to each scenario, $\Pr_s$, then we desire to identify a core plan that minimizes core costs, expected value of the adaptation costs, and expected value of operational costs. This is the objective function in (6).

The AEP problem expressed in (6) identifies a core investment plan $\boldsymbol{x}_{GT}$ that is "positioned" in investment space so as to minimize the cost of the core plus the expectation of the cost of adapting the core to all of the scenarios (including the operating cost in each scenario). In a sense, the identified core investment is centroidal to the deterministic investment for each scenario. Figure 1.5 illustrates three "investment trajectories," one corresponding to scenario 1, another for scenario 2, and a third, in blue, for the core. We also observe the yellow cylinders at each time $t = 1, 2, 3$ of the decision horizon; these yellow cylinders represent the adaptation necessary at those times to make the core feasible under each respective scenario.

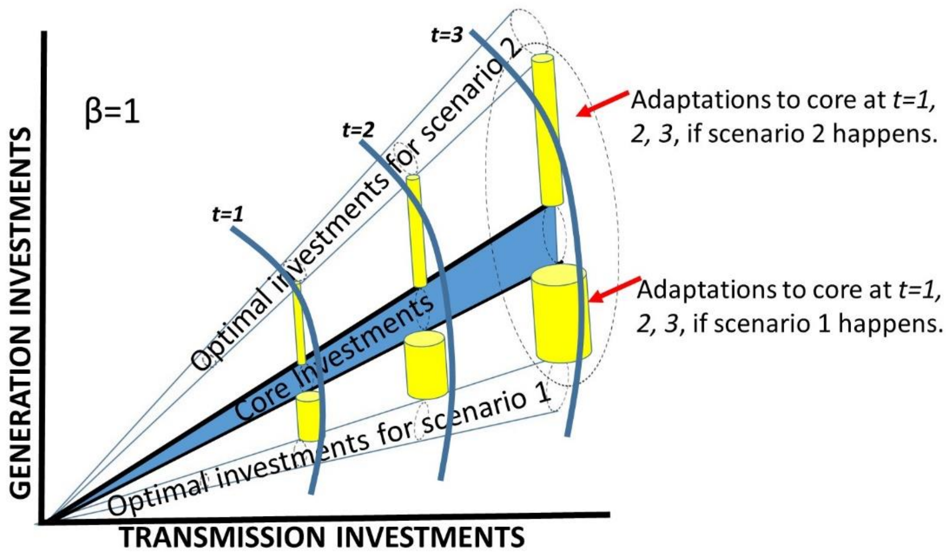


Figure 1.5: Relationship of core and adaptation investments

There are two remaining issues of significance to the conceptual understanding of AEP. The first is the operational cost. Equation (6) includes a probability-weighted operational cost expressed as a function of_ $\boldsymbol{x}_{GT}(t)+\Delta\boldsymbol{x}_{GTs}(t)$. This represents the cost of operating the power system under scenario *s*. We point out that the operational cost is not expressed as a function of $\boldsymbol{x}_{GT}(t)$ alone because the core is not a scenario in itself. In other words, there is no future realization of the uncertainties that correspond to the core; the only future realization of uncertainties that can actually occur are scenarios $s = 1, ..., N_s$.

The second remaining issue of significance to the conceptual understanding of AEP relates the value of investment made in the core (colored blue in Figure 1.5) to the value of investment made in the adaptation (colored yellow in Figure 1.5); it is a way to adjust the robustness of the solution to uncertainties to be encountered in the future. There are two extremes to consider:

- If a large amount of core investment is made, with very little adaptation, we are executing an expansion plan (the core) that will require very little adjustment in the future, independent of which scenario occurs. This is a highly robust, but expensive plan.
- Alternatively, if little core investment is made, with a great deal of adaptation, we are executing an expansion plan (the core) that will require a very large amount of adjustment in the future, independent of which scenario occurs. This plan has little robustness, but it is inexpensive.

We envision that the decision-maker would like to control the relationship between robustness and core costs, i.e., s/he would like to have a sort of lever that could be pushed in one direction to increase robustness and cost or in the other direction to decrease robustness and cost. We provide exactly such a lever through a multiplier β on the total adaptation costs, as indicated in (6). The influence of β is as follows. When β is small, adaptation costs appear to the optimization problem to be small relative to the core costs, and therefore adapting is more attractive than building core, and so solutions tend to have small cores and large adaptations. When β is large, adaptation costs appear to the optimization problem to be large relative to the core costs, and therefore adapting is less attractive than building core, and so solutions tend to have large cores and small adaptations. Figure 1.6 illustrates a solution for β=1 on the left and a solution for β=4 on the right.

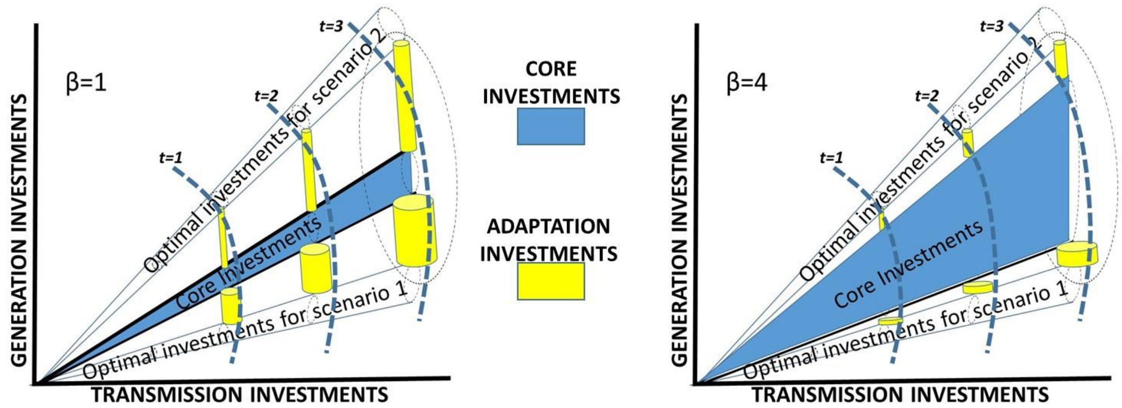


Figure 1.6: Different AEP solutions for different values of β

1.2.4 Comparing TSP and AEP

In this section, we identify the most significant differences between TSP and AEP. Articulation of these differences begins with referring to their defining illustrations, i.e., for TSP, Figure 1.3, and for AEP, Figure 1.5. By comparing these illustrations, we make the following observations:

1. *Essence of solution*: The essence of the solution to a TSP problem are the here and now variables corresponding to the choices made in the first decision period, $\boldsymbol{x}_{GT1}$ and the wait and see (recourse) variables corresponding to the choices made in the second decision period, $\boldsymbol{x}_{GT2s}, s = 1 \ldots N_s$. In contrast, the essence of the solution to an AEP problem are the choices made across all time for the core investment, $\boldsymbol{x}_{GT}(t), t = 1, \ldots N$, and the choices made across all time and for all scenarios for the adaptations $\Delta \boldsymbol{x}_{GTs}(t), t = 1, \ldots N,\ \ s = 1, \ldots, N_s$.
2. *Problem structure and computational intensity*: In TSP, the uncertainty is represented conditionally as branches; it can be conceptualized via a tree-like structure. This structure manifests itself computationally as a series of nested optimization problems, one for each decision period considered. Thus, computational intensity increases significantly with the number of decision periods. In AEP, the uncertainty is represented as a specified set of trajectories through the time intervals, one for each defined scenarios. Each scenario is explicitly specified through time. The uncertainty is brought into the problem via the presence of more than one such trajectory. Although use of many scenarios can result in large optimization problems, the problem remains a single optimization problem. One implication of this difference is that AEP can represent multiple decision periods while remaining computationally tractable. This observation has implications associated with computational intensity, as will be further described in this section.

3. *Actionable information*: In the TSP problem, the actionable information is the investment portfolio in the first investment period, $\boldsymbol{x}_{GT1}$; no recourse investments would be made until after the uncertainty is revealed. In the AEP problem, the actionable information is the core investment trajectory, $\boldsymbol{x}_{GT}(t), t = 1, \ldots N$, no adaptation investments would be made until after the uncertainty is revealed. We observe that the AEP core investment trajectory is analogous to the TSP here and now variables, and the AEP adaptation variables are analogous to the TSP wait and see variables.
4. *Application*: We view that the main application of TSP is *decision-making*. That is, TSP provides a good basis for making a decision on what to do *now*. In contrast, we view that the main application of AEP is *planning*. That is, AEP provides a good basis for identifying a sequence of investments through time. We believe the two approaches, TSP and AEP, are complementary, and it is useful to apply them both to the same planning problems so as to take advantage of their relative strengths.

There are several other differences between TSP and AEP that cannot be observed in the defining illustrations of Figure 1.3 and Figure 1.5. These differences are fully articulated in [14] and briefly summarized below.

- *Non-anticipativity*: SP imposes non-anticipative constraints in its formulation, meaning that all decisions (both here and now, and recourse) must be made without knowing the future, i.e., decisions must occur before future observations are made. This requirement is satisfied when scenarios sharing the same history (parameter values) are forced to share the same set of decisions. In AEP, scenario decisions (adaptations) do not share the same history in the sense that they share common parameter values. However, they do share the same history at each time step in regards to the core investment decisions being the same across every scenario. Consequently, we say that AEP has non-anticipative-like behavior at each time period[4].
- *Robustness parameter*: Adaption has a robustness parameter for controlling core investment against adaptations and thus robustness against core cost. It would not be difficult to implement a similar idea in TSP, weighting investments made in the second (recourse) decision period against decisions made in the first decision period. However, the effect may be different in that the non-anticipative constraints may change the effect a robustness parameter has on the relationship between here and now decisions and wait and see decisions. This is a research point that we have not yet investigated.
- *Algorithmic complexity*: It is shown in [14] that whereas TSP decision variables (and constraints) grow exponentially with time periods, those for AEP grow linearly with time periods. This is important since linear programming algorithms (e.g., simplex, barrier) have compute-times that increase with decision variables. Although we have found that two-stage TSP problems and their AEP counterparts generally have similar compute-times, our algorithmic complexity result suggests that multi-stage TSP problems would be far more computationally intense than their AEP counterparts.

[4] An alternative way to consider this issue is in terms of "memory," i.e., the ability of investment decisions in period *t* to start from, and depend on, the investment decisions made in period *t-1*. All TSP investments have memory meaning that an investment made in one time period is carried over to the next time period. AEP, on the other hand, has memory in its core investments but not in its adaptation investments.

1.3 System studied

The material in this section, Section 1.3, is condensed from [14]. A 312-bus representation of the US Western Interconnection (WI) was used in all analyses reported in this document. The system topology and network data was obtained from the Western Electric Coordinating Council (WECC) as a 2024 Transmission Expansion Planning Policy Committee (TEPPC) case [15] and then Kron-reduced [16].

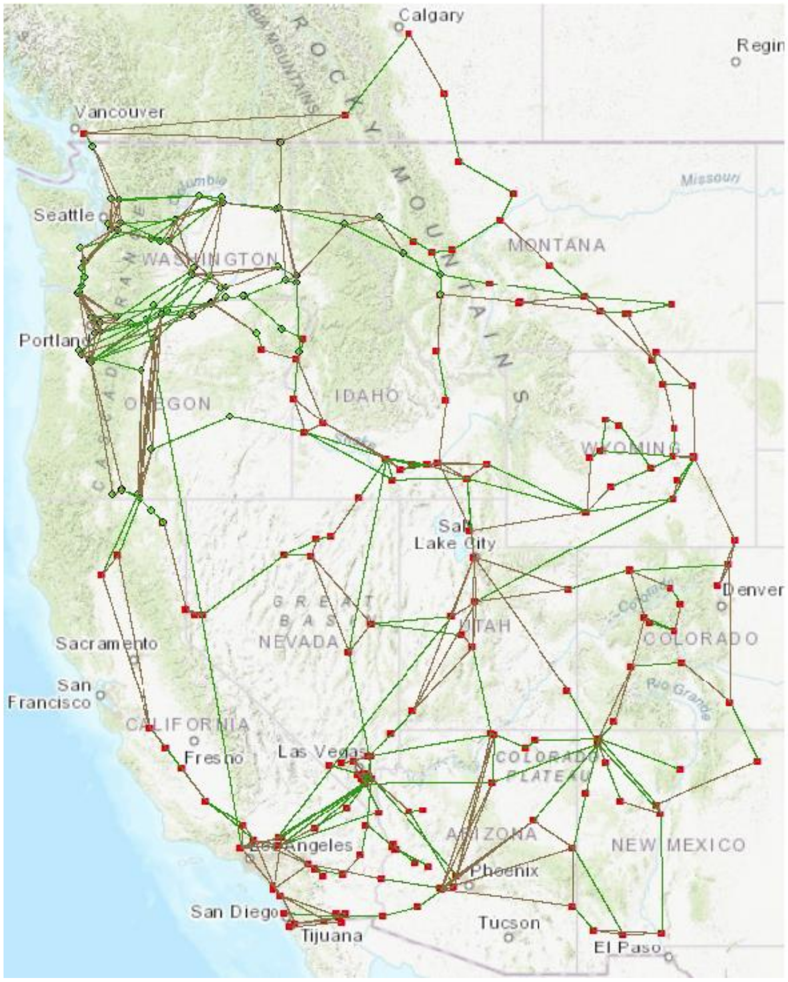


Figure 1.7: One-line diagram of system used in analyses of this report[5]

We briefly describe below what we refer to in this report as base conditions. These conditions were used in all analyses described in this report, unless otherwise indicated. For each of the below items described, additional detail may be found in [14].

[5] Service Layer Credits: Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, @ OpenStreetMap contributors, and the GIS User Community.

1.3.1 Operational blocks

To avoid computational intractability, it was necessary to reduce the number of operating conditions per year to 12, with each operating condition weighted by the number of hours per year it represents.

1.3.2 Wind and solar profiles

We refer to the power generation output on an hour-by-hour basis for wind and solar resources as their profiles. Wind and solar profiles for existing resources were obtained from the WECC 2024 TEPPC Common Case [15]. Wind profiles for candidate wind resources were obtained from the National Renewable Energy Laboratory (NREL) Wind Toolkit [17]. Solar profiles for candidate solar resources were obtained based on data extracted from the National Solar Radiation Database (NSRDB) [18] and converted to power production using the NREL System Advisor Model [19].

1.3.3 Generator costs

Generator operating cost data for existing generation resources were obtained from the WECC 2024 TEPPC Common Case [15]. Generator capital and operating cost data for candidate wind and solar were obtained from the NREL Annual Technology baseline workbook [20]. Generator capital and operating cost data for gas-fueled combined cycle and combustion turbine units were obtained from WECC's capital cost tool [21]. Generation fuel prices were taken from the 2024 TEPPC Common Case [15].

1.3.4 Candidate transmission

Transmission line costs are computed based on allowance for funds used during construction (AFUDC), length, ROW area land rent, terrain multipliers, and base conductor costs. Base conductor costs are adapted from the 2014 WECC Capital Costs for Transmission and Substations Report [22]. Addressing candidate transmission for a reduced model where many circuits are equivalents requires specialized treatment as described in [14].

1.3.5 End effects

End effects refer to the influence of economic decision-making under the unavoidable condition of finite-time decision horizons [23]. Without specialized treatment, end effects can inappropriately skew results, e.g., technologies having low capital costs but high operational costs appear inappropriately attractive for investment during the last years of the decision horizon because their high operating costs are only considered for that short time period between when they are built and when the decision-horizon ends. There are various ways to address this issue; here, we eliminated investment opportunities following the last year of the decision horizon but artificially extended the simulated operations for 30 years beyond the decision-horizon final year.

1.3.6 Simulation setup

The planning horizon for the simulation is 2018 to 2036 where investment decisions are made in the years 2018, 2024, and 2030. Investments become operational 6 years after the decision is made to build them.

1.3.7 Model formulation

Reference [14] provides a detailed expression of the model formulation, with explicit and careful definitions of all nomenclature used in the expression. The model is highly evolved in a number of ways, two of which are highlighted below.

- *Toggles to different forms*: Software switches enable the model to be configured as a TSP, an AEP, or a (deterministic) CEP.
- *Transmission line modeling*: Transmission lines may be modeled as existing or candidate. All existing lines are modeled using linearized "DC" power flow representation; although this is an approximation relative to AC power flow representation, the approximation retains the effect of line impedances. Candidate lines may be modeled using linearized "DC" power flow representation as well, but in this case, it is necessary to use the disjunctive representation, and the model becomes a mixed integer linear program. Alternatively, candidate lines may be modeled using "pipes," i.e., neglecting the effects of impedances; such a representation is referred to as a transportation model, and in this case, the overall model remains a linear program (LP). We refer to the model using impedance representation for existing lines and a transportation representation for candidate lines as a "hybrid" model. We use it heavily because it provides a reasonable tradeoff between model fidelity and compute time.

1.3.8 Distributed energy resources (DER)

The software is capable of representing distributed photovoltaics, demand response (DR), and energy efficiency (EE). Because build costs for DR and EE are low, they generally are built before other technologies; however, we place a limit on the amount of invested DER per year.

1.4 Motivation for compute time reduction

Co-optimized generation, transmission and distribution expansion planning problems under uncertainty (using either TSP or AEP) are computationally intense. In order to make these problems computationally tractable, several assumptions and simplifications are made to the models which contribute to faster compute times, but impact model fidelity. For example, as described in Section 1.3.7, there are various candidate transmission line representations which provide either low fidelity and low compute times (e.g., the transportation model), or high fidelity and high compute times (e.g., the disjunctive model), with the hybrid model offering a reasonable compromise level of both fidelity and compute time. Even so, compute times normally encountered by our 312-bus size models can be excessive. For example, a typical AEP run generates 5.13 million variables and 5.11 million constraints and requires 22 hours of compute-time to complete on a single core of an Intel 2.0Ghz E5-2650 processor. Such long compute times makes work difficult for the analyst since a planner is often required to make multiple runs, each time inspecting results and making input data changes. Lengthy computation impacts productivity and thus necessitates improvements to make the problem solve faster. And of course, these comments apply to the 312-bus model; normal sized planning datasets are from 60 to 200 times this size, ranging from 20,000 buses in the WI to 60,000 buses in the EI, rendering stochastic programming applications on them fully intractable. As a result, we desire to, first, understand the effect of various problem characteristics on compute time, and second, to devise ways to significantly reduce this compute time.

1.5 Structure of this report

The remainder of this report is organized as follows. In Section 2, we describe influences on computational intensity in expansion planning problems under uncertainty, and we provide numerical experiments to illustrate. In Section 3, we summarize some effective methods for achieving compute-time reduction, and we illustrate these methods with additional computational experiments. In particular, results from a shared memory parallel implementation using CPLEX barrier optimization and parallelized Benders decomposition are presented with order of magnitude speedup observed compared to serial implementation. The last section concludes.

2. Influences on compute time in expansion planning with uncertainty

Solving deterministic expansion planning problems are computationally expensive, and solving them under uncertainty via TSP or AEP is significantly more so. In this chapter, in Section 2.1, we classify and identify influences on computational intensity of expansion planning problems that arise from the problem structure and the chosen modeling fidelity. Section 2.2 summarizes results of numerical experiments performed to quantify the level that each of these influences have on compute-time. Section 2.3 describes approaches based on data compression and network reduction that can be used to mitigate the impact on compute-time of some of these influences.

