Decision-making under uncertainty for electric power system operation and expansion planning

by

Marc Barbar

B.S., The Pennsylvania State University (2017)
S.M., Massachusetts Institute of Technology (2019)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
Department of Electrical Engineering and Computer Science
May 13, 2022

Certified by .....
James L. Kirtley Jr.
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by .....
Ignacio J. Pérez-Arriaga
Visiting Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Decision-making under uncertainty for electric power system operation and expansion planning

by
Marc Barbar

Submitted to the Department of Electrical Engineering and Computer Science
on May 13, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Decision-making under uncertainty is required in a multiplicity of situations in power system operation and capacity expansion planning. This thesis investigates the drivers and impact of uncertainty on power system infrastructure planning and proposes several methods to design and operate a power system at granular modeling level. The focus of the thesis is on Emerging Markets and Developing Economy countries, specifically India and Nigeria. However, the work presented in this document can be adapted to other situations, in the power sector or elsewhere, that share similar traits. Moreover, incorporating uncertainty in generalized optimization models often yields inaccurate results due to the lack of precision in representing the problem. This thesis carefully examines a set of situations and presents appropriate decision-making under uncertainty frameworks that yield meaningful results. The thesis is divided into three parts: drivers of uncertainty, the impact of uncertainty, and accounting for uncertainty in electricity resource design, with applications to Emerging Markets and Developing Economy countries.

Thesis Supervisor: James L. Kirtley Jr.
Title: Professor of Electrical Engineering and Computer Science

Thesis Supervisor: Ignacio J. Pérez-Arriaga
Title: Visiting Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to acknowledge the sponsors of this research, the MIT Universal Energy Access Laboratory, MIT Energy Initiative's Low Carbon Energy Centers, and the Future of Storage Study. This work would not have been possible without the incredible teams of Abuja Electric Distribution Company and the Tata Power Delhi Distribution Company. The data provided was indispensable for my work.

I have received an enormous amount of help and guidance from many people. My department supervisor, Prof. James L. Kirtley, supported my research interest from the start and ensured that I was well supported throughout the way. Co-supervisor Prof. Ignacio J. Pérez-Arriaga steered me from the early stages of my problem formulation to the final presentation of ideas in this thesis. He has encouraged me to question and discover the hard problems for myself while keeping me on track.

Thesis committee member, Prof. Robert J. Stoner, always guaranteed financial support, pointed out the weak areas in my work, and questioned the implication of any produced result. Committee member Prof. George C. Verghese provided a unique platform for the sound delivery of this thesis. Committee member Prof. Fernando de Cuadra (of Comillas University) assisted me in the applied research on optimization methods. Committee member Dr. Dharik S. Mallapragada spent countless hours supporting my research and publications and pushed me to improve my critical thinking, research, writing, and presentation skills.

In my long journey at MIT, I had the opportunity to work with brilliant people who were critical to the completion of this thesis. Dr. Pablo Dueñas-Martinez was always available as a exceptional resource in the field of optimization methods. I had the privilege of working with Meia Alsup; her help with the data collection and interpretation was the first stepping stone of this thesis. Prof. John Deutch and Prof. Michael Golay supported my doctoral plans and offered unique teaching opportunities in the field of sustainable energy and energy policy history, which expanded my knowledge in the field and impacted the work presented in this thesis.

My peers and lifetime friends from MIT made this experience beautiful. Omar Kahil and Lama Aoudi have been there for me since I started the program and have kept me sane along the way. I do not think I could have finished this thesis without them.

My father Fadi, mother Shamma, and sister Julia, there was never anything I needed that they did not provide.

Thank you all.

Contents

1 Introduction **18**
1.1 Decision-making under uncertainty 20
1.2 Related work 21
1.3 Thesis objectives and contributions 23

2 Long-term demand growth in the electricity sector — case study of India **25**
2.1 Introduction 26
2.2 Demand forecasting 29
2.2.1 Input data processing 30
2.2.2 Business-as-usual model 32
2.2.3 Technology model 36
2.3 Technical validation 39
2.3.1 Back-testing 40
2.3.2 Cross-comparison 40
2.3.3 COVID-19 pandemic impact on year 2020 41
2.4 Discussion 42

3 Impact of uncertainty at distribution level — case study of Delhi **43**
3.1 Introduction 44
3.2 Real options for non-wire alternatives 47
3.2.1 NWA valuation criteria 48
3.2.2 System design optimization 49
3.2.3 Simulation 50
3.3 Delhi case-study input data 52
3.3.1 Demand scenarios 52
3.3.2 Distribution network data 55

3.3.3 Cost assumptions . . . 55
3.4 Results . . . 57
3.4.1 Single-network case — Delhi . . . 57
3.4.2 City-wide network simulation . . . 60
3.5 Discussion . . . 64

4 Impact of uncertainty at distribution level — case study of Abuja 66
4.1 Introduction . . . 67
4.2 Wuse Market case-study . . . 69
4.3 Optimization framework . . . 72
4.4 Results . . . 73
4.4.1 Existing system design . . . 74
4.4.2 Alternative design . . . 75
4.5 Discussion . . . 80

5 Impact of uncertainty at generation and transmission levels – case study of India 84
5.1 Introduction . . . 85
5.2 Scenario-based transmission modeling . . . 87
5.2.1 Demand-side scenario model . . . 87
5.2.2 Supply-side optimization model . . . 88
5.2.3 Resource cost and performance assumptions . . . 90
5.2.4 Renewable-resource supply curves . . . 91
5.2.5 System cost of electricity generation expansion . . . 91
5.3 Results . . . 92
5.3.1 Reference case . . . 92
5.3.2 Scenario analysis spanning supply, demand drivers, and policy 94
5.3.3 Impact of key supply and demand-side drivers . . . 95
5.3.4 Impact of distribution-level storage . . . 100
5.3.5 Technological vs. policy drivers to reduce new coal investments 102
5.4 Discussion . . . 104

6 Representative period selection for power system planning using autoencoder-based dimensionality-reduction 108
6.1 Introduction . . . 109
6.2 Autoencoder-based representative period selection . . . 113
6.2.1 CEM model, input and output data . . . 113
6.2.2 Dimensionality reduction - autoencoders . . . 114
6.2.3 Clustering algorithm . . . 115

6.2.4 Overall loss-function definition . . . 116
6.2.5 Error metrics . . . 118
6.3 Case-study description . . . 120
6.4 Results . . . 122
6.4.1 CEM results comparison - single case . . . 122
6.4.2 CEM results comparison- multiple case . . . 122
6.4.3 Impact of input data on CEM results . . . 125
6.4.4 Solution time comparison . . . 126
6.5 Discussion . . . 126

7 Learning-based multi-stage electricity resource capacity expansion planning under exogenous and endogenous uncertainties 130
7.1 Introduction . . . 131
7.1.1 Odisha design problem . . . 133
7.1.2 Tree-based modeling . . . 134
7.2 Regret MCTS . . . 138
7.2.1 Tree reduction . . . 139
7.2.2 Leaf-node regret . . . 140
7.2.3 Confidence bound . . . 142
7.2.4 Tree search . . . 142
7.2.5 Design and dispatch optimization . . . 144
7.2.6 Transition-matrices formulation . . . 144
7.3 Results . . . 145
7.4 Discussion . . . 149

8 Conclusion 151
8.1 Contributions of the thesis . . . 151
8.2 Limitations and future work . . . 153
8.3 Discussion . . . 155

A Supplementary information — Chapter 2 156
A.1 Data Records . . . 156

B Optimization model 162
B.1 Objective function . . . 162
B.2 Demand balance constraint with non-served energy . . . 163
B.3 Variable renewable energy availability constraint . . . 163
B.4 Storage constraints . . . 163
B.5 Thermal generation constraints . . . 164

B.5.1 Generation unit commitment 164
B.5.2 Generation minimum and maximum power output 165
B.5.3 Generation ramping limits 165
B.6 Network expansion 165
B.6.1 Direct-current power-flow approximation through line susceptance and voltage deviation 165
B.7 Non-negativity constraints 166
B.8 Optimization model nomenclature 166

C Supplementary information — Chapter 4 **169**

D Supplementary information — Chapter 5 **171**

E Supplementary information — Chapter 6 **191**

F Supplementary information — Chapter 7 **194**

G Code availability **196**
G.1 Demand forecasting model 196
G.2 Real options model 196
G.3 Design and dispatch optimization model 196
G.4 Autoencoder model 196
G.5 Regret MCTS model 197

List of Figures

2-1 State and regional level distribution of annual electricity 2050 for stable GDP growth, baseline cooling, and home electric vehicle (EV) charging scenario. 27
2-2 Summary results of India's electricity demand forecasting at national level with stable GDP growth, baseline cooling, and home electric vehicle (EV) charging. 27
2-3 2030 Load profile for Southern Region across three days in summer. Scenario: stable GDP growth, reference cooling, home electric vehicle (EV) charging. Base: projected electricity demand of lighting, appliance, and industrial equipment. E2W: electric two-wheelers (e.g., bikes and scooters), E3W: electric three-wheelers (e.g., tricycles and auto rickshaws), E4W: electric four-wheelers (e.g., passenger vehicles). 28
2-4 Simplified schematic of demand forecasting method: steps in gray, input data in green, models in orange, and outputs in blue. 30
2-5 Southern region back test annual demand growth given GDP projection 33
2-6 Southern region back test seasonal demand variation given weather data 33
2-7 Back test result for Southern Region regression model with and without the additional noise variation. 40
2-8 2020 year-to-date demand comparison with projections 42

3-1 Flowchart showing steps in the flexible valuation framework used for modeling storage as non-wire alternative. 48
3-2 Flowchart of flexible valuation policy iteration. 52
3-3 Historical electricity consumption per capita for India and China [1]. Per Capita Consumption is defined as gross electricity generation by all sources plus net import divided by mid year population and reported in kWh. 54
3-4 Stationary distribution result of MCMC simulation 55

3-5 Sample distribution network diagram: three main feeders: commercial (top), industrial (middle), residential (bottom). Bus number in **bold** and line number in *italic*. Size, length, parameters and topologies vary across the library of feeders analyzed [2]. 56
3-6 Hourly dispatch of NWA battery storage for one summer week load profile in 2030 for a distribution network from the city of Delhi. The demand scenario is based on the mid demand scenario. 58
3-7 Simulation of investment outcomes for single network case in Delhi, shown in Fig. 3-2, with simulated demand trajectory from the transition matrix. $y = 2020$, $Y = 20$, four timesteps p of 5 years intervals as per the input projected data [3]. 60
3-8 Approach for computing megacity-level NWA battery storage potential in Indian megacities. 61

4-1 GIS and low flow result of distribution network serving the Wuse Market (highlighted in red). 69
4-2 Existing system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages. 74
4-3 Existing system minigrid dispatch for year 2020 demand without simulated grid outage. 75
4-4 Greenfield system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages. 76
4-5 Greenfield system minigrid dispatch for year 2020 demand without simulated grid outage. Time starts at midnight on the first day. . . . 77
4-6 Greenfield system minigrid dispatch for year 2030 demand without simulated grid outage. 78
4-7 Greenfield system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages under low-cost storage assumptions. 78
4-8 Difference plot of system design given unscheduled long duration infrequent grid outages. 79
4-9 Difference plot of system design given unscheduled long duration infrequent grid outages under low-cost storage assumptions. 80

5-1 Regional distribution in generation and storage power capacity as well utilization trends for 2050. Regional transmission transfer capacity and its average utilization is shown. 95

5-2 Hourly generation dispatch and load profile for three days during summer (left) and winter (right) periods for 2050. Model outcomes based on reference case as defined in Table 5.3. Storage charging is shown in the "Load + charge" curve as well as by the negative generation for storage. Technology names and their respective abbreviations in Table D.4 96
5-3 Installed capacity (1^{st} row), annual energy generation (2^{nd} row), storage energy capacity (3^{rd} row) and annual CO_2 emissions (4^th row) for reference case (1^{st} column), as well as cases with alternative assumptions for battery storage capital cost (2^{nd} column), high AC efficiency (3^{rd} column) and gas prices (4^{th} column). Detailed assumptions for each case are provided in Table 5.1 and Table 5.2 97
5-4 Impact of distribution-level storage deployment on dispatched generation (1^{st} row), installed storage energy capacity (2^{nd} row) and annual CO_2 emissions (3^{rd} row) under the reference case (1^{st} column) and the low-cost storage (2^{nd} column) case. 101
5-5 Model outcomes for high AC efficiency-low cost case, defined by low battery storage capital cost, high AC efficiency and low NG price (1^{st} column) as well as impact of carbon price with and without scenario assumptions. Model outcomes include installed capacity (1^{st} row), annual energy generation (2^{nd} row), storage energy capacity (3^{rd} row) and annual CO_2 emissions (4^th row). Columns 2-4 highlight outcomes compared to the reference case for the following cases: a) high AC efficiency-low cost case (2^{nd} column), b) low carbon price case, where CO_2 price starts at 20$/tonne in 2030 and grows by 5% each year (3^{nd} column) and high AC efficiency-low cost + low carbon price scenario case (4^{rd} column). 106
5-6 System average cost of electricity generation expansion (SCOE) per modeling period for a range of cases evaluated in the study. Note that SCOE does not include fixed costs associated with existing generation and transmission assets in 2020. 107

6-1 Autoencoder architecture used in the RPS method. Encoder is constructed using three main layers: Convolution, Maximum Pooling and, Long short term memory (LSTM) [4]. Decoder is constructed using two main layers: Up sampling and deconvolution [4]. Clustering is performed in the latent with the representative periods identified from clustering subsequently decoded to produce input time series for the RCEM. 116
6-2 Three types autoencoder structures proposed. Type 1: input autoencoding only, Type 2: joint input and output autoencoding, Type 3: separate input and output autoencoding. 117
6-3 ERCOT electricity system. Single bus system (square green node); three bus network (square green node and blue triangle nodes); eight bus network is all inclusive. Line thickness highlights the average power transfer in 2018 between zones on a relative basis. 121
6-4 Difference in installed capacity and annual generation between reduced-space CEM (RCEM) results and full-space CEM results for three bus network. Results based on RCEM using $k = 8$ for ERCOT load year 2020, mid-range VRE technology cost, 1,000 \$/tonne CO_2 price. . . . 123
6-5 Violin plot featuring the kernel density estimator of the distribution of data points of the four weighted absolute error metrics (SCOE, NSE, generation, and capacity) grouped by RPS method for RCEM outcomes with different number of representative periods 4, 8, and 20. The data presented is based on the eight bus system discussed in section 6.3 and is consistent with the mean values presented in Table 6.2. 128
6-6 Box and swarm plot of optimized values of α and β parameters for Type 3 autoencoder across six simplified case studies described in section 6.2.3. Results based on RCEM outcomes for single bus system with 4, 8, and 20 representative periods for all 90 distinct cases as defined by parameter values in Appendix E Table E.1. 129
6-7 Runtime confidence band of reduced-space optimization and all autoencoder-based RPS for all sizes and scenarios across the single, three and eight bus systems. 129

7-1 Proposed Regret MCTS methodology flowchart. 138
7-2 Example of single-stage decision tree with one parent node and nine state nodes that are grouped together into three leaf-nodes. Color map refer to the same decision but different external event realizations. 139

7-3 Decision transition graph; D1: business-as-usual, D2: grid upgrades, D3: minigrid. . . . 145
7-4 Events transition graph; E1: nothing, E2: grid reinforcement, E3: independent minigrid. . . . 145
7-5 Three sequences example in decision tree. . . . 146
7-6 Three days dispatch result of minigrid design solution of optimization. 147
7-7 Resulting tree from selection-expansion-simulation only. . . . 148
7-8 Average Decision error relative to ground truth results of standard UCB, Bayesian UCB and Regret MCTS. . . . 149
7-9 Box plots of sampled results from interrupted tree search at 400, 600, 800 and 1,000 iterations of four decision-making methods: Deterministic UCB, Bayes and Regret for the four nodes (9x9x9x9) case. . . . 150

A-1 India's GDP curve-fit and forecasting to 2050 . . . 157
A-2 Residential survey categorized hourly demand profile . . . 158
A-3 Commercial survey categorized hourly demand profile . . . 158
A-4 Normalized sample charging profile schemes . . . 158
A-5 2050 cooling demand contribution to peak results comparison with IEA's Future of Cooling . . . 159
A-6 Cooling demand contribution to peak demand . . . 159
A-7 Results comparison with stated policy World Energy Outlook projections . . . 160
A-8 Results comparison with sustainable policy World Energy Outlook projections . . . 160
A-9 Results comparison with Brookings India 2030 projections . . . 161
A-10 Electric Vehicle demand results comparison with IEA's Global EV Outloook . . . 161

D-1 Reference case model outcomes with alternative assumptions about decadal renewables installation limits. Reference = decadal installation limits as shown in Table D.13. Half cap = decadal installation limits are 0.5 of the reference values. No cap = no decadal installation limits. . . . 172
D-2 Regional demand growth projections for India as per demand forecasting model results [3] . . . 173
D-3 Hourly generation dispatch for reference winter and summer load profiles for 2030. Technology names and their respective abbreviations in Supplementary Table D.4 . . . 173

D-4 Hourly generation dispatch for summer reference and high AC efficiency load profiles for 2050. Technology names and their respective abbreviations in Supplementary Table D.4 174
D-5 Hourly generation dispatch for summer reference and DLS load profiles for 2040. Technology names and their respective abbreviations in Supplementary Table D.4 174
D-6 Emissions intensity and Transmission expansion outcomes for modeled cases considered in the main text 175
D-7 Technology capacity factors across the scenarios over different modeling periods 176
D-8 Impact of Morning (top) and day (bottom) electric vehicle (EV) charging schemes relative to evening EV charging scheme under the reference case 177
D-9 Storage power and energy capacity deployment trends in the reference and sensitivity cases 178
D-10 Yearly average DLS impact on supply and demand per hour for 2030 179
D-11 Hourly load profile by month in 2040 across various demand scenarios considered here 180
D-12 Indicative GenX capacity expansion optimization model run time with respect to number of clustered weeks. Outputs based on reference case. 181
D-13 Generation capacity difference relative to 20 representative weeks for the reference case. 182
D-14 Renewable resources supply curve calculation flowchart 182
D-15 Deployable solar and wind resources potential maps 183
D-16 Reference case frequency of hourly load variation 184

E-1 Installed capacity (top) and annual generation (bottom) for reduced-space CEM (RCEM) results and full-space CEM results for three bus network. Results based on RCEM using $k = 8$ for ERCOT load year 2020, mid-range VRE technology cost, 1,000 \$/tonne CO_2 price. . . . 192

List of Tables

2.1 Input Data Sources 30
2.2 GDP projections curve fit results. x is the year, y in the GDP growth rate. A, B, and C are parameters resulting from fitting the curves to the historical state-level GDP data. R-square is the average goodness of fit of the curve under parameters A, B, and C to the historical GDP data of all considered states in India. 31
2.3 Business-as-usual Regression R-squared consumption results 41
2.4 Business-as-usual Southern Region consumption Regression performance of select parameters 41

3.1 Projected peak demand (GW) under the baseline and high-AC-efficiency scenarios as per Chapter 2 assuming stable GDP growth for the city state of Delhi. Further details in [3] 45
3.2 Proposed transition Matrix 54
3.3 Input cost assumptions for the model. Sources [5, 6]. Peaks hours are defined to be between 8 PM and 12 AM. 56
3.4 Optimized storage sizing in the single network case for Delhi and cost outcomes. Results correspond to three demand growth scenarios (low, mid, high) in 2030 for Delhi using sample network of Fig. 3-5 and cost assumptions of Table 3.3. CAPEX is the annualized capital cost of energy and charge, OPEX is the fixed and variable operation and maintenance costs for one year. VAR is charging cost of the battery storage system given the tariff schedule. 58
3.5 Annualized investment cost (AIC) results for traditional network upgrades and NWA battery storage with deferred network upgrades for a deferral period $p = 5$. Results for case of single network in Delhi and three demand growth scenarios (low, mid, high) for 2030. 59

3.6 MDP simulation results for the case of single network in Delhi across all transitions in Table 3.2 for three demand growth scenarios (low, mid, high) for 2030. Results are annualized USD values. Negative values indicate that NWA battery storage is more expensive than traditional upgrades and vice versa. 59
3.7 Flexible valuation framework results for megacity-level NWA battery storage analysis under mid-range storage capital cost projections . . . 62
3.8 Flexible valuation framework total results of megacity-level NWA battery storage analysis. Initial investment in 2020 with 10 year deferral period for periods 2030 and 2040. Distribution network line useful life is set to 30 years and battery storage useful life is set to 15 years. . . 63
3.9 Storage cost impact on outputs of the flexible valuation framework applied to the four Indian megacities, for year 2030. Low, mid and high storage capital cost assumptions are sourced from [5]. Breakeven costs are 261 USD/kWh and 227 USD/kW for energy and power respectively. 64

4.1 Definition of frequent, infrequent, long, and short outages through number of yearly occurrences and duration of outages for simulated cases based on discussion with Abuja Electric Distribution Company 71
4.2 Capital cost assumptions for various resources. All costs in 2018 dollars, unless otherwise noted. Solar costs assume DC to AC ratio of 1.34 [5, 7]. 71
4.3 Operation and maintenance cost assumptions for various resources. All costs in 2018 dollars, unless otherwise noted. Solar costs assume DC to AC ratio of 1.34 and diesel generator heat rate is fixed to 0.85 gallons per hours per kW [8, 7, 5]. 72
4.4 Low cost storage scenario CAPEX and OPEX costs projections [5]. . 72
4.5 Total discounted cost, system cost of electricity per period and discounted results from optimization of the seven scenarios given a discount rate of 9% and total demand for the Wuse Market of 2,950 MWh in 2020 with 5% annual demand growth for periods 2025 and 2030. . 81

5.1 Demand estimates from bottom-up forecasting model [3] that is used as inputs to the supply-side modeling. 89

5.2 Capital cost assumptions for various resources. All costs in 2018 dollars and sourced from NREL annual technology baseline 2020 [9], unless otherwise noted. Wind, solar PV and gas generation capital costs have been de-rated by 72%, 51%, and 70% respectively to account for estimated capital cost differences for these resources between U.S. and India, as per central technology cost for 2019 from [10]. Solar costs assume DC to AC ratio of 1.34[9] 93
5.3 Scenario definition (see Table 5.2 for detailed costs) 94

6.1 Mean values of error metric defined in Eq. 6.9 - 6.13 to quantify performance of different representative period selection (RPS) methods for the 1-bus, 3-bus and 8-bus networks. Row-wise color mapping: red is highest and green is lowest value across the row. 124
6.2 Mean error metric results of different representative period selection (RPS) methods grouped by cluster number k for the 8-bus system. Row-wise color mapping: red is highest and green is lowest value across the row. 125

7.1 Design options 141
7.2 Decisions and events sequences. 146
7.3 Margin of regret (ρ_g) and asset regret (Υ_g^S) values for leaf-nodes $S1$ and $S2$ for the three defined sequences. 147

A.1 Output data headers descriptor 157

B.1 Set nomenclature of the electricity resource capacity expansion model. 166
B.2 Index nomenclature of the electricity resource capacity expansion model. 167
B.3 Parameter nomenclature of the electricity resource capacity expansion model. 167
B.4 Variable nomenclature of the electricity resource capacity expansion model. 168

C.1 Capital expenditure results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030. 169
C.2 Annual fuel cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030. 170
C.3 Annual fixed operation and maintenance cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030. . . . 170
C.4 Annual variable operation and maintenance cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030. 170

D.1 DLS system cost of electricity comparison . . . 171
D.2 Carbon price scenarios . . . 181
D.3 Abbreviations for Eqn. D.1 . . . 183
D.4 Modeled technology abbreviations . . . 185
D.5 System assumptions . . . 185
D.6 Storage duration comparison across scenarios and modeled periods . . 185
D.7 Zonal power transfer limits (Annex 7.1 of [11]). Zonal definitions as shown in Chapter 2 Fig. 2-1. . . . 186
D.8 Regional thermal power existing capacity and parameters [9, 5, 12, 13, 14, 15, 16] . . . 187
D.9 National thermal power parameters [9, 17, 5, 10, 12, 18, 19, 20] 187
D.10 Hydro power existing capacity and parameters [9, 17] . . . 188
D.11 Existing Variable Renewable Energy [9, 5, 12, 13, 14, 15, 16] 188
D.12 Gompertz curve fitting results . . . 190
D.13 Decadal VRE installation limit (MW) for the reference case 190

E.1 Input data considered in statistical analysis. Complete description of each parametric scenario, including description of low, medium and high technology cost assumptions. . . . 191
E.2 Autoencoder parameters. Encoder maximum pooling layer size is divisible by flattened time series of one period to satisfy dimensionality reduction to cluster latent representation by k. . . . 193
E.3 Standard deviation of capacity error metric grouped by cluster number k for the eight bus system. Row-wise color mapping: red is highest and green is lowest. . . . 193

F.1 Capital expenditure costs for each node for the three node case study. 194
F.2 Operational expenditure costs for each node for the three node case study . . . 194
F.3 Decision-dependent external event realization probabilities for leaf nodes 1 and 2. . . . 195
F.4 Decision and external event dependent load projections for leaf nodes 1 and 2. . . . 195

Chapter 1

Introduction

This thesis addresses the problem of dealing with uncertainty factors that complicate infrastructure planning in power systems, and the design of optimization models that handle those uncertainty factors. In particular, this thesis focuses on uncertainty aspects that are more prevalent in Emerging Market and Developing Economy (EMDE) countries.

The power sectors of various EMDE countries are very different, but some common traits define the boundaries of the problems addressed in this thesis. The power sector in EMDE countries requires particular attention to power delivery reliability, cost recovery mechanisms, and regulatory frameworks, on top of legal and monetary risks [21]. As a result, specific EMDE-country features introduce substantial uncertainty in electricity demand. Furthermore, due to challenges in implementing effective cost recovery mechanisms, investors must also face significant uncertainty that impacts generation and network designs.

All these factors render power system design a significant challenge in EMDE countries. Due to long capital recovery periods of large, long-lived assets and the substantial uncertainties, the design of adequate electric grid planning methods has not made as much progress as in industrialized countries. In recent history, governmental bodies and electric utilities in EMDE countries have primarily been concerned with deploying minimum-cost resources to maintain acceptable system reliability with a short-term perspective. As a result, they have frequently avoided appropriately addressing the longer-term design problem.