2.1 Influences from problem structure and modeling granularity

Figure 2.1 illustrates two classes of influences on expansion planning problems under uncertainty: problem structure and modeling fidelity, on the one hand, and the nature of the solution, on the other. Problem structure and modeling fidelity capture features associated the particular power system representation; the nature of solution captures features associated with the way the problem is solved. We address the first class of influences in the remainder of this section, describing the various features identified on the left-hand-side of Figure 2.1. The second class of influences, on the right-hand-side of Figure 2.1, will be described in Section 3.

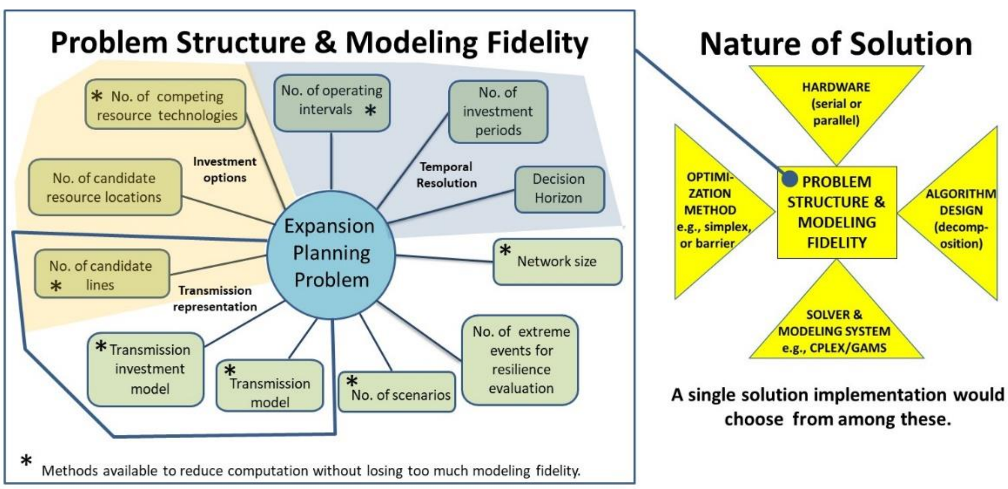


Figure 2.1: Influences on computational intensity in expansion planning problems

The left-hand-side of Figure 2.1 identifies 11 influences, some of which are grouped because of their similar nature. We describe these influences below, beginning with network size and moving counterclockwise around the figure.

2.1.1 Network size

One may represent the network in varying degrees of granularity, in terms of aggregation of buses, generators, loads, and circuits. Planning models for the Western Interconnection typically have 20,000 to 30,000 buses; planning models for the Eastern Interconnection may have over 70,000 buses. Models of these sizes are intractable for direct application of today's state-of-art deterministic CEP tools; stochastic applications such as TSP and AEP are unthinkable. There are two basic compression/reduction approaches to address this: network reduction and grid partition. These approaches are described in Section 2.3.

2.1.2 Temporal resolution

This influence involves three separate features of an expansion planning problem: decision horizon, number of investment periods, and number of operating intervals. Increasing temporal resolution increases computation because each operating interval is a distinct operating condition on which the power flow equations must be imposed (resulting in associated auxiliary variables of generation, angle, and circuit flow). Each of the three features associated with temporal resolution is described in what follows.

Decision horizon

The decision horizon is the time period over which investment decisions may be made. Typical decision horizons are between 10 and 40 years, although it is entirely possible to make them shorter or longer to satisfy a particular need. Each year in a decision horizon should be modeled, although different modeling approaches may be applied to different years.

Number of investment periods

The highest fidelity choice regarding the number of investment periods is to make every year and investment period. However, doing so increases computational intensity, and so it is typical to reduce the number of investment periods to every other year or every third year. Years that are not investment periods simply become "operational years," which means that they are simulated operationally and associated operational costs are computed and used within the optimization, but no investments are permitted during that year.

Number of operating intervals

For maximum fidelity, the number of distinct operating intervals simulated is 8760 times the number of years in the decision horizon. However, this makes the expansion planning problem intractable, and so it is necessary to reduce this number significantly. There are three general approaches for doing this. One approach which maximizes computational efficiency but at the lowest level of modeling fidelity is to select "typical hours." For example, if one were to identify four typical hours per season, modeling four seasons would result in 16 operating conditions per year. To account for intertemporal (e.g., unit commitment) constraints, one needs to select "typical days" or "typical weeks." The assumption in any of these approaches is that all unselected conditions are adequately represented by the "typical" conditions.

There are two basic approaches to improving temporal resolution while maintaining acceptable modeling fidelity: use an external production cost model and deploy decomposition with high performance computing (HPC). These approaches are described in Section 2.3.

2.1.3 Investment options

There are two types of investment options: resources (including DER) and lines. In this section, we focus only on the first of these (resources). We address the second of these (lines) in Section 2.1.4 because its impact on computational intensity is intimately related to the selected transmission model and the selected transmission investment model. There are two main features related to resources that affect computational intensity: number of competing resource technologies and number of candidate resource locations.

Number of competing resource technologies

An essential choice for designing an expansion planning study is to determine what types of resource technologies are to be competed within the tool. Typically, this decision is made based on an understanding of what might possibly be economically attractive. For example, the first GEP programs of the 1970's always included nuclear, coal, and gas-fueled combustion turbines. It is very typical in today's environment to include, at a minimum, wind, solar, and gas-fueled combined cycle and combustion turbine plants, with energy efficiency and demand response programs (which can be considered to be resources). It is quickly becoming of high interest to also include storage. Of course, one may include all of these together with many other resource options including, for example, biopower, geothermal, wave energy, concentrated solar thermal. In some studies deploying specialized constraints, it may be of interest to include some or all of them. However, if there is no particular reason for including them, because of their high capital costs, they are unlikely to be ever selected. In such cases, they should not be included because doing so increases compute time.

Number of candidate resource locations

If the expansion planning tool represents transmission[6], and thus multiple buses, then candidate resources are represented locationally. The number of resource technologies when multiplied by the number of candidate resource locations gives the number of resource-related decision variables. Increasing decision variables increases computational intensity. The most effective way to limit this influence is to restrict the resource technologies and candidate resource locations to only those technologies and locations for which selection is conceivable based on the constraints employed for the study of interest. Such information may be identified after performing the expansion planning calculation under a few different conditions; an effective way to initially do this is to make "one-year runs," i.e., runs where the decision horizon is limited to a single year (and of course that single year must be an investment year). One-year runs have constraints and equations corresponding to only one year, so that the number of constraints and the number of decision variables is only *1/N* of that found in a run having an *N*-year decision horizon.

2.1.4 Transmission representation

There are three attributes associated with transmission representation: the transmission model, the transmission investment model, and the number of candidate lines.

[6] The simplest form of a GEP represents all resource technologies at a single busbar, and so such a problem is a one-bus problem. One may still represent the cost of each resource technology option based on its location, but otherwise, there is no locational influence of the generation on network operation.

Transmission model

The transmission model refers to the manner in which transmission is represented via the power flow constraints. A decision on this feature must be made if transmission is to be represented, even independent of whether transmission investment is to be allowed. As indicated in Section 1.3.7, there are two transmission models that can be used[7]: a transportation model or an impedance model. Use of the transportation model significantly increases computation speed. However, this speed is obtained at the cost of significant loss of fidelity in that the flows in a transportation model are entirely dictated by economics and line capacities, i.e., they are not influenced by electric circuit behavior imposed by Kirchhoff's voltage law and influenced by each line's impedance. The alternative is the transmission line representation of the "DC power flow" model, which depends on setting line resistance to zero, setting per unit voltages to 1.0, and assuming angular difference across each line is "small," so that the approximation $\sin(\theta_j - \theta_k) = \theta_j - \theta_k$ (in radians) holds for any line connecting buses *j* and *k*. Such an approximation is reasonably good for computing real power flows under "normal" (i.e., unstressed) conditions.

Transmission investment model

GEP formulations that use the DC power flow transmission model remain linear programs, but TEP and CEP formulations become mixed integer nonlinear programs (MINLP). This is because of the power flow equality relations $P_{jk} = (B_{jk}^{\text{exist}} + Z_{jk} B_{jk}^{\text{invest}})(\theta_j - \theta_k)$ where the binary decision variable Z_{jk} (which indicates whether the transmission investment between buses *j* and *k* should be made or not) multiplies the angle variables θ_j and θ_k (here, B_{jk}^{exist} is the susceptance of the existing line, B_{jk}^{invest} is the susceptance of the new, i.e., invested, line, and P_{jk} is the power flowing along the existing and new *j* to *k* line).

MINLP problems are very difficult to solve, and so we must avoid the nonlinear condition to have any hope of maintaining tractability in solving a TEP or CEP. There are five different transmission investment models which allow to do so, in order of increasing computational intensity and increasing model fidelity.

1. *Transportation model*: Here, we use a transportation model for the transmission lines. In this case, the transmission investment model is simply the transmission line capacities. No impedances are represented in either existing or invested lines. The optimization problem is a linear program.
2. *Constant impedance model*: Here, we use a DC power flow transmission model, but the transmission investment model is (as in the transportation model) is simply the transmission line capacities. This model becomes increasingly less accurate as transmission is invested because the transmission capacities expand but line impedances are not decreased to reflect the additional transmission. The influence of KVL, i.e., the impedances, although present,

[7] A third one can be considered: AC power flow constraints. However, representing AC power flow constraints would make the expansion planning problem nonlinear and nonconvex. Given the already very intense computational challenges associated with solving expansion planning problems, we accept that at this point of the R&D path for this technology, using AC power flow equations for expansion planning is an unwise approach.

becomes more and more skewed as more and more transmission is invested. The optimization problem is a linear program.

3. *Hybrid model*: This transmission investment model uses two transmission models: the constant impedance model is used for existing lines, and the transportation model is used for invested lines. Impedances and capacities of existing lines remain constant. Invested lines, as transportation models, are expanded via their capacities. In essence, this model maintains the existing transmission system as an impedance model (subject to the "DC power flow" approximation) and adds new transmission as "pipes" which may be thought of as HVDC lines. The optimization problem is a linear program.
4. *Iterative model*: This model iterates using the constant impedance model described above, but after each iteration, the impedances of invested lines are adjusted to match their new capacities. The optimization problem solved in each iteration of this model is a linear program. Experience with this model indicates it requires between 3 and 6 iterations to converge.
5. *Disjunctive model*: In this model, the flow equalities used include one for the existing circuit (terminating at buses *j* and *k*, numbered circuit *b*) $P_b = B_b^{exist}(\theta_j - \theta_k)$, and one for a parallel invested circuit (also terminating at buses *j* and *k*, numbered circuit *b+1*), $-M(1-Z_{b+1}) \le P_{b+1} - B_{b+1}^{\text{invest}}(\theta_j - \theta_k) \le M(1-Z_{b+1})$. Here, Z_{b+1} is a binary investment variable, M is a large positive number so that, if investment is not made ($Z_{b+1}=0$), the middle term becomes unconstrained and there is no relation enforced between flow P_{b+1} and $B_{b+1}^{\text{invest}}(\theta_j - \theta_k)$; if investment is made ($Z_{b+1}=1$), the middle term becomes constrained from above and below to zero, imposing equality to zero, and $P_{b+1} = B_{b+1}^{\text{invest}}(\theta_j - \theta_k)$ is enforced. The optimization model is a mixed integer linear program (MILP). This formulation is equivalent to the MINLP of the full model, but it avoids the nonlinearity, and standard MILP solvers are available to handle it. However, TEPs and CEPs when modeled this way are computationally intensive.

The most commonly chosen model is the hybrid model because it is a linear program, and at least for reduced (relative to those used in industry) network representations, it can usually be solved in hours or days. However, the hybrid model incurs some loss of modeling fidelity; the iterative approach is a promising way to reach the disjunctive model's fidelity without having to solve integer programs.

Number of candidate lines

Each candidate line adds a decision variable, a flow equality, and a flow inequality, and so limiting the number of candidate lines is attractive; this is particularly important when using the disjunctive model as each candidate line contributes an additional binary variable and thus increases the dimensionality of the integer program. In Section 2.1.3, we indicated that "one-year runs" are effective ways to limit the number of candidate resource locations; one-year runs are likewise effective ways to limit the number of candidate lines.

2.1.5 Number of scenarios

In handling uncertainty, a scenario is a possible future defined by the assignment of a value to each uncertain parameter. For example, consider having the following seven uncertain parameters: load growth; hydro production; natural gas price; CO_2 emissions price; transmission cost; solar PV capital cost; wind capital cost. In the simplest case, consider that each uncertain parameter may take either of two values: high, or low. This means there are $2^7=128$ different combinations of these parameters. Because in TSP and AEP, each scenario requires its own operational evaluation, increasing the number of scenarios is computationally costly. Therefore, scenario reduction is commonly performed in order to identify a subsect of the total scenario-space that is representative of that space. Scenario reduction approaches are described in Section 2.3.2; in addition, it is the primary focus of the work described in the Part I report for this project.

2.1.6 Number of extreme events for resilience evaluation

Modeling resilience within an expansion planning program is a relatively new topic that has only recently received attention. We have made some effort to investigate this for designing electric infrastructure accounting for hurricane events in Puerto Rico [24]. As this area matures, it is likely that expansion planning programs will need to simulate extreme events. Inclusion of extreme events, for resilience evaluation, within the CEP can significantly increase computation time.

2.2 Numerical experiments

To quantify the difference in computational cost between various parameters affecting simulation time, numerical experiments were conducted to identify the major computational bottlenecks in adaptive expansion planning problems. In this work, the same computing resource was used for all experiments: an HPC cluster with each node having two 2.0 GHz 8-core Intel E5 2560 processors (so 16 cores per node), and 128GB memory. The modeling environment GAMS is used to generate CPLEX source code for a linear program that solves using the barrier method to a duality gap of 10^{-6}, keeping all other parameters at their basecase value, except for the parameter being investigated. Basecase reference simulations are described in Section 1.3 and are summarized here as follows: 357-bus WECC system, 12 operating blocks, 3 investment periods and 8 scenarios for an 18 year planning horizon. Numerical experiments pertaining to the effect of various influences are summarized in Table 2.1. These influences include network size, temporal resolution (decision horizon, number of investment intervals, and number of operating intervals); investment options (number of competing resource technologies[8]); transmission representation (three different transmission investment models, and also number of candidate lines); and number of scenarios. Inspection of this table indicates that temporal resolution, number of scenarios, network size and changing the transmission representation from a hybrid to disjunctive model (making it an MILP) have the largest influence on compute-time. In fact, using the disjunctive model increases computational intensity to a level where it can be called computationally intractable. Using the data given for network size, and assuming compute-time grows linearly with number of buses (it actually grows exponentially), compute-time for 20,000 and 70,000 bus models would be 6 days and 22 days, respectively.

[8] Number of competing resource locations was not investigated because we know that its influence is less than that of network size.

Table 2.1: Summary of numerical experiments characterizing influences on compute time

Case ID	Influence Tested	Change made relative to base condition	Total compute-time, hrs	Δ compute-time, hrs
0	Base condition	None	1.9	N/A
1	Network size	Changed number of buses from 312 to 99	0.3	-1.6
	Temporal resolution			
2	Decision horizon	Changed from 18 to 9 years.	1.85	-0.05
3	No. of investment periods	Changed from 3 (every 6 years) to 6 (every 3 years)	24.0	+22.1
4	No. of operating intervals	Changed from 12 blocks/year to 16 blocks/year	6.0	+4.1
	Investment options			
5	No. of competing resource technologies	Changed from 20 to 13 (removing all technologies which were not getting invested in)	1.9	0
	Transmission representation			
6	Transmission investment model	Changed from hybrid to transportation model	1.3	-0.6
7	Transmission investment model	Changed from hybrid to constant impedance model	1.65	-0.25
8	Transmission investment model	Changed from hybrid to disjunctive model	>24	>24
9	Number of scenarios	Changed from 8 to 12	4.1	+2.2

2.3 Data compression/reduction methods for reducing compute time

Figure 2.2 characterizes approaches for reducing compute time of an expansion planning application. Three approaches are illustrated: data compression/reduction, decomposition approaches, and parallelized HPC. In this section, we focus on the first of these, data compression/reduction, as these approaches generally address the influences from problem structure and modeling granularity described in Sections 2.1 and 2.2. The other two approaches, decomposition and high-performance computing, are addressed in Section 3.

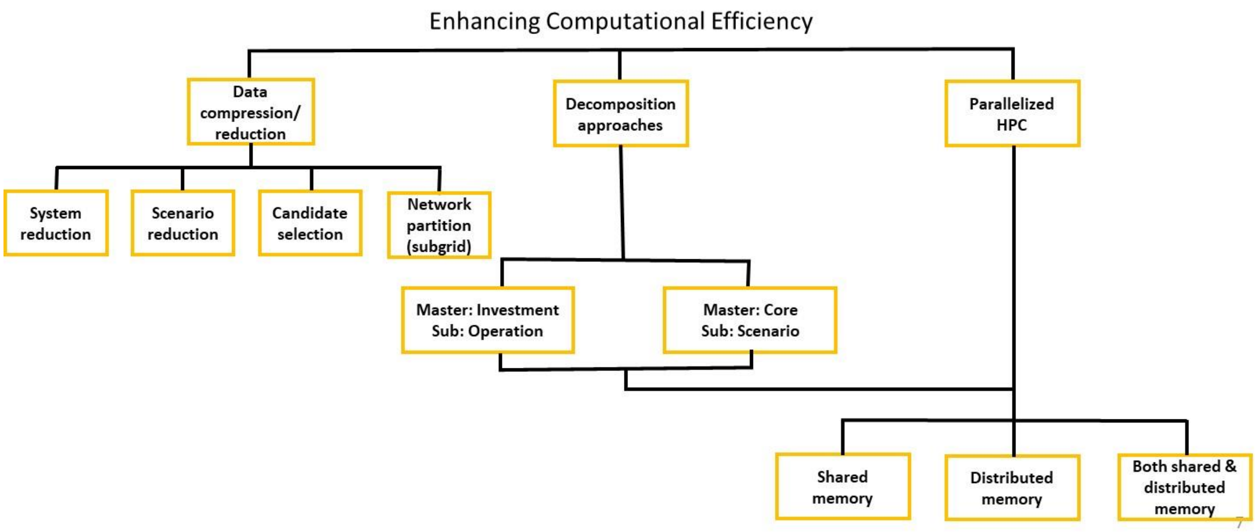


Figure 2.2: Mitigation measures for reducing compute time

2.3.1 System reduction

Motivated by the impact of network size on compute-time observed in Section 2.2, it is typical in performing expansion planning to use a reduced equivalent model of hundreds of buses, rather than a standard size industry-grade planning model of many thousands of buses (e.g., a WI model will have between 20,000 and 30,000 buses, and an EI model will have over 70,000 buses). There are various methods for developing network equivalents, but the most common method is Kron reduction, originally proposed in [25], matured in [26], and well described in [27], adjusted to heuristically move generation resources to retained buses without losing their identity.

Although network reduction effectively provides a model of appropriate size for use in expansion planning programs, the resulting investment portfolio is identified in terms of buses and lines that are equivalents, rather than actual, existing buses and lines. Further analysis of these investments in an industry size planning model (power flow or production cost) requires that these investments be translated to the full-scale model. We refer to the application that does this as the translation application, which can be thought of as an inversion of the reduction process. This inversion process is performed via two-steps. In step 1, the invested generation resources are distributed from their equivalent buses to actual buses using various heuristics. In step 2, an optimization is solved on the large-scale model (which already has the invested generation resources) by iteratively updating a TEP constant impedance transmission investment model, where in each update, line impedances are modified to be consistent with the invested transmission capacity.

2.3.2 Scenario reduction

There are two general scenario reduction approaches: input approach and output approach. In the input approach, a clustering algorithm such as k-medoids is applied to the set of uncertainties in terms of the actual values those uncertainties may take. In the output approach, each possible combination of uncertain input parameters (e.g., for seven parameters, with each one having two possible values they can take, there are 2^7=128 combinations) is used to deterministically generate an expansion plan, and then a clustering algorithm is applied to the set of plans in terms of the actual investments that are made in each. For the reduced model described in Section 1.3, with seven uncertain parameters, we found we needed to reduce the number of scenarios from 128 to between 6 and 12 in order to achieve acceptable fidelity with reasonable compute-time. As mentioned in Section 2.1.5, the Part I of this project's reporting describes an advanced method of scenario reduction based on an iterative stochastic method.

2.3.3 Candidate selection

This involves reducing the number of candidate lines, candidate generation technologies, and candidate generation locations, to reduce the number of decision variables, and hence decrease computational complexity. As described in Section 2.1.3, one approach to reducing candidate lines, resource technologies, or resource locations is to use "one-year runs" with all possible lines, resource technologies, and locations modeled, and choose candidates based on where and what was invested. Because one-year runs only represent a single year, they are very fast. However, because they represent the conditions of only a single year, this approach can potentially under-select. An improvement, implemented in [28], is to make one-year runs for the conditions of year 1, year *N*/2, and year *N*, and then select candidates lines, resource technologies, and resource locations as the union of all lines, technologies, and resource locations invested across the three one-year runs.

2.3.4 Network partition (subgrid)

A common issue in co-optimized expansion planning (CEP) is that the region of interest for performing expansion planning is a part of a larger system; thus there is an internal system and an external system. The analyst wants to identify investments for the internal system with an intraregional study, accounting for the external system only insofar as the external system affects the investments identified in the internal system. A network partition approach for doing this was developed under funding by the Bonneville Power Administration as reported in [29], also called the subgrid approach. The remaining material in this subsection is an adaptation of that reporting.

A 2-step procedure is illustrated in Figure 2.3. In step 1 a low-fidelity but fast linear program (using the hybrid transmission investment model) on the entire interconnection is deployed to determine reasonable tie line flows between the external and internal system. In step 2, the tie line flows connecting to the external system are fixed to their step 1 values, and a high-fidelity mixed integer linear program (the disjunctive model) is solved on just the internal system. Computational tractability is further improved in step 2 in two ways: (1) lines not receiving investments in step 1 are eliminated from the list of expansion candidates; (2) continuous line investments are translated into their binary disjunctive representation (the binary disjunctive representation, developed in [30], is a refinement on the disjunctive transmission investment model that increases computational efficiency when there is possibility for investing in more than one line between buses *j* and *k*). Both of these improvements serve to reduce binary decision variables in step 2.

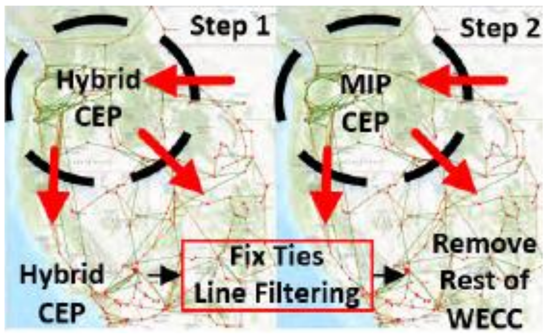


Figure 2.3: Two-step grid partition procedure[9]

This approach was applied to the network described in Section 1.3 using a deterministic disjunctive model for the full network and for the subgrid partition; compute time for a case with 20 operating blocks per year reduced from 12.1 hours when the entire network was solved using the disjunctive model to 2.0 hours when the subgrid approach was applied. A similar strategy was employed for an AEP implementation (a linear program), and compute time reduced from 2.86 hours to 0.31 hours.

[9] (Service Layer Credits: Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community).

3. Compute time reduction

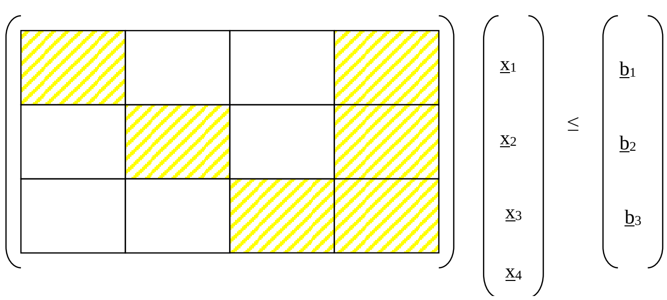
In this chapter, we describe efforts towards investigating the use of decomposition and parallelization on HPC to mitigate the computational intensity of expansion planning problems under uncertainty, as indicated in the middle and right-hand parts of Figure 2.2. We begin this chapter by describing in Section 3.1 some overriding features of this work. Section 3.2 provides an analytic formulation of the AEP problem using Bender decomposition. In Section 3.3, we summarize HPC architectures and associated parallelized designs along with computational results. Promising strategies not investigated are discussed in Section 3.4.

3.1 Selection of solution strategy

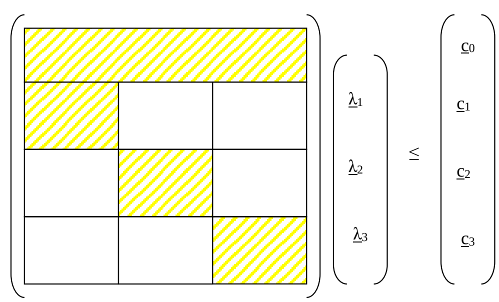
There are six features of this work necessary to understand the selection of a solution strategy and the context in which this selection was performed in this project. The first five of these features are captured on the right-hand-side of Figure 2.1: algorithm design; hardware; optimization method; modeling system and solver; and dependence on problem structure and modeling fidelity. A sixth feature is a direct consequence of the first five: complexity associated with identification of solution approaches.

3.1.1 Algorithm design

Optimization problems may be solved directly without regard to the problem's structure. However, many problems have certain kinds of structure that allows them to be solved with algorithms that partition the problem into smaller parts. This structure is commonly referred to as a block-angular structure, and if a problem has it, it is observable in the constraint matrix. It typically appears in one of two ways: either in terms of linking variables or in terms of linking constraints, as illustrated in Figure 3.1. Here, the left-hand structure has a set of variables (in this case, $\underline{x}_4$) that occur in every constraint, but otherwise, the constraints are decoupled. Problems with this structure are amenable to decomposition via the method of Benders [31]. In the words of A. Geoffrion [32], "J.F. Benders devised a clever approach for exploiting the structure of mathematical programming problems with complicating variables (variables which, when temporarily fixed, render the remaining optimization problem considerably more tractable)." The right-hand structure has a set of constraints (in this case, the top row) which utilize all variables, but otherwise, the constraints are decoupled. Problems with this structure are amendable to decomposition via the method of Dantzig and Wolfe [33].



Constraint matrix: block angular structure with linking variables



Constraint matrix: block angular structure with linking constraints

Figure 3.1: Illustration of problem structures amenable to decomposition

The underlying principle associated with decomposition is, if the problem's compute time is proportional to an exponential function of the problem size, then it can be faster to solve many problems of small size, the sub-problems, than it is to solve one problem of large size. Therefore, decomposition procedures enable computational efficiencies on serial computers (without parallelization). However, such problems are naturally parallelizable in that each sub-problem may be solved independent of the other sub-problem's solutions, and so decomposed *and* parallelized solutions offer efficiencies from two different, but interdependent means.

The constraint matrix for expansion planning problems under uncertainty offers block-angular structure in two ways. First, the time periods are typically independent or close to independent, a feature that suggests expansion planning problems may be decomposed into a master investment problem and operational sub-problems. This was indeed the approach taken in an early effort by Bloom[10] to decompose a GEP [34, 35, 36]. In addition, expansion planning problems solved under uncertainty can exhibit block angular structure as a result of being able to solve scenarios independently; reference [37] recognized this and implemented a progressive hedging decomposition approach. Others have consider Lagrangian relaxation [38] where the complicating constraints are removed from the constraint equations and then dualized, i.e., added to the objective function, with a penalty term (the Lagrange multiplier) proportional to the amount of violation of the dualized constraints [39]. There are other interesting algorithmic ways to pursue solution to the AEP problem. For example, hybrid decomposition methods employ more than one strategy [40]. Nested decomposition deploys two decompositions (possibly, but not necessarily different decomposition methods), with one operating on the sub-problem of the other [41], e.g., the higher level decomposition may operate on scenarios and the lower-level decomposition may operate on the time periods for each scenario.

3.1.2 Hardware

Hardware used in parallelization may be either shared memory or distributed memory. A shared memory architecture, illustrated in Figure 3.2, consists of multiple processors which read and write asynchronously into a common pool of memory, i.e., a common address space. In this architecture, all processors have the ability to access all available memory such that each processor can compute in parallel with the others while sharing the same memory resource. However, if data at a memory location is changed by a particular processor, it affects the data seen at that location by all other processors. Race conditions, i.e., when multiple processors access the same memory location simultaneously, can result in faulty processing. This architecture is highly attractive for parallelization because, since memory is shared, it requires no additional effort to facilitate data communication between parallelized tasks.

[10] Jeremy Bloom, a Ph. D. student at MIT, working with M. Caramanis under contract with EPRI, developed what is now a heavily used GEP program in industry called EGEAS.

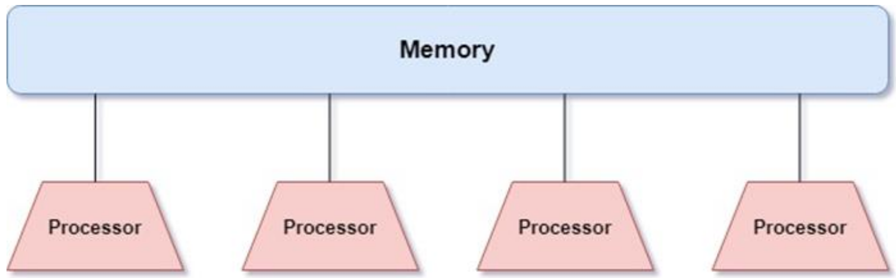


Figure 3.2: Shared memory architecture

On the other hand, a distributed memory architecture employs multiple processors each of which has access to their own local memory but cannot access the memory belonging to other processors. Thus, if a particular processor while performing its own task needs to coordinate with another processor which is performing its own task, one processor should send (i.e., "ask"), the other receive ("respond"), in some formalized way to initiate the passage of information between the two. The formalized communication capability is referred to as message passing interface (MPI) and has now matured to a library of subroutines that are standardized; they are publicly available on the MPI forum (www.mpi-forum.org/docs/).

Shared- and distributed-memory architectures may be combined, and usually are, resulting in a hybrid shared/distributed memory architecture. In this architecture, a network connects multiple 'nodes' each consisting of several 'cores' or processors with a common shared memory resource within each node. The shared memory component refers to the multiple cores and common memory within each node. The distributed memory component refers to multiple nodes connected by a network, where memory in a particular node is not explicitly visible to other nodes. Message passing is necessary to communicate between multiple nodes, which increases scalability of computation, but also significantly increases programming complexity.

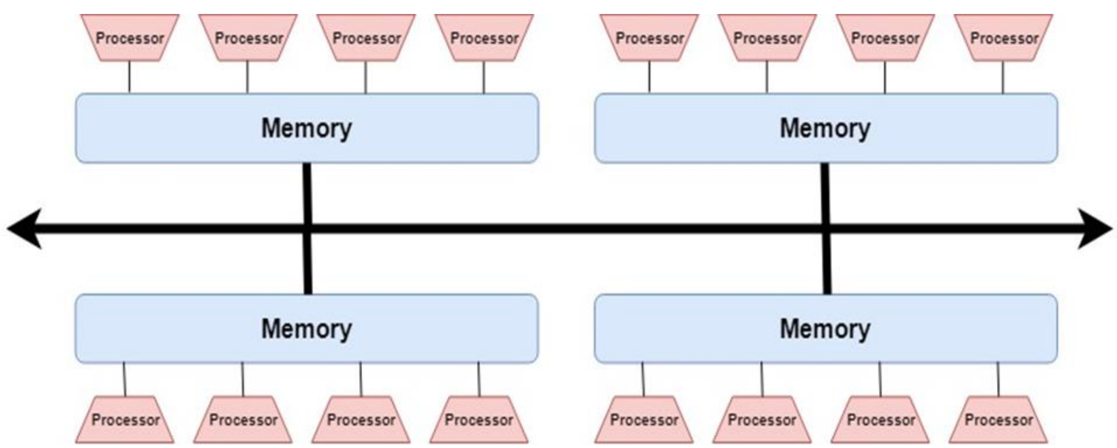


Figure 3.3: Hybrid shared/distributed memory architecture

3.1.3 Optimization methods

Our main interest in this project has been in solving the AEP problem, which we usually formulate as a linear program, though it can be formulated as a mixed integer linear program depending on whether the disjunctive transmission investment model or hybrid transmission investment model is used for the core transmission investments. The two optimization methods most commonly used for solving linear programs are simplex and barrier. In contrast to the simplex method which visits corner points (points at the boundary of the feasible space), the barrier method is an interior point method that reaches the optimum through a sequence of points starting from and remaining within the interior of the feasible region. The optimization method most commonly used for solving mixed integer linear programs is the branch and cut method, which employs branch and bound, solving a linear program at each node, and deploying cutting planes to tighten the linear programming relaxations. In results presented on MILP solutions in Section 3.3, we refer to "percent gap," which is the difference between the objective function lower bound (identified as the solution to the linear program when all integer variables are relaxed to be continuous variables) and the objective function upper bound (identified as the best integer solution found so far in the process of solving the problem).

3.1.4 Modeling system and solver

The modeling system we have used in this project is the Generalized Algebraic Modeling System (GAMS), in which we develop the optimization problem specification. GAMS then translates the optimization problem specification to a source code that is read by the particular solver; we have used the CPLEX solver in all of our work.

The CPLEX solver offers both simplex and barrier methods of solving linear programs, but the barrier method is also available in a shared memory parallelized mode. In addition, GAMS offers the ability to deploy a shared/distributed memory architecture with MPI calls embedded within the GAMS optimization specification.

3.1.5 Problem structure and modeling fidelity

Section 2 of the report described the influences on computational intensity in expansion planning problems under uncertainty, focusing on influences from problem structure and modeling fidelity. In particular, Section 2.3 addressed ways to reduce computational intensity using reduction and data compression. We point out here that there is interdependency between the problem structure, modeling fidelity, and reduction/data compression methods used, on the one hand, and the effectiveness of a particular decomposition/parallelization approach. This is the essence of what Figure 2.1 is intended to communicate.

3.1.6 Complexity associated with identification of solution approaches

Identifying the "best" overall solution strategy, in terms of modeling structure and modeling fidelity, algorithm, modeling system/solver, and hardware, is a difficult problem, because the choices that must be made are numerous, and there is significant interdependency between them.

The question is, "Is there a close-to-systematic way to identify the 'best' solution implementation?" To help guide us in "efficient strategy selection," we have developed the chart in Figure 3.4, where the top of the chart indicates what we think are low reward/low effort strategies, and the bottom of the chart indicates what we think are high reward, high effort strategies. The approach, then, is to use the low effort strategies to help inform us of which high effort strategies seem reasonable to pursue[11].

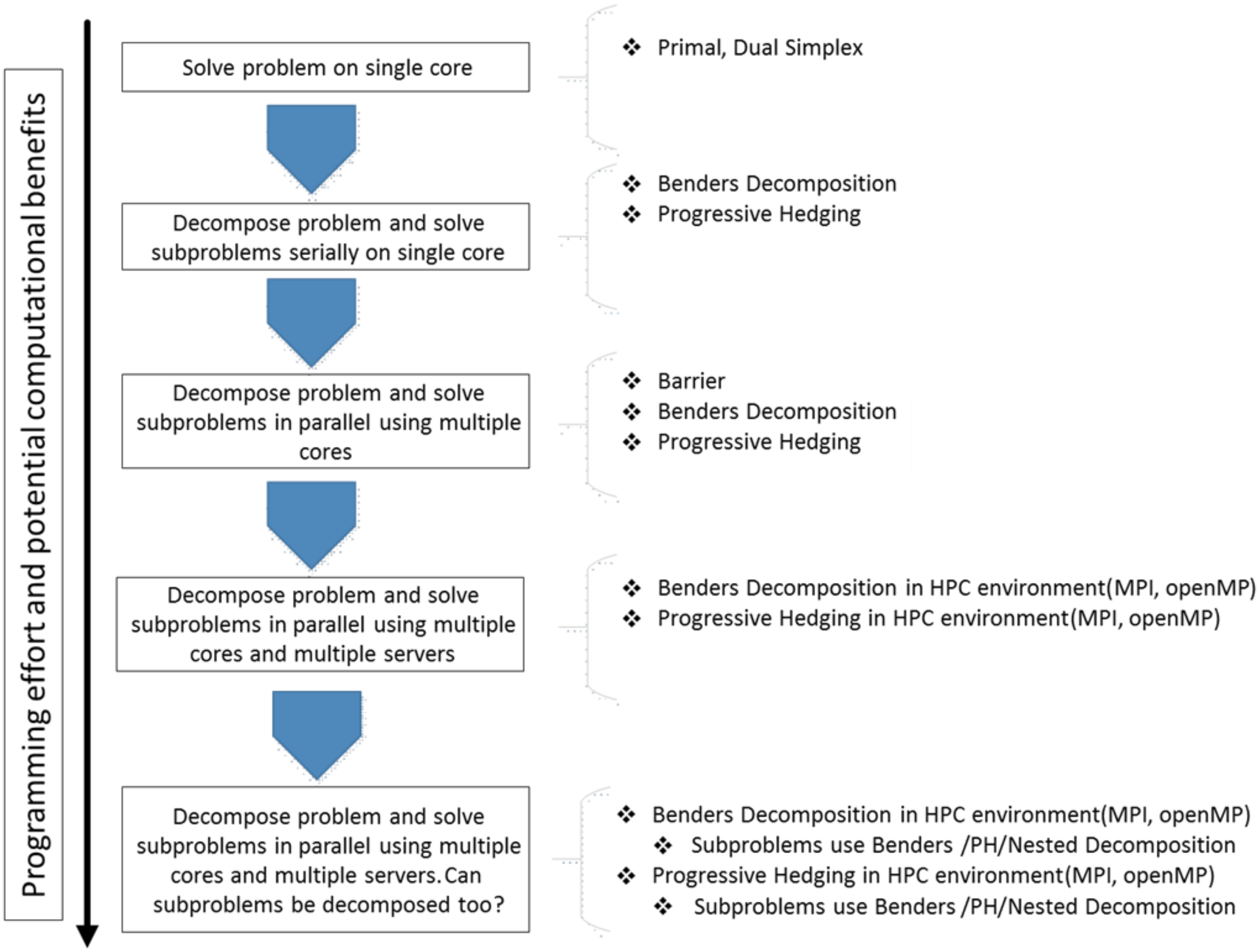


Figure 3.4: Efficient strategy selection

After the first step (baseline), Benders decomposition was applied to the problem to break it into parallelizable problems which could then be run on a distributed memory HPC cluster. A shared memory approach was implemented using the parallel Barrier optimizer in CPLEX. Our work, in terms of decomposition and HPC architecture, are presented in Section 3.2 and Section 3.3 along with results.