Bringing reform to the power sector in EMDE countries requires attention to all three

components: power delivery, cost recovery, and regulatory framework. Ensuring reliable delivery of power is a major challenge in EMDE countries. Upstream failures in the bulk power system — large generation and transmission infrastructure — and in the medium- and high-voltage distribution network leave large and medium-size commercial and industrial customers stranded without power. For instance, the Wuse Market customers of Abuja Electric in Nigeria understand this situation very well, and have sought alternative solutions from a minigrid developer to maintain electric power for their business activities. In large cities such as Delhi, highly congested streets constrain the local utility from improving network quality; as a result, many households in Delhi find themselves without access to electricity.

Rural communities in Indian states such as Odisha require little power presently but necessitate long distribution lines. Voltage quality at the end of the feeder is often poor, to the extent that some isolated rural communities choose not to use it for household activities, and commercial and industrial loads resort to diesel generators. Access to capital is a primary concern for government bodies and the local utilities. These entities hesitate to commit significant capital for long-term planning of the electric grid, a multi-decadal problem, due to the immediate need to fix current technical issues on the grid, unfavorable policy, and uncertainty in demand realization. Finally, current policies can prevent innovative solutions from materializing and achieving sustainable and quality electricity access. For example, Distributed Energy Resources (DER) may require policy updates to expand their deployment further and drive out the use of pollutant backup diesel generators.

The situations described above present a class of decision-making problems related to electricity resource design. The uncertainty in demand growth, technology cost decline, or other major external events complicate the development of efficient mathematical and numerical programs that solve the design-under-uncertainty problems for a particular system and multi-stage least-cost system planning. Solving such a class of problems enables insight into investment strategies that ensure the least-cost, reliable, clean electricity resource design, particularly in EMDE countries.

This thesis focuses on approaches to address some of these problems in two EMDE countries: India and Nigeria which present two major EMDE-related power system problems: network-constrained and generation-constrained systems, respectively. While India and Nigeria receive much attention, the author recognizes that they are not representative of all EMDE countries and their particular challenges in power system design and their specific uncertainties.

1.1 Decision-making under uncertainty

Decision-making in electricity systems must consider the characteristics of EMDE countries from an engineering (power delivery), policy (regulation), and financial (cost recovery) perspective. Decisions such as the adoption of on- versus off-grid generation supply, grid expansion, maintenance, or investment in renewables supported by battery storage, cannot be made without considering the relevant uncertainties.

Uncertainty causes the inability to foretell consequences precisely. Two types of uncertainties are relevant to decision-making: exogenous and endogenous. Exogenous uncertainty is independent of the decision process (e.g., weather forecast, technology cost), while endogenous uncertainty is decision-dependent (e.g., demand growth that depends on supply reliability, generation technology cost, and tariff structures). In both types of uncertainties, decision-making is modeled as an agent that acts based on observations from the environment. The agent then chooses an action based on expected return minimization/maximization criteria. There are three classes of methods for decision-making processes of multi-stage problems [22]:

Exhaustive methods

Exhaustive or explicit programming is the simplest and most direct way to design a decision-making process. The agent is conceived to anticipate all different scenarios and to take the most favorable action. Explicit programming such as brute-force and computationally tractable decision trees may work well for simple problems but quickly place a significant burden on providing a complete strategy considering all possible scenarios.

Optimization methods

In the search for an optimal decision, uncertainty can be reduced to risk evaluation by assigning future scenarios a probability of realization. Probability-based deterministic optimization of different scenarios remains a prime method of choice in designing multi-stage decision-making processes. Stochastic optimizations further expand deterministic ones by introducing risk in either a mathematical program's objective function or constraints. Unfortunately, there are no multi-stage optimization methods that work well for all problems. The decision-sequence space is affected by the curse of dimensionality: the size of the space grows exponentially with the number of stages, the number of possible outcomes, and the size of the decision space at each stage.

Learning-based methods

Supervised inference models based on collected data are used to approximate a value function for improving a decision-making process. Various challenges in data collection and value function approximation reduce the practical usage of such methods. Multi-stage models aim to find a sequence of decisions that minimizes an expected cost function. Decisions and random events (discrete or continuous) affect the value of future outcomes. Reinforcement Learning methods aim to understand the impact of a decision on the sequence of decisions and the policy objective. Often, approximations using domain-knowledge of the problem effectively address the challenge of exploration versus exploitation in sequenced-based learning.

This thesis is a collection of work on modeling, optimization, and heuristic planning of electricity resources, with applications in EMDE countries. First, demand growth is explored in detail, as it is one of the major uncertainties of electricity resource planning in EMDE countries. Second, the demand growth uncertainty and technology cost projections are modeled as exogenous uncertainties in optimization models to solve the electricity resource design problem. Finally, after identifying the dimensionality problem that uncertainty creates, a time-series clustering method is used in a domain-knowledge empowered learning-based method to solve exogenous and endogenous uncertainties that govern the same electricity resource design problem.

1.2 Related work

Proper usage of decision-making-under-uncertainty methods in the design of electricity supply infrastructure requires careful attention to the drivers of the uncertainty being considered. From a techno-economic perspective, demand growth and technology cost are substantial uncertainties when planning reliable power systems in EMDE countries. Technology cost uncertainty includes various components ranging from raw material cost to supply chain economics. While regulatory uncertainty in EMDE countries is an essential aspect of decision-making, this thesis limits its focus to the techno-economic uncertainties of demand and external events that impact the planning, such as technology cost.

Several studies have been carried out to provide insight on electricity demand as a single value projection [23, 24, 25, 26, 27, 28, 29]. Data-driven forecasting models alone are insufficient because new load patterns cannot be inferred from historical data. This necessitates bottom-up electricity demand forecasting. Several bottom-up methods project appliance-level consumption [30, 31, 32] and urban demand [33, 34,

35]. However, such methods are either generic to energy systems and not focused on electricity, or are too data-intensive since they rely on bottom-up data aggregation. In the context of EMDE countries, a blend between forecasting models and bottom-up projections is appropriate for projecting demand growth scenarios.

There is significant literature on introducing uncertainty to deterministic decision-making models for electricity system design in the form of options analysis [36, 37, 38]. However, these studies are either concerned with bulk power system design and ignore distribution-level planning or are not concerned with the temporal nature of local DER operations. Moreover, these studies focus on DER that are a flexible and dispatchable resource (diesel generators) or energy-producing (rooftop solar) as opposed to energy-limited resources, such as battery storage. Generation design and dispatch models that account for battery storage are generally constrained to deterministic formulations [39, 40, 41, 42, 43, 44].

In the context of EMDE countries, innovative business, regulatory and legal frameworks such as the Integrated Distribution Framework [45, 46] are necessary to improve and reform the power sector [21]. These frameworks bring forward flexible strategies such as under-the-grid minigrids [47, 48, 49, 50], and deployment of battery storage as a non-wire alternative [51]. This necessitates decision-making tools that consider the decision-dependent uncertainties in planning flexible strategies.

To solve the design and dispatch problem of electricity resource planning, the temporal vectors (e.g., demand profile and variable renewable energy availability profile) must be effectively represented by reduced-order vectors to speed up the optimization run time to solve the multi-stage problem in a tractable way. Time-series clustering has been heavily discussed in the literature in the context of electricity resource planning [52, 53, 54]. Several studies have explored the problem of time-series clustering for the design of energy systems [55]. [56, 57] focus on operational pattern similarity in temporal data to aggregate generation unit dispatch models without integration in the overall electricity system. [58] presents a clustering method for energy system design but at a high computational burden, without addressing the scalability of the method. On the other hand, data-driven approaches [59, 60] have been proven to be effective in time-series aggregation of only prelabeled data.

This thesis considers power system decision-making under uncertainty as a multi-stage stochastic problem (MSSP). In MSSP, uncertainty is characterized as either exogenous (decision-independent) or endogenous (decision-dependent) [61]. Several MSSP models are developed with exogenous uncertainties [62, 63, 64, 65, 66, 67]. This thesis addresses both exogenous and endogenous uncertainties where the deci-

sions influence the realizations of uncertain events or parameters by altering their underlying probability [61]. Little work is done on this particular type of uncertainty in multi-stage stochastic resource design and dispatch problems [68, 69, 70]. To the author's knowledge, there is no prior work on multi-stage stochastic problems with endogenous uncertainties applied to the problem of power system design and dispatch.

1.3 Thesis objectives and contributions

This thesis addresses three main questions related to decision-making under uncertainty in the context of electricity resource design in EMDE countries.

1. *What are the uncertain factors that can impact the design of the different segments of the power supply?* Among the different factors identified, this thesis analyses in detail demand growth (in the context of India) and explores how some uncertain parameters characterize its future evolution, with important implications on the potential design pathways.
2. *How do these uncertain factors impact electric grid planning at the various stages — generation, transmission, and distribution?* This thesis presents a comparative analysis of various transmission- and distribution-level electricity generation and network designs, to examine the impact of the identified uncertain factors in situations relevant to India and Nigeria.
3. *How to handle uncertain factors in particular investment planning of electric grid infrastructure?* This thesis addresses the multi-stage design problem of on-grid versus off-grid electricity connection in the context of rural communities in India, subject to significant uncertainty.

The body of this thesis is organized according to the above-mentioned questions:

Question 1

The first question is addressed in Chapter 2, where an electricity demand projection methodology for India is developed, and the data is used throughout this thesis. Potential demand trajectories are quantified, and subsequently, the uncertainty in demand growth at India's state and regional levels, which will be used in Chapters 3 and 5. This thesis shows how hourly consumption patterns of electricity demand are subject to change due to the introduction of new loads. In the context of EMDE

countries, incorporating demand growth uncertainty in generation and network designs (addressed in the subsequent questions) must first be supported by appropriate projections.

Question 2

The second question is addressed in Chapters 3, 4, and 5, where optimization-based models are developed to deal with exogenous uncertainties. Chapter 3 presents a framework based on real options to optimize distribution network planning using battery storage in the case of the city-state of Delhi while considering the uncertainty in demand using data from Chapter 2. Chapter 4 presents the minigrid of the Wuse Market customers in Nigeria using the real options optimization framework of Chapter 3. Chapter 5 expands the analysis of demand uncertainty to the transmission level of the electric grid design and dispatch problem. Various designs are presented based on assumptions regarding demand growth, cost, and policy trajectories.

Question 3

The final question of this thesis is addressed in Chapters 6 and 7, where learning-based methods are developed to improve the computational complexity required to solve decision-making under exogenous and endogenous uncertainties. Efficient time-domain reduction techniques are presented in Chapter 6, allowing faster exploration of power system design solutions without deteriorating the result of the reduced-order optimization. Chapter 7 presents a learning-based decision-making methodology for solving multi-stage decision-dependent stochastic problems using the time-domain reduction method of Chapter 6. We apply the method of Chapter 7 to a simplified problem of under-the-grid minigrid design considering grid supply maintenance and uncertainty in rural communities in Odisha, India.

Finally, Chapter 8 gathers the conclusions of the thesis. Taken as a whole, the work presented in this thesis demonstrates novel frameworks for addressing uncertainty in electricity system design using DER local generation, under-the-grid-minigrids, and grid supply. This thesis addresses the deployment of clean energy resources, emphasizing the role of battery storage and pathways towards decarbonization without compromising the electricity system reliability. While this thesis deals with applications of electricity resource planning in EMDE countries, principally, the models and methods developed in this thesis may also be applied to other contexts of planning under uncertainty.

Chapter 2

Long-term demand growth in the electricity sector — case study of India

Question 1: *What are the uncertain factors that can impact the design of the different segments of the power supply?*

The first question of this thesis' objectives is addressed in this chapter by focusing on the Indian power sector. India is expected to witness rapid growth in electricity use over the next three decades. Here, we introduce a custom regression model to project electricity consumption in India over the coming decades to project growth in existing appliance use. Additionally, we include a bottom-up estimate of electricity consumption stemming from areas that do not have local historical data to learn from, this model is applicable to any electricity demand driver (e.g., electric cooking). For this thesis, we focus on two major drivers of load growth: air conditioning and vehicle electrification. The model projections are available at a customizable level of spatial aggregation at an hourly temporal resolution, making them useful for long-term electricity infrastructure planning studies. Furthermore, the approach is used to develop electricity consumption data sets spanning various technology adoption and growth scenarios up to the year 2050 in five-year increments. The data aim to provide a range of scenarios for India's demand growth given new technology adoption.

With long-term hourly demand projections serving as an essential input for electricity infrastructure modeling, this data investigation enables the subsequent work of

this thesis on decision-making under uncertainty. The demand growth scenarios produced here are later used as exogenous and endogenous uncertainty in the electricity resource design problem.

2.1 Introduction

Many assessments of future electricity demand in India project large increases in electricity consumption from adoption of air conditioning technologies in the buildings sector over the next two decades [23, 26, 25]. This large growth is likely to make India among the top nations in terms of electricity consumption, implying that technology choices related to energy consumption and production in India are likely to play a significant impact on global climate change mitigation efforts. Additionally, the Indian government has been pushing for the transportation sector's electrification, starting with two- and three-wheel vehicles,which is further likely to increase overall electricity demand. As of 2020 in India, there are 152,000 registered electric vehicles [26]. Air conditioning (AC) related electricity demand accounted for 32.7 TWh, contributing to less than 2.5% of the total demand in 2019 [25]. However, both air conditioning and transport electrification are anticipated to introduce structural changes in the temporal and spatial trends in electricity consumption patterns, that has important ramifications for long-term resource planning for the electricity sector [29]. This chapter presents an bottom-up approach to estimate electricity consumption in India for various scenarios of technology and policy adoption. We focus on providing aggregated consumption estimates as well as spatio-temporally resolved consumption profiles that would be relevant for regional and national electricity system planning studies. The approach enables quantifying the impact of various growth and technology adoption scenarios on quantity and pattern in electricity consumption. The datasets detailed in this chapter include annual energy consumption at India's state, regional, and national levels as visualized in Fig. 2-1, as well as underlying consumption profiles at an hourly time resolution. The annual energy consumption is forecasted on a five-year increment to 2050. Fig. 2-2 shows one scenario of national electricity demand forecast. In addition to the snapshot of annual consumption, hourly load profiles are developed at the same resolution as seen in Fig. 2-3.

The forecasting is divided into two steps: business-as-usual and technology. Business-as-usual is a statistical model that infers data it can be trained on i.e. historical electricity demand. The technology model is a bottom-up approach that adds new loads to the total demand. Among new loads, we focus on residential and commercial

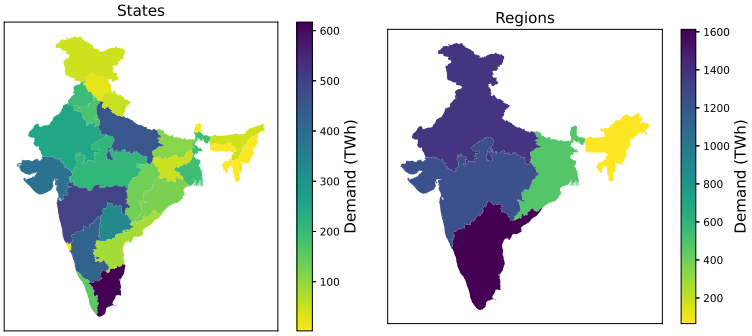

Figure 2-1: State and regional level distribution of annual electricity 2050 for stable GDP growth, baseline cooling, and home electric vehicle (EV) charging scenario.

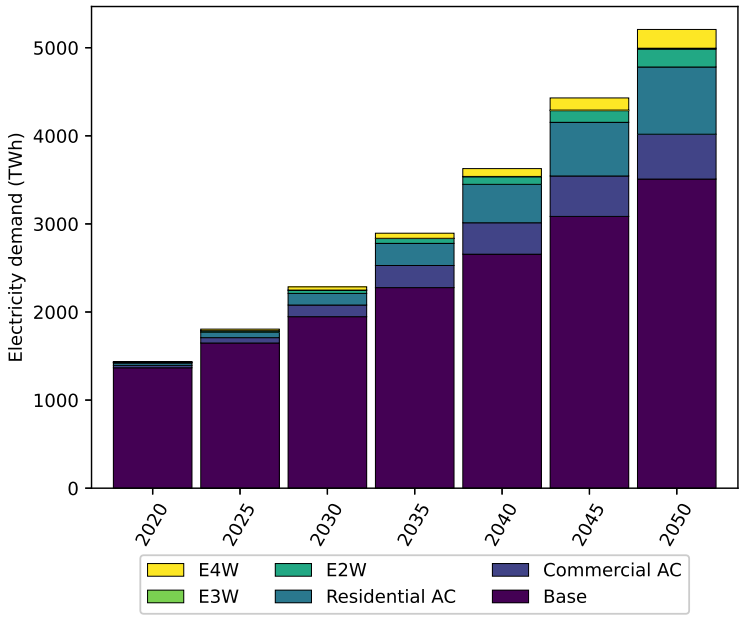

Figure 2-2: Summary results of India's electricity demand forecasting at national level with stable GDP growth, baseline cooling, and home electric vehicle (EV) charging.

cooling as well as various electric vehicles (EV). Some key insights from cooling

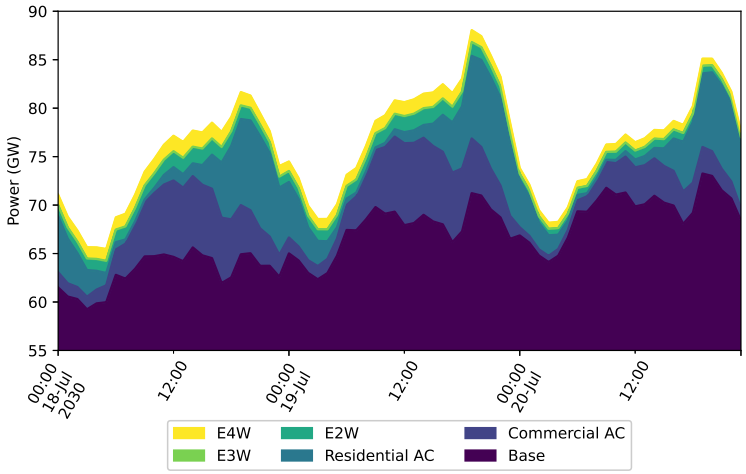

Figure 2-3: 2030 Load profile for Southern Region across three days in summer. Scenario: stable GDP growth, reference cooling, home electric vehicle (EV) charging. Base: projected electricity demand of lighting, appliance, and industrial equipment. E2W: electric two-wheelers (e.g., bikes and scooters), E3W: electric three-wheelers (e.g., tricycles and auto rickshaws), E4W: electric four-wheelers (e.g., passenger vehicles).

[25] and EV [26] studies highlighting peak demand development motivate the need for demand forecasting at the hourly resolution. Cooling demand due to mainly short time constant split unit air conditioning installation in India is expected to increase the peak to mean ratio (also sometimes referred to the "peakiness") of electricity demand in India as well as shift the timing of peak demand from evenings to midnight [25]. While electric vehicles do not constitute a large portion of the total demand, certain charging schemes can contribute significantly to the peak demand [26]. Numerous energy demand forecasts for India have recently been published as decadal snapshots [23, 29, 28], however granularity of demand at an hourly resolution has not been presented in these studies. Our approach enables quantifying the impact of different technology and structural elements, such as adopting energy efficient vs. baseline cooling technology or work-place charging vs. home charging for EVs, on the hourly electricity consumption profiles. These insights and the accompanying data sets are essential to carry out generation and transmission expansion as well as distribution network planning. Thus, they are essential to a sustainable energy infrastructure development in the Indian context.

Similar to other forecasting studies, we model Gross Domestic Product (GDP) growth

[71] to be the main econometric driver of the business-as-usual demand forecasting, and thus three scenarios are introduced: slow, stable, and rapid GDP growth. We examine two AC load scenarios: high AC efficiency and baseline equipment as per the International Energy Agency's Future of Cooling study [25]. Finally, we evaluate three EV charging mechanisms: home, work, and public charging. This totals the number of data sets spanning three input dimensions to 18 scenarios. Technology adoption growth has been correlated with economic growth under the assumption that new technologies are adopted faster when the economy is growing faster and vice versa. We present two cooling scenarios to highlight the difference in energy-efficient and regular air conditioning units and bring attention to the need for policy and programs that favor energy-efficient cooling unit sales. Furthermore, we present various EV charging mechanisms to inspect the demand impacts that electric vehicle charging can have on the electric grid at different times. The produced data can be used as input to electricity infrastructure planning both at the distribution and transmission level.

2.2 Demand forecasting

Fig. 2-4 illustrates the major steps of our proposed demand forecasting approach. We use two models to estimate future electricity demand in India. In the first model, we use a linear regression model to project daily peak and consumption on a regional basis; this is the business-as-usual scenario. We then add natural variation to the projections by finding the error between the training data and results and scaling it to every region based on seasonality. Then we fit the projected peak and total consumption to an annual hourly load profile for 2015 [72] featuring an evening peak . In the second model — technology model — we take AC and EV adoption into account as an additive component on top of the business-as-usual predictions. We restrict the bottom-up modeling to AC and EV loads to the increased interest in space-cooling demand and electric vehicle adoption in India [25, 73]. However, bottom-up modeling of new loads can be expanded to other types of loads (e.g., electric cooking). GDP data, which is an independent variable in the model, is chosen to be the main driver of growth of the business-as-usual scenario as well as technology adoption rates. The input data used are publicly available and are referenced in Table 2.1.

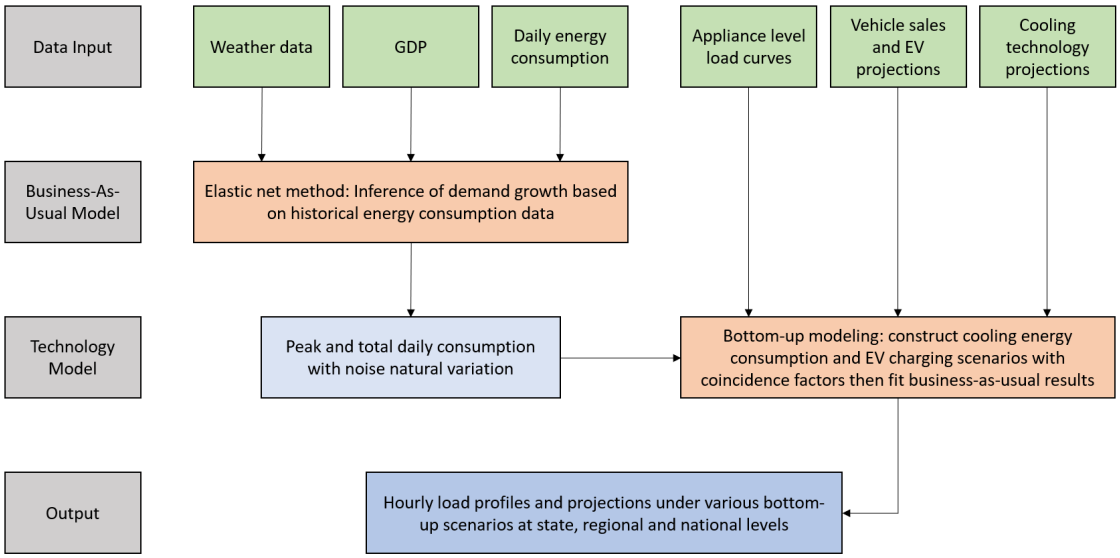

Figure 2-4: Simplified schematic of demand forecasting method: steps in gray, input data in green, models in orange, and outputs in blue.

Table 2.1: Input Data Sources

Data	**Source**
State-wise Historical GDP	Ministry of Statistics and Programme Implementation [71]
Vehicle Sales and Registration	Society of indian Automobile Manufacturers [74]
Air Conditioning Stock and Capacity	International Energy Agency [25]
Load Profile	United States Agency for International Development [75]
State-level sector-wise energy consumption	Power System Operation Corporation [76, 77]

2.2.1 Input data processing

Although GDP is widely used for forecasting energy demand, it is specifically essential in the case of India, where economic growth is expected to ramp up over the next few decades similar to the recent trends in China [78]. We based our demand forecast on GDP projections from a PricewaterhouseCoopers (PwC) report [79], that

projected India's GDP to grow from 3.6 trillion in 2020 to reach 28 trillion USD in 2050. Considering the historical national GDP data for India starting in 1990, we fit and project an exponential curve for rapid growth and a Gompertz curve [1] for slow growth [80] as detailed in Table 2.2. We use PwC's projections to define the stable GDP growth scenario. Curve fitting and projection results are illustrated in Fig. A-1. The rapid growth scenario produces an annual average growth rate of 9.5% , PwC's growth rates start at 7.8% for the first projected decade and ends at 6.2 % in the final projected decade. The slow growth scenario starts at 7.2% growth rate in the first projected decade and ends at 3.9% in the final projected decade. To break down the regional energy consumption projections to state level we use the ratio of GDP per capita of the corresponding state to the GDP per capita of the region it is in. For each GDP growth scenario, we fit the same functions given state-wise data to produce GDP forecast at the same resolution. GDP per capita at state-level is computed using the projected GDP data and state level population projections [81].

Slow	**Rapid**
Gompertz Growth Curve	Exponential Growth Curve
$y = Ae^{-e^{\frac{\mu e(B-x)}{A}+1}} + C$	$y = Ae^{Bx} + C$
A = 7×10^{-29}	A = 1×10^{-64}
B = 4.64×10^{-2}	B = 8.7×10^{-2}
C=0	C=0
R-squared = 0.949669	R-squared = 0.989361

Table 2.2: GDP projections curve fit results. x is the year, y in the GDP growth rate. A, B, and C are parameters resulting from fitting the curves to the historical state-level GDP data. R-square is the average goodness of fit of the curve under parameters A, B, and C to the historical GDP data of all considered states in India.

GDP dependence and limitation

Relating growth in electricity demand to GDP is a strong generalization, however it is not a novel one in the case of India. Strong correlation between economic growth and energy consumption has been established in the Indian context in this study and other studies [82] given data from the past two decades [71]. We recognize that GDP as a metric of economic growth has several limitations particularly related to projecting

[1] A Gompertz curve is a sigmoid function that describes growth as being slowest at the start and end of a defined time period.

how economic growth is distributed among society within a state or nation. This may be the strongest limitation of the data we are presenting in the manuscript. However, lack of historical record and long-term projections of alternative open-access economic data at the desired spatio-temporal resolution limits the development of a framework to project energy consumption with other metrics. While GDP and energy consumption growths may differ in the long-run, there is an evident correlation between the two that can be used to estimate long-run energy consumption growth. Deviating away from linear regression may yield better results, however, data scarcity is again a limitation to the development of more complex models. Furthermore, this manuscript motivates the need for more bottom-up projections and not just regression models because historical consumption cannot infer consumption trends from new demand sources such as cooling and EVs.

Additionally, since the Future of Cooling study by the International Energy Agency relies on GDP forecasts developed by the International Monetary Fund[25], we elect to use a similar metric. We intentionally develop a large bandwidth of projection scenarios to mitigate the limitation of an individual snapshot representing a singular assumption. The motivation behind presenting the described results is ability to compare different scenarios and post-analyze the demand growth and the trade-offs. To produce a large bandwidth of growth scenarios we needed to use a straightforward metric that has enough historical data to produce various fitted curves for projections.

2.2.2 Business-as-usual model

The business as usual projections are modeled with a linear regression considering weather and economic growth features. The historical daily peak and total consumption for each electric grid were obtained from the Power System Operation Corporation (POSOCO) for 2014-2019 [76]. The GDP used in the model was obtained, as explained in the previous section. Weather data was secured from the NASA Merra-2 data set [83]. The choice of features for the regression model is limited to GDP and weather variation due to the limitation in availability of data, both historical and future projections, at the desired spatial and temporal resolution. GDP is identified as a long-term parameter driving growth in year over year demand projections as highlighted in Fig. 2-5. Sampled weather data of 50 random locations per state are identified as a short-term parameter driving seasonal variation within a year's demand projections as highlighted in Fig. 2-6. Previous parametric analysis on these features and their coefficient for short and long term demand forecasting in both time and frequency domain [84] reinforce their use as features for the business-as-usual regression model. We present detailed outcomes for the Southern region,

with further details available in [84].

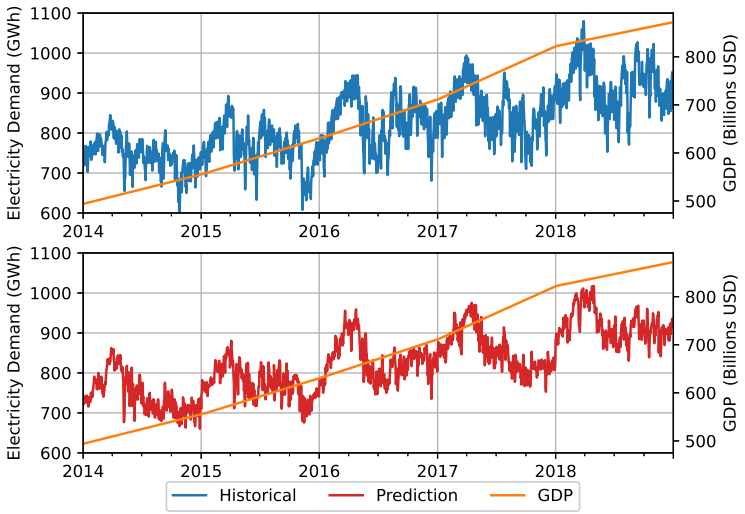

Figure 2-5: Southern region back test annual demand growth given GDP projection

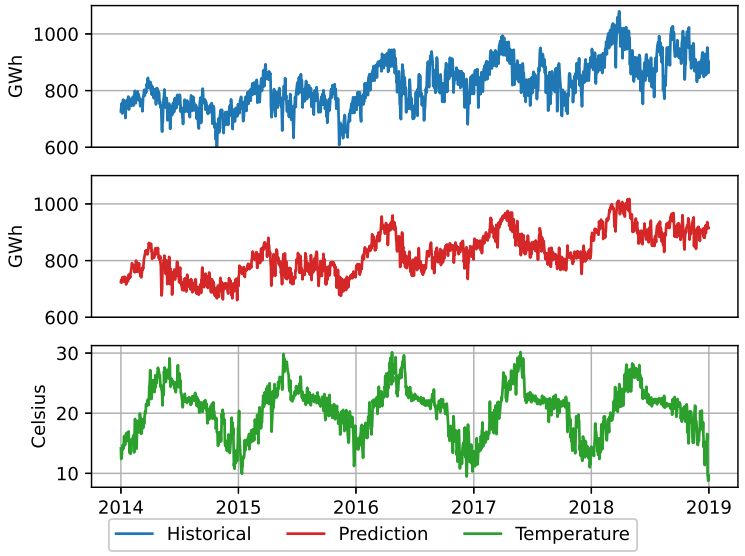

Figure 2-6: Southern region back test seasonal demand variation given weather data

NASA Merra 2 data acquisition

For each of the five electric grid demand regions highlighted in right panel of Fig. 2-1, the largest cities in each region were identified using population data made available by the United Nations[85]. Then, the city's latitude and longitude were used to pull down the corresponding environmental data from the Nasa Merra-2 data set. The cities used for each of the five regions are listed here:

- Northern: Delhi, Jaipur, Lucknow, Kanpur, Ghaziabad, Ludhiana, Agra
- Western: Mumbai, Ahmadabad, Surat, Pune, Nagpur, Thane, Bhopal, Indore, Pimpri-Chinchwad
- Eastern: Kolkata, Patna, Ranchi (Howrah was ignored because the environmental factors are the same as Kolkata)
- Southern: Hyderabad, Bangalore, Chennai, Visakhapatnam, Coimbatore, Vijayawada, Madurai
- Northeast: Guwahati, Agartala, Imphal

From the NASA set, 11 variables were included for each city: specific humidity, temperature, eastward wind, and northward wind (all 2m above the surface and 10m above the surface - eight total variables), precipitable ice water, precipitable liquid water, and precipitable water vapor. In particular, the instantaneous two-dimensional collection "inst1_2d_asm_Nx (M2I1NXASM)" from NASA was used. Detailed descriptions of these variables are available in the Merra-2 file specification provided by NASA [83]. The environmental variables available from the NASA MERRA-2 dataset were given on an hourly basis. The daily minimum, daily, maximum, and daily average was calculated for each of the 11 variables for each day.

Forecasts

The business-as-usual demand forecasting problem was divided into ten separate problems, corresponding to one problem each peak and total consumption for each of the five regional grids shown in Figure 2-1. To ensure the model would not overfit the data, the model was trained with a regularized regression method (Elastic Net) that linearly combines L_1 and L_2 penalties [86] to fit the results, and validated on the held out 2019 data. Elastic Net expands the basic ordinary-least squared (OLS) equation to include to terms: lasso and ridge. In lasso regression, we add a penalty (L_1) that equals the absolute sum of the coefficients of the basic OLS equation. In ridge regression, we add a penalty (L_2) that equals the square of the magnitude of

the coefficients of the basic OLS equation. Lasso method tends to set coefficients of the regression to absolute zero, while ridge method never does. An L_1 ratio (Lasso) of 0.9 was chosen to minimize error in 2019 as the validation set. Then all of the models were trained with 0.9 L_1 ratio on the full dataset.

Addition of natural variation

This step aimed to match the statistical characteristics of an actual load year with the projected year. 2019 was used to derive the differences. Natural variation was estimated by a distribution characterized by the mean and standard deviation of the differences (in absolute value). Then, a natural variation adjustment was added to that day (with a random true/false bit for positive or negative variation). The noise was calculated for each region and peak demand and daily consumption separately. This part of the process is non-deterministic and replication of the results requires using the same natural variation vector used in our projections.

Hourly profiles

The statistical inference model presented above forecasts daily consumption driven by state-level economic parameters and weather data. The produced projections are at a daily resolution. We downscaled the data to hourly load profiles based on the 2015 hourly load profile data [72]. The result of the regression model is at regional level, breaking it down state-wise is pro-rated based on state-wise to region-wise GDP per capita projections ratios for the respective year. To do so, we tag each day of the year by the month it corresponds to and whether it is a weekday or weekend. We cluster demand for each hour by month and day. Each hour of the day then has its own cluster of demand data from 2015 based on the assumption that the same hour of the day for a given month and the same day type will exhibit similar demand behavior. This biases the construction of the profiles to demand patterns from 2015 only, which would result in multi-decadal projections fitted to 2015 patterns only. To minimize the impact of this bias, we use the historical weather data[83] of the testing data years (2014-2019) for each day to simulate daily temperature variations that are reflected in higher or lower demand. We sample weather data for each day and compare it to 2015, and subsequently use the normalized difference to scale the demand on a daily basis. Finally, we sample demand for each hour of the year from the corresponding cluster (defined by month and weekend or weekday) and scale it accordingly. Constructing the hourly load profile and fitting them to match the projected daily consumption and the projected daily peak demand then becomes a trivial exercise of sampling and fitting from the corresponding clusters and weather

data space. The 2015 hourly demand data used in this study is documented in detail in [72] and has been used in projecting demand for supply-side modeling efforts [87]. Limited availability of complete hourly data at state and regional level in India biases the hourly profiles to the 2015 datasets. However, the business-as-usual projections are for existing demands composed mainly of lighting and appliance at the residential level and large daytime loads at the commercial level [75]. Our approach implicitly assumes that energy consumption trends for these loads will follow historical patterns and therefore sampling from a given year with post-processed noise variation can yield reasonable results.

Impact of climate change on business-as-usual demand

As per the International Energy Agency (IEA) World Energy Outlook (WEO) 2019 [88] only 5% of households in India currently own air conditioning units and 2.6% of commercial building energy use is from space cooling. Historically, electricity consumption in India has been driven by lighting and appliances in the residential sector [75] with commercial and industrial sector contributing via larger daytime loads. Since cooling demand is not historically available in the data that the business-as-usual regression model is learning from, there is no parametric value to projecting increase in temperatures since there is no evident correlation between temperature increase and lighting or appliance use. Moreover, since space cooling is a small percentage of current electricity demand in India, no major trends can be identified given the limited daily training data that is being used for the business-as-usual regression. It is then safe to assume that weather remains constant for the business-as-usual demand.

2.2.3 Technology model

Since a regression model can only produce forecasts of data it can learn from, additional bottom-up processing must be carried out to get a full picture of India's demand in the future. We identify trends and data points at the state level of the country to build a regional profile as well as the national one.

Cooling

Cooling is divided into two main categories: residential and commercial. The ratio of commercial to residential consumption is computed from state-level data [89] and is used as the ratio of commercial to residential cooling demand. Using the IEA's baseline and efficient cooling projections from the Future of Cooling study [25], we

use the annual sales and unit types to calculate the energy consumption and growth rate at a national level and pro-rate it down to state level given GDP per capita. Surveyed hourly demand profiles [75] are indicators of behavioral cooling energy consumption patterns as exemplified in Fig. A-2 and A-3. The survey produce various profiles given climate seasons, household income and size. We apply a time-domain convolution of these profiles to generate a representative profile for each state for the various climates and seasons.

We can generate the air conditioning demand profiles for two weather seasons (winter and summer) by convolution of the sample profiles to generate a smooth aggregated demand profile. Moreover, coincidence factors must be applied to properly estimate the simultaneity of the demand and its peak. Two coincidence factors are identified: weekday and weekend, values are extracted from a Reference Network Model Toolkit [90]. We break down the national cooling demand to residential and commercial at state level by identifying state-level sector size and growth trends. Scaling the profiles to match the projected cooling energy demand produces hourly energy consumption profiles from residential and commercial cooling. Aggregating the appropriate states together will produce the same results at the regional level.

More importantly, the IEA's future of cooling study [25] stresses the usage of Cooling Degree Days (CDD) to project cooling demand dependency on temperature. The unit consumption pattern and projections of capacity for India's share of global cooling demand is based on growth in electrification, urbanization as well as Purchasing Power Parity. The IEA future of cooling study estimates that a 1-degree Celsius increase in decadal average temperature in 2050 will to lead to 25% more CDD and a 2-degree Celsius increase will lead to 50% more CDD. Climate change impacts are considered in the unit sales and energy consumption data used from the IEA's future of cooling study. In our analysis, we use IEA's 50% increase in CDD to model cooling demand in 2050. For prior periods, we interpolate CDD between 2018 and 2050 to model cooling demand. The increase in CDD and the addition of noise variation are introduced for the purpose of modeling the projected increase in peak demand due to climate change. We follow the IEA's unit sales and energy consumption projections to project space-cooling demand in the bottom-up modeling. Specifically, this analysis does not consider frequency nor forecast of extreme weather events because they are of short duration and therefore would not have much impact when integrated over time. Moreover, identifying outlier extreme events is a high noise-to-signal ratio exercise especially when projecting load growth 30 years into the future.

Electric vehicles

The second component of the technology model projects EV demand in India. The data presented here considered electric two, three, and four-wheel vehicles. Two-wheelers, being the dominating vehicle in terms of annual sales in India [74], are expected to be electrified first, followed by the three-wheelers and regular cars [91]. The Indian government has set a goal of converting 100% of two-wheeler sales and 30% of all vehicle sales to electric by 2030 [92], so the starting point is vehicle sales at the state level [74]. Using the regression equations of the corresponding GDP growth scenarios, we can project car sales with the EV targets by 2030 met in the rapid growth scenario. From vehicle sales and conversion rates, we get an estimate of the number of EV that will require charging. From a market survey on the average commute distance of vehicles in urban areas and rural areas [91], long and short-range battery capacity and EV energy can be estimated. We introduce a mix of EV sales starting with short-range as the dominant market product and shifting to long-range, a market-dominant market in 2050. This trends reflects the current economic competitiveness of short-range EVs vs. existing internal combustion engine vehicles as well as the long-term competitiveness of long-range EVs with declining battery costs.

Similar to the construction of the cooling profiles, a coincidence factor must be implemented, so as to not over-predict peak EV charging demand. Since this is a new consumption behavior and given the relatively small batteries of two-wheelers, it is assumed that every vehicle needs to charge every other day on average for urban drivers and every day for rural ones based on preference survey data [93]. Three-wheelers, which are predominantly used for commercial purposes are assumed to be charged every day [93]. This yields an average daily consumption from EV charging. As shown in Fig. A-4, three different charging profiles — home, work, public – are identified in an EV pilot project study in Mexico City [94]. We recognize that Mexico and India differ greatly in many socio-economic aspects, however the pilot project reported in Mexico — and specifically in densely populated Mexico City — is an EMDE country EV adoption project that is used a starting point for vehicle charging behavior. The different hourly EV charging profiles collected were for a pilot project to deploy electric two-wheelers and small sedans in the metropolitan area of Mexico City. This presents two synergies enabling the usage of the charging profiles in India. Under the assumptions that EV deployment will be more prevalent in urban areas in India with initial conversion of smaller vehicles (two-wheelers and three-wheelers), the charging data collected [94] is a suitable fit for potential EV charging schemes in India. Energy consumption is computed from vehicle sales,

projections, and electrification conversion. That calculated number is then fitted under the chosen charging profile. Time series rolling forward and backward in time is applied to the profiles to smoothen the peakiness of the total constructed hourly time series.

Data Dependence

The technology model relies heavily on surveyed data to produce the representative hourly profiles for cooling and electric vehicle demands at state levels. This is indeed a limitation, and our projections assume that future technology adopters will behave just like initial adopters. In the absence of a better alternative at a similar spatial and temporal resolution, the bottom-up modeling effort provides a reasonable estimate of temporal patterns expected from these new demand sources. For the hourly sample cooling profiles, the main assumption is that cooling demand consumption is only dependent on weather patterns and econometric patterns. Specifically, we apply a weighted sum convolution of the income level cooling profiles based on the states' GDP per capita ranking. For the total cooling demand at national level, we depend on the air cooling unit sales projection as well as break down of unit energy consumption under baseline and efficient scenarios of the IEA's Future of Cooling report [25]. We pro-rate residential cooling at state level using the GDP per capita projections. For commercial cooling we use the state-wise sector growth trends [77]. A sanity check for this break down is to sum both residential and commercial state-wise cooling demand and compare to the IEA's all India cooling demand annual electricity consumption projections to 2050. The difference is highlighted in Fig. A-5 and A-6. Regarding the EV profiles, while there are alternative choices of charging schemes, we identified the synergies with the Berkeley study [94] to be best reflective of the bookend EV charging scenarios across India.

2.3 Technical validation

The Business-as-usual statistical model is validated using standard statistical metrics when backtesting is applied. Further details on the backtesting are available elsewhere [84]. For the technology model, we compare our estimates to the IEA's WEO [23, 88, 95, 96] and Brookings India [28]. Furthermore, our projections compare favorably against the EV projections to the IEA's Global Electric Vehicle Outlook 2020 [26].

2.3.1 Back-testing

Daily consumption and peak are projected for all five regions, we show the daily consumption back tests of the Southern Region in Fig. 2-7. It is important to note that the regression model captures the organic growth of the historical demand as well as the seasonal variation in demand but is not accurate at predicting daily variation. This shortcoming can be attributed to the small training dataset that is available. To compensate for this short-coming, we add additional noise variation as discussed earlier in the Methods section. We compare the R-squared value of the regression only versus the regression and noise time series as shown in Table 2.3. Additionally, selected parameter performance metrics of the model for the Southern Region are presented in Table 2.4. The model's independent variables are the 2 meters and 10 meters elevation historic temperature and humidity data for the selected cities and GDP data for the state. Various weather parameters will have a higher coefficient than GDP since the latter is not as granular as a metric but will still be factored in for longer term growth as interpreted by its Fourier component [84].

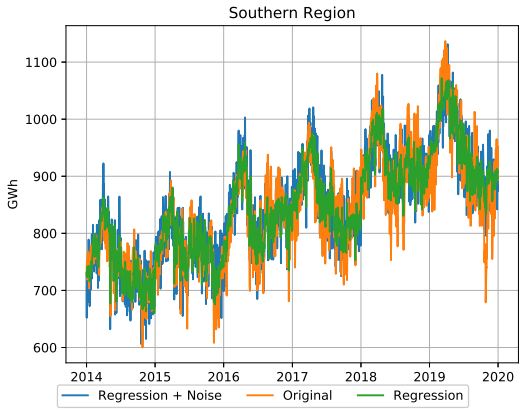

Figure 2-7: Back test result for Southern Region regression model with and without the additional noise variation.

2.3.2 Cross-comparison

Fig. A-7 and A-8 compare the forecasting results to the WEO 2020 projections of India's Energy Demand to 2040. Our band of projections is notably wider due to the large number of scenarios that are combined to forecast energy demand. We further compare our results to Brookings India's study in Fig. A-9. We also compare our electric vehicle projections to those of the Global EV Outlook in Fig. A-10. Finally,

Region	Regression	Regression + Noise
Eastern	0.709	0.798
Northeastern	0.608	0.722
Northern	0.691	0.784
Southern	0.744	0.825
Western	0.680	0.778

Table 2.3: Business-as-usual Regression R-squared consumption results

Parameter	Coefficient	Standard Error	t value	P > | t |
GDP	13.3630	34.663	0.386	0.700
Bangalore t2m max	931.0014	353.052	2.637	0.008
Chennai h2m min	1.736e+04	1.65e+04	1.050	0.294
Hyderabad t10m max	-502.7828	356.558	-1.410	0.159
Vijayawada h10m min	-2.229e+04	7671.803	-2.905	0.004

Table 2.4: Business-as-usual Southern Region consumption Regression performance of select parameters

we compare our air conditioning demand contribution to the peak demand to the Future of Cooling study in Fig. A-5.

2.3.3 COVID-19 pandemic impact on year 2020

The COVID-19 pandemic has drastically affected the global population in various ways. Energy consumption dropped severely as people were advised to stay at home. While it is not possible to project such "Black Swan" events from historical data, their long-term effects can be modeled as delayed growth under various recovery schemes. Fig. 2-8 shows that our projections for the month of January 2020 align with the realized demand, which is prior to the global outbreak of COVID-19. Evidently, there is a strong mismatch in the following months as the outbreak developed into a global pandemic. However, in the later part of the year, signs of recovery are noticed where the historical daily consumption once again reaches projected levels.

The impact of extreme events on energy consumption are difficult to predict at a granular level. Our projections are at a five year increment so that such yearly variations are smoothed out and the regression towards the mean phenomenon is observed. Moreover, the recovery from extreme events and their long-term impact can depend on many factors: economic, social, scientific and more. Without modeling

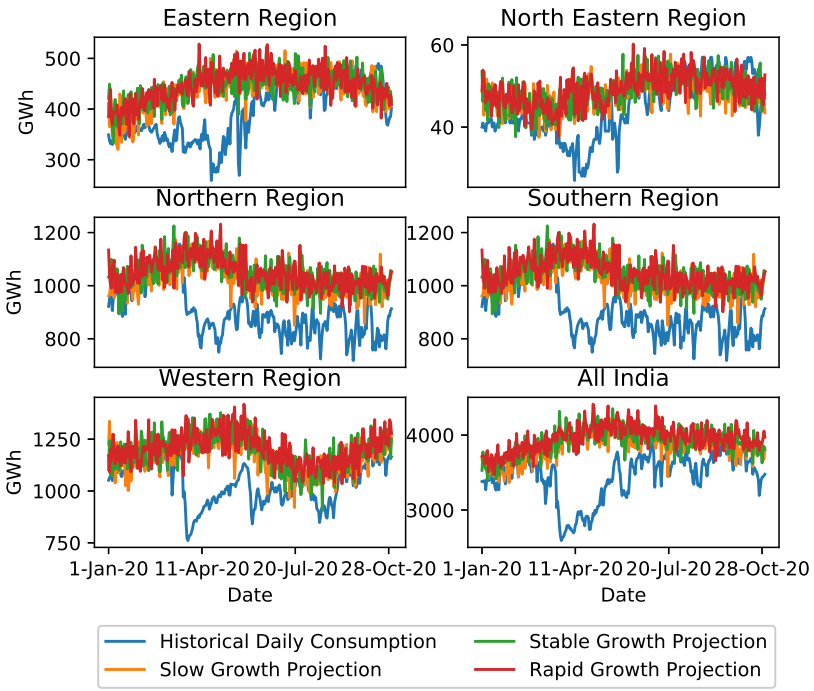

Figure 2-8: 2020 year-to-date demand comparison with projections

those events in detail, projected growth can model the long-term average growth rate. In case of a negative extreme event, a smaller growth rate can model the long-term impact caused by the slow down. Similarly, a positive extreme event can be modeled as larger growth rate to include the long-term impact by the rapid growth. With signals of a fast recovery in total daily consumption for most regions, we elected to disregard projections that model long-term COVID-19 pandemic impact to avoid confirmation bias. Moreover, there is little data to support projections modeling a long-term impact on Indian energy consumption. We believe that the model and data presented in this chapter are valid beyond the COVID-19 pandemic.

2.4 Discussion

The data of the various demand growth scenarios presented in this chapter is used for modeling demand growth uncertainty in the multi-stage electricity resource design problem throughout this thesis.

Chapter 3

Impact of uncertainty at distribution level — case study of Delhi

Question 2: *How do these uncertain factors impact electric grid planning at the distribution stage?*

In the previous Chapter 2, we introduced a methodology to quantify the wide range of demand growth scenarios that India can potentially witness by mid-century. This chapter parametrizes these scenarios as exogenous uncertainty of electricity demand growth. Here, we focus on the impact of demand growth uncertainty on planning the distribution grid of peculiar cases of megacities in India.

The growing demand for electricity in EMDE countries such as India is causing loading and congestion problems on distribution networks, particularly in urban locations, adversely impacting sustainable development and economic growth. Electric utilities in these economies face unique constraints regarding raising the capital required to upgrade their congested networks. Battery storage has emerged as a non-wire alternative (NWA) to feeder upgrades. This chapter presents a flexible valuation framework for battery storage use in distribution networks and its application in the context of EMDE countries distribution network planning. We evaluate the value of storage as an NWA using a multi-stage decision-making process that combines system optimization with Markov-decision processes (MDP) to identify the least-cost network upgrade strategy under demand growth exogenous uncertainty.

When applied to feeders in Delhi, India, the approach highlights the cost-effectiveness of battery storage to manage load growth while deferring network investments.

Across the low, medium, and high battery storage capital cost projections for 2030, we estimate that 18 to 29 GWh of battery storage capacity could be deployed to defer 11,752 to 15,914 km of medium voltage distribution feeder lines that are loaded at 60% or more of their ampere capacity in 2030, resulting in 12 to 16% capital cost savings. Interestingly, the lowering storage capital costs do not always increase NWA storage deployment due to network capacity constraints limiting opportunities for off-peak storage charging.

3.1 Introduction

Investments in electrical distribution networks tend to be *lumpy* [97] since they require large capital commitment initially and involve significant economies of scale because the assets have long lifetimes (20 to 40 years). Consequently, long-term distribution network planning is often necessary to identify the timing and size of investments needed to meet future demand reliably and cost-effectively while maximizing asset utilization. With declining cost of Li-ion battery energy storage (referred as battery storage here on), there is growing interest to consider its use as a non-wires alternative (NWA) to defer expensive distribution network upgrades and serve rapidly growing peak demand within electricity distribution networks. The modularity of battery storage as well as its flexibility both in terms of location and speed of deployment are in stark contrast to the attributes of conventional network investments and thus represent a potentially valuable option to be considered in network planning. The role for storage as a flexible investment option [98] is particularly relevant for loaded urban distribution systems in megacities in fast-growing Emerging Market and Developing Economy (EMDE) countries, such as Cairo (Egypt), Delhi (India) and Jakarta (Indonesia), because of several factors [99, 100]. First, many of these cities are experiencing rapid electricity demand growth, due to growing adoption of air conditioners (AC) for space cooling [3, 32] that contributes to network congestion through increasing peak electricity demand [25]. Second, distribution companies in many of EMDE countries are often financially constrained [101] and have to contend with relatively high cost of capital to finance network investments. Third, the premium on land use and geographical constraints in some of these megacities could result in further network investments (reconductoring and upgrading lines) to be operationally challenging or infeasible [102, 103]. Here, we analyze the optimal sizing and placement of battery storage and its economic value as an NWA at the primary feeder level in urban electricity distribution networks of Indian megacities such as Delhi.

India's electricity demand is projected to more than double by mid-century [28], primarily from increasing electricity use for space cooling in the buildings sector and to a lesser extent, by electrification of transportation [25, 27]. Much of the growth in energy demand is concentrated in megacities like Delhi, where 55% of electricity use is residential, which is more than double the national average (24%) [3]. Distribution companies in Delhi are witnessing a level of growth in cooling demand that is capable of overloading the network equipment - for example, feeder data for 2018 indicates that 28% of feeders were loaded at 60% or more on an ampere capacity basis. As of 2020, Delhi's peak power demand was 6.7 GW [104], and long-term demand projections from our prior work [3] suggest wide variation in possible outcomes depending energy needed for space cooling related electricity demand (see Table 3.1). This wide range of possible future outcomes creates significant uncertainty for investment planning in the distribution networks.