[11] Although both Benders decomposition and Progressive Hedging are presented as alternatives in the chart, only Benders decomposition was implemented for this project because of the presence of pre-defined functionality within GAMS, which was not available for Progressive Hedging (however, there is an programming environment, PySP, which could be used in future work to implement it).

3.2 Benders decomposition for LP

Figure 3.5 provides a high-level illustration of our implementation of the AEP problem with Benders decomposition. This implementation decomposes by scenario, implying that the adaptations are computed within the sub-problems for a given core investment.

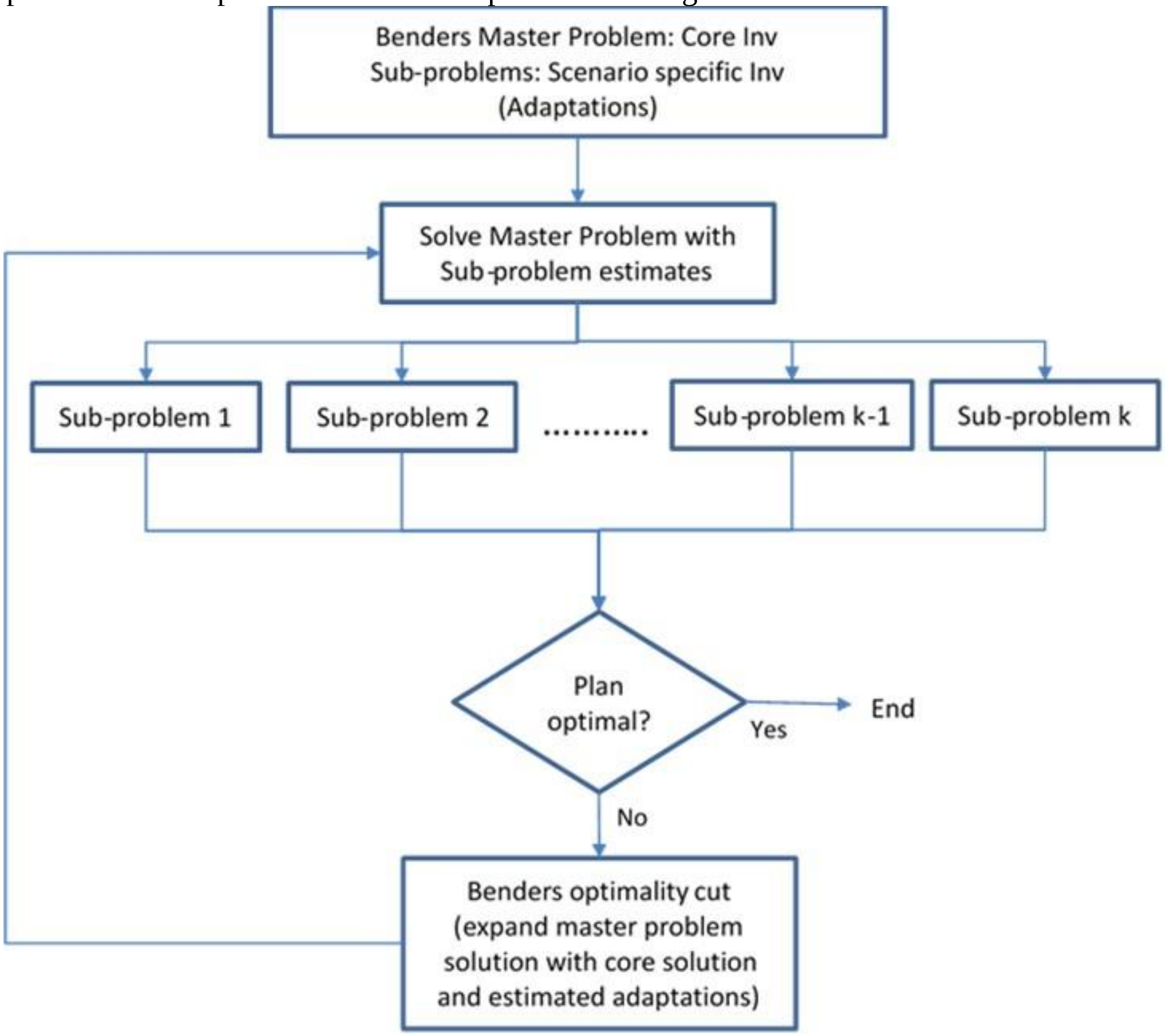


Figure 3.5: Implementation of AEP decomposed by Benders

Within the AEP problem, transmission investments occur in the core and in the adaptations. In both cases, we may implement either the hybrid model or the disjunctive model, and we need not implement the same model in both. Thus we will test two different AEP models. In both models, we represent the adaptation line investments using a hybrid model. The difference in both models lies in how core line investments are modeled - in one model, we represent the core line investments using the hybrid model (making this a linear program or LP). But in the other model, we represent the core line investments using the disjunctive model (making this an MILP – we will call this model Core MIP or CMIP). In both models, we compare the computational intensity with and without Benders decomposition, with results presented in Section 3.3.1.

3.2.1 AEP – general problem formulation

The Benders AEP implementation illustrated by flowchart in Figure 3.5 is described below. This formulation is given as a linear program[12]. When solving this problem without decomposition we refer to this as the Core-MIP or CMIP formulation. When solving this problem with Benders decomposition we refer to this as the Benders Core-MIP or BC-MIP formulation. We would like to be able to model both core and adaptive investments as binary, using the disjunctive approach. However, this would be computationally prohibitive, as it would require integer variables in both the master and subproblems. Furthermore, since the core is regarded as the solution to the AEP, we only end up using the core in implementing the resulting plan.

Indices

t : time
e' : Equations where parameters $a_{e,t}$=0 and $b_{e,t,s}$= 0
e'' : Equations where at least one of $a_{e,t}\neq 0$ and $b_{e,t,s}\neq 0$
e : Equations
s : Scenario
i : iteration through Benders

Sets

E' : Set of equations where parameters $a_{e,t}$=0 and $b_{e,t,s}$= 0 (master problem equations)
E'' : Set of equations where at least one of $a_{e,t}\neq 0$ and $b_{e,t,s}\neq 0$ (subproblem equations)
$E= E'\cap E''$: Set of all equations
T : Set of all time periods t
S : Set of all scenarios s

Parameters

I_t : Investment cost of core investment ΔC_t
$I_{t,s}$: Investment cost of scenario specific investment $\Delta C'_{t,s}$

Decision Variables

ΔC_t : Core investment
$\Delta C'_{t,s}$: Scenario specific investment
$Ops_{t,s}$: Operational costs at time t, scenario s
$y_{t\,s}$: Generic decision variable in model constraints

Parameters

$a_{e,t}$: Constraint coefficients of ΔC_t
$b_{e,t,s}$: Constraint coefficients of $\Delta C'_{t,s}$
$d_{e,t,s}$: Constraint coefficients of $y_{e,t,b,s}$
E_e : Constraint constant value for equation e

[12] However, the implementation can be easily converted into a mixed integer implementation (deploying the disjunctive transmission investment model) by changing the core line investment decision variables type in the master program from continuous to binary.

Problem statement: The AEP optimization problem is stated differently from that given in Chapter 1. The intent of the Chapter 1 statement is to communicate the conceptual understanding of the AEP. The intent of this statement is to facilitate the explanation of representing the AEP via the Benders decomposition. We here give the high-level expression of the AEP problem.

$$min \sum_t I_t \Delta C_t + \beta \sum_{t,s} P_s I_{t,s} \Delta C'_{t,s} + \beta \sum_{t,s} P_s Ops_{t,s} \quad (1)$$

Subject to

$$\sum_t a_{e,t} \Delta C_t + \sum_{t,s} b_{e,t,s} \Delta C'_{t,s} + \sum_{t,s,b} d_{e,t,s} y_{t,s} \leq E_e, \quad \forall\, e \in E \quad (2)$$

We provide additional comments in regard to the objective and the constraints

- **Objective:** The objective minimizes cost for the entire planning horizon. The first term is the cost of the core investments. The second term is the probability-weighted cost of the adaptive investments scaled by β while the third term is the probability-weighted operational costs scaled by β.
- **Constraints:** This is a generic representation of the constraints. It indicates that constraints may contain core decision variables ΔC_t, adaptive decision variables $\Delta C'_{t,s}$, and a number of other decision variables $y_{t,s}$. The subscript e indicates the constraint number. A simplification to the formulation is in the use of subscripts on ΔC_t, $\Delta C'_{t,s}$ with only t or t,s. In fact this only allows for one capacity investment. An additional index could be used to indicate there are many capacity investments. The same is true of $y_{t,s}$ as this is a generic non-capacity based decision variable. An additional index could be added to account for more than one generic non-capacity based decision variable are present.

3.2.2 AEP - Benders decomposition formulation

The general AEP problem formulation of Section 3.2.1 is, in this section, expressed in a form that facilitates its solution via Benders decomposition. Below we express the master problem, the subproblem, and the optimality cuts of the master problem generated in each iteration.

Master problem i – MP$^{(i)}$ (Solve for optimal values $\Delta \boldsymbol{C_t^*}$ and $\boldsymbol{R_s^*}$):
The master problem minimizes the cost of core capacity subject to resource capacity constraints on the core capacity.

$$\zeta = min \sum_t I_t \Delta C_t + \sum_s R_s \ \ where\ \boldsymbol{objMaster} = \boldsymbol{\zeta} - \sum_{\boldsymbol{s}} \boldsymbol{R_s} \quad (3)$$

$$\sum_t a_{e',t} \Delta C_{,t} \leq E_{e'}, \quad \forall e' \in E' \quad (4)$$

$$Optimality\ Cuts \quad (5)$$

Subproblem i for scenario each scenario s' – (SP^{i,s}):
The subproblems fix their core capacity values to those found in the master problem, and then minimize the cost of adaptive investments plus operational costs, subject to constraints (7)-(8).

$$R_{s'} = \ min\left(\beta \sum_{t,s'} P_{s'} I_{t,s'} \Delta C'_{t,s'} + \beta \sum_{t,s'} P_{s'} Ops_{t,s'}\right) \ where\ \boldsymbol{objSub} = \sum_{\boldsymbol{s}} \boldsymbol{R_s} \quad (6)$$

$$\sum_t a_{e'',t} \Delta C_t + \sum_{t,s} b_{e'',t,s} \Delta C'_{t,s} + \sum_{t,s,b} d_{e'',t,s} y_{t,s} \leq E_{e''}, \quad \forall e'' \in E'' \quad (7)$$

$\Delta C_t = \Delta C_t^*$ *where* ΔC_t^* *is the solution to the master (*$\pi_{t,s}$ *corresponds to subproblem duals)* (8)

Optimality Cut – $OC^{i,s}$:
Additional nomenclature used to express the optimality cuts are as follows:
i : (index) : iteration through through the master/subproblem solves
$\Delta C_{t,s}^{(i)}$ (Parameter) : Core values found in *i^th^* iteration of subproblem *s* (actually fixed to master problem values)
$Ops_{t,s}^{(i)}$ (Parameter) : Operational costs found in i^{th} iteration of subproblem *s*
$\pi_{t,s}^{(i)}$: Dual values found in i^{th} iteration of subproblem *s*

The optimality cuts are expressed in (9).

$$R_s \geq \beta \sum_{t,s} P_s I_{t,s} \Delta C_{t,s}^{(i)} + \beta \sum_{t,s} P_s Ops_{t,s}^{(i)} + \sum_{t,s} \pi_{t,s}^{(i)} (\Delta C_t - \Delta C_t^{(i)}) \qquad (9)$$

The optimality cuts that are added to the master problem (5) after each subproblem solve. One optimality cut is generated for each subproblem at every iteration. Once, inserted into the master problem, the decision variables within this equation include R_s and ΔC_t. All other values within the optimality cut of (8) when placed in the master problem (3-5) are parameters. A schematic showing the solution strategy and structure of the Benders decomposition code as applied to AEP is presented in Figure 3.6.

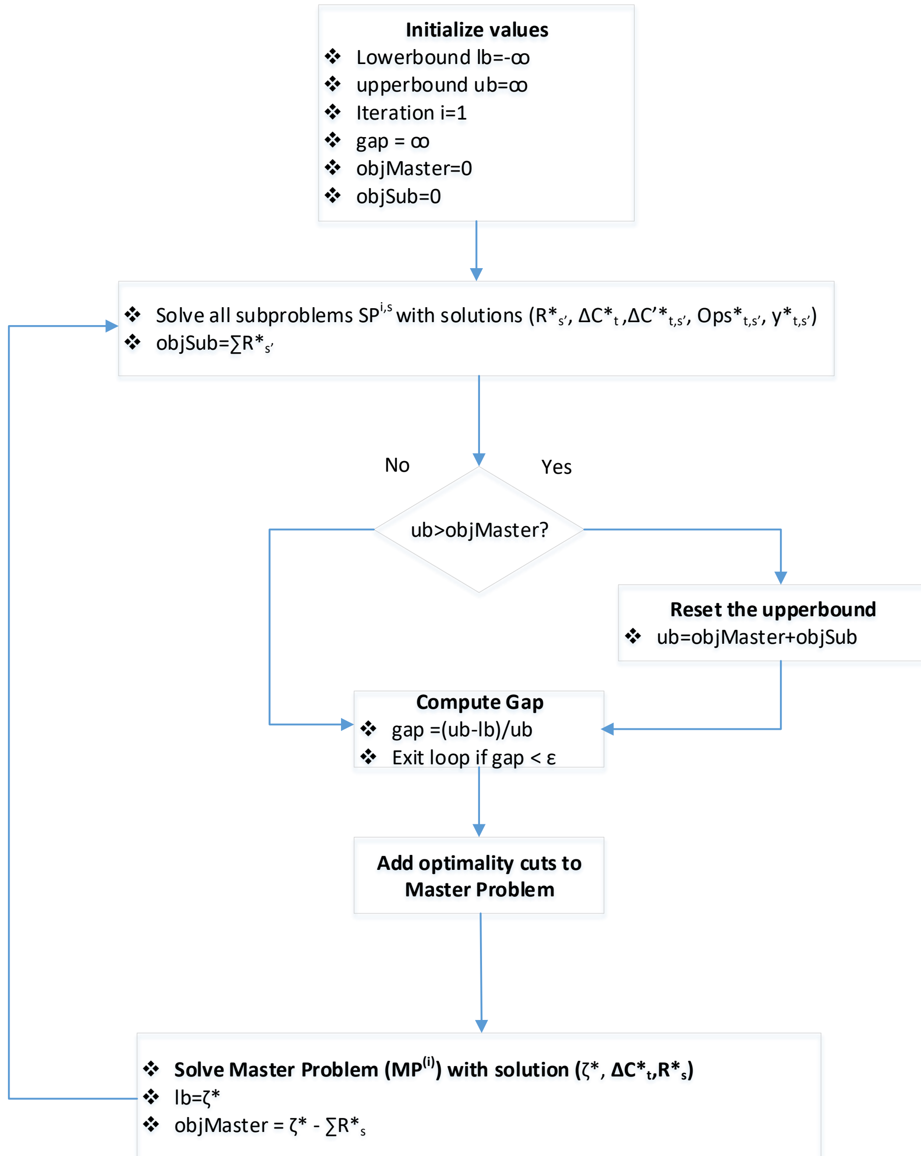


Figure 3.6: Flowchart schematic detailing Benders decomposition as applied to AEP

3.3 Reduction in compute time using shared memory HPC

Results from a shared memory parallel Benders decomposition implementation of the AEP are presented in Section 3.3.1, and results from a shared memory parallel solve using CPLEX Barrier optimizer are presented in Section 3.3.2. Additionally, an ongoing effort to implement the decomposed AEP problem using Benders on a hybrid distributed-shared memory HPC cluster is described in Section 3.3.3.

3.3.1 Shared memory parallel Benders decomposition

We show three results in this section. Figure 3.7 shows the progression of a parallel Benders LP solution in terms of percent gap and upper and lower bounds on the solution. Figure 3.8 and Figure 3.9 compare the CMIP (without Benders) vs BC-MIP (with Benders) solution time. All simulations are run in a shared memory parallel computer on the HPC CyEnce/Condo cluster at Iowa State University, consisting of two 2.0GHz 8-core Intel E5 2560 processors per node (yielding a total of 16 cores per node) with 128GB memory.

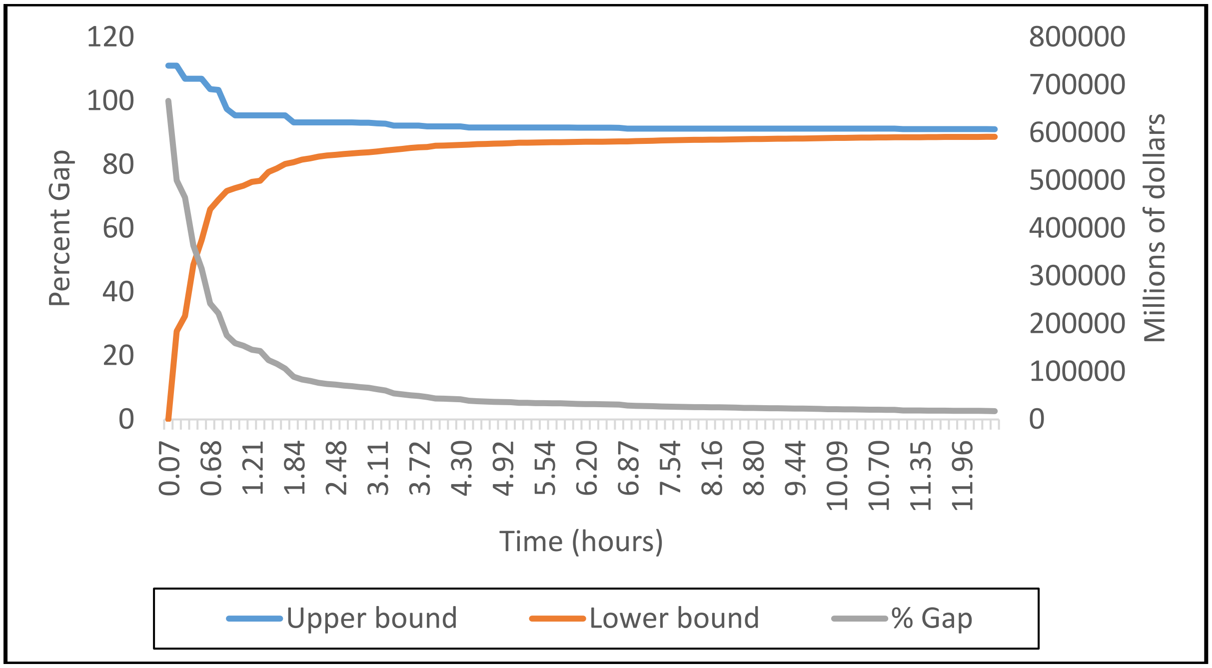


Figure 3.7: Parallel Benders LP

In Figure 3.7, Benders decomposition is applied to the 8 scenario LP version of the AEP problem, with all variables being continuous and no integer variables. Compared to the baseline time of 22 hours for a serial solution to solve to optimality (without decomposition), it takes about 6 hours for the parallel Benders LP to reach a 5% gap and about 11 hours for it to reach a 2.5% gap at 100 iterations.