Table 3.1: Projected peak demand (GW) under the baseline and high-AC-efficiency scenarios as per Chapter 2 assuming stable GDP growth for the city state of Delhi. Further details in [3]

	High AC efficiency	**Baseline**
2020	6.7	6.7
2030	12.7	15.2
2040	25	36.7
2050	34	63.8

Historically, distribution companies have not considered demand uncertainty in their long-term network planning, but have instead resorted to deterministic net present value methodologies [97, 51]. Demand forecasts enable a comparative assessment of program implementation such as efficiency, policy, and technology under various scenarios. Probabilistic forecasting and flexible planning may be most useful in situations when the magnitude of future outcomes exhibit wide variations, which is the case for the peak electricity demand projections for a city like Delhi (Table 3.1). Accounting for uncertainty in investment planning may particularly be important when contemplating the use of distributed energy resources (DER) as an alternative to grid expansion, owing to the modularity of DER technologies and their speed of deployment. Until recently, most DER in EMDE countries have been in the form of diesel generators deployed near large commercial and industrial (C& I) loads[105]. However, declining costs for Li-ion battery storage[5] make it a more attractive option. Moreover, battery storage provides the added advantage of not creating local air pollution, a major environmental externality in most EMDE megacities. Furthermore,

depending on the energy source used for battery charging, the carbon footprint of energy discharged from battery storage is lower than the diesel generation commonly used in EMDE countries to meet peak demand[106].

Distribution network investment planning has been previously assessed in several studies. For example, in [36], the authors compare the deployment of low-cost diesel generation in rural communities of Latin America to network reinforcement costs while accounting for uncertainty in electricity prices. Another study [38] addresses the value of options analysis for DER under future technology cost uncertainty. Authors of another study[37] proposes a flexible investment strategy for renewables incorporation and least-cost system design at transmission level. Such studies that consider uncertainty in electricity system decision making are either concerned with bulk power system design and ignore distribution level planning, or are not concerned with the temporal nature of local DER operations, since their assumed DER are flexible and dispatchable resource (diesel generators, hydro) or energy producing (rooftop solar PV). Moreover, frameworks developed for analyzing role for DER generators are not as informative when evaluating the role for an energy-limited resource such as battery storage that also implicitly couples multiple periods of network operations through its charging-discharging patterns. While regulatory frameworks are being developed for battery storage at distribution level [107], generation design and dispatch models which can assess the different roles of battery storage on a network are generally constrained to deterministic formulations [39, 40, 41, 42, 43, 44] and do not typically consider the impact of long-term uncertainty in different factors influencing planning decisions, such as demand, technology costs and policy evolution. Other studies have considered temporal variability of grid operations and battery storage sizing and operations in the systematic planning of distribution network [108, 109, 110]. Multi-stage stochastic programming approaches have also been applied to the problem of distribution network planning, wherein grid operations were modeled using representative periods and investment in energy storage was not considered as a model variable [111, 112]. In summary, studies considering demand or other types of uncertainties in distribution network planning have typically relied on a limited temporal resolution of system operations and thus may under-value energy storage's ability to alleviate network constraints by shifting generation over time.

Here, we develop a financial valuation framework for battery storage use in distribution network planning that accounts for: a) design and dispatch of battery storage on the electric distribution grid subject to network and operational constraints consider hourly system operations over the year and the b) impact of long-term uncer-

tainty from demand growth on the value and timing of battery storage deployment as a NWA. Our approach is based on combining a linear programming based optimization with a Markov Decision Process (MDP) based simulation that provides a transparent mechanism for evaluating the role for battery storage as a NWA. We illustrate the value of our framework through investigating the potential for battery storage in distribution networks prevalent in Delhi and other megacities in India. Although our analysis is based on available feeder conditions in Delhi, this approach offers general insights about the conditions under which it is economically viable to defer network investment by deploying battery storage. Moreover, while our method can readily incorporate investment in other types of DERs like rooftop PV, we only consider battery storage NWA for the case study of distribution feeders in India due to its novelty as a technology, modularity and, minimal space footprint, which is an important practical consideration for a highly congested urban megacity. We also ignored diesel generators due to their CO_2 and air pollution related externalities. In summary, the key contributions of this chapter are:

1. The development of an hourly network expansion optimization model at the distribution level that is used to compare the relative cost-effectiveness of battery storage NWA and traditional network upgrades while accounting for independent sizing of battery storage power and energy capacity and operations.
2. The application of the optimization model in a multi-stage simulation-based environment to evaluate the value of battery storage as an NWA under long-term demand growth uncertainty and various battery storage technology cost scenarios.
3. Quantifying the potential for using battery storage as an NWA in the distribution networks for four major Indian megacities.

This chapter is structured as follows: Section 3.2 describes the methods, including the key assumptions and structure of the proposed model; Section 3.3 details the Delhi case study input data we use; Section 3.4 describes results for Delhi and other megacities in India; in Section 3.5, we discuss the cost implications for the Indian case study.

3.2 Real options for non-wire alternatives

We break down the model for evaluating battery storage as a NWA into three categories: financial, technical, and probabilistic. Therefore, our overall approach is divided into three broad steps. Step 1 (section 3.2.1) presents the valuation criteria

for battery storage to be financially feasible as a NWA given network parameters and demand projections as modeled in Fig. 3-1 Box 1. Battery storage is utilized as a flexible investment to defer large network upgrades. Thus, battery storage costs and deferred network upgrade costs must be cheaper than traditional network upgrades, otherwise battery storage is not financially feasible. We must also solve for the battery storage system, location, size and operation on the network. In this context, step 2 (Section 3.2.2) involves the system design optimization with hourly dispatch to size and place battery storage as a NWA on a network. A simulation of various demand growth trajectories converges to a single decision for network investment planning under demand growth uncertainty. Step 3 (Section 3.2.3) describes a Markov decision process (MDP) of exploring the various system designs from section 3.2.2 that satisfy the battery storage NWA valuation criteria of section 3.2.1 to identify a least-cost network investment planning framework that adequately considers battery storage. Fig. 3-1 highlights the overall flexible valuation framework. The overall objective of the flexible valuation framework is to explore near-term solutions using storage as an asset to minimize the overall capital expenditure on the system by taking into consideration long-term demand growth uncertainty.

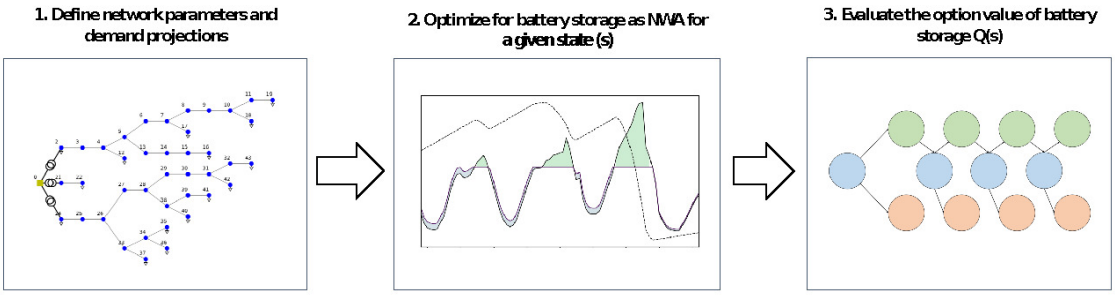

Figure 3-1: Flowchart showing steps in the flexible valuation framework used for modeling storage as non-wire alternative.

3.2.1 NWA valuation criteria

Distribution companies may defer long-term investments by deploying battery storage to meet their short-term peak demand needs and mitigate short-term financial commitments. For a battery storage NWA system to be beneficial, the net present cost of deferring traditional network investment must outweigh the battery storage system cost. This condition is shown in Eq. 3.1. Given a planning horizon starting in period y_0 and ending in y_n where n is the number of years to consider demand

growth uncertainty, the valuation criteria considers capital (I) at time t and the total fixed (F) and variable (V) costs during the planning horizon $[y_0, y_n]$. Three cost structures are compared: traditional network upgrades l, battery storage b and deferred network upgrades d. The right-hand size of line 1 of Eq. 3.1 refers to the costs of traditional network upgrades l over the entire planning horizon $[y_0, y_n]$. Lines 2 and 3 of Eq. 3.1 splits the planning horizon into two subsets: $[y_0, p]$ and $[p, y_n]$. The first subset refers to periods in which battery storage is used as a NWA and network upgrades are deferred until period p. The second subset refers to the remainder of the planning horizon starting at p and ending at y_n. All three incurred costs (I, F, V) are considered for battery storage and deferred network upgrades during their respective lifetimes. Variable costs for battery storage include charging cost at the available wholesale electricity tariff.

$$\begin{aligned} O_t^s(p) = C_{l,t}^I + \sum_{t \in [y_0, y_n]} (C_{l,t}^F + C_{l,t}^V) \\ - C_{b,t}^I + \sum_{t \in [y_0, p]} (C_{b,t}^F + C_{b,t}^V) \\ - C_{d,t}^I + \sum_{t \in [p, y_n]} (C_{d,t}^F + C_{d,t}^V) \end{aligned} \tag{3.1}$$

For a given demand growth scenario s,$O_t^s(p)$ in Eq. 3.1 defines the cost of deferring traditional network investments l by p periods by installing battery storage b at time t and subsequently upgrading the network (d) at period p. This refers to the option value of battery storage NWA for network deferrals. If $O_t(p) < 0$ then battery storage is a financially feasible NWA for a set of costs, deferral period p and planning horizon.

3.2.2 System design optimization

The system design optimization stage evaluates the cost-optimal location, sizing and dispatch of storage subject to operational constraints. This is achieved by formulating and solving a linear program for capacity expansion and dispatch [113] of a power system network [114] as described in Appendix B. The model objective is to minimize the total system cost which includes annualized resource expansion (generation, storage, networks) and, operational costs as described in Eq. B.1. The operational constraints are: 1) balance of system at the hourly level (Eq. B.2), 2)

time-dependent capacity constraints for generation resources (Eq. B.3), 3) battery storage state of charging, energy, power capacity limits and degradation (Eq. B.4 - B.9, Eq. B.1), 4) generation unit commitment (Eq. B.10 - B.14), 5) generation minimum and maximum power (Eq. B.15, B.16), 6) generation ramping limits (Eq. B.17, B.18), 7) direct current power flow approximation through line susceptance and voltage deviation (Eq. B.19), 8) network flow limits (Eq. B.20, B.21) and, 9) non-negativity constraints (Eq. B.22).

For the purpose of placing and sizing battery storage on the system in the context of distribution networks in India, we restrict the linear program of Appendix B in the following ways: a)we consider storage deployment exclusively at the feeder node of the network, b) we do not conside existing or investment in distributed generation in the network since the overall objective is to relieve congested lines in highly dense urban cities with minimal space for DERs like PV and c) we do not consider the battery storage's ability to inject power upstream since battery storage feed-in tariffs were not yet established in India when this study was carried out [115]. We assume that there is enough upstream generation (from the transmission system) to meet the demand on the feeder, hence the maximum available generation capacity is greater than peak demand. Additionally, the minimum upstream generation supply and ramping limits are set to meet the minimum demand and time step change in load. The optimization model has three investment variables, namely storage capacity, storage power, and network line capacity upgrades. Non-served energy is included in the objective function to allow for the possibility of feasible solution via load shedding. This is a single-stage optimization so the design of the system is based on the inputted annualized investment costs and demand for a given time period only. Under this formulation, the supply-demand balance of Eq. 3.1 is enforced while respecting storage capacity and network flow constraints. The optimization identifies the capacity of battery storage to be deployed, only if it is cost-optimal for the current stage and consistent with the valuation criteria defined in section 3.2.1. In particular, storage dispatch must adhere to network flow constraints both during charging and discharging periods. The model is formulated in Pyomo [116] and solved using the mathematical programming solver CPLEX [117].

3.2.3 Simulation

A Markov decision process (MDP) is a sequential decision problem for an observable and stochastic environment with a Markovian transition model and discounted rewards. It consists of a set of states, a set of actions, a transition model, and a reward function. The sequence of decisions in the distribution network planning problem

can be modeled as an MDP given the uncertainty in demand growth. Each state has a value that is calculated using the system design optimization described in Section 3.2.2. At each state there are two possible actions: traditional network upgrades or battery storage NWA upgrades. The transition matrix $P_{i,j}$ models conditional probabilities of growth in electricity demand from state i to state j. Policy iteration is a solution for MDP involving two steps: (1) estimating value function for a given policy, (2) using the estimated value function to find a better policy. Given possible recursion of the two-step process, the value function (Q) that is generated under a policy D, which maps every state to a decision is expressed in Eq. 3.2 where $F(s)$ is the cost resulting from the system design optimization at state s and decision variable x, $P(s, s')$ is the probability of transitioning from state s to s' and γ is a discount rate. The MDP iterates through various sequences of actions (policies) to find the highest reward value (as per Eq. 3.2) given a state transition matrix.

$$Q^D(s) = F(s) + \gamma \sum_{s'} P(s, s') Q^D(x, s') \tag{3.2}$$

To simplify the iterative process of the MDP, we restrict the action of battery storage NWA upgrade to future states in which $O_t^{s'}(p) < 0$, i.e. use of battery storage as NWA is justified (see Eq. 3.1). Otherwise, the simulation terminates with traditional network upgrade action as seen in Fig. 3-2. Therefore, a decision to take the battery storage NWA action is only possible when D of Eq. 3.3 is less than 0, which refers to the expected value of option across all states. The MDP can thus be translated to the flexible valuation framework as: $F(s)$ is the total system design cost at stage s (current stage) resulting from the system design optimization of section 3.2.2, $P(s, s')$ is the probability of transitioning from state s to s', D is the policy option between traditional network upgrade and battery storage NWA, x is the quantity of upgrades (traditional or storage) that is optimized using the system design optimization model described earlier (and available in Appendix B), and Q is the new system cost given x at stage s'.

$$D_s(p) = \sum_{s'} P_{s,s'} \cdot O_t^{s'}(p) \tag{3.3}$$

The MDP explores the multi-stage stochastic decision making process by iterating over a chain of policy decisions given the available options at each stage. In other words, this is a discounted sum of real options evaluated in a chain of decisions. At each stage, when the battery storage NWA option is feasible (Eq. 3.1) is satisfied),

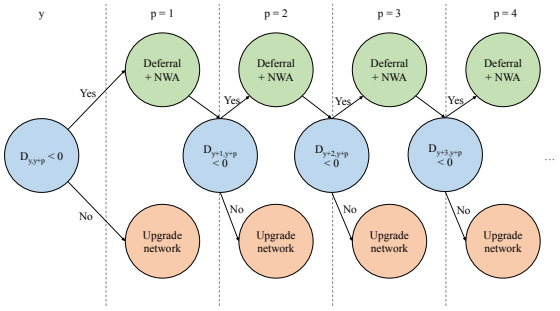

Figure 3-2: Flowchart of flexible valuation policy iteration.

the battery storage system is placed and sized through the system design optimization (see Appendix B). This process is repeated for the various trajectories. The resulting system cost of a policy D from s to s' under a transition probability $P(s, s')$ and decision x to add storage or upgrade network given the existing system cost F at stage s, populates the MDP Eq. 3.2. Note that a policy is only explored when Eq. 3.3 is satisfied, meaning that on expectation it is possible to place storage that satisfies the flexible valuation Eq. 3.1. If storage is not viable (i.e. Eq. 3.1 is not satisfied), the policy is therefore to expand the network which ends the MDP iteration. Each case takes less than 1 second to build and solve on a personal computer making it efficient to solve numerous times to get the total annualized investment cost for a stage s and plug it in as the value function $Q(s)$ of the MDP Eq. 3.2.

3.3 Delhi case-study input data

We demonstrate the value of the flexible valuation framework described above through a case study of network planning for megacities in India, such as Delhi, till 2040 under demand uncertainty.

3.3.1 Demand scenarios

The demand scenarios used as inputs to the network planning problem are adapted from the results of a previously documented demand forecasting model [3] that produces hourly resolved electricity demand data at the state-level for various technology and growth scenarios. From the dataset of scenarios developed previously [3], we select three scenarios to define low, medium and high demand outcomes: 1) high AC efficiency coupled with stable GDP growth, 2) baseline AC efficiency and stable

GDP growth and 3) high AC efficiency and rapid GDP growth. [1]. The high AC efficiency scenario translates into less electricity demand on a distribution network and is therefore considered the *low* growth scenario. The baseline efficiency under stable GDP growth is considered the *mid* growth scenario since it models a business-as-usual outcome. The *high* growth case is selected as rapid GDP growth, meaning strong economic growth and spending power, which leads to higher electricity demand growth but also higher AC efficiency since spending power is higher enabling stronger sales of efficient AC units.

Given the three electricity demand scenarios up to 2050 (low, mid, high), we use Monte Carlo Markov Chain (MCMC) to estimate the stationary distribution of demand growth in India by sampling from electricity demand consumption data from China. While China has achieved faster growth than India in the past four decades, as seen in Fig. 3-3, it is anticipated that India will experience high growth over the coming decades [27] . Moreover, Fig. 3-3 positions India two decades behind China in electricity consumption per capita as of 2019. India's projected electricity demand, primarily due to space cooling, is strongly compared to China's electricity consumption trend over the past two decades [25]. With India's consumption growth trend following China's, we use the MCMC to produce an estimate distribution of electricity demand growth by sampling from the Chinese electricity consumption data (Fig. 3-3) as the prior information on future electricity consumption in India.

MCMC simulations enable estimating a stationary posterior distribution. For every generated random value x, a transition kernel is used to assess the parameters of the desired distribution. The transition kernel is split into two steps: a proposal step and an acceptance/rejection step. Given the samples used to approximate the prior distribution, a proposal distribution, and an acceptance criteria, the MCMC iterates until the posterior distribution is stationary. The proposal distribution used in the simulation is a Gompertz distribution since Gompertz growth curves are often used to model long-term growth described as a sigmoid function which describes growth as being slowest at the start and end of a given time period [3, 118]. We define the acceptance criteria as the log error between the prior samples and the proposal distribution. Running the MCMC simulation yields a stationary distribution of demand

[1]The baseline AC efficiency scenario corresponds to electricity sales projections based on presently available AC units and a high-efficiency scenario that projects adoption of efficient AC units as defined by a recent study [25]. As of 2018, the sales-weighted average Seasonal Energy Efficiency Ratio (SEER) for AC units in India was 3 and the global average was 4 [25]. Under the baseline scenario, the gap between India's SEER and the global average is maintained. Under the high-efficiency scenario, India's SEER rating is projected to reach 8.5 by 2050 [25]

growth as seen in Fig. 3-4.

The result of the MCMC is a continuous distribution of projected electricity demand growth in India. We assume that every projected period of demand growth is an independent event that is sampled from the probability distribution derived by the MCMC. Therefore we construct a transition matrix of low, mid and high growth scenarios as defined in [3]. Low growth is assumed to be less than 5%, high growth is assumed to be larger than 8%, with mid growth referring to the in-between range. For two independent event with probability distribution, the joint probability distribution that falls within a range of values is a bivariate distribution. The bivariate distributions defined by the demand growth ranges are bucketed into a table. This table serves as the transition matrix in the MDP. Table 3.2 refers to the expected values of each bivariate distribution defined by the joint probability of column and row events. During the MDP, we sample from this table to simulate a chain of events.

Table 3.2: Proposed transition Matrix

	Low	**Mid**	**High**
Low	0.34	0.33	0.33
Mid	0.38	0.32	0.3
High	0.2	0.8	0

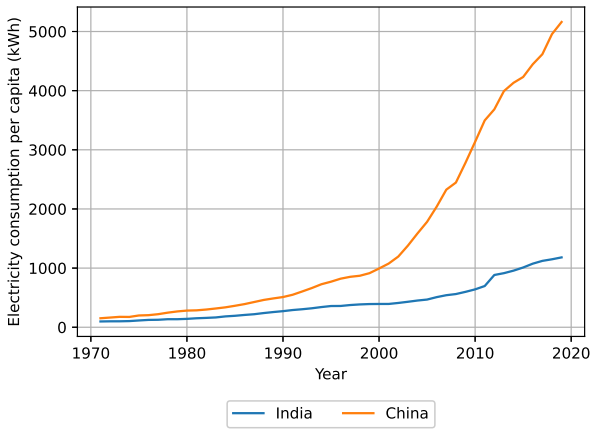

Figure 3-3: Historical electricity consumption per capita for India and China [1]. Per Capita Consumption is defined as gross electricity generation by all sources plus net import divided by mid year population and reported in kWh.

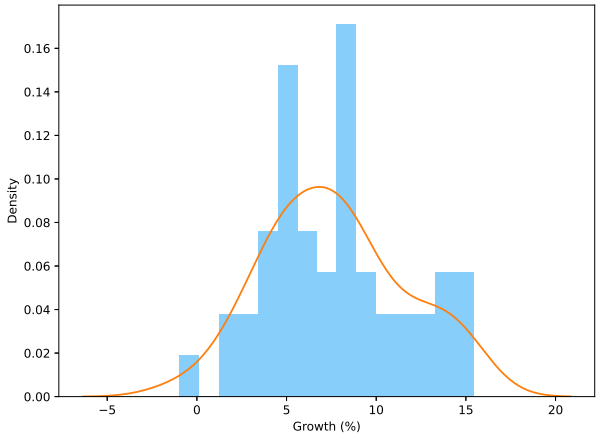

Figure 3-4: Stationary distribution result of MCMC simulation

3.3.2 Distribution network data

We apply the flexible valuation framework method to a benchmark medium voltage distribution network [119] and adapt the network equipment — line, transformer, voltage, current — to match the data of primary distribution (33 kV and 11 kV) networks operated by Tata Power Delhi Distribution Limited [2]. As an illustrative example, we model a 1 MW distribution network that is divided into three main feeders leaving the distribution substation: residential, commercial, and industrial as seen in Fig. 3-5. Using 2018 loading reports [2], the substation is initially loaded at 50% of the rated 1 MW capacity. We assume a loading limit of 90% of the rated capacity of the network, in this case, the loading limit is 900 kW.

3.3.3 Cost assumptions

Relevant cost inputs used in the modeling are presented in Table 3.3. Storage life is set to 15 years as per other storage assessment reports [5, 120, 6]. We model battery storage degradation (C^d in Eq. B.1) as a 1.46% [121] per annum energy capacity capital expenditure cost premium. Specifically, we assumed costs for battery storage NWA to be the same as transmission-level storage. As discussed later on, higher storage cost assumptions will reduce the option value of battery storage as an NWA. Since storage is an energy-dependent resource as described in the system design optimization (section 3.2.2 and Appendix B Eq. B.2), battery charging (ψ) will be from upstream generation which has a variable cost (C^V) which varies throughout the day as per Table 3.3. Therefore any variable cost incurred to charge the battery

storage is associated with the cost of battery storage NWA as per the system design optimization.

Table 3.3: Input cost assumptions for the model. Sources [5, 6]. Peaks hours are defined to be between 8 PM and 12 AM.

	2030	**2040**
Energy Cost (USD/kWh)	168	147
Power Cost (USD/kW)	146	128
O&M Cost (USD/kW-yr)	20	18
New line (USD/km)	350,000	350,000
Reconductoring (USD/km)	650,000	650,000
Off-peak Tariff (USD/MWh)	55	55
Peak Tariff (USD/MWh)	90	90
Discount rate γ (%)	9	9

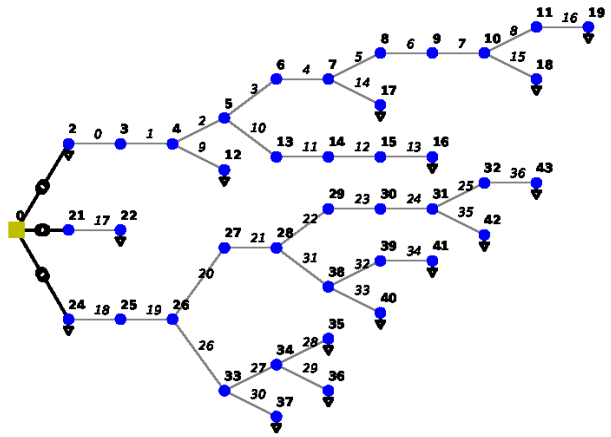

Figure 3-5: Sample distribution network diagram: three main feeders: commercial (top), industrial (middle), residential (bottom). Bus number in **bold** and line number in *italic*. Size, length, parameters and topologies vary across the library of feeders analyzed [2].

3.4 Results

We first present the result of a single policy iteration of the MDP for a demand growth trajectory that is sampled from the stationary transition matrix. The resulting model outcome will demonstrate the process through which storage is sized and the policy D is evaluated as per the flowchart of Fig. 3-2. Subsequently, we present the result of the full MDP, aggregated at the city-level for the four Indian megacities, Delhi, Mumbai, Bengaluru, and Kolkata. The input data that vary by the city are demand projections and distribution network characteristics sourced from [3, 35]. We evaluate battery storage NWA options for a deferral period of 5 years in the single network case. We further expand the deferral window to 10 years when we evaluate the Indian megacities cases.

3.4.1 Single-network case — Delhi

We detail the results of storage design on the distribution network for the demand trajectory based on the mid-level demand projection [3] in 2030. As previously mentioned in the input data, the initial feeder capacity is 1 MW and it is 50% loaded in 2018. The system design optimization yielded a solution of 1.5 MWh battery storage capacity with 380 kW power (5.5 hours duration) on bus 28 of the feeder (node on residential trunk feeder in the network as seen in Fig. 3-5). Note that although the system design optimization was constrained to only one location for storage deployment, it is possible to relax the problem to allow for multiple storage systems on the network may be cost-optimal. Fig. 3-6 presents the dispatch of the installed battery storage with respect to the substation's hourly load profile. From the perspective of the substation, the peak demand does not exceed 900 kW (loading capacity of the network) since the battery storage discharge is used to satisfy some of the demand during peak demand periods.

The optimization is repeated for the three possible demand trajectories (low, mid, high). The results are shown in Table 3.4. To further illustrate the results of the system design optimization, we solve two variants of the valuation framework: one with only traditional network upgrade and another with battery storage NWA + deferred investments. The results of the annualized investment cost for each demand projection's system design solution are shown in Table 3.5. We define the annualized investment cost (AIC) as the total cost of an option (CAPEX + OPEX) with annualization performed using a discount rate of 9% [122]. Battery storage NWA AIC includes the variable charging cost during off-peak hours. In the case where storage is not included in the optimization as a decision variable (traditional network upgrade

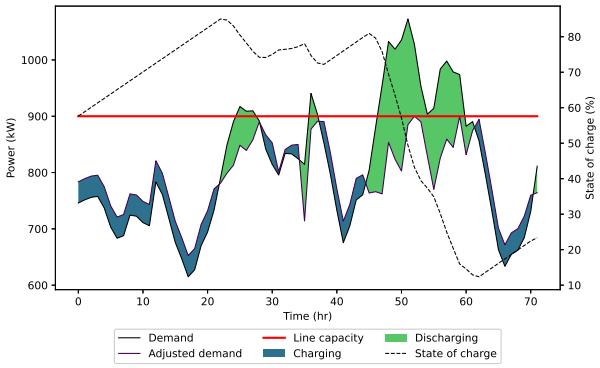

Figure 3-6: Hourly dispatch of NWA battery storage for one summer week load profile in 2030 for a distribution network from the city of Delhi. The demand scenario is based on the mid demand scenario.

only), the flow balance constraints (see Appendix B) result in 2,3 and 4 kilometers of line upgrades required under the low, mid and high demand growth scenarios, respectively.

Table 3.4: Optimized storage sizing in the single network case for Delhi and cost outcomes. Results correspond to three demand growth scenarios (low, mid, high) in 2030 for Delhi using sample network of Fig. 3-5 and cost assumptions of Table 3.3. CAPEX is the annualized capital cost of energy and charge, OPEX is the fixed and variable operation and maintenance costs for one year. VAR is charging cost of the battery storage system given the tariff schedule.

	Low	**Mid**	**High**
Power (kW)	300	380	420
Capacity (kWh)	1,200	1,520	1,680
CAPEX (USD)	8,358	10,586	11,701
OPEX (USD)	6,000	7,600	8,400
VAR (USD)	4,950	8,778	12,474

Using the transition matrix produced by the MCMC (Table 3.2), we run a large number of iterations of the MDP until the chain of policy is stationary. To further elaborate on the results, Table 3.6 highlights the various option costs O and thus possible policy decisions D (from Eq. 3.1 and 3.3). Note that the results of the high scenario are negative, that is also reflected in the AIC comparison of Table 3.5 where the battery storage NWA AIC under high growth is more expensive than

Table 3.5: Annualized investment cost (AIC) results for traditional network upgrades and NWA battery storage with deferred network upgrades for a deferral period $p = 5$. Results for case of single network in Delhi and three demand growth scenarios (low, mid, high) for 2030.

Traditional network upgrade AIC (USD)		
Low	Mid	High
14,673	22,009	29,345
NWA battery storage AIC (USD)		
Low	Mid	High
12,969	19,453	29,937

the traditional network upgrade AIC in 2030. In Table 3.6, we note that the option value is positive for the low and mid demand growth scenarios, implying that deferral of network upgrades by battery storage NWA is cheaper than traditional network upgrades on expectation (Table 3.5). On the other hand, given high demand growth, it is cheaper to upgrade the network immediately. The flexibility of battery storage NWA allows the utility to adopt a "wait and see" strategy and benefit from lower than anticipated growth to defer upgrades. For this reason, the option value of low projection is the highest, and the high one is the lowest (negative) as seen in Table 3.6.

Table 3.6: MDP simulation results for the case of single network in Delhi across all transitions in Table 3.2 for three demand growth scenarios (low, mid, high) for 2030. Results are annualized USD values. Negative values indicate that NWA battery storage is more expensive than traditional upgrades and vice versa.

	NWA battery storage AIC			**Policy value**
	Low	Mid	High	
Low	(1,361)	36,328	74,018	35,692
Mid	36,328	15,864	5,632	20,679
High	(32,057)	5,632	43,322	(24,519)

Iterating through the various policies D at various stages and exploring the available flexible options O yields a stationary chain of decisions given various demand growth trajectories. Fig. 3-7 illustrates the final result of the full MDP given the expected demand growth trajectory that is simulated from the transition matrix. In the earlier periods of the modeling, there is a large uncertainty on the level of demand in the final

period (2050). Therefore, when a policy needs to be chosen (2030) due to network overload, the MDP places battery storage as opposed to upgrading the network. In later periods, the uncertainty of demand in 2050 yields an early retirement of the battery storage system and a traditional network upgrade that satisfies the demand growth until the end of the simulation period (2045).

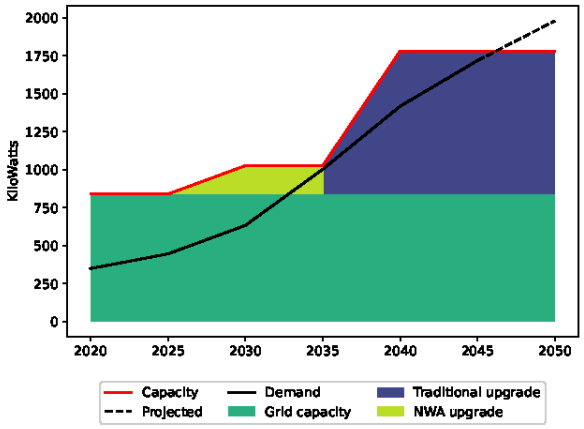

Figure 3-7: Simulation of investment outcomes for single network case in Delhi, shown in Fig. 3-2, with simulated demand trajectory from the transition matrix. $y = 2020$, $Y = 20$, four timesteps p of 5 years intervals as per the input projected data [3].

The early retirement of storage is taken into consideration when it is installed in 2030 since the policy would not have been an option if Eq. 3.1 was not satisfied. The option value differs based on technology costs, transition probabilities, demand growth, and desired deferral periods. Evidently, a higher cost of storage yields results that favor traditional network upgrades. On the other hand, lower network reconductoring and new line costs will also favor traditional network upgrades. For the particular case of Delhi, storage costs were chosen from mid-range projections of their respective periods [5] and the network upgrade costs were collected from benchmark surveys as well as historic upgrade costs of local distribution utility [2, 6].

3.4.2 City-wide network simulation

We apply the flexible valuation framework to estimate battery storage NWA across select megacities in India (Bengaluru, Delhi, Kolkata, Mumbai) that collectively accounted for 52 TWh of electricity consumption in 2018 with an estimated 72,763

circuit kilometers of distribution lines at 33 and 11 kV serving dense urban areas [123, 124] (see Table 3.7). We apply the flexible valuation framework to four cities that vary by population size, circuit kilometer length, load profiles, and location. For each city, we apply the above-described method to study nine representative feeders identified. The representative feeders were identified based on clustering of the library of urban feeders (and their respective hourly demand profiles) for Delhi provided by TPDDL [2]. Each representative feeder is characterized by:

1. Loading percentage varying from 40 to 80% of ampere capacity [2]
2. Represented demand, defined as the hourly load profile modeled on the feeder, which varies by megacity according to available survey data [35]
3. Serviced demand, defined as the total annual demand (MWh) that the distribution network feeders service with the same loading percentage
4. Serviced circuit kilometers, defined as the total circuit kilometers (km) that are at the corresponding loading percentage

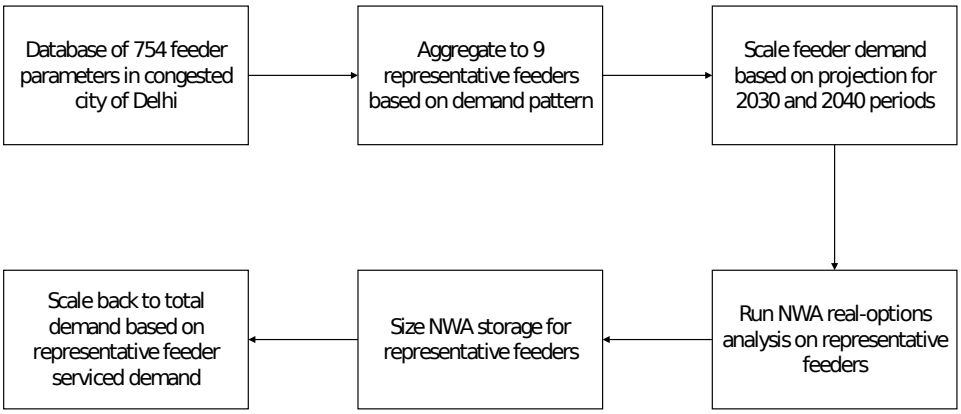

Figure 3-8: Approach for computing megacity-level NWA battery storage potential in Indian megacities.

We assume a similar distribution in feeder loading for the other megacities as indi-

cated by the data for Delhi. We estimate the circuit kilometers that each of the nine feeders represents in every megacity based on the respective data on the serviced demand and the calculated ratio of serviced demand to circuit kilometers available for these feeders in Delhi. The flexible valuation framework is applied on all the representative feeders of the four selected megacities by using the appropriate demand projections in 2030 and 2040 [3] and the same transition matrix (Table 3.2). Network investment costs are calculated based on the circuit kilometer length of each representative feeder in each megacity. The resulting battery storage NWA capacity for the representative feeders is scaled using the calculated ratio of each feeder's serviced demand to represented demand. To explore the impact of storage technology cost on the flexible valuation framework, we consider alternative cost trajectories for battery storage [5] as well as calculate the cost of storage where battery storage NWA is no longer competitive with traditional network upgrades (referred as the break-even storage capital cost). Finally, as noted earlier, since electricity is mostly contracted in Delhi and other cities in India, we do not consider the value of energy arbitrage where battery storage NWA can sell excess energy back to the grid since as of 2020, there was no tariff schedule to accommodate such transactions.

Table 3.7: Flexible valuation framework results for megacity-level NWA battery storage analysis under mid-range storage capital cost projections

	Year	**Bengaluru**	**Delhi**	**Kolkata**	**Mumbai**
Demand (TWh)	2018	10	23	4	15
NWA battery storage (GWh)	2030	3	14	1	11
	2040	15	50	35	40
Overloaded lines (km)	2030	1,265	6,093	792	12,224
	2040	1,467	7,070	919	14,184

Under the mid-growth scenario, we estimate that 20,373 km of 72,763 km of the four megacities' distribution networks will be overloaded by 2030. Under the same demand growth projection, an additional 23,640 km will be overloaded by 2040 [3, 123]. Applying the flexible valuation framework to the representative feeders and scaling the total demand each feeder represents, we estimate that 29 and 140 GWh of battery storage NWA could be cost-effectively deployed across the four megacities in 2030 and 2040, respectively (see Table 3.7). This would defer 15,914 km of network upgrades for 2030 and an additional 18,127 km for 2040. Table 3.8 highlights two option costs: (1) flexible budget through storage and deferred upgrades: total annualized cost of storage installed on the feeders and annualized deferred network upgrade costs (after battery storage NWA use is exhausted) and (2) the total an-

nualized traditional network upgrade cost. Table 3.8 highlight the total budgets of the flexible valuation framework and traditional network upgrades across the four megacities. Total budget is calculated by summing the annualized investment cost of each year over a 30 years horizon. For traditional network investment, the total budget is the annualized investment cost of network upgrades times 30. For the flexible valuation framework, the total budget is the annualized investment cost of battery storage NWA for the deferral period and the annualized investment cost of deferred network investment for the rest of the planning horizon. Deploying battery storage NWA before traditional network upgrades produces capital cost savings of 16 and 15 % in 2030 and 2040, respectively, on a total budget basis. More battery storage NWA is deployed per unit kilometer in 2040 than in 2030 due to the expected increase in the concentration of load during the peak hours as space cooling drives electricity demand growth [3]. Battery storage NWA is carried over from one stage to the next (stored in the value of F of Eq. 3.2) existing system cost and remains available as long as it is dispatchable and network upgrades can be deferred. The resulting useful life of battery storage NWA ranges between 5 and 10 years.

Table 3.8: Flexible valuation framework total results of megacity-level NWA battery storage analysis. Initial investment in 2020 with 10 year deferral period for periods 2030 and 2040. Distribution network line useful life is set to 30 years and battery storage useful life is set to 15 years.

Cost (in Millions of 2020 USD)	**2030**	**2040**
Annualized storage cost	\$207	\$261
Annualized deferred upgrades costs	\$76	\$136
Annualized traditional upgrades costs	\$117	\$133
Total flexible budget	\$2,932	\$5,324
Total traditional budget	\$3,503	\$6,266

We evaluate the flexible valuation framework under the low and high cost storage scenarios (Table 3.9). The result of battery storage NWA and deferred upgrades is the same in the low-cost and the mid-cost scenarios, which indicates that the binding constraint is dispatch — i.e. the availability of off-peak network capacity throughout the day to charge the battery storage for peak hours discharge. This finding suggests that battery storage NWA may not be viable for the networks that are initially heavily loaded. Under the high storage cost scenario, we estimate that cost-effective battery storage NWA deployment would defer 11,752 km and 13,717 km of network upgrades in 2030 and 2040, respectively. Consequently, capital cost savings drop to 12% and 10% for the respective periods. Not surprisingly, the higher cost of storage

implies less economic battery storage NWA. We further increase the cost of storage energy to 261 USD/kWh and power to 227 USD/kW and find that battery storage is eliminated and all overloaded lines are traditionally upgraded without any deferrals.

Table 3.9: Storage cost impact on outputs of the flexible valuation framework applied to the four Indian megacities, for year 2030. Low, mid and high storage capital cost assumptions are sourced from [5]. Breakeven costs are 261 USD/kWh and 227 USD/kW for energy and power respectively.

	Low	**Mid**	**High**
Energy (USD/kWh)	116	168	236
Power (USD/kW)	101	146	205
NWA battery storage (GWh)	29	29	18
Deferred lines (km)	15,914	15,914	11,752

As detailed earlier, battery storage NWA is driven by capital investment savings for utilities rather than the competitiveness of storage as a resource. The attractiveness of battery storage NWA has a proliferation potential in network-constrained environments where utilities have short-term financial commitments. Our results show that up to 29 GWh of battery storage capacity can serve as NWA to shift up to 7 GW of peak demand for a total of 140 hours in 2030 and up to 35 GW of peak demand for a total of 183 hours in 2040. This indicates that up to 338 and 741 GWh of peak electricity consumption can be shifted in 2030 and 2040 respectively. India's total electricity demand is projected to be 2.3 and 3.5 TWh with 347 and 626 GW of peak demand in 2030 and 2040 respectively, under the mid-range growth scenario [3]. If adopted at scale, the load-shifting potential of battery storage NWA can impact the dispatch of generators on the bulk-power system. Specifically, the battery must be charged in the day, and in EMDE where coal [9] is the dominant baseload generation, the long-term cost and environmental benefits may not outweigh the short-term cost benefits at the distribution level.

3.5 Discussion

Utilities in EMDE are primarily concerned with capital allocation owing to the high cost of financing. In the case of network equipment, we define the capital utilization rate (CUR) as the ratio of equipment loading in a given period $W_{t,y}$ to network capacity M_t. Based on the MDP simulation for mid-growth demand trajectory of the single network case in Delhi, we estimate a higher CUR for the storage and deferred

network investment(59%) as opposed to the case of traditional network investment (53%). Additionally, utilities in EMDE face shorter-term financial commitments [101] due to a lack of long-term loan availability. Improving CUR will therefore serve utilities better to recover their investment and fulfill their financial commitments.

The flexible valuation framework presents an approach that combines system design optimization with multi-stage decision making under uncertainty for distribution network planning. The strategy assesses the feasibility of short-duration battery storage as an alternative to network upgrades given the uncertainty in demand growth. The simulation detailed for a single distribution network in the Delhi case shows that storage can shave the peak demand and thus prevent the network from overloading, which enables the deferral of the lumpy network upgrades to future periods when capital is cheaper (discounted) and uncertainty is lower. We compute the scaled-up effects of this strategy by applying it to the case of distribution networks across four Indian megacities, that result in an estimated 29 GWh and 140 GWh of storage capacity deployment in 2030 and 2040, respectively. We find that under reasonable cost assumptions for battery storage, high uncertainty of demand growth and high cost of capital, installing battery storage NWA and deferring traditional networks to the future is cost-effective. The flexible valuation framework enables utilities to adopt a *wait and see* strategy with smaller initial investment costs when there is high uncertainty about future demand growth.

Our framework solves for electricity resource design of a distribution network problem by considering the exogenous uncertainty of demand growth and battery storage technology cost. We further expand our assessment of the impact of uncertainty on the electricity resource design problem to supply-constrained case where battery storage non-wire alternatives are of minimal use since it is an energy-dependent resource.

Chapter 4

Impact of uncertainty at distribution level — case study of Abuja

Question 2: *How do these uncertain factors impact electric grid planning at the distribution stage?*

This chapter further extends the application of the real options framework developed in Chapter 3 to the supply-constrained case of distribution planning in Abuja, Nigeria. Previously, the real options framework was applied to the network-constrained cases in megacities in India. The framework perceives demand growth as an exogenous uncertainty and plans the network design through battery storage non-wire alternative. In the case presented in this chapter, the real options framework is extended to an under-the-grid minigrid design with upstream grid supply availability as an exogenous uncertainty.

Poor reliability of power supply to the end customers connected to the last mile of the main distribution grid is incentivizing the deployment of local reinforcements known as "under-the-grid minigrids". Such minigrids are emerging to provide power to commercial and industrial customers, to boost electricity access and economic growth. However, improved grid supply prematurely terminates the financial viability of a minigrid. This paper presents a minigrid design method under grid supply uncertainty to ensure electricity access for commercial and industrial customers whose objective is minimum net present cost and financial viability for the minigrid developer. Furthermore, we consider the uncertainty through simulated scenarios of grid-related outages that have historically pushed consumers to resort to diesel generators for power. Finally, our approach evaluates the trade-off between diesel and

battery storage to mitigate power outages and ensure minimum cost supply.

4.1 Introduction

Minigrids have been considered as one of the solutions to achieve universal electricity access along with the main grid expansion and solar home systems in developing countries. While grid expansion has been considered as the premium option, some end-customers connected to the last mile of the distribution network experience poor reliability of power supply. This is mainly caused by the insufficient grid generation capacity or the lack of adequate transmission and distribution infrastructure. This state of affairs in the grid infrastructure in developing countries stems from a cumulative funding gap resulting from utilities operating under financial duress [125]. Furthermore, considering demand growth, an extra burden is placed on the already constrained grid infrastructure. In this situation, under-the-grid minigrids may serve as a complement to the grid to ensure sustainable and reliable power supply to these customers.

While the "under-the-grid minigrids" concept is not new, the research on the design of such minigrids has been focusing on minigrids interconnected to reliable grids [126, 127, 128, 129, 130, 131]. In [126], the focus was on storing the energy during low grid energy prices and discharging during the peak demand through an economic optimization technique for battery management to reduce the operating cost of a grid-connected microgrid. In [127] a multi-microgrid interconnection scheme to address power fluctuation caused by PV integration in the distribution system is proposed by optimizing the capacity of the energy storage system as well. In [128] a model for microgrid planning under uncertainty/errors in energy demand, generation, and grid energy price forecasts was proposed. As in [126] there is no consideration of uncertainty in grid power reliability. [129] investigates a unit commitment problem based on the cost-benefit analysis and here-and-now approach for optimal battery storage system sizing in microgrids considering wind power stochastic behavior. The scheduling of DGs is done based on the maximization of minigrid's total revenue in grid-connected mode by exchanging power with the grid and minimization of minigrid's total costs in off-grid operation mode. In [130] a microgrid planning model is developed considering various incentives such as renewable energy investment-based incentives, tax benefits, and grid ancillary services. This study focuses on the impact of financial incentives on microgrid design. In [131] provides a sizing algorithm that transforms an existing distribution network into a sustainable autonomous feeder-based microgrid. The focus of the study was to improve the voltage profile and

reduction of losses using DG integration and network reconfiguration.

While a great deal of research has been undertaken on the optimal design of minigrids that can operate in grid-connected mode, particularly in developed countries, the motivation is more on (i) reducing grid energy costs through self-consumption, (ii) improving grid power quality such as voltage profile and reduction of losses, or (iii) ensuring continuous power supply when the main grid fails following a high-impact-low-probability event such as storm, earthquake, etc. Unlike, the above cases, this chapter deals with grid-connected minigrids, in Emerging Market and Developing Economy countries, with the motivation of providing reliable power supply to grid-connected customers with poor power supply reliability. Designing such a minigrid requires the consideration of the uncertainty of grid power availability because of the randomness of power outages. We incorporate uncertainty in grid supply availability through reporting the mode of various scenarios of a deterministic optimization.

The scope of this chapter will focus on the commercial minigrid case study in the city of Abuja in Nigeria. The minigrid serves the Wuse Market traders which have small shops in the open-air market space in the Wuse neighborhood of Nigeria. The main grid supply to the market is unreliable, inconsistent, and of low voltage quality. First, we examine the commercial opportunity for an interconnected minigrid developer when the customers are willing to pay for better service in the market. Then, we evaluate the future scenarios of least-cost electricity generation design for the minigrid under grid-supply uncertainty. Additionally, we quantify the role of battery storage to further proliferate solar PV as a clean distributed energy resource (DER) instead of traditional backup diesel generation. Our analysis considers interconnected minigrids in the context of the supply-constrained environment of Abuja, where the uncertainty regarding the duration and timing of grid availability to meet demand is a key driver for storage use. Finally, we investigate hybrid generation designs under various grid outage simulations to assess the interconnected minigrid model in a supply-constrained environment such as Nigeria, where a failure to meet load at the distribution level is generally due to under-supply at the bulk power system level.

The rest of the chapter is organized as follows: section 4.2 presents the situation in the Wuse Market in Abuja, Nigeria. Section 4.3 details the least-cost optimization tool used to solve the electricity dispatch of the interconnected minigrid given grid supply uncertainty. Section 4.4 presents the results of the interconnected minigrid strategy with further discussion in section 4.5.

4.2 Wuse Market case-study

The Nigerian distribution utility Abuja Electric Distribution Company (AEDC) recently introduced a Distributed Energy Solutions and Strategy for AEDC (DESSA) to provide reliable power supply at a lower cost than the power from diesel generators to willing customers by efficiently combining local backup provided by a third party and grid service – when available [132]. Under the DESSA program, the regulator permits the distributor to contract with a third party to supply power to its customers within a portion of its service territory for an agreed portion of the day at a tariff negotiated separately between those customers and the third-party supplier subject to regulatory approval. At other times of the day, the distributor, AEDC, supplies the same customers at the regulated tariff and must compensate the third party if the grid supply is unavailable. Depending on the predictability and duration of outages, and the temporal character of the load, the third party generally seeks to design a hybrid generator that minimizes the sum of his fixed and operating costs. This generally involves a combination of PV with battery storage, plus a diesel generator. Under present conditions, the third-party supplier is charged a distributed use of system (DUOS) fee for using the grid network infrastructure. AEDC does not currently have a program to buy excess generation from third-party developers.

Figure 4-1: GIS and low flow result of distribution network serving the Wuse Market (highlighted in red).

The Wuse Market is an open-air merchandise and food market home to over 2,155 small businesses as seen in Fig. 4-1. The 2019 peak demand was 993 kW occurring between the hours of 1 pm and 4 pm and a modest nighttime load when the market is closed, provided by a shared cold storage room. The annual electricity generation supplied by the grid to Wuse Market was 10% of the total annual demand of the market in 2018 [8]. This indicates the poor reliability of grid power supply to the

market. The interconnected minigrid developer invested 2 million USD in capital expenditure with less than 70% of the cost attributed to the generation assets: solar PV, batteries, power converters, and a diesel generator). The developer plans to install 1 MW of solar PV nameplate capacity (installed on the rooftop of the market and potentially the parking), 1.2 MWh of lithium-ion short-duration battery storage, and a 1 MW backup diesel generator to serve the Market. This interconnected minigrid project is expected to allow the market to extend its operational hours from 6 PM to 9 PM and eliminate the use of over 3,000 small gasoline and diesel generators connected directly to the shops [48]. This will significantly increase productivity and reduce the air and noise pollution in the market.

The minigrid developer will lease the distribution assets from AEDC under a Distribution Use of System (DUOS) payment of 12 Naira per kWh (0.03 USD/kWh). AEDC collects a DUOS payment for every kWh the minigrid developer supplies the market. The minigrid is connected to the grid at the 11 kV primary distribution network of AEDC and can therefore purchase electricity from the utility. Under the terms of the agreement [48], AEDC must supply 2,177 MWh of electricity during the priority hours (7 to 10 AM and 4 to 8 PM) of the day. In other words, AEDC is obligated to supply 60% of the priority hours electricity demand of the market. The negotiated tariff that the minigrid developer pays AEDC to purchase electricity ranges between 0.13 and 0.17 USD/kWh. If AEDC fails to meet its supply requirement, the developer is requested to compensate for the generation shortage without getting charged the DUOS fee.

Under the Tripartite Agreement [133], a 20-year agreement has been signed between the three parties: AEDC, minigrid developer, and the Wuse Market Traders Association. The minigrid developer assumes AEDC's full license obligations for the Wuse market retail electricity sales. The minigrid developer also assumes responsibility for the construction of the minigrid system as well as its operation and maintenance. Prepaid smart meters will be used to guarantee payment to the developer so there is no risk in payment collection. Presently, the Wuse Market traders are paying a price of 0.38 USD/kWh for electricity. Under the Tripartite Agreement, that price is expected to drop. Furthermore, AEDC estimates that its total revenue from the Wuse market will go up by 70% [48]. However, with upstream grid-supply uncertainty AEDC will find itself cutting its revenue since it has to waive the DUOS charge. Additionally, the minigrid developer will require more diesel backup generation that is costlier than the wholesale electricity from the grid. Under such conditions, the traders of the Wuse Market may find themselves served by unreliable supply if the minigrid is inadequately designed.

Table 4.1: Definition of frequent, infrequent, long, and short outages through number of yearly occurrences and duration of outages for simulated cases based on discussion with Abuja Electric Distribution Company

Count (per year)	**2020**	**2025**	**2030**
Frequent	100 - 200	50 - 100	0
Infrequent	50 - 100	0 - 50	0
Duration (hours per year)	2020	2025	2030
Long	24 - 12	12 - 6	0
Short	6 - 3	3 - 0	0

Table 4.2: Capital cost assumptions for various resources. All costs in 2018 dollars, unless otherwise noted. Solar costs assume DC to AC ratio of 1.34 [5, 7].