Figure 3.8 compares, for a 2-scenario problem, performance of the C-MIP implementation (without Benders – we call this C-MIP normal solve) and of the BC-MIP implementation (we call this Benders C-MIP), where the left hand side column represents C-MIP normal solve and the right hand side column represents Benders C-MIP. We observe that during the 3.5 hours of the simulation, Benders C-MIP makes steady progress while reducing its gap to ~ 5%. On the other hand, the C-MIP normal solve makes almost no progress until about 3.5 hours into the simulation when it drops from ~99% gap to 0.5% gap after finding a particularly good integer feasible solution. Thus even though Benders C-MIP is faster for the first 3 hours and makes more progress during this time, the C-MIP normal solve catches up later in the simulation.

Despite C-MIP normal solve outperforming Benders C-MIP, it is satisfying to see that the C-MIP normal solve 0.5% gap upper and lower bound falls within the Benders C-MIP ~5% gap upper and lower bounds after 3.5 hours. This helps validate that the Benders C-MIP is working properly and that both the formulations' codes are comparable.

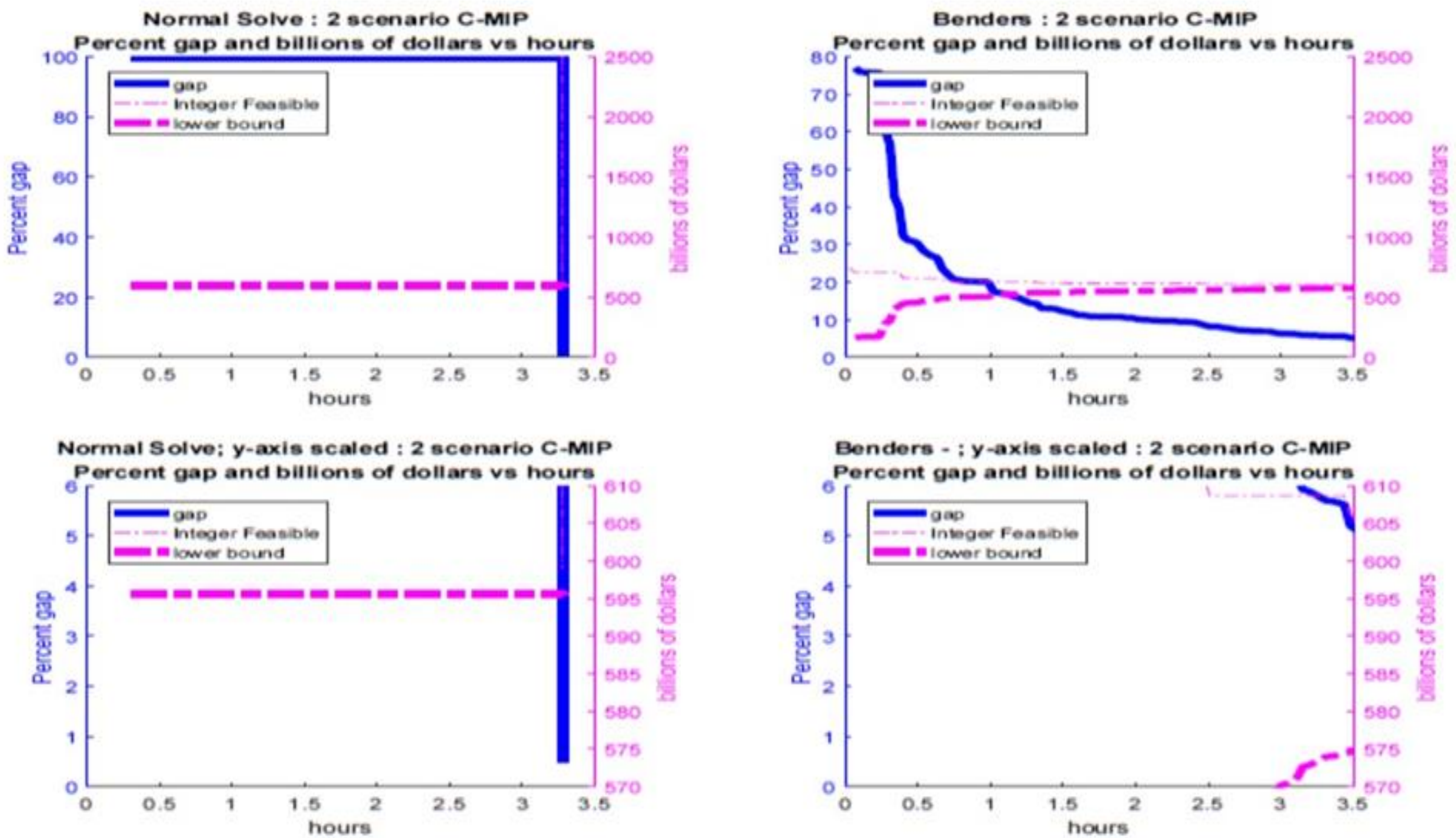


Figure 3.8: Comparing 2-scenario CMIP with and without Benders

While the result of Figure 3.8 does not show the Benders C-MIP as faster than the C-MIP normal solve, it does motivate a second test shown in Figure 3.9. This test is identical to that of Figure 3.8, except it uses 8 scenarios instead of 2. Given the results of Figure 3.8, we might expect that for an increase in continuous decision variables in both CMIP normal solve and Benders C-MIP, the Benders C-MIP might start outperforming the C-MIP normal solve. This is because each of the LP's solved at each step of the branch and bound algorithm in the C-MIP will take longer. This is exactly what moving from 2 scenarios to 8 scenarios does. The 8 scenario problem has exactly the same number of integer decision variables but substantially more continuous decision variables. However, because Benders C-MIP processes the sub-problems in parallel (which is where the additional continuous decision variables are placed), the increase in scenarios has little effect on the Benders C-MIP.

We observe in Figure 3.9 that the Benders C-MIP far outperforms C-MIP in terms of finding a reasonable integer feasible solution (upper bound) as well as reducing the MIP gap. While the C-MIP normal solve barely budges from 100% gap in ~24 hours, the Benders C-MIP attains a ~ 2% gap in 24 hours. It is important to note in this result that the upper bound is the integer feasible solution in both the left hand and right hand columns of Figure 3.9. Thus, even though the C-MIP normal solve obtains a better lower bound than the Benders C-MIP, the lower bound is not an integer feasible solution (as indicated in Section 3.1.3, the lower bound is the solution to the problem where all integers are relaxed to be continuous variables).

Results from the experiments conducted by implementing a MIP version of the AEP problem (with and without Benders decomposition) clearly indicate that solving an MIP brings with it much greater computational complexity than an equivalent LP, which is a well-known fact. However, it is interesting to note that within the MIP experiments, implementing Benders decomposition helps to find an integer feasible solution as the number of scenarios increases, as indicated in Figure 3.9.

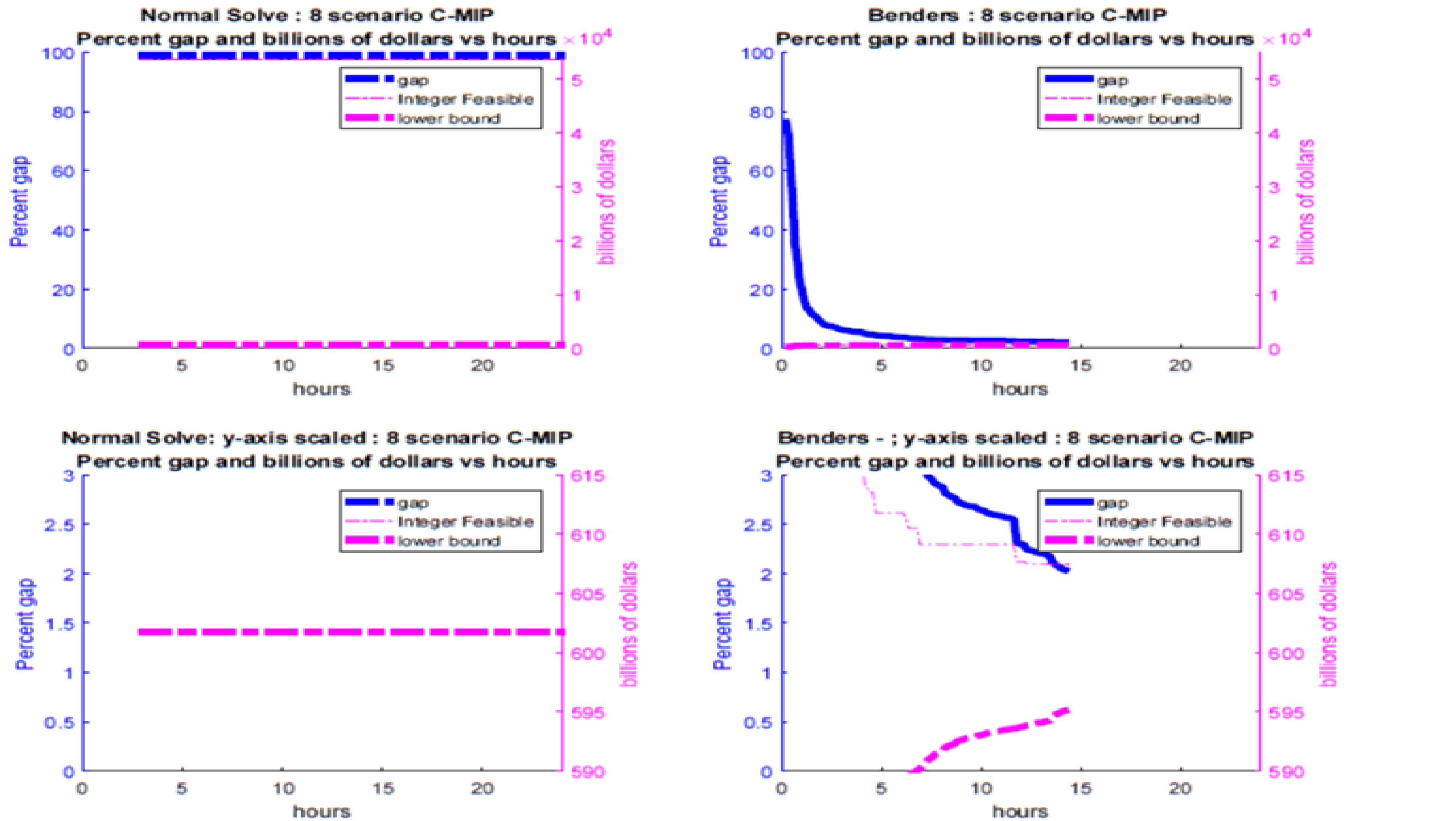


Figure 3.9: Comparing 8-scenario C-MIP with and without Benders

3.3.2 Shared memory parallel solve using CPLEX Barrier optimizer

Figure 3.10 shows results plotting total computation time against the number of cores used in a shared memory implementation using the CPLEX parallel Barrier optimizer. The results are shown for the same AEP problem applied to the BPA system used in the baseline serial computation (1 core) with the only change being the increased number of processors used in the computation. All simulations are run in a shared memory parallel computer on the HPC CyEnce/Condo cluster at Iowa State University, consisting of two 2.0GHz 8-core Intel E5 2560 processors per node (yielding a total of 16 cores per node) with 128GB memory.

From a serial implementation time of approximately 22 hours on 1 core, an order of magnitude reduction in computation time is observed when increasing number of cores. At the maximum number of cores, 16, the parallelized shared memory implementation takes 1.91 hours, a significant improvement.

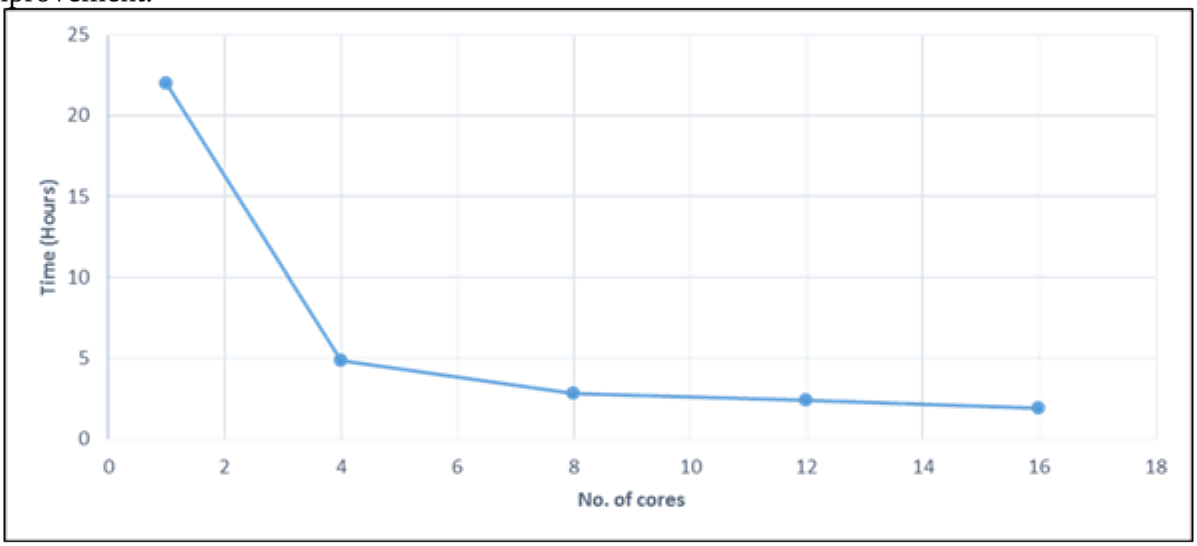


Figure 3.10: Shared memory parallel solve using CPLEX barrier – plotting total compute time vs. number of cores used

Computation times were also noted for varying number of cores, with an excellent improvement from 22 hours to 5 hours when leveraging parallel computing using just 4 cores, and to 4 hours when using 8 cores. Beyond this point, the benefit per core decreases as the number of cores are increased, because of limitations imposed by the serial part of the problem being solved (consistent with Amdahl's Law).

3.3.3 Hybrid distributed-shared memory parallel solve using MPI

Although the CyEnce/Condo HPC cluster at Iowa State University can be used in a shared memory parallelized form, as described in Section 3.3.2, it is actually intended for use as a hybrid distributed-shared memory cluster. Having observed a significant reduction in total compute time with the shared memory parallel implementation, a hybrid implementation would be expected to perform better owing to more number of nodes available for parallel computation, as long as the communication latency due to data transfer between the nodes does not outweigh the advantages of faster computation in parallel.

Benders decomposition, as described in Section 3.2, enables the use of a hybrid HPC architecture by breaking down the AEP problem into a master problem and multiple sub-problems which are independent of each other, and can thus be solved simultaneously in parallel. In section 3.4.1, the results presented are from a shared memory implementation of the Benders decomposition problem as applied to AEP. The same formulation can be converted to run on a hybrid HPC cluster by embedding Message Passing Interface (MPI) code.

The particular modeling system we are using, GAMS, has built-in support for using MPI. This built-in support can be accessed in two ways: (1) embedded MPI code within GAMS (2) Python package mpi4py interfaced with GAMS. In our efforts to employ the hybrid HPC architecture for the AEP Benders decomposition problem, we chose the embedded MPI code within GAMS (based on the spbenders example in the GAMS model library) in order to consolidate all the code within one modeling platform. Figure 3.11 shows the computation flow using MPI in GAMS, as applied to the AEP Benders decomposition problem. Although the structure of the problem remains the same (as described in Section 3.2), the principal difference lies in the way the sub-problems are handled, as described in the next paragraph.

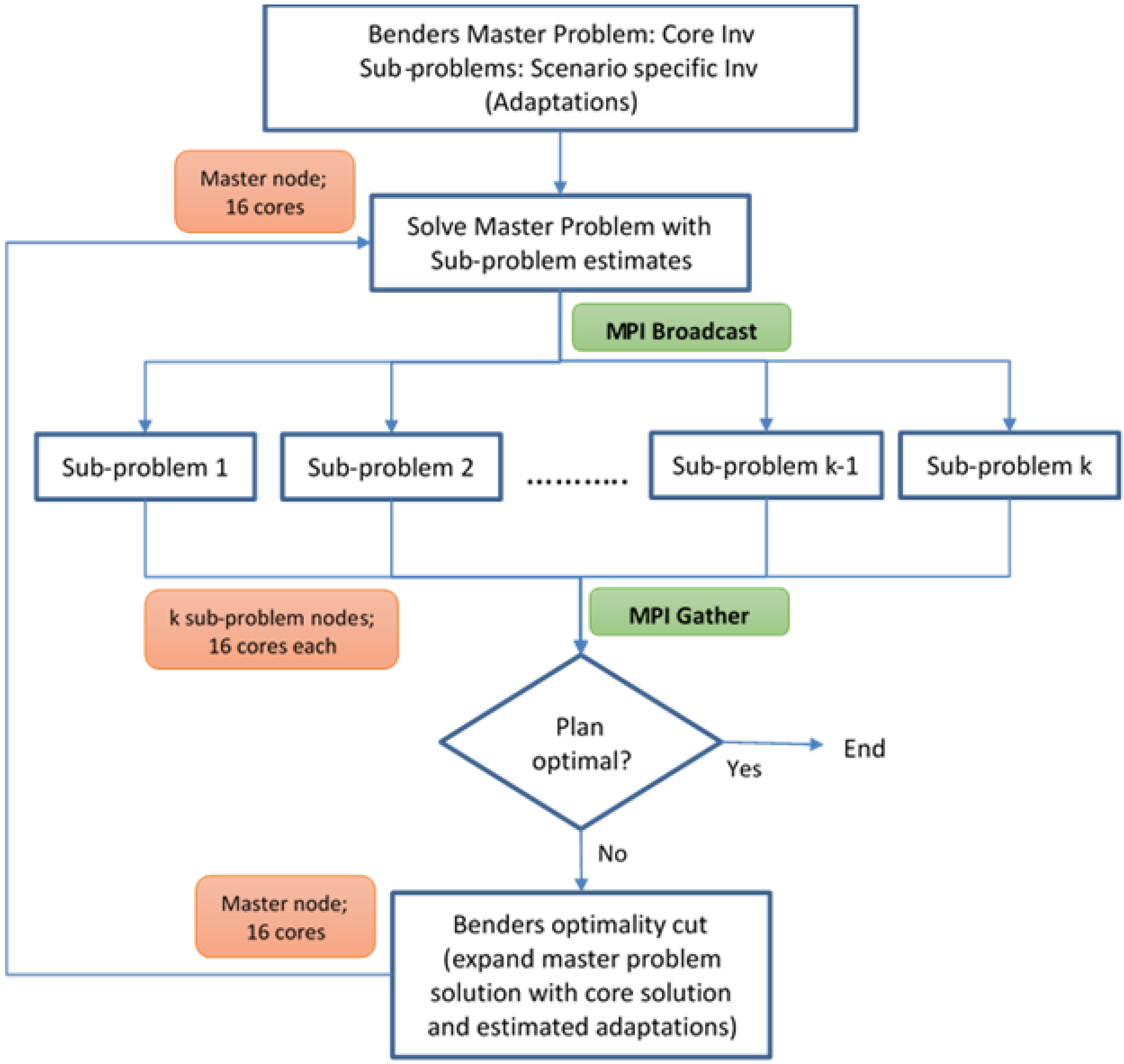


Figure 3.11: Flow of computation using MPI embedded in GAMS for hybrid distributed-shared parallel HPC implementation of Benders decomposition applied to AEP

In the shared memory implementation, the sub-problems were spawned as multiple threads, each using a specified number of cores for computation. In a basic hybrid implementation, the MPI command to start computation, while calling GAMS, spawns *k*+1 different copies of the Benders decomposition model (where *k* is the number of sub-problems), with the MPI environment variable PMI_RANK deciding what role each part of the particular instance plays. The GAMS job with PMI_RANK=0 implements the master problem on one node (core investments and sub-problem estimates), with inherent shared memory parallel computing capabilities using parallelized CPLEX Barrier solver. GAMS jobs with PMI_RANK=1 to *k* implements the sub-problems on *k* different nodes, with inherent shared memory parallel computing capabilities of their own. The GAMS MPI embedded code facilitates the communication of the master variables received by each of the *k* nodes through an MPI Broadcast command, and at the end of one iteration of the hybrid parallel solve, MPI Gather command is used to communicate the sub-problem variables and cut information. Since this is a work in progress, results are expected to show compute times within the range of the shared memory parallel implementation using CPLEX barrier solver.

3.4 Other solution strategies

We intend that we will very soon complete our work on the hybrid distributed/shared memory parallel solver using MPI, and we expect to obtain solve times for the model size we are investigating to be less than one hour. This is a significantly improvement over day-long solve times. However, this does not enable AEP application on large-scale industry-size models between 20,000 and 70,000 buses. To achieve this, we intend to work along in two different directions, as described below.

1. *Further efforts in paralleled HPC*: Here, we will investigate nested decomposition methods, where, for example, we decompose by both scenarios (as done in this project) and operating conditions. Progressive hedging also is a promising direction.
2. *Internal/external production simulation*: Because the production simulation (PS) function is the most compute-intensive, it may make sense to provide a higher-fidelity PS application external to the expansion planning application. We have already developed a high-level design of this concept and we hope to begin working on it very soon.

4. Conclusions

It is important to consider uncertainty in expansion planning since the infrastructure investments will differ when compared to deterministic models which consider only one scenario, and these differences matter in terms of potential economic savings while planning for an uncertain future. However, these problems are computationally intense, especially when considering uncertainty, but are highly useful in exploring the future. This project has shown that adaptive expansion planning problems are parallelizable, and that the use of decomposition techniques to harness the power of parallel computing is indeed promising.

Numerical experiments performed in this project have clearly demonstrated how a modest increase in some of the factors which affect computational intensity of expansion planning problems can have an almost detrimental effect on the total compute time. With serial computation, improving model fidelity by reducing assumptions almost always made the problem intractable. However, using decomposition and parallel computing applications, this project has been able to show that in the same time it takes to run a serial implementation of the problem, parallel computing can enable a much higher level of model fidelity and accuracy, while also solving problems of the same model fidelity as the existing serial implementation approximately 10 times faster.

As with any other research project, there remain several areas of interest for future work. They include (but are not limited to) using different decomposition methods such as nested decomposition (by scenarios and operating blocks, or by time period and geographical region) to further break down the problem, exploring hybrid decomposition techniques (Benders decomposition and Lagrangian relaxation) and Progressive Hedging (for which tools are available such as PySP).

References

[1] International Atomic Energy Agency, “Expansion planning for electrical generating systems: a guidebook,” 1984, available at www-pub.iaea.org/MTCD/publications/PDF/TRS1/TRS241_Web.pdf.

[2] P. Maloney, P. Liu, Q. Xu, J. McCalley, B. Hobbs, S. Daubenberger, A. Johnson, and S. Williams, “Wind capacity growth in the Northwest United States: cooptimized versus sequential generation and transmission planning,” Wind Engineering, 1-23, DOI 10.1177/0309524X188149, January 2019.

[3] E. Spyrou; J. Ho; B. Hobbs; R. Johnson; J. McCalley, “What are the Benefits of Co-optimizing Transmission and Generation Investment? Eastern Interconnection Case Study” IEEE Transactions on Power Systems, Year: 2017, Vol: 32, Issue: 6, pp. 4265-4277, DOI: 10.1109/TPWRS.2017.2660249.

[4] J. McCalley, P. Maloney, P. Liu, B. Hobbs, and Q. Xu, “Cooptimization and anticipative planning methods for bulk transmission and resource planning under long-run uncertainties,” Final Project Report, submitted to the Bonneville Power Administration, September 28, 2018.

[5] S. Lemos-Cano and J. McCalley, “Co-optimized analysis and design of electric and natural gas infrastructure,” Energies, May, 2019, 12, 2012; doi:10.3390/en12102012.

[6] E. Ibanez, and J. McCalley, “Multiobjective evolutionary algorithm for long-term planning of the national energy and transportation systems,” *Energy Systems Journal*, Vol. 2, Is. 2 (2011), pp 151-169.

[7] K. Oikonomou, M. Parvania, and R. Khatami, "Optimal Demand Response Scheduling for Water Distribution Systems," IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 5112-5122, 2018.

[8] K. Oikonomou and M. Parvania, "Optimal Coordination of Water Distribution Energy Flexibility With Power Systems Operation," IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 1101-1110, 2019.

[9] A. Orths, C. Anderson, T. Brown, J. Mulhern, D. Pubdianto, B. Ernst, M. O’Malley, J. McCalley, and G. Strbac, “Flexibility from Energy Systems Integration,” to appear in IEEE Power and Energy Magazine.

[10] F. Munoz, B. Hobbs, J. Ho, and S. Kasina, “An Engineering-Economic Approach to Transmission Planning Under Market and Regulatory Uncertainties: WECC Case Study,” IEEE Transactions on Power Systems, Vol. 29, No. 1, January, 2014.

[11] A. Liu, Q. Zheng, J. Ho, *V. Krishnan, B. Hobbs, M. Shahidehpour, and J. McCalley, “Co-optimization of Transmission and Other Supply Resources," NARUC Project No. 3316T5, prepared for the Eastern Interconnection States Planning Council (EISPC), September 1, 2013.

[12] CAISO, “2011-2012 Transmission Plan,” California ISO, March 2012.

[13] MISO, “Regional Generation Outlet Study,” Midwest ISO, November 2010.

[14] P. Maloney, “Methods for cooptimization planning and plan validation under uncertainty,” Ph. D. dissertation, Iowa State University, 2019.

[15] WECC TEPPC 2024 Datasets, https://www.wecc.org/SystemAdequacyPlanning/Pages/Datasets.aspx

[16] Y. Zhu and D. Tylavsky, “An optimization-based dc-network reduction method,” IEEE Transactions on Power Systems, 3(3):2509-2517, 2017.

[17] C. Draxl, B. Hodge, A. Clifton, and J. McClan, "Overview and meteorological validation of the wind integration national dataset toolkit," Technical Report NREL/TO-5000-61740, 2015.
[18] A. Habte, M. Sengupta, and A. Lopez, "Evaluation of the national solar radiation database (NSRDB): 1998-2015, technical report nrel/tp-5d00-6722, 2017.
[19] N. Blair, A. Dobos, J. Freeman, T. Neises, M. Wagner, T. Ferguson, and S. Janzou, "System advisor model, SAM 2014.1.14: general description," Technical Report, NREL/TP-6A20-61019, 2014.
[20] NREL Annual Technology Baseline Workbook. https://atb.nrel.gov/electricity/archives.html, 2016.
[21] WECC, 2014 TEPPC Generation Capital Cost Calculator. https://www.wecc.biz/Reliability/E3 WECC ProForma FINAL.xlsm.
[22] WECC, 2014 TEPPCC Transmission Capital Cost Report, available at www.wecc.biz/Reliability/2014_TEPPC_TransCapCostCalculator.xlsx.
[23] V. Krishnan and J. McCalley, "Building foresight in long-term infrastructure planning using end-effect mitigation models," DOI 10.1109/JSYST.2015.2410758, IEEE Systems Journal, Vol. 11, Issue 4, 2017, pp. 2040-2051.
[24] C. Newlun, A. Figueroa, J. McCalley, A. Kimber, and E. O'Neill-Carrillo, "Co-optimized expansion planning to enhance electrical system resilience in Puerto Rico," to appear in Proc. of the North American Power Symposium, September, 2019, Wichita, Kansas.
[25] G. Kron, "Tensor Analysis of Networks," Wiley, 1939.
[26] R. Podmore and A. Germond, "Development of Dynamic Equivalents for Transient Stability Studies," Final Report, EPRI EL-456, 1977.
[27] J. McCalley, J. Dorsey, Z. Qu, J. Luini, and J. Filippi, "A new methodology for determining transmission capacity margin in electric power systems," IEEE Transactions on Power Systems, Vol. 6, Issue 3, 1991.
[28] A. Figueroa, "Opportunities and benefits for increasing transmission capacity between the US eastern and western interconnections," Ph. D. Dissertation, Iowa State University, 2017.
[29] J. McCalley, P. Maloney, P. Liu, B. Hobbs, and Q. Xu, "Cooptimization and anticipative planning methods for bulk transmission and resource planning under long-run uncertainties," Final report provided to the Bonneville Power Administration.
[30] Y. Li and J. McCalley, "An innovative disjunctive model for value-based bulk transmission expansion planning," Electric Power Systems Research, vol. 143, pp. 7-13, 2017/02/01/ 2017.
[31] J. Benders, "Partitioning procedures for solving mixed variables programming problems," Numerische Mathematics, 4, 238-252, 1962.
[32] A. Geoffrion, "Generalized benders decomposition", Journal of Optimization Theory and Applications, vol. 10, no. 4, pp. 237–260, Oct. 1972.
[33] G. Dantzig and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, 8: 101–111. doi:10.1287/opre.8.1.101, 1960.
[34] J. Bloom, "Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Benders' Decomposition," Operations Research, Vol. 31, No. 1, January-February 1983.
[35] J. Bloom and M. Caramanis, "Long-Range Generation Planning Using Generalized Benders' Decomposition: Implementation and Experience," Operations Research, Vol. 32, No. 2, March-April, 1984.

[36] J. Bloom, “Long-Range Generation Planning Using Decomposition and Probabilistic Simulation,” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 4 April 1982.

[37] J. Watson and D. Woodruff, "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," Computational Management Science, vol. 8, no. 4, pp. 355-370, 2011/11/01 2011.

[38] M. Guignard, "Lagreangean Relaxation," Sociedad de Estadistica e Investigacion Operativa, MadridDec. 2003, Available: https://perso.ensta-paristech.fr/~diam/ro/online/Monique.Guignard-top11201.pdf.

[39] I. Grossman and B. Tarhan, “Tutorial on Lagrangean Decomposition: Theory and Applications,” presentation slides, available at http://egon.cheme.cmu.edu/ewo/docs/EWOLagrangeanGrossmann.pdf.

[40] T. Roy, "A Cross Decomposition Algorithm for Capacitated Facility Location," *Operations Research*, vol. 34, no. 1, pp. 2-184, Jan.-Feb. 1986.

[41] C. Lara, D. Mallapragada, D. Papageorgiou, A. Venkatesh, and I. Grossmann, "Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm," European Journal of Operational Research, vol. 271, no. 3, pp. 1037-1054, 2018/12/16/ 2018.

Part III

Optimization-Based Methodology for Identifying Candidate Transmission Expansion Scenarios

Christopher L. DeMarco
Jonathan Snodgrass, Graduate Student

University of Wisconsin-Madison

For information about this project, contact

Christopher L. DeMarco
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
1415 Engineering Dr.
Madison, WI 53706
Phone: 608-262-5546
Email: cdemarco@wisc.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, AZ 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Notice Concerning Copyright Material

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Ideally, Transmission Expansion Planning (TEP) seeks to identify transmission lines and other equipment whose addition to the transmission system improves system operation, reliability, and economics. Ideally, one might seek to optimize with respect to appropriate criteria, over the "decision space of possibilities," potentially consisting of all plausible line and equipment upgrades. As noted elsewhere in this report, the problem is extremely challenging due in part to the range of criteria to be considered in optimizing (even approximately) over the multiyear future time horizon appropriate to long-lived transmission capital equipment, while accounting for the many uncertainties inherent in such a decision problem. However, the uncertainty and long time horizon is not the only part of the challenge. The TEP also presents a huge decision space, imposing a computational cost that grows rapidly with network size. Consider a "green field" problem, starting a network of *n* buses: one is confronted with {*n choose 2*} possible pairs for originating bus and terminating bus when adding just one new transmission line. While it is certainly true that the large number of social and land use constraints confronting transmission line routing will greatly reduce this number, and expansion decision that allow for multiple line and equipment upgrades in a large network still presents a huge number of credible transmission expansion plan possibilities to be considered. The goal in this portion of the project is to develop a computationally efficient but approximate optimization formulation of the TEP problem, that will allow a preliminary, first search over all possible transmission additions. From this approximate optimization, one can then dramatically reduce the search space. This computationally efficient step produces a much smaller set of candidate scenarios, those that ranked highly in the approximate optimization, that then serve as a computationally tractable search space for more complete (and more computationally costly) analysis.

Traditionally transmission expansion was handled by vertically integrated utility companies, who would present their proposals to the public service commission (PSC) or public utility commission (PUC) of the states within their operating territory. In the traditional vertically integrated environment, both generation and transmission aspects of system planning were managed within the same organization, the planning projects could be coordinated and information could be freely shared.

However, with the introduction of competitive power system markets, the previously vertically integrated utilities were split into 3 groups: generating companies (or independent power producers, IPPs), regional transmission operators (RTOs), and load serving entities (LSEs). These three groups of entities are not permitted to share information as they previous did during the utility monopoly days, thus increasing the complexity of transmission expansion. Additionally, the inclusion of renewable energy resources has introduced increased levels of variability into the electric grid, so transmission planning must account for greater uncertainly in the nature, location, and time variability of generation. This is highly relevant to any optimization formulation of the transmission planning problem, because even a rough approximate optimization formulation should ensure power delivery feasibility for the plan; i.e., that the proposed transmission system can successfully serve the projected future load scenarios, with the projected future generation. One measure of the degree of approximation versus completeness in the analysis is the degree to which accurate power flow modeling is employed in judging the feasibility of power delivery. In the most efficient but least accurate approximation, one may require only system wide power balance, with sum of generation matching load (perhaps considering necessary reserve, and/or treating an approximation of losses). In the most detailed representation, one might insist upon a full AC power flow representation, perhaps with consideration of *n-1* contingency criteria, optimized at some relatively fine time sampling rate (e.g., hourly), over perhaps a 20 year time horizon that includes the considerations of capital cost expenditures associated with a given expansion scenario.

Given these complexities and the size of the transmission networks, it is clearly computationally intractable to consider all potential transmission upgrades, and create scenarios for all possible locations and capacities of generation resources, all while treating the full details of security constrained AC optimal power flow. Thus, various modeling approximations and relaxations will be employed here. Again, the goal will be to provide a computationally tractable tool that can approximately explore the space of all possible transmission upgrades, and from that space extract a much smaller set of plausible expansion plans to pass on to other, more detailed analyses.

1.2 Overview of the Problem

Fully and rigorously formulated, transmission expansion planning (TEP) presents a large-scale mixed integer non-linear programming (optimization) problem (MINLP). Currently, no tractable methods exist for solving large-scale MILNPs in a reasonable amount of time, so a combination of one or more of the following is utilized: simplifications to the power flow model [1, 2], relaxations of other transmission expansion optimization constraints [3], or implementation of various heuristic algorithms [4, 5]. The authors' method falls in to the first category of using a simplified version of the power flow equations.

In this portion of the project, we seek to introduce a tractable optimization approach for transmission expansion. The approach here uses a sequence of linear programs, starting from a complete network graph (i.e., consideration of *every possible* bus-to-bus interconnection). Any existing lines in the network are treated as fixed, known paths in the graph, typically not to be changed (unless equipment retirements are also part of the decision to be made). All other possible paths, that at the outset include every possible bus-to-bus interconnection, represent candidate additions. Each pass through the "inner loop" linear program solves a minimum-cost flow problem, and as a by-product, identifies low-impact, low-flow paths that may be deleted from further consideration. Note that this inner loop only imposes active power balance at each bus, without consideration of the "branch relation" that dictates flow through a line as a function for voltage difference across the line. In the dc circuit analogy that is the basis of the DC power flow approximation, one may say that the inner loop linear program gains efficiency by imposing only KCL constraints, relaxing KVL and branch relation constraints. The inner loop is then followed by a DC power flow, performed sequentially with updates of line parameters to move toward a solution that maintains active power flow on each line commensurate with its geographic length, and with selection from a discrete list of allowable voltage levels, and associated conductor type/spacing (the latter determining effective susceptance for the DC power flow). The set of representative voltage levels and conductor type/tower type/spacing was assembled through compiling of these statistics for many parts of the U.S., based on publically reported data in Federal Energy Regulatory Commission's (FERC) Form 1.

A brief overview of the new transmission expansion algorithm is as follows. First the researchers or entity conducting the study must assemble the required load cases of interest, along with descriptions of available generation (location, min/max power constraints, and piecewise linear of quadratic cost curve). Note that in the approximate analysis conducted, there are no inter-temporal constraints considered, so that each load/available generation cases is a single snapshot; the transmission optimization is performed with the constraints associated with each snapshot, so that the resulting network is DC power flow feasible for every snapshot. For the transmission network, an approximate line length in miles must be identified for every possible candidate line. For studies in this project, these candidate line length values were obtained by applying a fixed multiplier to the straight line mileage between the originating and the terminating bus.

A key point in maintaining tractability for the inner loop linear program is the very simple objective function used. The magnitude of MW flow on each line is weighted, with a simple selection of weighting factor being that of the line length in miles. In this case, the objective being minimized in the inner loop can be thought of as "MegaWatt-Miles," summed across every line in the network. If more detailed information is available, the weighting factors among lines can be further differentiate by the relative per mile cost of construction on a given transmission path

After iteratively running the linear program, a new network topology is generated that includes either new transmission lines to be added to an existing system, or an entirely new greenfield system. Then, a series of sequential DC power flow calculations are performed to select the voltage level and transmission line parameters of the candidate transmission lines.

2. Transmission Expansion Planning Data

2.1 Geographic Data

The first step in the transmission expansion algorithm considered here is to select the geographic region of study, with GPS coordinates for buses that may then be used to estimate candidate transmission line path mileage for every line, as described above. For numerical experiments using the algorithm to create wholly synthetic networks, it was observed that there exists close agreement between the number of census tracts and number of electrical substations across the United States as a whole, and reasonably close agreement on a state-by-state basis. Therefore, in cases for which exist substation GPS location data was not used (due to CEII concerns in publicly available research literature, such real-world GPS data was not used here), US census bureau tract data was used to break each state up into sub-areas to assign substations across a given region of the United States. For the census tracts that included bodies of water or sensitive environmental areas, the centroid calculations were modified to avoid placing the substation in these areas. 0 shows the results of the Geographic Information Systems (GIS) work to site hypothetical substations, with each small dot representing a plausible but synthetic substation location used in test studies of the algorithms here.