	2020	**2025**	**2030**
PV ($/kW AC)	1,354	1,095	836
Li-ion storage - energy ($/kWh)	299	206	168
Li-ion storage - power ($/kW AC)	260	179	146
Diesel generator ($/kW)	400	400	400

Taking into consideration the existing plans for the interconnected minigrid, we evaluate the system's dispatch behavior and stress-test it against various unscheduled grid outage scenarios to identify the change in the minigrid's operation (see Table 4.1). We further relax the design of the system to identify a greenfield cost-optimal strategy that takes into consideration the environment of the Wuse Market under the Nigerian Electric Regulatory Commission (NERC). We use the same negotiated tariffs [8, 48], local diesel fuel cost of (2.2 USD/gallon) [134] and diesel generation cost per kW from an open-source catalog of components [7]. Solar PV and lithium-ion battery storage costs figures are extracted from the National Renewable Energy Laboratory's Annual Technology Baseline report [5] under mid-range cost projections (see Tables 4.2 and 4.3). We further investigate a minigrid system design under a low-cost storage scenario as defined in Table 4.4. The electricity demand of Wuse Market is expected to grow by 5% annually as per AEDC [8].

Table 4.3: Operation and maintenance cost assumptions for various resources. All costs in 2018 dollars, unless otherwise noted. Solar costs assume DC to AC ratio of 1.34 and diesel generator heat rate is fixed to 0.85 gallons per hours per kW [8, 7, 5].

	2020	**2025**	**2030**
PV ($/kW-yr AC)	16	13	10
Storage ($/kW-yr)	36	25	20
Diesel generator ($/kW-yr)	100	100	100
Diesel fuel cost ($/gal)	2.2	2.2	2.2
AEDC rate ($/kWh)	0.15	0.17	0.19
DUOS ($/kWh)	0.03	0.04	0.05

Table 4.4: Low cost storage scenario CAPEX and OPEX costs projections [5].

	2020	**2025**	**2030**
Energy CAPEX ($/kWh)	247	160	116
Power CAPEX ($/kW AC)	215	139	101
OPEX ($/kW-yr)	30	20	14

4.3 Optimization framework

An optimization model evaluates the cost-optimal design and dispatch of the generation mix subject to operational constraints. This is achieved by formulating and solving a linear program to solve for capacity expansion and dispatch [113] of a power system network [114] as described in Appendix B [135]. The model objective is to minimize the total system cost which includes annualized resource expansion (generation, storage, networks) and, operational costs as described in Eq. B.1. The operational constraints are: 1) balance of system at the hourly level (Eq. B.2), 2) time-dependent capacity constraints for generation resources (Eq. B.3), 3) battery storage state of charging, energy, power capacity limits and degradation (Eq. B.4 - B.9, Eq. B.1), 4) generation unit commitment (Eq. B.10 - B.14), 5) generation minimum and maximum power (Eq. B.15, B.16), 6) generation ramping limits (Eq. B.17, B.18), 7) direct current power flow approximation through line susceptance and voltage deviation (Eq. B.19), 8) network flow limits (Eq. B.20, B.21) and, 9) non-negativity constraints (Eq. B.22).

We include the following charges in the model as variable costs: a) DUOS charge, referring to the cost the developer pays the utility when using the distribution network on a kWh basis, and b) developer tariff, which refers to the price at which

the developer buys power from the utility on a kWh basis to either supply the consumers or charge the battery storage system. Grid availability is constructed as an hourly time series by including a given randomly simulated frequency and duration of outages. For each investment period, the solution of the resulting linear program outputs a scenario-based analysis of deterministic solution given the prescribed grid availability.

This optimization jointly solves for the demand balance as well as generation design and dispatch to form an under-the-grid minigrid. To explore the evolution of the minigrid generation expansion planning over time, a multi-period deterministic optimization is carried out over three investment periods – 2020, 2025, and 2030 - whereby the design from one period (2020) is used as the initial condition for the next period (2025) with the ability to add or retire capacity as required to meet the load specified for that period. To take into consideration the uncertainty in grid supply availability, we initially solve for the design of the minigrid without any unscheduled outages to find the initial design. Then, we use the initial design in two sets of new optimizations with varying time windows of dispatch. The first set considers the initial design as an upper bound to the design variables and the a lower bound in the second set. Considering each full year dispatch, we perform a Monte Carlo simulation. In the simulation, we slice the time series into windows ranging between 4 to 12 hours (minimum to maximum battery storage visibility) and roll the time window for each investment period. At each windows, we collect two designs that are minimally and maximally bounded by the initial design. The collected results form a scenario-based analysis of the optimization to assess the impact of unscheduled grid supply outages on the initial design. We report the mode of the collected results.

4.4 Results

We present the results of the multi-stage optimization for the design and dispatch of the Wuse market minigrid. Demand growth projection is set to 5% annually as per AEDC's projections [8], we assume a similar growth rate to AEDC's grid supply commitment to the Wuse Market constrained to the same priority hours (7 to 10 AM and 4 to 8 PM). There are two sets of results: the 2020 initial design constrained to the existing system as described in section 4.2 and an alternative least-cost design that is unconstrained by existing assets. Both sets of results include additional generation expansion plans under least-cost economic dispatch for future periods (2025, 2030), given demand growth. Additionally, both sets of results include AEDC's scheduled grid outages in the 2020 and 2050 periods as per section 4.2.

To study the impact of uncertain grid outages that are unscheduled, a series of unscheduled outages is generated by performing a Monte Carlo simulation using the ensemble of grid availability profiles described in Table 4.1. We report the most frequently occurring system design for the ensemble. All cost results are reported in Appendix C.

4.4.1 Existing system design

The existing system is fixed to a 1 MW solar nameplate capacity, 1 MW diesel generator, a 1.2 MWh battery storage system, and a daily grid supply of 2.2 MWh during the priority hours. Fig. 4-2. Additionally, an economic dispatch plot for three days in Fig. 4-3. Diesel generation is the dominant resource of choice complemented by grid supply from AEDC. Since solar is fixed to 1 MW, solar generation is dispatched to meet as much as half of the daytime peak demand. Storage remains a secondary resource that is charged by grid supply during the nighttime hours (when upstream generation is least constrained).

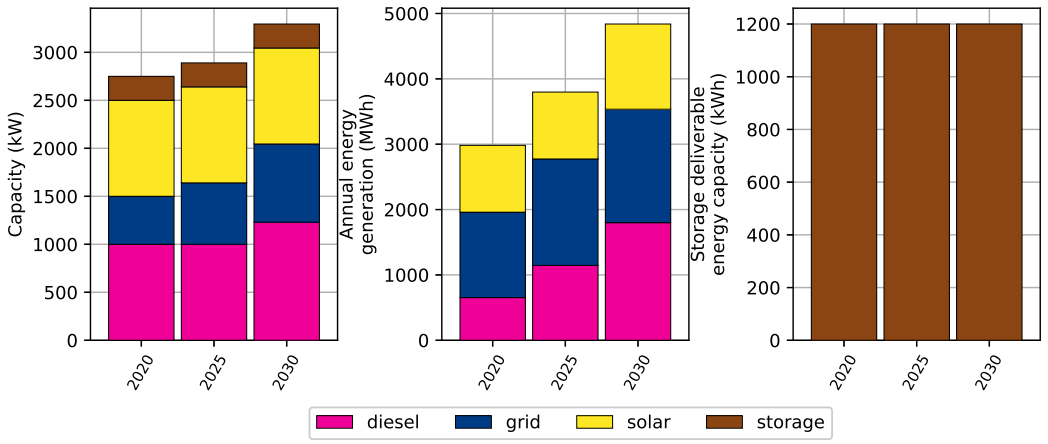

Figure 4-2: Existing system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages.

Under 5% annual demand growth and grid supply generation availability, the existing system design does not need additional assets until 2030. In 2025, the excess demand from the previous period is fully supplied by diesel generation and additional grid supply. Furthermore, in 2030, demand growth will necessitate an additional 230 kW generation capacity according to the least-cost system design with mid-range battery storage costs. The design resulting from the optimization model is the same for both

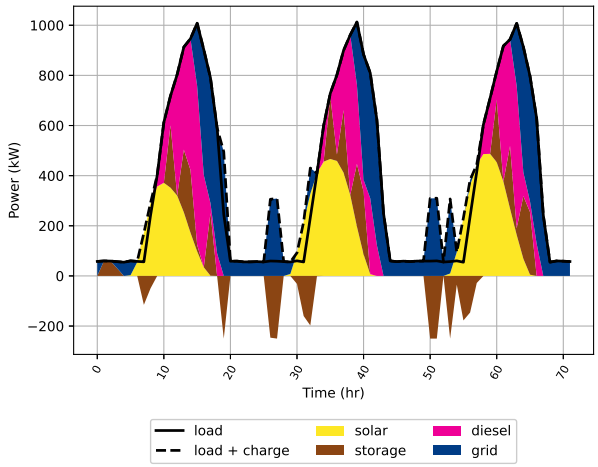

Figure 4-3: Existing system minigrid dispatch for year 2020 demand without simulated grid outage.

mid-range and low-cost storage projections as per Table 4.4. The CAPEX annuity is 286,000 USD given a 9% discount rate and diesel fuel costs of 192,000 USD for the modeled the year 2020.

Under the existing system design in 2020, modeled unscheduled upstream outages do not impact the dispatch and there is no non-served energy as defined in Eq. B.2. There is enough diesel capacity to meet any scheduled outage at any time in the day. Even in the 2025 modeled period under various unscheduled grid outage simulations, diesel generation is available to meet the demand. Storage does not play a role and is not considered in future optimal design given the large existing diesel generator (1 MW) that can replace grid supply outage.

4.4.2 Alternative design

We present alternative designs that are not constrained by the existing system as reported in [48]. All the alternative designs are greenfield meaning they are not constrained by existing infrastructure. We present six different greenfield design cases with simulated outages as per Table 4.1, technology cost structures as per Tables 4.2 and 4.3, and low cost battery storage costs as per Table 4.4: 1) reference case greenfield system (GF), 2) reference case greenfield system under infrequent and short grid outages (GF O1), 3) reference case greenfield system under frequent and long grid outages (GF O2), 4) low-cost storage greenfield system (GF LS), 5) low-

cost storage greenfield system under infrequent and short grid outages (GF LS O1), and 6) low-cost storage greenfield system under frequent and long grid outages (GF LS O2). These six cases are compared to the existing system (ES) case on a design, dispatch, and cost object basis. The reference case assumes no unscheduled grid outages and therefore is optimized to be the greenfield least-cost economic dispatch. We simulate grid outages given the range of frequency and duration of outages as per Table 4.1. We report the best-case (short-infrequent) and the worst-case (long-frequent) scenarios of grid outages. Furthermore, we evaluate the role of battery storage by considering low-cost storage projections [5] as per Table 4.4.

The reference case (GF) yields 1.5 MW of solar nameplate capacity, an 800 kWh short-duration battery storage system, and a 600 kW diesel generator as seen in Fig. 4-4. While capital expenditure is higher in GF than it is in ES, the total system cost is cheaper due to lower operational and fuel costs (see Tables C.1, C.2, C.3, C.4). Fig. 4-5 shows the minigrid dispatch to meet the demand of the Wuse Market. The main difference in dispatches between GF and ES in 2020 suggests that the expansion of solar PV can reduce dependence on diesel generation to meet the daytime load of the Wuse Market.

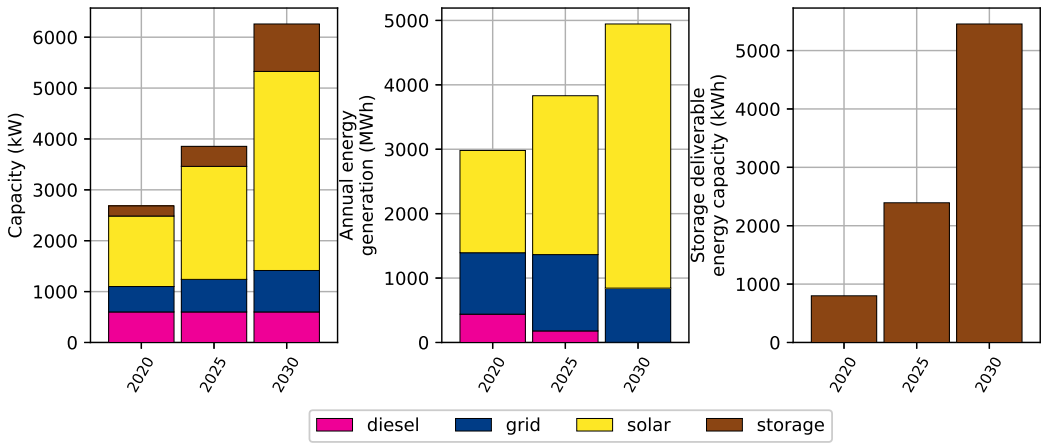

Figure 4-4: Greenfield system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages.

Fig. 4-4 illustrates a growing investment in battery storage from 2020 to 2030 due to expanded solar capacity and grid reliability. Specifically, 2.3 MWh of battery storage is installed on the system in 2025 as opposed to ES where 1.2 MWh is installed in 2020. Battery storage plays two roles in the scenario-based analysis

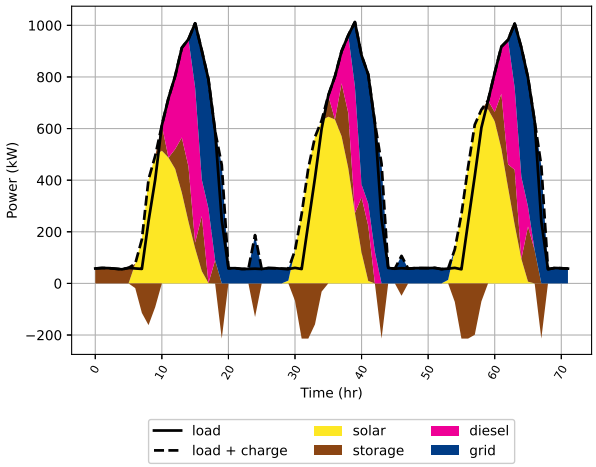

Figure 4-5: Greenfield system minigrid dispatch for year 2020 demand without simulated grid outage. Time starts at midnight on the first day.

of the deterministic optimization: complementing solar PV and arbitrage with the grid supply when electricity prices of AEDC vary between 0.13 USD/kWh and 0.17 USD/kWh as mentioned in section 4.2. With grid outages being less frequent in the future modeled periods 2025 and 2030, battery storage becomes even more attractive because of 1) declining technology cost and 2) the reliability of the grid to charge the batteries. In GF, diesel generation is reduced over time and is only occasionally used during peak hours of low solar potential and grid unavailability (Fig. 4-6).

Under low-cost storage (Table 4.4), the economic design and dispatch results of the case (GF LS) are similar to the reference case (GF) as seen in Fig. 4-7. Declining costs of storage further reduce dependence on diesel generation starting in 2025. Moreover, while grid supply capacity is projected to expand in 2025 and 2030, low-cost storage with solar PV becomes more competitive than the AEDC tariff schedule which pushes the Wuse Market interconnected minigrid towards islanded operation mode.

To understand how grid outages impact the role of storage, we investigate the system sizing and operation with uncertain grid supply, i.e., when generation outages are not scheduled. We randomly simulate a variety of annual frequency and duration of outages as per Table 4.1. The best-case scenario is infrequent-short duration outages and the worst-case scenario is frequent-long duration outages. We report results given reference case costs and low-cost storage costs. The four sets of results: reference greenfield system infrequent-short outages (GF O1), reference greenfield

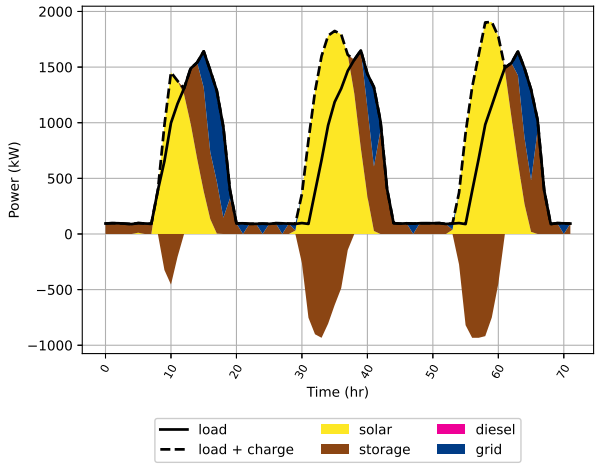

Figure 4-6: Greenfield system minigrid dispatch for year 2030 demand without simulated grid outage.

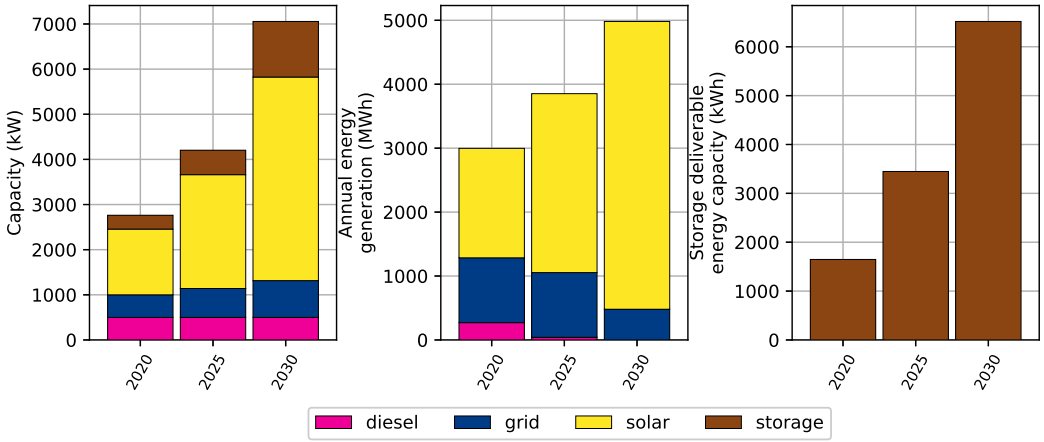

Figure 4-7: Greenfield system design with scenario-based analysis of the deterministic least-cost future design for the Wuse Market minigrid without simulated outages under low-cost storage assumptions.

system frequent-long outages (GF O2), low-cost storage greenfield system infrequent-short outages (GF LS O1) and, low-cost storage greenfield system frequent-long outages (GF LS O2) are presented in Fig. 4-8 and 4-9.

In all cases, unscheduled grid outages will require further diesel capacity as a backup generator. This is evident due to the low capital cost nature of diesel generators.

Infrequent outages have the strongest impact on storage since the battery management system is not optimized to handle the variability of dispatch requirements due to grid unavailability. Regardless of the cost level, storage capacity is cut in half and the generation is compensated by diesel capacity (Fig. 4-8 and 4-9). Therefore, storage capacity has an inverse relationship with the frequency of unscheduled grid outages. However, in the cases where outages are the longest, the system tends to shift towards off-grid operation and start to rely more storage capacity powered by solar PV. A short-duration battery storage system cannot create value from energy arbitrage due to prolonged grid outages but further pushes the minigrid to the islanded mode of operation. The effects are similar but of smaller magnitude in the low-cost storage cases.

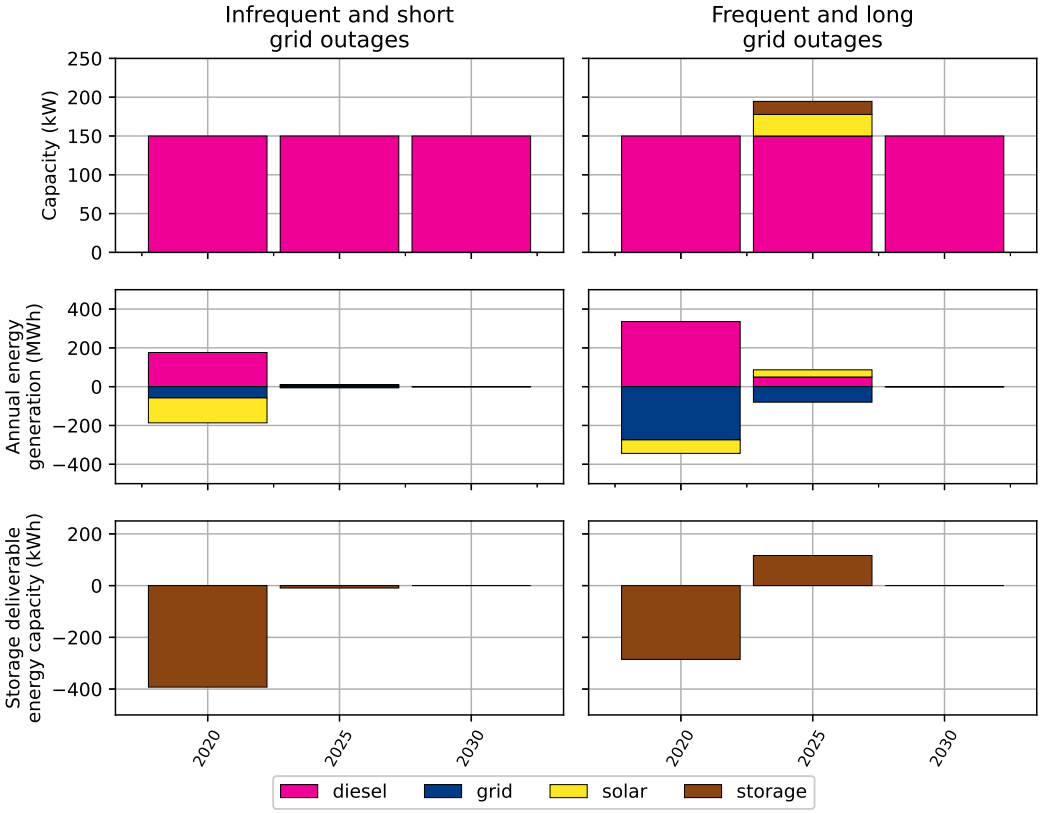

Figure 4-8: Difference plot of system design given unscheduled long duration infrequent grid outages.

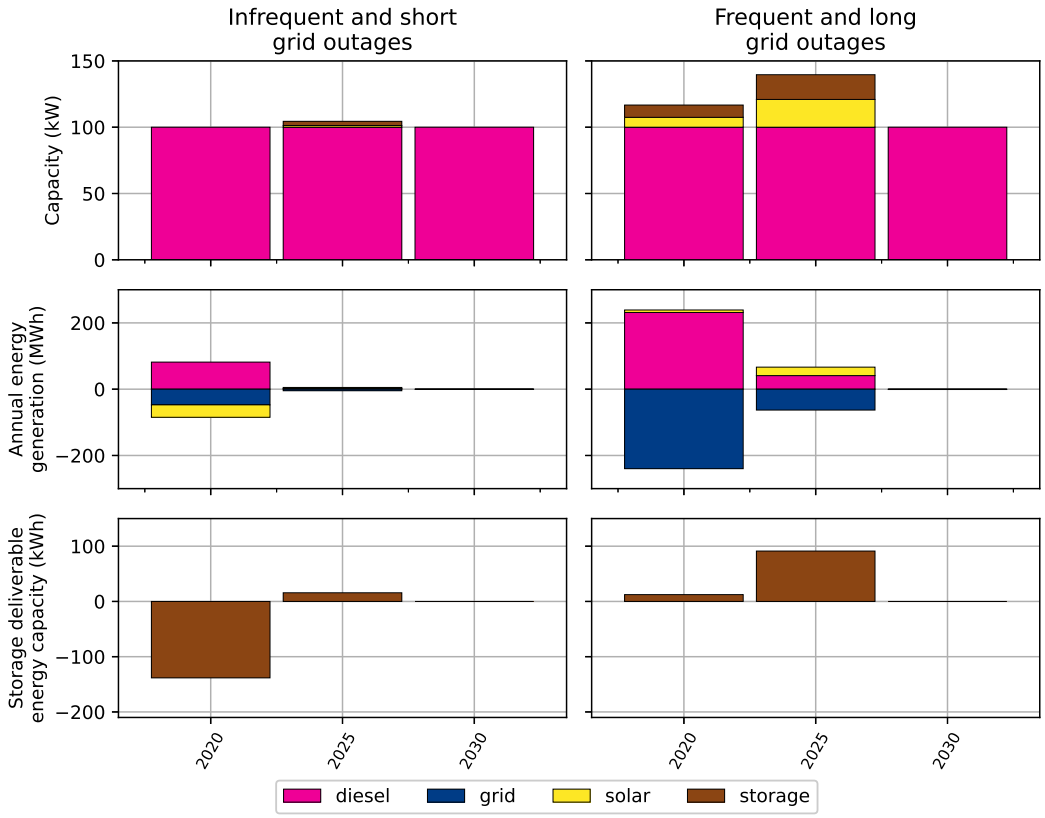

Figure 4-9: Difference plot of system design given unscheduled long duration infrequent grid outages under low-cost storage assumptions.

4.5 Discussion

We presented three cases of design and dispatch solutions for the Wuse Market interconnected minigrid:

- Existing system (ES)
- Greenfield system (GF)
- Greenfield low-cost storage system (GF LS)

We further expanded on the greenfield system cases to consider unscheduled grid outages: infrequent-short (O1), frequent-long (O2). No grid outage simulation cases were expanded on the ES case because 1 MW is initially placed on the system which can compensate for any type of grid outage in both periods 2020 and 2030.

Table 4.5: Total discounted cost, system cost of electricity per period and discounted results from optimization of the seven scenarios given a discount rate of 9% and total demand for the Wuse Market of 2,950 MWh in 2020 with 5% annual demand growth for periods 2025 and 2030.

	Total Cost	**SCOE (USD/kWh)**			**Total SCOE**
		2020	**2025**	**2030**	
ES	$ 7,308,640	$ 0.23	$ 0.11	$ 0.08	$ 0.41
GF	$ 5,745,664	$ 0.21	$ 0.08	$ 0.04	$ 0.34
GF O1	$ 5,857,772	$ 0.21	$ 0.09	$ 0.04	$ 0.35
GF O2	$ 5,932,908	$ 0.22	$ 0.09	$ 0.04	$ 0.35
GF LS	$ 5,229,396	$ 0.21	$ 0.07	$ 0.03	$ 0.32
GF LS O1	$ 5,283,547	$ 0.21	$ 0.08	$ 0.03	$ 0.32
GF LS O2	$ 5,364,154	$ 0.22	$ 0.07	$ 0.03	$ 0.32

$$\text{SCOE} = \frac{\text{Total annualized cost}}{\text{Total demand}} \tag{4.1}$$

The existing system dramatically reduces the system cost of the Wuse Market from the blended system cost of electricity (SCOE) of 0.38 USD/kWh [48] to 0.23 USD/kWh (Eq. 4.1) in 2020, as seen in Table 4.5. However, under the existing system, there would be no incentive to expand solar PV or storage capacity in future periods. 36% of the Wuse Market interconnected minigrid demand in 2030 would be diesel generation.