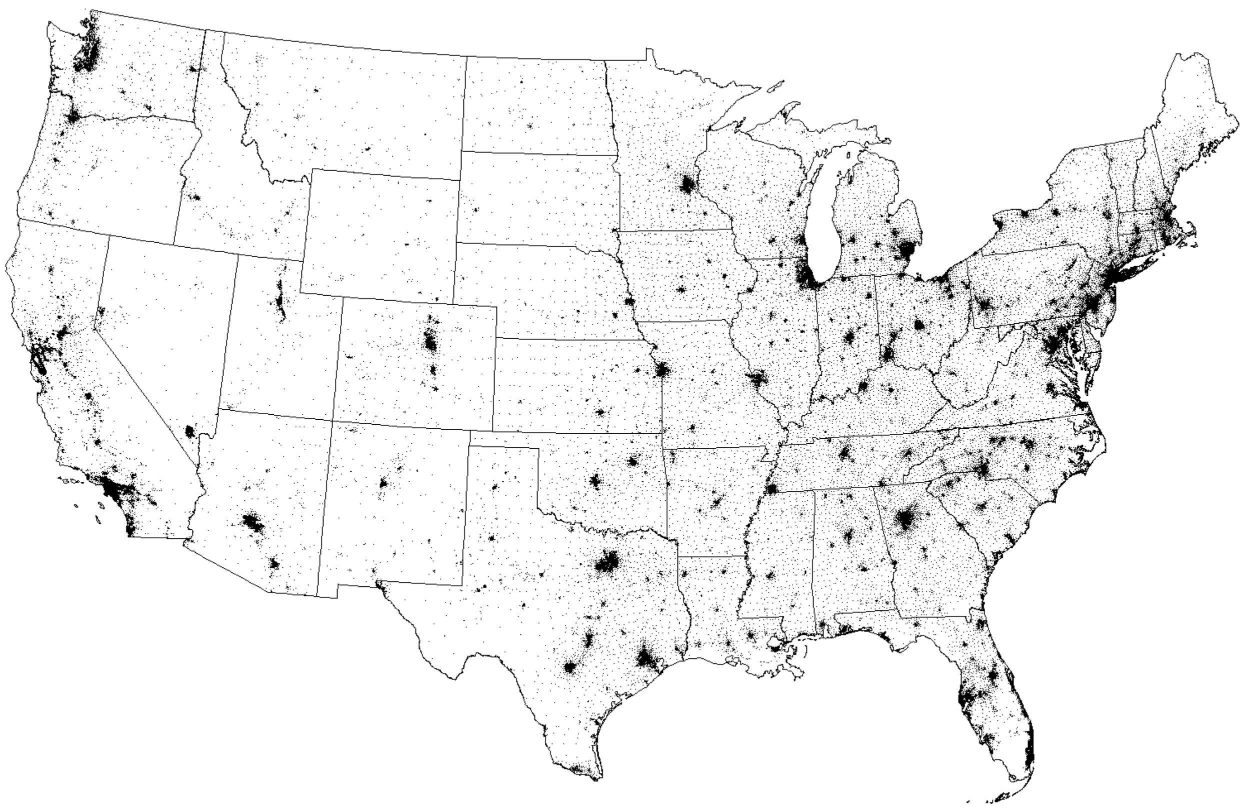


Figure Transmission Expansion Planning Data.1 Synthetic substation locations across continental U.S, for algorithm testing

2.2 Creation of Load Data TEP Algorithm Testing

Next, in order to test the algorithms with non-proprietary data, synthetic load demand was assigned to each of the synthetic substation locations. Load demand data was created using a combination of US census bureau data, land use databases (0 and 0 below), peak electrical loads by state or region, load profiles by region and load profiles by load class. Peak load data was collected for each state. As a "zero order" approximation, load then apportioned to each synthetic substation within the based on the relevant census tract's percentage of the state's population. For more advanced studies performed, percentage of loads in each class (residential, commercial, industrial) was estimated for each census tract by land use data associated with that census tract. The percentage of total load that was apportioned to each census tract was determined by the land use intensity database. Finally, hourly load data for the entire year was created using the load class information and hourly load profiles obtained from publicly available data as published by the major U.S. ISOs. The flowchart for this process is shown below in 0

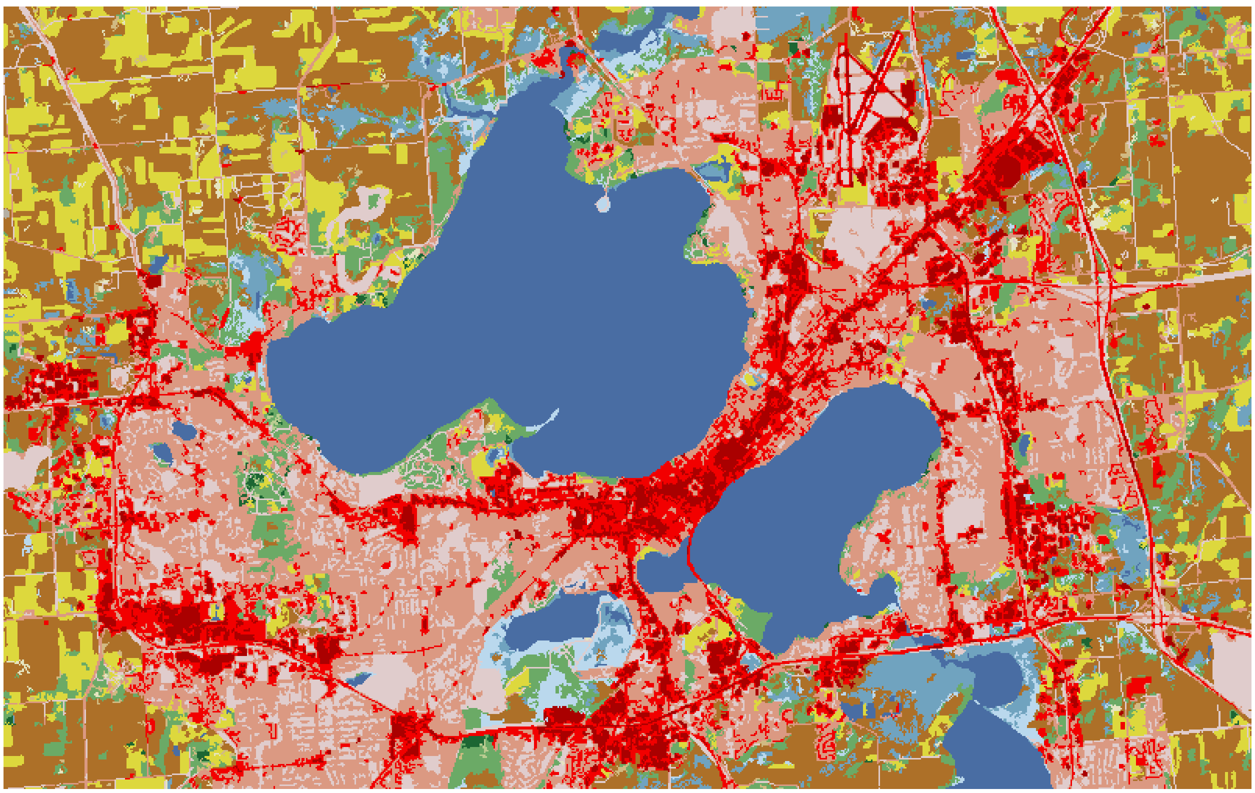


Figure Transmission Expansion Planning Data.2 National land cover database 2011 (NLCD)

Figure Transmission Expansion Planning Data.3 Key for the NLCD figure

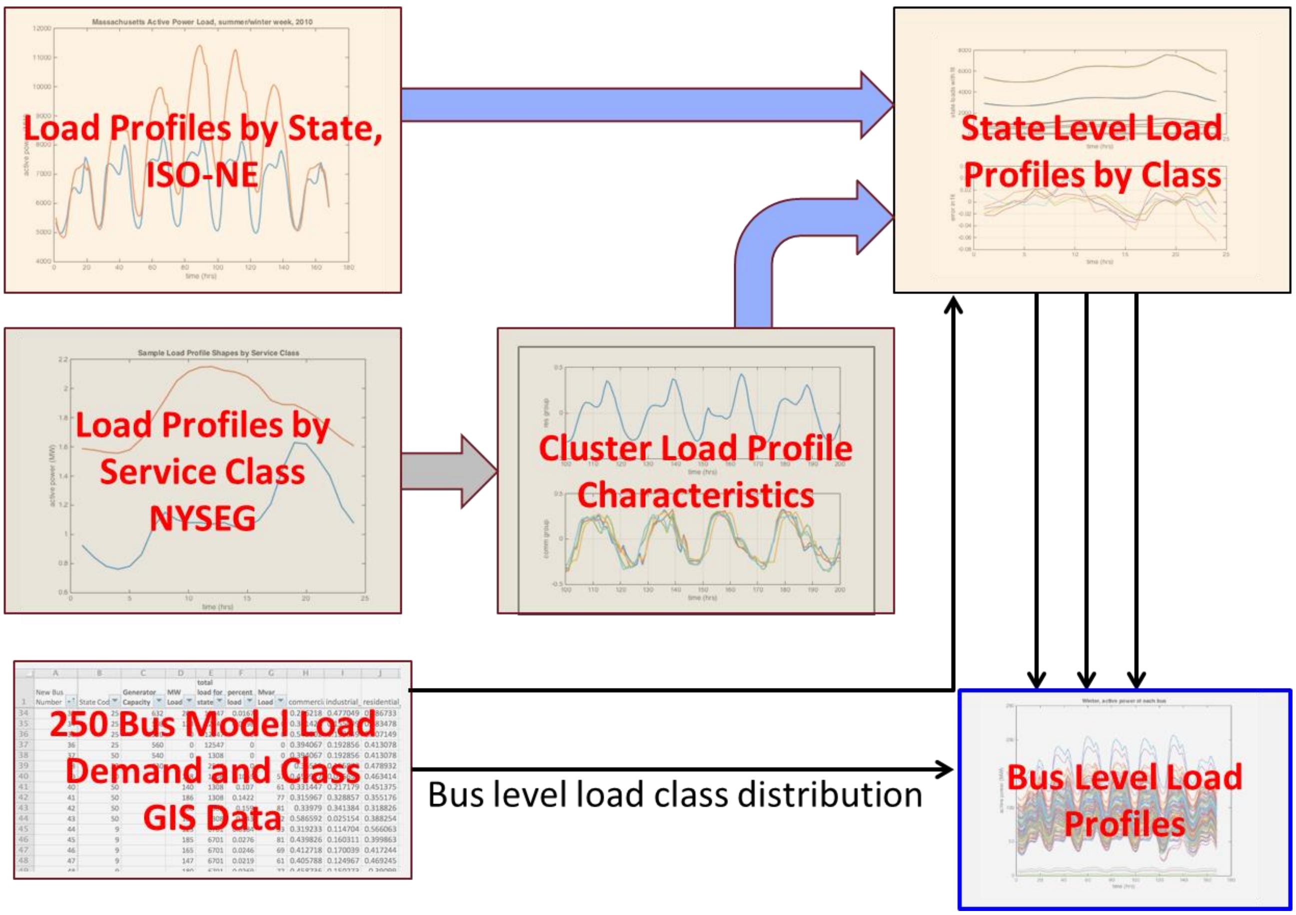


Figure Transmission Expansion Planning Data.4 Graphical flowchart for the creation of bus level load profiles

2.3 Generation Data

The next step was to create/collect generation data, and assign (for the purposes of studies here) plausible synthetic cost curves to each generator. Clearly, for real-world application of these algorithms, generator costs would be obtained from bid data.

For the testing of algorithms here, the generation data employed utilized the publicly available EIA 860 database [7] of all power generating stations in the United States, visualized below in 0. The database includes data on generator location, fuel type, capacity, minimum output levels, power factor, and other data. Since the 2010 census and land use data was utilized in creating the synthetic transmission systems, the 2015 EIA data was used, and “rolled back” to the 2010 operating point. Thus, the 2015 data was used, and generators built after 2010 were removed from the system, and generators retired between 2010 and 2015 were re-added to the system. The generators retired between 2010 and 2015, as well as some newer wind units, were missing minimum load and power factor information. The existing generation data was averaged as the basis for interpolation to create realistic data to fill in the missing quantities. Again, it should be stressed that tis collection of generation data was solely for the purpose of creating reasonably realistic test cases for the TEP algorithms being developed; perfect agreement to real-world data was not sought.

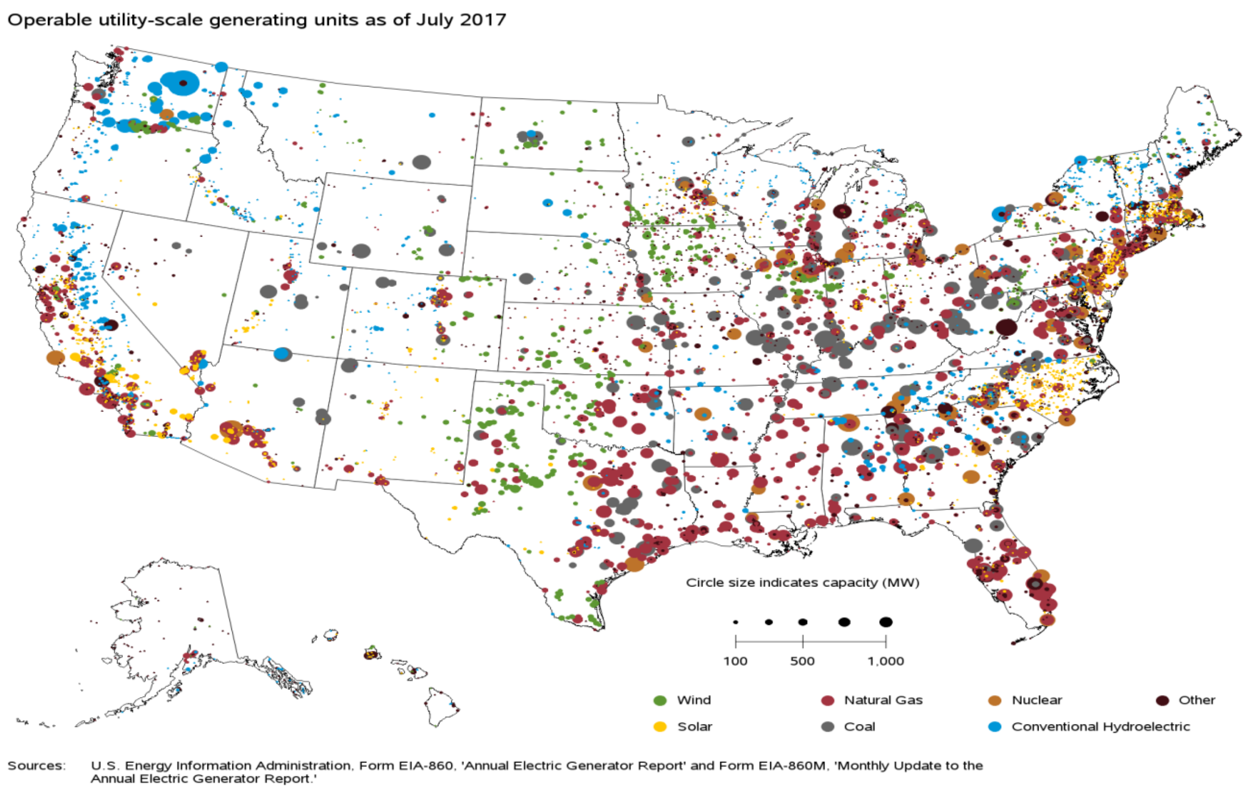


Figure Transmission Expansion Planning Data.5 Utility-scale generating units by location, size and fuel type as of July 2017

Finally, synthetic cost curves were created and assigned to each generating unit. For this step, the generators of the same technology and fuel type at the same geographic location were clustered so that the capacity of the resulting plants were within the range of the existing or synthetic heat rate curves, or publicly available generator offer curves. This data was then assigned to the generators, and linear regression was used to create a quadratic curve that best fit the data.

2.4 Existing Transmission Line Statistics

The next step in creating plausible scenarios of a TEP problem uses statistical data on existing networks, to create reasonable ranges to guide the assignment of line voltage level and electrical parameters. In the work here, these targets were created from publicly available databases of transmission system network information. First, the FERC Form 1 database [8] was used to gather or calculate statistics on the existing transmission system including conductor size (in kcmil), individual and aggregate transmission line lengths and capacities, and phase-to-phase spacing (geometric mean distance or GMD).

The Form 1 already had line lengths for each individual transmission line or line segment, but the data was not well organized; undergraduate research assistants in the project converted the data into a format that was easier to analyze. 0 shows the results of one of the analyses performed on the FERC Form 1 data, and 0 gives an example of the total transmission mileage in two states. Next, Tyler and Andrew calculated potential GMD values for each tower type at each voltage level. This was accomplished by matching the tower type and voltage level given in the Form 1 to typical conductor spacing values found in [9-12].

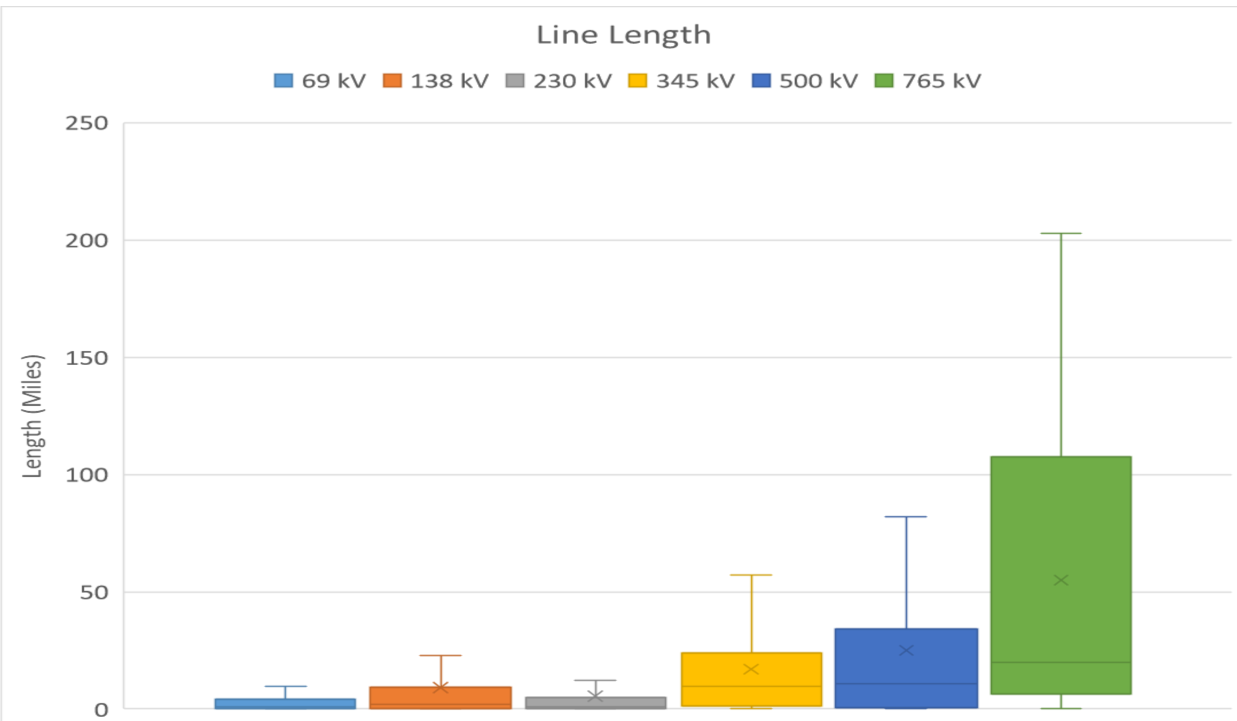


Figure Transmission Expansion Planning Data.6 Ranges of transmission line lengths by voltage level for a subset of the PJM network

Table Transmission Expansion Planning Data.1 Total transmission line miles by voltage class for Tennessee and Virginia

State	Voltage Class	Miles
TN	69	202
	138	3758
	230	3009
	345	8
	500	1464
	765	334
VA	69	202
	138	3758
	230	3009
	345	8
	500	1464
	765	334

Ampacity for each conductor type in the FERC Form 1 database was estimated. In particular, using manufacturer datasheets from Southwire and reference handbooks such as [13], the conductor size in kcmil and conductor type (ACSR, ACSS, etc) was used to identify the ampacity of the conductor and calculate the associated MVA rating at each voltage level.