The greenfield set of results presented under the different cost of storage and grid outage simulations have a lower system cost of electricity as per Table 4.5. We define Total SCOE as the discounted sum of the SCOE per modeled period given a discount rate (9%). Table 4.5 shows that while SCOE in 2020 between the various cases is similar, there is a large gap in subsequent periods. The gap in SCOE can be explained by looking at the breakdown of costs of the SCOE: capital cost in Table C.1, fuel costs in Table C.2, fixed O&M costs in Table C.3 and variable O&M costs in Table C.4 in Appendix C present the least-cost solution for every case.

Interconnected minigrids present a better solution than the existing status quo at the Wuse Market in Abuja Nigeria. However, it is difficult to assess that the Tripartite Agreement is a win-win-win scenario long term. While in the short-term the Wuse Market traders are benefiting from reduced costs, there is little to suggest that cost would decline in the future. The interconnected minigrid developer invested in a

large diesel generator that alone can serve the peak demand of the minigrid and therefore has little incentive to invest in alternative generation (solar PV + storage). Diesel generation complements AEDC's grid supply without competing with it. The role of storage and expanded solar PV in interconnected minigrids is limited from the perspective of the minigrid developer and AEDC.

While solar PV and storage do not compete with backup diesel generation that complements the grid, they are cost-competitive with the grid itself. Regulation in Nigeria and many EMDE countries is not expanded when it comes to DERs, but we show that storage complemented by solar PV at both reference and low-cost figures can eliminate diesel generation, reduce dependence on upstream generation and provide a lower system cost of electricity to the Wuse Market minigrid developer. This presents an opportunity for AEDC to further expand its non-wire alternative network upgrade strategy to include battery storage as an option and increase VRE penetration at the distribution level. Without a reliable and cheap grid supply, interconnected minigrids are a win-win-semi-win situation with the latter being the Wuse Market traders. The objective is clean and reliable access to electricity and under present circumstances islanded minigrids are a cheaper solution than interconnected minigrids from the perspective of the customer.

The results reported in this chapter are particular to the cost structure of the Wuse Market case study in Abuja, Nigeria. Moreover the results are reported from a scenario-based analysis of a deterministic optimization to emulate the reality of dispatch behavior. We use such a methodology and report the mode of the various designs in an effort to emulate the reality of dispatch behavior where there may be foresight on the grid supply availability for the next several hours (4 to 12 hours considered). Additionally no constraints on the environmental impact of diesel has been applied which can increase the variable cost of the generators and alter the design to favor solar PV and battery storage. Finally, the methodology of scenario-based deterministic optimization is not only applicable to grid supply availability in this peculiar case of minigrid-under-the-grid design. Such an optimization-based methodology can be applied to various decision-making under uncertainty situations where the uncertainty is exogenous and can be predicted to a certain extend (e.g., foresight over the next several hours of dispatch).

At this point in the thesis, we have explored two distribution-level electricity resource design cases under endogenous uncertainties of demand growth, technology cost, and grid supply availability. The peculiar situations in India (network-constrained) and Nigeria (supply-constrained) led to developing a multi-stage optimization framework using real options to solve for a long-term least-cost electricity resource design of the

distribution stage of the electric grid. We further expand this uncertainty assessment to the generation and transmission level with application to the Indian bulk power system. The subsequent chapter highlights the range of findings from scenario-based analysis of the deterministic optimization and sheds light on the scale of the problem of multi-stage electricity resource design under uncertainty.

Chapter 5

Impact of uncertainty at generation and transmission levels – case study of India

Question 2: *How do these uncertain factors impact electric grid planning at the generation and transmission stages?*

After investigating the impact of uncertainty on distribution planning in chapters 3 and 4, this chapter extends the assessment of exogenous uncertainty on electric grid planning to the generation and transmission stages. We use the regional level demand scenarios produced in Chapter 2 and assess the potential pathways of the Indian bulk power system with consideration of high penetration of distribution-level storage, as presented in Chapter 3. Finally, this chapter exposes the computational complexity of solving granular electricity resource optimizations when considering exogenous uncertainty.

Global energy sector decarbonization efforts are contingent on technology choices for energy production and end-use in Emerging Market and Developing Economy (EMDE) countries such as India (e.g., electric cooking, space cooling, and electric vehicles). Here, we use an integrated demand-supply framework to quantify the impacts of demand growth and temporal patterns on long-term electricity system evolution. Under projected renewables and Li-ion storage cost declines, our supply-demand modeling points to renewables contributing substantially (46-67%) to meet annual electricity demand in India by 2030. However, without appropriate policy measures to phase out existing coal generation, even such rapid adoption of renewable

energy coupled with one or more technological levers such as low-cost energy storage and demand-side measures such as setting aggressive AC efficiency standards and deploying distribution-level storage are insufficient to reduce annual CO_2 emissions in 2050 vs. 2020 because of the relatively higher growth rate of projected electricity demand over this period.

5.1 Introduction

Electricity generation in EMDE countries over the next few decades will significantly impact global greenhouse gas (GHG) emissions as access and economic development increase electricity demand in these regions[99, 100, 32]. Notably, because many of these countries are located in hot climate zones, they are expected to see growing energy demand for space cooling, which in many cases, is likely to be exacerbated by climate change impacts [3, 136, 137]. For EMDE countries as a whole, more investigative planning of energy infrastructure from both the demand [138, 139] and supply [140] perspective is warranted to ensure efficient use of limited capital and alignment with global mid-century climate mitigation goals. Of these countries, India stands out since it already ranked 3^{rd} in terms of CO_2 emissions in 2018 [24], owing to its large population and reliance on coal for primary energy (44% of primary energy demand in 2019 [10]), and in particular for electricity generation (72% of supply [10]). Yet, these national statistics belies the lower annual per-capita primary energy consumption in India (23 million btu (MMBtu)) compared to other high-income countries like United States (310 MMBtu) and Germany (165 MMBtu) in 2018 [10]. By one estimate, growth-driven energy consumption could result in India's final energy use in 2040 being 81% greater than in 2019 [10], with demand for electricity growing much faster in this scenario than other forms of energy, at 161% [10]. While decadal electricity demand growth projections for developed countries such as the U.S. is driven primarily by the electrification of transportation [141], in India and many other EMDE countries with hot climates, the building sector is projected to dominate electricity demand growth over this period, primarily due to the widespread adoption of air conditioning (AC) systems [10]. As compared to other new sources of demand (e.g., EVs), the relative inflexibility and timing of AC use means that it will not only increase aggregate demand but also change the temporal load shape and impact peak consumption. For example, a recent study estimates that space cooling could contribute as much as 45% of peak electricity demand in India by 2050 compared to 10% in 2016 [25]. In the short-term, increases in peak demand, which tends to occur after sunset [142], will likely be met with relatively high emissions intensity coal-based electricity generation [143], owing to its dominant share of supply

today. Assessing pathways for grid decarbonization in the Indian context and other similar regions, therefore requires a granular study of the temporal patterns of AC demand in conjunction with dynamics of electricity generation.

Recently, several studies have analyzed the operation and long-term evolution of India's bulk power system between 2030 and 2050 at different levels of granularity in representing grid operations, existing generation, evolution of demand profile, and investments in new generation, storage and transmission [144, 145, 142, 9, 146, 147, 148, 149, 150]. Some studies model grid operations for various generation capacity scenarios in 2030 to quantify the operational feasibility of different levels of variable renewable energy (VRE) penetration and the flexibility provided by coal and hydro generation as well as new battery storage to integrate VRE generation [144, 146, 142]. Other studies model the long-term evolution of India's electrical grid (to 2050) [151, 9, 147, 145] subject to approximations regarding spatial and temporal variations in demand and VRE resource characterization and its impact on capacity investment. Notably, the temporal resolution of grid operation in these studies range from a few time periods ($<$50) per year at a state-level spatial resolution [9, 151] to hourly operations at the regional level [147, 145]. In addition, some studies use simplified modeling based on a single resource profile per region [147] while others use detailed representations of VRE resource availability that may include land availability [145] and transmission interconnection costs [9]. Some studies also model investment over multiple periods and the temporal evolution of the power system from 2020 to 2050 [151, 9]. A key finding across many of these studies is the cost-effectiveness of VRE generation deployment in the future Indian electricity system ($>$50% as much as 80% of annual generation [145]). However, none of these investment planning studies account for the structural changes in the electricity consumption profile for electricity use over time, resulting from factors such as AC adoption, and to a lesser extent EV adoption. Therefore, these studies present an incomplete picture of long-term evolution of the power system in India.

Our contribution improves upon prior work by developing a holistic framework for assessing the impact of supply *and* demand-side drivers on the long-term evolution of the power sector in India and other EMDE countries. This supply-demand interaction is addressed by combining bottom-up demand forecasting with high temporal resolution capacity expansion modeling (CEM) that uses high spatial resolution VRE resource availability and detailed representation of hourly grid operations [41]. The bottom-up demand forecasting model, documented in detail in Chapter 2 and [3], captures the growth of business-as-usual components as well as new components, namely AC and EVs, in estimating electricity demand at the state-level in future

years at an hourly resolution. This granularity enables us to explore the system impact of demand-side interventions, such as improved AC efficiency standards, alternative EV charging schedules as well as the potential impact of distribution level energy storage (DLS) deployment to manage congestion in the local distribution system. The resulting regional demand profiles are subsequently used as inputs to a multi-period power system CEM that considers grid operations at an hourly resolution, to evaluate the least-cost trajectory of power system investment and operation in India from 2020 to 2050. This framework, (see Chapter 2 Fig. 2-4), is used to address the following questions in this study: a) How do various demand-side drivers (AC, EV load growth) impact the evolution of India's power system in terms of generation capacity mix and CO_2 emissions when factoring interactions with long-term supply-side factors such as natural gas (NG) prices, VRE availability and energy storage capital costs? and b) How does AC demand growth impact the need for energy storage, at distribution and transmission levels, and both existing and new coal generation under various technology and policy scenarios?

5.2 Scenario-based transmission modeling

5.2.1 Demand-side scenario model

The alternative electricity demand scenarios evaluated here are developed using a previously documented open-source model [3] that uses separate approaches to estimate future electricity demand for existing end-uses ("business-as-usual" model) as well as demands from emerging end-uses such as ACs and EVs ("technology model"). Electricity demand from existing end-uses is estimated for future periods using a regression model that is trained on historical regional electricity demand available for 2012-2019[76] at the daily resolution and hourly demand for 2015 [147]. In addition, this model incorporates weather data at daily resolution and GDP forecasts at monthly resolution to incorporate seasonal trends and long-term growth respectively.

The technology model enables a bottom-up approach to estimate demand from new loads, which in this study relates to space cooling in residential and commercial buildings as well as EV charging [3]. The model relies on AC sales data projection as well as types of units being sold to meet the expected space cooling demand. Two AC scenarios were considered: a baseline scenario with electricity sales projections based on currently available AC units and a high efficiency scenario that assumes preferential adoption of efficient AC units as defined by a recent study [25], which considers a scenario where the global average Seasonal Energy Efficiency Ratio

(SEER) rating of ACs grows from 4 in 2018 to 8.5 in 2050. AC efficiency, as reflected in SEER ratings, may differ greatly between the United States and India due to the types of AC units installed. While heating, ventilation, and air conditioning (HVAC) systems use efficient cooling methods such as variable refrigeration in the U.S. context, their system cost is high for the Indian market where less efficient split units are expected to be installed. Although high AC efficiency scenario is defined here based on improvements in SEER ratings in the Indian context, it can also be viewed as the outcome of other building sector interventions like passive cooling that reduce overall electricity demand to achieve a similar level of thermal comfort as in the baseline scenario. As of 2018, by comparison, the sales-weighted average SEER for ACs in India was 3 [25]. Residential and commercial AC demand growth was estimated at the state level and then aggregated to the regional level [3] to be input to the supply-side optimization model. For EVs, the technology model uses vehicle sales data and government goals for EV sales targets in future years [73] to estimate EV charging demand. Additionally, hourly projections of EV charging demand at the regional level were derived for each decade after applying a 1D-convolution to survey data related to typical charging patterns in an EMDE countries settings [3]. As compared to AC demand, electricity demand from EV charging is projected to be relatively modest, both in terms of annual consumption and in terms of contribution to peak demand, as seen in Table 5.1. The reference electricity demand projection for our analysis is estimated assuming stable GDP growth, baseline AC efficiency, evening EV charging scheme. Evening EV charging is predominant in other EMDE countries such as Mexico [94] and therefore is chosen as the schedule for the reference case.

5.2.2 Supply-side optimization model

We use a multi-period version of the power system CEM [148, 150, 152], GenX [41] to evaluate the least-cost investment and operation of the Indian power system under alternate technology, demand and policy scenarios. The GenX model is open-source [113].

For this study, GenX is configured as a multi-period investment planning model with four investment periods (2020, 2030, 2040, 2050) and hourly representation of grid operations. We solve the resulting linear programming (LP) model using a dual dynamic programming (DDP) algorithm that makes this problem computationally tractable by decomposing the problem into individual sub-problems per investment period and subsequently uses information from solution of the model in future investment periods ("Forward Pass") to adjust investment decisions in previous periods

Table 5.1: Demand estimates from bottom-up forecasting model [3] that is used as inputs to the supply-side modeling.

		2020	**2030**	**2040**	**2050**
Reference case	Peak demand (GW)	197	347	626	901
	Annual demand (TWh)	1,421	2,282	3,523	4,773
High AC efficiency case	Peak demand (GW)	197	317	501	677
	Annual demand (TWh)	1,421	2,207	3,205	4,199
DLS case	Peak demand (GW)	197	341	600	901
	Annual demand (TWh)	1,421	2,282	3,523	4,773
EV charging case	Peak demand (GW)	197	345	624	897
	Annual demand (TWh)	1,421	2,282	3,523	4,773
AC contribution to peak demand	Baseline efficiency	4%	15%	32%	42%
	High efficiency	4%	10%	17%	19%
EV contribution to peak demand	Evening charging	1%	4%	6%	10%
	Morning charging	1%	3%	5%	9%

("Backward Pass") [153]. For each investment period, the model includes the following grid operations constraints: a) flexibility limits of thermal power plant operations via linearized unit commitment constraints [154, 155], b) supply-demand balance at each hourly time step and each zone, with power flow associated with linear losses and transfer capacity limits between zones, c) modeling hydro power plant operation to adhere to available information on inflows and reservoir capacity [147] and d) modeling other storage resources with inter-temporal storage balance constraints as well as capacity constraints on maximum rate of charging and discharging.

These operational constraints are modeled over 20 representative weeks of grid operation, that are selected from an single year of load data based on 2015 weather patterns, VRE and hydro resource profiles (more details discussed later on) via k-means based clustering [20, 156]. The operations over the 20 representative weeks are scaled up to estimate annual operation cost and other operational metrics of interest such as VRE curtailment and CO_2 emissions. The choice of 20 representative weeks was made to balance accuracy of capturing intra-annual variability in load and VRE, hydro resource availability as well as computational run times for the multi-period CEM (see Fig. D-12). Additionally, Fig. D-13 highlights the relative error in capacity outcomes compared to the model with 20 representative weeks as the baseline.

We represent the Indian grid using five separate balancing regions (North, West,

South, East and Northeast) defined by the grid operator [104], with region-specific load profiles developed for each investment period based on the above-mentioned demand-forecasting model [3] - example outcomes are shown in Chapter 2 Fig. 2-1. The power flows between these regions are modeled based on a simplified network representation that enforces power exchange limits between the regions (see Table D.7). For 2020, these power limits are derived from the grid operator [147, 104]. These limits may be expanded with additional transmission investment in future periods.

5.2.3 Resource cost and performance assumptions

GenX models operations over four periods, 2020, 2030, 2040 and 2050, with investment in new resources (nuclear, VRE, coal, NG combined cycle gas turbine (CCGT), NG open cycle gas turbine (OCGT), battery energy storage and transmission) considered in the last three periods. The key cost assumptions of the generation and storage resources are summarized in Table 5.2, where we specifically account for technology investment costs specific to the Indian context. For example, based on data from IEA [10], we derate U.S. centric capital cost projections [5] of wind and solar by approximately 70% and 50% to account for historical differences in capital cost between U.S. and India. We characterize 2020 power generation capacity as well as their operational flexibility based on the documentation of the Regional Energy Deployment Model-India [9] as well as prior studies [147, 157]. For simplicity, we do not distinguish between the operational characteristics (e.g. minimum stable power and heat rate) of supercritical and subcritical coal power plant resources within a zone, but distinguish between heat rates of existing thermal power plants across zones [157]. Major system assumptions including fuel costs and value of lost load are denoted in Table D.5. Existing hydro power plants are classified as either reservoirs plants that can flexibly adjust their output vs. run-of-river resource that do not have any flexibility in their output and hence are treated as must-run resources [157]. The hourly inflows, reservoir capacity for hydro power generation are derived from a prior study [147]. The model incorporates both lifetime-based and economic-based retirement of generation and storage resources. For existing resources, particularly coal and NG, we estimate cumulative lifetime based retirements by 2030, 2040 and 2050 by zone that represent a minimum amount of capacity to be retired by those time frames, based on data from [9, 157]. Because of the assumption of perfect foresight of future technology cost, VRE resource availability, demand, and policy, the model strategically may choose to retire more than the prescribed minimum capacity if it can lead to reduction in the total system cost over the modeling horizon.

5.2.4 Renewable-resource supply curves

Similar to other power system planning studies [158, 9, 159], GenX uses supply curves to model the investment in VRE resources that account for variation in the VRE resource in terms of resource quality, interconnection cost and total deployable capacity within each zone. This supply curve is developed starting from the spatially-resolved (10 km^2) wind and solar resource data for 2014, available from the Renewable Energy Potential Model (reV) [12] using a sequence of steps, described in detail in Fig. D-14. The resulting spatially-resolved capacity potential for wind and solar PV are illustrated in Fig. D-15, where each site is associated with a unique interconnection cost and hourly capacity factor (CF) profile. To simplify the representation of resource variation in GenX, we aggregate PV and wind resource in each zone into 3 bins that are generated by clustering sites (using k-means) based on their levelized cost of energy (LCOE). Here, the LCOE for each site is computed using site-specific CF and interconnection costs as well as capital costs and Fixed O&M costs. The resulting parameter inputs for each resource bin per zone in the GenX model [157] include: a) hourly CF, computed as the weighted average CF for sites within each bin, where the weights correspond to the available area for VRE deployment associated with that site, b) total available capacity per bin, computed based on $32MW/km^2$ and $4MW/km^2$ for spatial density of solar and wind resources respectively [9] and c) weighted average annualized interconnection cost associated with each bin.

In addition, we also impose installation limits for total wind and solar PV capacity deployed per investment period to account for potential constraints owing to supply-chain and labor resource limits. These constraints are derived based on fitting a Gompertz function growth curve to trends in VRE capacity deployments seen in the Chinese context (see Table D.12). Table D.13 highlights the imposed decadal VRE installation limits. Fig. D-1 demonstrates the impact of alternative assumptions about the installation limits (0.5X of the reference case and no installation limits) on generation capacity. Fig. D-1 highlights that installation limits primarily impact VRE deployment, mainly wind, in 2030 and 2040 but are less impactful in 2050 in the reference case, where value decline in VRE generation is major driver for capacity installation decisions. Further details are provided in Appendix D section D.

5.2.5 System cost of electricity generation expansion

System average cost of electricity (SCOE), defined in Eq. D.1 is often used to to quantify the cost impacts of various technology and policy drivers. In the context of

multi-period investment planning model, we define SCOE for each modeled period as the ratio of total annual system cost for the year divided by the total demand served in that modeled period. Total annual system cost includes operating cost, both fixed and variable, and annualized investment cost for the period. The latter includes investment cost of: a) resources deployed in the current period and b) resource invested in prior periods that have not yet reached their modeled lifetime and hence are accruing fixed costs related to their investment. We do not include any investment costs associated with existing generation or transmission assets as of 2020, but consider fixed operating costs for existing generation. Operational costs are calculated for the model period only based on scaling up the hourly operation costs for the modeled 20 representative weeks using their hourly weights. Since total system cost does not include unpaid investments costs of existing generation and transmission, the SCOE estimated in this study are not reflective of electricity prices for a given case but are indicative of the cost of generation expansion and are thus used for comparison of case results. Complete SCOE formulas are provided in the Supporting Information (SI).

5.3 Results

5.3.1 Reference case

We first evaluate model outcomes for a reference case that provides an internally consistent point of comparison to explore the impact of alternative technological and policy assumptions. For this study, the reference case is defined based on (see Table 5.3): a) electricity demand projections using baseline AC efficiency and evening EV charging assumptions [3], b) technology capital cost following the "reference" trajectory, adapted from [5], in (see Table 5.2), c) decadal VRE installation limits derived from VRE installation trends in the Chinese context (see Methods and Table D.13) d) NG prices held constant at $11/MMBtu throughout the model horizon and e) no carbon policy.

In the reference case, we find 362 GW of VRE capacity by 2030, corresponding to an annual average installation rate that is 3.7 times the average capacity additions in 2010-2019 (see Fig. D-1). Due to disparities in VRE resource quality and land availability, VRE generation capacity is predominantly deployed in the Southern and Western regions (see Fig. D-2). This is accompanied by deployment of 57 GW of new coal capacity by 2030, that along with available thermal and hydro resources, is operated flexibly to integrate the installed VRE generation with <5% VRE cur-

Table 5.2: Capital cost assumptions for various resources. All costs in 2018 dollars and sourced from NREL annual technology baseline 2020 [9], unless otherwise noted. Wind, solar PV and gas generation capital costs have been de-rated by 72%, 51%, and 70% respectively to account for estimated capital cost differences for these resources between U.S. and India, as per central technology cost for 2019 from [10]. Solar costs assume DC to AC ratio of 1.34[9]

Resource & Units	**Scenario**	**Capital Costs**		
		2030	**2040**	**2050**
PV ($/kW AC)	Reference	558	407	369
Wind ($/kW AC)	Reference	995	843	754
Li-ion storage - energy ($/kWh)	Reference	206	168	136
	Low cost	160	105	82
Li-ion storage - power ($/kW AC)	Reference	179	137	119
	Low cost	139	92	72
CCGT ($/kW)	Reference	706	675	655
OCGT ($/kW)	Reference	647	616	598
Nuclear [19] ($/kW)	Reference		2,800	
Coal [10] ($/kW)	Reference		1,200	
Biomass [19] ($/kW)	Reference		864	
Inter-regional transmission ($/MW-km)[9]	Reference		312	

tailment (see Fig. D-1), as illustrated in Fig. 5-2 and Fig. D-3. New coal capacity is predominantly installed in later periods and is concentrated in the Northern region, which has the second highest electricity demand in 2020 (427 TWh) and a 5.7% projected growth rate (see Fig. D-2). Consequently, by 2050, the Northern region holds 46% of the national coal capacity in the reference case. The role for new thermal generation, mostly coal, only becomes important as demand increases further by mid-century and VRE growth plateaus due to its decline in value with increasing penetration [5]. In generation terms, this means that India could see VRE contributing over 59-66% (see Fig. D-1) of annual generation in 2050, depending on the annual VRE installation rate, which we assume is limited in the reference case (see Table D.13). Under the reference case, annual CO_2 emissions decline by 20% from 2020 to 2040, but then rebounding by 48% from the 2020 level as demand increases further by 2050.

Li-ion battery storage is not found to be cost-competitive until 2040. Deployment is mainly in regions with high solar PV penetration as seen in Fig. 5-1 (e.g., North),

due to an increasingly stronger "duck" curve[160] resulting from rising solar output and rising evening demand (see Fig. D-3). Battery storage is dispatched to meet evening peak demand (see Fig. 5-2), with an average storage duration (i.e. total installed energy capacity divided by the total installed power per modeling period) less than 4 and a half hours in 2050. The abundant VRE resources in the Western and Southern regions and high demand in the Northern region also leads to transmission expansion in the South-West-North corridor of 77 GW by 2050, which corresponds to a 65% increase in the transfer capacity relative to 2020 as illustrated in Fig. 5-1.

5.3.2 Scenario analysis spanning supply, demand drivers, and policy

We evaluated several alternative scenarios to systematically quantify the impact of various factors on electricity system evolution in the Indian context. These include: 1) high AC efficiency and low implied space cooling demand as detailed in section 5.2, 2) alternative (morning, day time) EV charging schedules, 3) low capital cost for grid-scale Li-ion energy storage (see Table 5.2), 4) deployment of DLS storage 4) low NG prices ($8/MMBtu), and 5) a moderate CO_2 policy starting at $20/tonne and increasing to approach $50/tonne by 2050 (see Table D.2). Table 5.3 summarizes the scenario names (columns) and their definition along various dimensions (rows).

Table 5.3: Scenario definition (see Table 5.2 for detailed costs)

Parameter	**AC**		**EV**			**Storage**		**DLS**	**Gas ($/MMBtu)**		**CO_2 Policy ($/t.)**	
Scenario	Baseline	High efficiency	Evening	Morning	Day	Reference	Low-cost		11	8	None	2030:20 2050:53
Reference	X		X			X					X	
High AC efficiency		X	X								X	
Low-cost	X		X				X				X	
Low gas price	X		X			X				X	X	
Morning charge	X			X		X			X		X	
Day charge	X				X	X			X		X	
Carbon price	X		X			X			X			X
DLS	X		X			X		X	X		X	
High AC efficiency -low cost		X	X				X		X		X	
High AC efficiency -low cost -carbon price		X	X				X		X			X
DLS and low-cost	X		X				X	X	X		X	

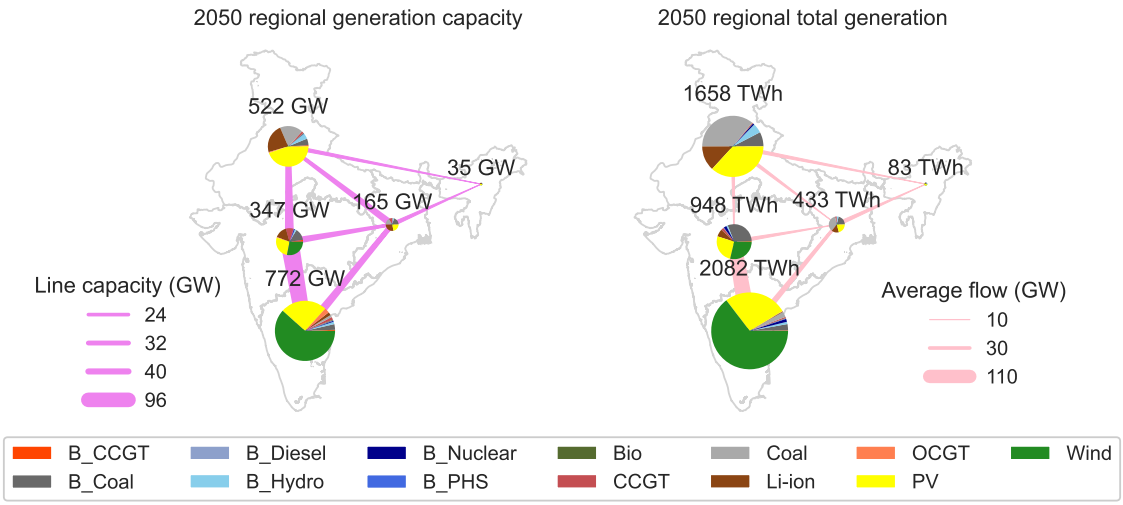

Figure 5-1: Regional distribution in generation and storage power capacity as well utilization trends for 2050. Regional transmission transfer capacity and its average utilization is shown.