Using all this information, a modest sized set of potential transmission line conductors was identified, with distinct sets assigned by each large state, or by regional area. 0 below illustrates an excerpt from the 138kV transmission sub-network lookup table.

Table Transmission Expansion Planning Data.2 Excerpt from the 138 kV transmission sub-network lookup table

Capacity (MVA)	**Voltage (kV)**	**conductor size (kcmil)**	**conductors per phase**	**Ra**	**Xa**	**Xa'**
205.0	138.0	715.0	1.0	0.128	0.399	0.092
219.0	138.0	795.0	1.0	0.119	0.403	0.0917
230.0	138.0	900.0	1.0	0.106	0.399	0.0907
238.0	138.0	954.0	1.0	0.099	0.395	0.0897

3. Network Topology Creation

After all of the data has been collected, the linear programming portion of the algorithm is run to determine the network topology.

3.1 Problem Formulation

As noted above, one of the computational challenges facing transmission expansion problems is the large search space of potential line additions; up to $\binom{n}{2}$ transmission lines that could be added to an existing transmission system, including adding lines to existing right of ways. For even modest numbers of buses n, the number of potential transmission paths becomes extremely large. If the full AC power flow model is used, the problem quickly becomes intractable. The work here uses a network flow model and begins from a case that includes all of the potential transmission paths in a complete graph, following the algorithm suggested by Garver [14]. we First note that the objective function of "MW-mile" flows descried earlier is a simply a weighted sum of absolute values of line flows; in optimization terminology, this constitutes a weighted L-1 norm, on the vector of MW line flows. The outer loop updates these weights each time at each iteration, before re-running the inner loop linear program with a new objective function. Viewing Garver's 1970's work from a modern perspective, this would be termed an "L-1 norm regularizer." In recent years, it has been widely recognized that such L-1 regularizers tend to enforce a sparse solution. Moreover, it is well recognized that with suitable "splitting" of each decision variable into the two variables representing its positive part an its negative part, the L-1 norm objective function becomes linear and (with linear constraints) is amenable to solution by linear program.

In the context of our inner loop linear program, the sparsity of the solution associated with a large percentage of the MW line flows being (near) zero at an LP solution point. Oversimplifying slightly, the outer loop deletes from consideration lines with near zero flow. As will be described below, viewed in greater detail the outer loop is more cautious, and at first only penalizes lines with low flow, in case they should become valuable and carry a higher flow in a later iteration. With the core computation of the inner loop being a linear program, the approach is scalable to very large problems. The authors have successfully implemented a modified version of Garver's algorithm to create, from ground up, a wholly new 30,000 bus synthetic network.

The linear programming problem proposed by Garver [14] and further described by Coffrin et. al [3] is summarized in the equation below.

$$\min_{p} c^T |p|$$

Subject to

$$|\mathrm{p}_{\mathrm{existing,k}}| \leq \text{line thermal limit } \forall\, k$$

$$Ap = \mathrm{P}^0$$

$$p = \begin{bmatrix} p_{existing} \\ p_{potential} \end{bmatrix}$$

Where $p_{existing}$and $p_{potential}$ are the flows on the existing transmission lines and the potential paths respectively. c is the relative weight applied to MW flow on each existing or potential transmission line. Transmission thermal line limits are enforced for existing transmission lines. For simplicity's sake, the authors followed Garver's recommendation and used the length of the line as an estimate of the relative cost of building new transmission lines, and weighted potential lines at 5 times the weight of existing lines. The MVA line flow limits are enforced for the existing transmission lines while the potential paths are given an arbitrarily high limit. $Ap = P^0$ represents the KCL equations, where A is the network incidence matrix and P^0 is the known, calculated vector of net power injections to each bus, with generation being positive injection and load negative injection.

When used for transmission expansion, the optimization formulation includes both existing transmission lines and all potential paths in a network. The potential paths parallel to existing lines are used to model increasing the capacity of an existing transmission line, i.e. upgrading its capacity.

In order to implement the optimization problem using the Matlab connector for CPLEXthe objective function is converted into a standard form problem. x is doubled a split into positive and negative flow, and A becomes a directed incidence matrix. i.e.

$$\tilde{p} = \begin{bmatrix} p^+ \\ P_- \end{bmatrix}$$
$$\tilde{A} = [A \quad -A]$$

3.2 Results of Initial Garver Algorithm

After each successful solution of the linear program, the potential paths with the highest flows are added to the system and become existing lines. The potential paths still remain in place, but a new line with an MVA line flow limit with corresponding voltage level, impedance and charging susceptance is added to the system. Modifying Garver's recommendation to only add one line of exactly 100 MVA corresponding to the potential path with the highest flow, the authors instead add lines to between 2 and 16 of the potential paths with the highest flows, using 125% of the flow on the potential path as the initial capacity of the new lines. The conductor lookup table X is utilized to determine the R, X and B parameters of the transmission line given the capacity specified from the linear programming result. This sequential linear programming process is repeated until only the existing lines have resulting flows and no power is flowing on the potential paths.

The result is shown below in 0. Notice that the results of algorithm at this stage do not yet identify voltage level for the candidate lines; this stage may be considered primarily focused on choice of topology. This is remedied in the next section: DC OPF feasibility

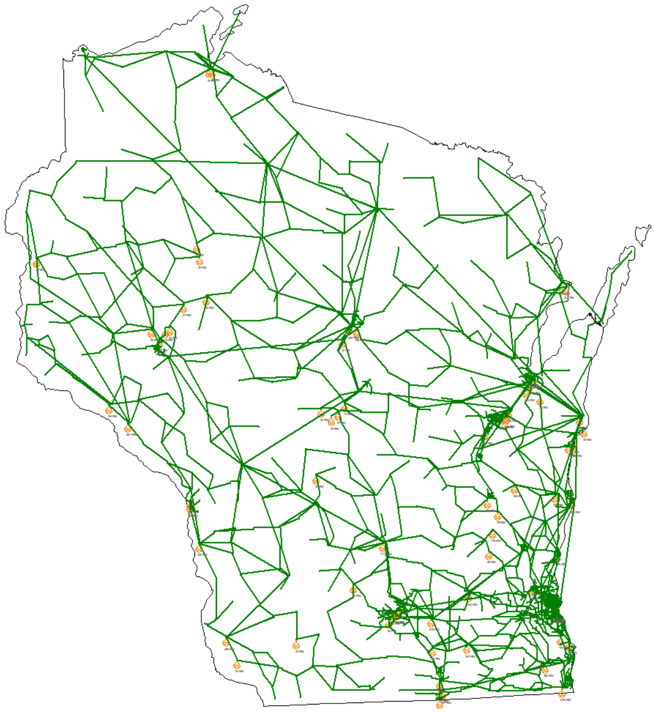


Figure Network Topology Creation.7 One line of initial transmission corridors of 1664 bus synthetic Wisconsin model

4. DC OPF Feasibility

Since the network flow model of the transmission system is a relaxed version of the DC power flow, a network that is feasible given a set of power injections and withdrawals with the network flow model will not necessarily be feasible when modeled with DC power flow. Thus, a set of sequential DC power flow calculations is performed, with a small subset of overloaded or substantially underutilized lines being incrementally upgraded or downgraded until the number of overloaded transmission lines is small. When used for transmission expansion, only one full sequential run of the DC power flow is needed, since transmission lines are only added to each voltage level separately.

4.1 Assigning Voltage Levels

After each successive run of the DC power flow, several transmission lines that are either overloaded or substantially underutilized are either upgraded (the conductor is increased to the next largest entry in the lookup table) or downsized (the conductor is decreased to the next smallest entry in the lookup table). The conductor lookup table that is utilized contains a finely discretized range of conductor MVA ratings that span all voltage levels that are present in the FERC Form 1 for the given region. Thus, sometimes when a transmission line's MVA rating is increased, the line moves to the next highest voltage level. This process of sequential DC power flow followed by incrementally changing conductors is repeated until the majority of transmission lines no longer have overloads. After this point, the voltage levels that are assigned to the transmission lines become the voltages of the buses that are present at the substations at the sending and receiving ends of the transmission lines. However, this results in network topologies that are fragmented and disconnected, since voltage levels are only assigned to each line individually and not to a collection of lines or buses. Thus, additional lines need to be added to the overall synthetic system to reconnect each voltage level.

4.2 Reconnecting each Voltage Level

To reconnect the voltage levels, each individual voltage is passed back to the Garver algorithm described in section 3 to add additional lines and create connected networks at each voltage level. To create larger networks, the algorithm mimics historical transmission network design by first creating multiple smaller regional networks, such as the ones that are RTOs operate. After assigning voltage levels to these networks the algorithm interconnects the different regional networks across at the same voltage levels. Since sometimes there is a large geographic distance between sub-graphs of each voltage level, the graph is partitioned until each sub-network can be connected using lines that are less than the maximum length specified in the FERC Form 1. To accomplish this, the Fiedler eigenvector of a weighted Laplacian matrix is used to partition the network into two sub-networks following the methodology described in [15], with further sub-partitioning performed as needed. The weighted Laplacian matrix is created using a combination of the line capacities and a bus-to-bus distance matrix to create a rank 1 matrix.

The results of the partition and reconnecting stage of the algorithm are show below in 0 and 0. 0 illustrates how the partitioning algorithm created 4 sub-networks within the 138 kV transmission system, that were then re-connected to form the one-line shown. 0 shows a single 345 kV network, which was not partitioned because the network is small enough as to make this step unnecessary.

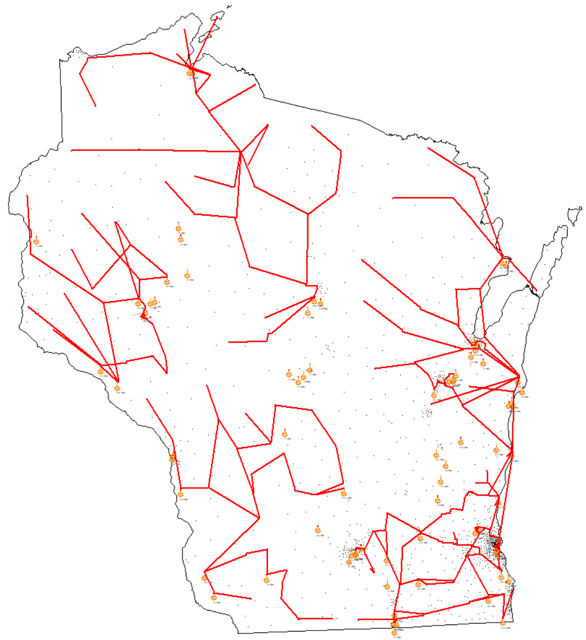


Figure DC OPF Feasibility.8 4, 138 kV sub-networks of the Wisconsin 1664 bus model

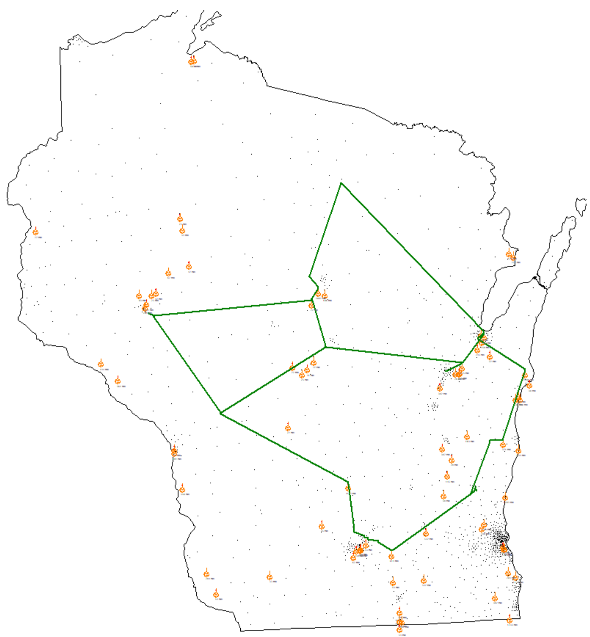


Figure DC OPF Feasibility.9 345 kV sub-network of the Wisconsin 1164 bus model

4.3 Reaching DC OPF Feasibility

After the sub-networks have been reconnected, the sequential DC power flow is re-run until no line overloads exist in the system under the DC power flow model. This is accomplished by iteratively run the DC power flow on the entire network and incrementally adjust line characteristics until there are only a few overloaded transmission lines in the network.

The authors are currently implementing scaling factors on the DC power flow to match the distribution of line flow ratings in the synthetic system with that of the distribution of line ratings in the FERC Form 715 [16]. In addition, the authors are working on matching the total capacity found in the FERC Form 1, as defined as the product of the capacity of each line, multiplied by the length of the line, summed over all lines. This quantity is measured in GVA-miles.

4.4 Conclusions

This portion of the project has developed an optimization approach to generating candidate scenarios for transmission expansion, based on an underlying linear program representation of the problem. It employs publicly available statistical data on existing transmission infrastructure in the United States. At relatively low computational cost it generates a wide variety of candidate line additions that use realistic selection of line parameters and voltage levels, while minimizing a weighted mileage-based cost function on the added lines, subject to constraints that require a feasible DC power flow solution for each resulting candidate network.

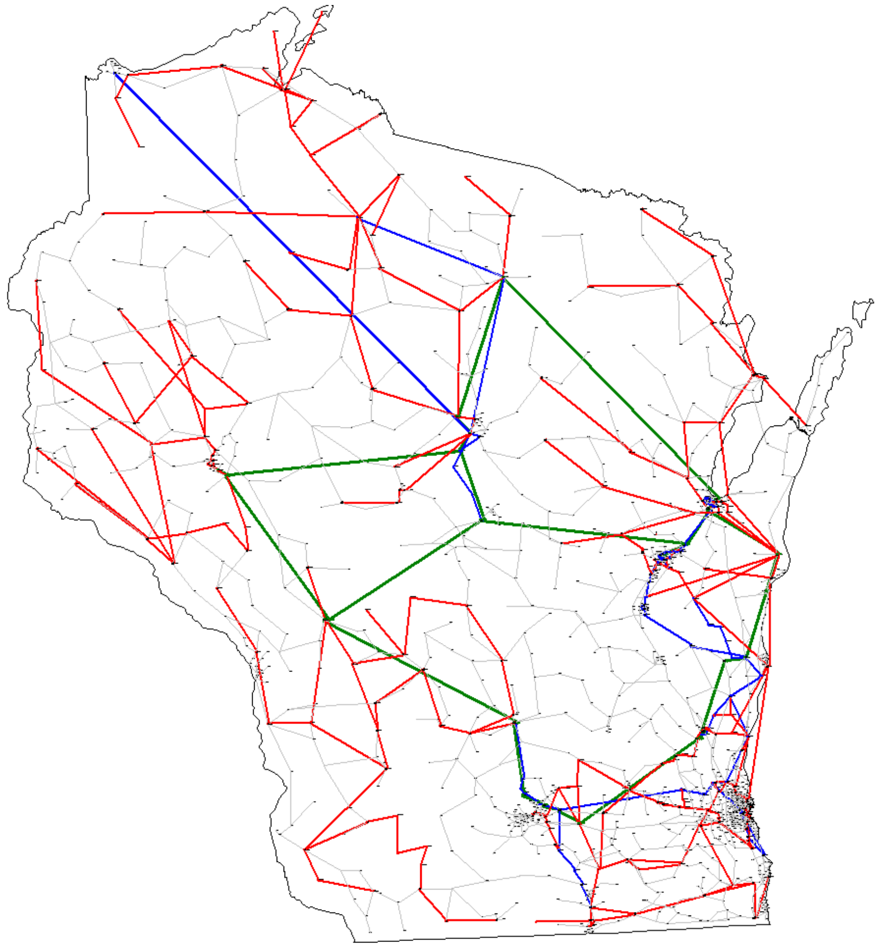


Figure DC OPF Feasibility.10 One-Line of the full Wisconsin 1664 bus network model

Finally, 0 shows the finished product of a test application of transmission expansion algorithm over the geographic footprint of the state of Wisconsin.

References

[1] D. K. Molzahn and I. A. Hiskens, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends in Electric Energy Systems,* vol. 4, pp. 1-221, 2019.

[2] R. Romero, A. Monticelli, A. Garcia, and S. Haffner, "Test systems and mathematical models for transmission network expansion planning," *IEE Proceedings - Generation, Transmission and Distribution,* vol. 149, pp. 27-36, 2002.

[3] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Network flow and copper plate relaxations for AC transmission systems," *Proceedings of the 2016 Power Systems Computation Conference (PSCC)*, 20-24 June 2016, Genoa, Italy.

[4] G. Latorre, R. D. Cruz, J. M. Areiza, and A. Villegas, "Classification of publications and models on transmission expansion planning," *IEEE Transactions on Power Systems,* vol. 18, pp. 938-946, 2003.

[5] C. W. Lee, S. Ng, J. Zhong, and F. Wu, "Transmission Expansion Planning From Past to Future," *Proceedings of Power Systems Conference and Exposition*, Oct. 2006, Atlanta, GA.

[6] University of Washington, Seattle, Power Systems Test Case Archive, R. Christie-moderator,, "118 bus power flow test case," 1993; available at https://labs.ece.uw.edu/pstca/

[7] E. US EIA, "Form EIA-860 detailed data, 2015 Form EIA-860 Data," ed: Schedule.

[8] F. FERC, "No. 1: Annual Report of Major Electric Utility," F. E. R. Commission, Ed., ed, 2010.

[9] L. O. Barthold, R. E. Clayton, I. S. Grant, V. J. Longo, J. R. Stewart, and D. D. Wilson, *Transmission line reference book: 115--138 kV compact line design*: Electric Power Research Institute, Inc.,Palo Alto, CA, 1978.

[10] General Electric Company. Large Transformer Division., General Electric Company. Energy Systems and Technology Division., and Electric Power Research Institute., *Transmission line reference book, 345 kv and above*, Second edition. ed. Palo Alto, Calif.: The Institute, 1982.

[11] USDA, *Electric transmission specifications and drawings, 115 kV through 230 kV*. Washington, D.C.: U.S. Dept. of Agriculture, Rural Utilities Service, 1998.

[12] C. Westinghouse Electric, *Electrical transmission and distribution reference book*, 4th Edition ed., 1950.

[13] Aluminum Association., *Aluminum electrical conductor handbook*, Second edition. ed. Washington, D.C. (818 Connecticut Ave., Washington 20006): Aluminum Association, 1982.

[14] L. Garver, "Transmission Network Estimation Using Linear Programming," *IEEE Transactions on Power Apparatus and Systems,* vol. PAS-89, pp. 1688-1697, 1970.

[15] C. L. DeMarco and J. J. Wasner , "A generalized eigenvalue perturbation approach to coherency," pp. 611-617, Proc. IEEE Conference on Control Applications, Albany, NY, Sept. 1995.

[16] F. FERC, "No. 715 - Annual Transmission Planning and Evaluation Report," Federal Energy Regulatory Commission, Ed.