5.3.3 Impact of key supply and demand-side drivers

Fig. 5-3 and Fig. D-8 highlight the impact of four main technological parameters spanning demand and supply, on the least-cost evolution of the Indian grid. Specifically, we evaluate model outcomes based on alternative assumptions for each parameter and compare them to the reference case.

While the reference case assumes mid cost projections for the U.S. context [5], the low-cost storage scenario follows the low-cost projections from the same reference. The low-cost storage scenario may be more plausible outcome for India, if storage follows trends similar to VRE in terms of cost differences between U.S. and India. In the low battery storage cost case[5], storage power and energy capacity increases by 424 GW (174% increase) and 3,625 GWh (332% increase), which enables 30% more VRE generation in 2050 compared to the reference case. This results in 54% lower annual CO_2 emissions compared to the reference case in 2050, the highest reduction among the considered parameter sensitivities (3% reduction for low gas price case and 4% increase for the high AC efficiency case in Fig. 5-3). This is largely due to the increased competitiveness of VRE, especially solar, which reduces new coal installations by 91% compared to the reference case by 2050. Low-cost storage reinforces the deployment of VRE, and leads to VRE supplying 65% of annual generation in 2050, accompanied by transmission level storage of average duration

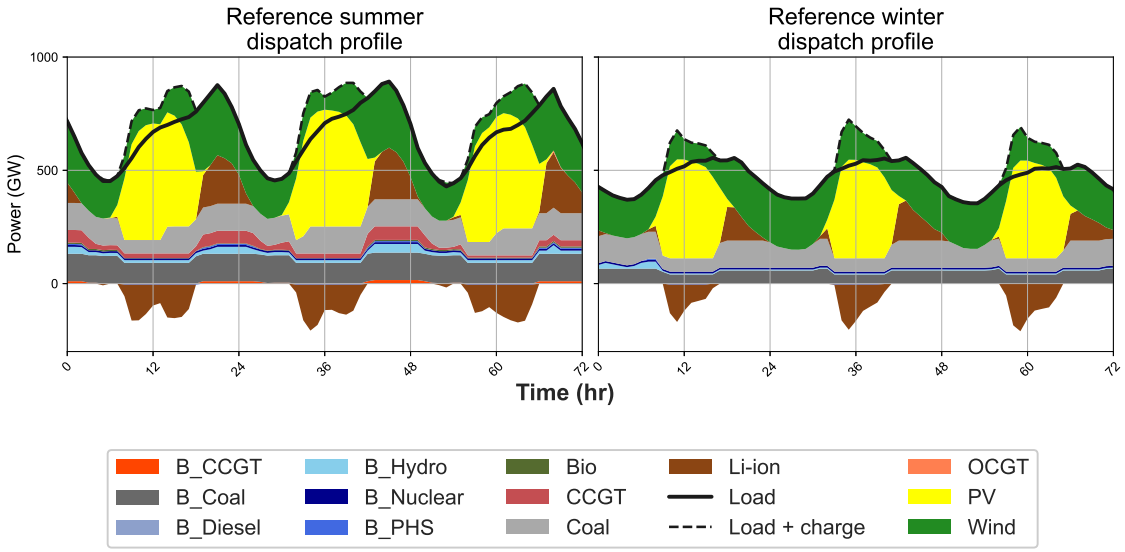

Figure 5-2: Hourly generation dispatch and load profile for three days during summer (left) and winter (right) periods for 2050. Model outcomes based on reference case as defined in Table 5.3. Storage charging is shown in the "Load + charge" curve as well as by the negative generation for storage. Technology names and their respective abbreviations in Table D.4

under 7 hours. This is achieved with system average cost of electricity in 2040 and 2050 being 22% and 39% lower than the reference case (see Fig. 5-6) and 92% less transmission expansion capacity by 2050 compared to the reference case (see Fig. D-6). We note that among the supply and demand drivers considered, storage cost are the important factor affecting annual CO_2 emissions in 2050. Moreover, low-cost storage eliminates the need to build new coal capacity particularly in 2040 and 2050 where we note mass deployment of storage as an enabling additional VRE generation.

Under the reference case, AC demand growth is projected to contribute 43% to peak demand in 2050, thereby creating the need for peaking generation capacity. Combined cycle gas turbines (CCGT) and open cycle gas turbines (OCGT) are best fit for peaking generation due to their greater operational flexibility compared to coal power plants and lower capital costs [157]. Moreover, given the relatively high cost of NG fuel vs. coal (see Table D.5), NG generation capacity is deployed but utilized sparingly, with annual capacity utilization for CCGT and OCGT plants at 5% and 3%, respectively, in 2040 under the reference case (see Fig. D-7). Because of this, the deployment of new NG generation capacity is closely tied to AC demand

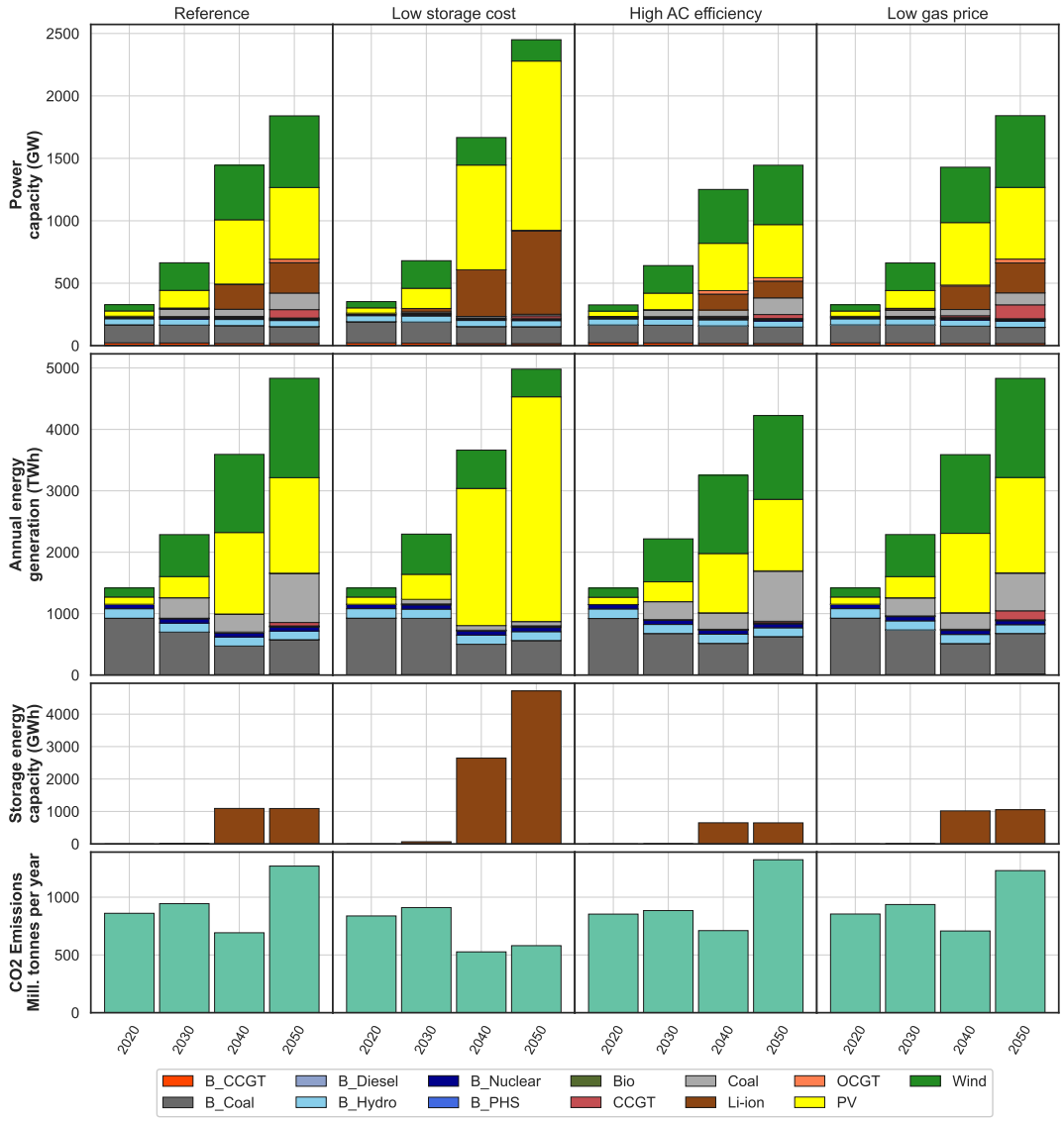

Figure 5-3: Installed capacity (1^{st} row), annual energy generation (2^{nd} row), storage energy capacity (3^{rd} row) and annual CO_2 emissions (4^th row) for reference case (1^{st} column), as well as cases with alternative assumptions for battery storage capital cost (2^{nd} column), high AC efficiency (3^{rd} column) and gas prices (4^{th} column). Detailed assumptions for each case are provided in Table 5.1 and Table 5.2

growth, with the high AC efficiency scenario virtually eliminating the need for new NG capacity (see Fig.5-3, column 3 vs. 4) in 2050. At the same time, low NG

prices improve the economic viability of NG generation, leading to higher CCGT capacity deployment and utilization vs. the reference case (see Fig. D-7) in 2050 that also reduces new coal capacity deployment by 28%. Low NG prices erode coal generation without significantly changing VRE deployment, owing to the operational flexibility of gas generation that, along with storage, complements integration of VRE generation at similar level of curtailment (7% in 2040). This explains why the low NG price case has 3% lower annual CO_2 emissions in 2050 compared to the reference case. It is important to note that India is both gas resource and infrastructure constrained, and the modeled low NG price case reflects the global average liquefied natural gas price over the past decade [161]. However, it does not consider the cost of the required NG transmission infrastructure (pipelines, compressor stations), nor the implied costs that may obtain from siting constraints.

We use a bottom-up demand forecasting model [3] to evaluate electricity demand under a high AC efficiency scenario in which India's average SEER rating trails the efficient global weighted average (8.5) by 15% as opposed to being 36% behind in the baseline global weighted average SEER (6.2) in 2050 [25]. Fig. 5-3 highlights the supply-side impacts of the modeled AC demand, where we account for regional disparities in AC adoption, which depend on climate and population size, giving rise to significant variation in estimated electricity demand for space cooling [3]. The high AC efficiency case in Fig. 5-3 shows a 22% and 13% decrease in installed capacity and generation, respectively. AC demand accounts for over 40% of peak demand in summer 2050, occurring during evening hours (8 PM to 12 AM), under the reference case. However, in the high AC efficiency case, AC demand accounts for less than 20% in the high AC efficiency case in summer 2050 (Table 5.1). In addition to reducing capacity and generation requirements, the reduction in AC demand also results in a flatter demand profile that has two further supply-side impacts: a) it reduces the value for peaking generation provided by NG power plants and battery storage and b) reduces the value of solar generation in serving demand in the day time as well as indirectly meeting evening peak demand via battery storage (see Fig. D-4). The impact of high AC efficiency on CO_2 emissions is most notable in 2030 when storage is not yet cost competitive. High AC efficiency is responsible for 7% annual CO_2 emissions reduction in 2030 when compared to the reference case and slightly lower system average cost of electricity (Fig. 5-6). Absent a carbon emissions constraint or low-cost storage, a flatter demand profile, however, leaves more room for baseload generation provided by existing coal and further investment in new coal generation in 2040 and 2050. Fig. D-3 compares the grid dispatch over a summer week for the reference case and high AC efficiency case, which suggests that peak demand is a key driver of value for solar and by association storage in the system.

In a regime when peak demand is reduced compared to the reference case and with the existing coal assets that India currently holds in its generation portfolio, the role of storage is not as significant and therefore less solar is built to charge storage. When less solar is built out, the optimization model pivots to more coal generation that is not affected by seasonality. This explain why high AC efficiency does not improve grid CO_2 emissions intensity of the subsequent modeled periods (see Fig. D-6). We further note that high AC efficiency on its own does not contribute to cumulative (i.e. summed over all model periods) emissions reduction. Cumulative emissions for the reference case is 3,766 million tonnes while the high AC efficiency case results in 3,772 million tonnes of CO_2 emissions. Therefore, while implementing demand-side efficiencies has clear positive outcomes, both in terms of cost and CO_2 emissions, in the short run (2030), a sustainable and continuous supply-side effort needs to complement it in the long-term to ensure emissions mitigation and cost-effectiveness.

Another possible demand driver is EV charging, which unlike AC use, also has the potential to offer flexibility to the system [141]. The reference case assumes EV charging predominantly takes place during evening hours (7 PM to 12 AM), which is generally consistent with residential EV charging schemes [94]. If instead, EV charging is predominantly shifted to morning hours (i.e., 5 AM to 10 AM), reflective of mixed charging infrastructure deployment, it reduces the contribution of EV charging to peak demand to 6% vs. 10% in the reference case [3]. Fig. D-8 (top) quantifies the grid impacts of morning relative to evening EV charging (reference case). As EV demand grows over time, charging demand during the evening hours can be met via short-duration battery storage that is charged during day time hours when solar generation is prevalent. However, if EV demand were to occur in the early morning hours, Fig. D-8 (bottom) highlights that deployment of overnight energy storage to discharge in the morning is not cost-effective and instead coal deployment is favored. The net impact is that morning EV charging schemes favor coal deployment over VRE and storage and result in a 2% higher system average cost of electricity in the three investment periods as well as 3% higher annual CO_2 emissions in 2050. As one might expect, aligning EV charging with periods of high solar irradiation gives rise to more solar with little or no storage, and no new coal generation. In this case, installed capacity of coal is reduced by 2% while solar and wind capacity is increased by 2%, and resulting in a less than 2% reduction in annual CO_2 emissions.

5.3.4 Impact of distribution-level storage

The projected growth in peak demand can not only drive investments in centralized generation and transmission capacity, but also at the distribution network level. For the latter, battery storage is increasingly viewed as viable non-wire alternative (NWA) network relief mechanism that can allow for deferring network upgrades with large financial impacts due to high cost of capital [36, 38, 162]. Deployment of battery storage to partially offset peak demand within the distribution system modifies the demand profile seen by the transmission system owing to timing and duration of battery charging and discharging. We compute such a "transmission level" demand and the accompanying distribution-level storage (DLS) deployment based on outcomes from modeling the operation and sizing of storage in urban distribution feeders in the Indian context using real options and linear optimization framework, described in Chapter 3. Because DLS is only deployed when network deferrals are economic [51], i.e. the present value of investments in battery storage is less than that of investments in network upgrades, their impact on the transmission system can be captured via the modified transmission level demand (demand + DLS charging - DLS discharging, see Fig. D-11) without representing DLS's capital or operating cost. Across the regions, cost-optimal DLS sizing points to an average storage duration (ratio of energy capacity to power capacity) of 2 to 6 hours that is consistent with the duration of the overloading peak demand as well as available off-peak charging hours without violating distribution network capacity constraints. In 2030, a total of 93 peak hours were shaved with deployment of DLS storage of 29 GWh nationally for the reference case demand scenario.

From the transmission system perspective, DLS, when deployed, reduces peak demand that occurs during evening hours, while increasing demand by charging during off-peak hours. Fig. D-10 highlights the temporal changes in transmission level demand from DLS discharging during evening hours and charging during earlier hours in the day (7 AM to 11 AM) when solar availability is not maximized. DLS operation aims to minimize peak demand and network upgrades and thus spreads out the charging over several hours rather than maximize charging during periods of abundant low-marginal cost supply from resources like solar. Consequently, Figure 5-4 shows that DLS deployment tends to shift the installed capacity mix to favor wind that has high capacity factors at night and early morning hours as well as coal resources over solar PV and battery storage. By 2050, demand growth has sufficiently materialized and DLS is no longer cost-effective as an alternative to network upgrades, and is consequently retired. The correlation between storage and peak demand is most pronounced under the low-cost storage case with DLS deployment

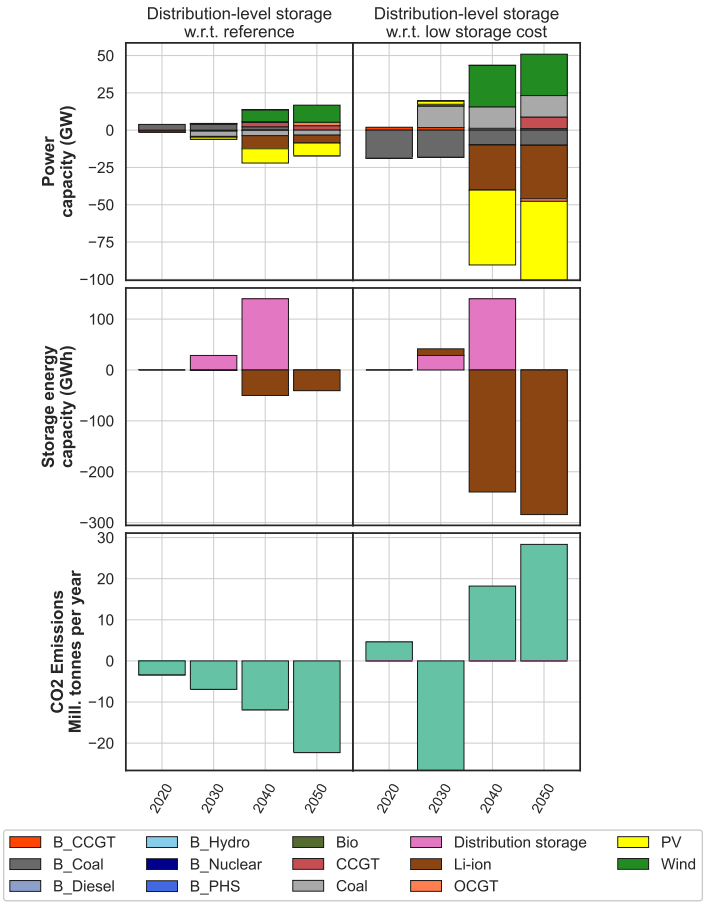

Figure 5-4: Impact of distribution-level storage deployment on dispatched generation (1^{st} row), installed storage energy capacity (2^{nd} row) and annual CO_2 emissions (3^{rd} row) under the reference case (1^{st} column) and the low-cost storage (2^{nd} column) case.

(Fig. 5-4 2^{nd} column), where the DLS enabled peak shifting has a trickle down effect on the generation design, since less storage is needed for peak hours discharging and therefore less solar capacity is installed to charge up the storage. Since demand is being met by either alternative VRE with less intra-day variability (i.e. wind) or coal, the role for solar + storage at scale is thus reduced with the presence of DLS, even under the assumption of low-cost storage. Over 90 TWh of additional wind capacity is installed with respect to the low-cost storage case (see Fig. 5-3 2^{nd} column) but also more coal since DLS results in flattening the demand profile. Overall, factors

like DLS and AC efficiency improvements are demand-side peak shifting or reducing mechanisms that, depending on the cost of storage, can indirectly lead to increase (reference case) or decrease (low-cost storage) coal generation. At the distribution level, DLS or high AC efficiency are clearly a cost saving mechanism that helps distribution companies minimize capital investment[163, 36]. When aggregating DLS to transmission-level for national planning considerations, the overall system cost of electricity does not improve compared to the reference case (see Table D.1 and accompanying discussion) and therefore necessitates further attention to trade-offs between transmission and distribution DERs [164]. It should be noted the impact of DLS deployments modeled here are relatively small when compared against the impacts of AC efficiency improvements as well as the total capacity deployed on the system in the reference case (Fig. 5-3).

5.3.5 Technological vs. policy drivers to reduce new coal investments

The outcomes of the individual technology cases point to possible strategies for minimizing future investment in stranded coal generation under global climate mitigation goals. This raises the question whether a combination of these approaches will be most beneficial for coal reduction. The high AC efficiency-low cost case, as defined in Fig. 5-5, highlights the collective impact of low-cost storage, low NG prices and high AC efficiency on power system evolution, where we see the combined effects of these supply- and demand-side drivers. As discussed above, low NG prices and high AC efficiency favor fossil generation (gas and coal, respectively) over solar and battery storage to meet peak demand compared to the reference case, while low-cost storage increases deployment of solar and storage. Collectively, in the high AC efficiency-low cost case, these factors lead to a 112% increase in need for energy storage power capacity compared to the reference case (Fig. 5-5 column 2) by 2050, while energy capacity deployment increase by 244% in 2050 compared to the reference case (see Fig. D-9). The flatter demand profile (on account of high AC efficiency) and lower energy storage costs, increases duration of storage deployed as compared to the reference case as well as low storage cost case (see Table D.6). Overall, this case leads to a 50% reduction in annual CO_2 emissions over the reference case in 2050 and the lowest cumulative emissions (i.e. sum of all modeled annual emissions) of any individual case considered here. Annual CO_2 emissions reductions in 2030 vs. reference case are attributed to high AC efficiency. Further down the line, emissions savings vs. the reference case are primarily attributed to low-cost storage that supports solar integration. Still, CO_2 emissions in 2050 are 1.3 times higher than the low-cost

storage case, in part because high AC efficiency leads to a flatter load profile that reduces the value of solar relative to coal generation, all else remaining equal. Flatter demand profile also reduces the need for peaking NG generation (Fig. 5-5 column 2) compared to the reference case.

Although such a technology-focused strategy can substantially reduce new coal investment, existing coal capacity [157] remains operational up to mid-century in the high AC efficiency-low cost case and contributes 16% of total generation even in 2050. This suggests that demand- and supply-side mechanisms to reduce emissions are insufficient for deep decarbonization of the grid, and additional policy measures may be needed. As an example policy measure, Fig. 5-5 explores the impact of a CO_2 price that starts at 20 \$/tonne in 2030, and increases by 5% over time, approaching 50 \$/tonne by 2050 (see Table D.2 — this price trajectory is adapted from another publication analyzing India's energy system [165]). There is precedent for carbon pricing in India. Since 2010, India has imposed a tax on coal production, which has increased steeply from INR 50 (\$0.70)/tonne coal in 2010 to INR 400 (\$5.61)/tonne coal since 2016. This practice is included in India's nationally determined contribution under the Paris agreement [166]. As compared to the reference case (see Fig. 5-5), the expectation of a rising CO_2 price leads to reduced utilization and early retirement of existing coal, discourages investments in new coal, and favors increased investment in low-carbon generation, mainly VRE and storage. Fig. 5-6 highlights that the cost impacts of a carbon policy are greatest in 2030 (21% higher system cost of electricity vs. the reference case) when existing coal supplies 30% of annual demand. However, in 2040 and 2050, a combination of factors, including declining coal utilization, investment in VRE and storage vs. new coal reduces the system annual cost of electricity by 20% (see Fig. 5-6), while reducing CO_2 emissions by 86% compared to the reference case. It is evident, therefore, that the initial electricity cost increase resulting from a carbon price can be mitigated over time by combining it with technological measures of the high AC efficiency-low cost case as shown in Fig. 5-6, primarily attributable to the low-cost of storage. The technology + policy approach reduce costs by 36% and 47% compared to the reference case for 2040 and 2050 respectively, and paves the way for grid decarbonization by mid-century through reducing CO_2 emissions by 97% compared to the reference case, retiring all of the existing coal in 2050 and reducing emissions intensity to 8 gCO_2/kWh (see Fig. 5-5 column 4).

5.4 Discussion

India represents one of many countries that will have to contend with electricity system CO_2 emissions impacts of rapidly growing electricity demand from space cooling and to a lesser extent, EV adoption, over the next 3 decades given its dependence on coal (953 GWh in 2020). Here, we present an analytical framework for evaluating the implication of such demand drivers in conjunction with other supply drivers on cost-optimal pathways to electricity system decarbonization by mid-century. We demonstrate this framework through a detailed assessment of India's electricity system, where we find, as have other recent studies [144], that large amounts of VRE generation, enabled by storage, are an important feature of a least-cost expansion of the electricity supply over the next 3 decades. However, growth and changing temporal patterns in electricity demand, driven by AC use, are projected to outstrip growth in cost-optimal VRE generation in our reference case and could lead to 48% higher CO_2 emissions in 2050 vs. 2020 levels. NG plays a marginal role, i.e. during extreme peak demand hours when solar and storage are exhausted, in the generation portfolio of the Indian electricity system. Installation of large capacity of low utilization NG turbines will marginally reduce coal dispatch without any significant impact on emissions, as long as prevailing NG price trends continue. Moreover, relatively lower NG prices that are within expectations of long-term liquefied natural gas prices, can only partially substitute new coal capacity by NG, but cannot displace existing coal capacity.

High AC efficiency reduces total generation and storage capacity and produces short-term CO_2 emissions and cost reductions, but without further incentives to restrict coal generation, is not expected to contribute to long-term emissions reduction efforts. Among the technology cost scenarios we have modelled, the cost of storage has the greatest impact on long-term CO_2 reductions. Storage complements solar generation by time-shifting either generation (transmission level storage) or demand (DLS). We further note that even small time shifts in demand drivers can have an incremental effect at bulk power system level as seen by EV and DLS. While we show that both demand and technological measures are valuable in reducing dependence on coal and increasing VRE penetration, complementing those measures with an incremental carbon price, or an equivalent measure, is the clearest pathway for deep decarbonization of India's electricity sector. With forty Indian companies already committed to an internal carbon price [167], the policy pathway is a pragmatic and well within reach solution that could place India as a global leader in VRE and grid-scale battery storage deployment.

Although these findings are based on a study of India's electricity system, many aspects can be generalized to other parts of the world, where similar supply- and demand-side factors persist (e.g. Indonesia, Nigeria, Vietnam). As an example, India's growth due to increased cooling demand can also be seen in Sub-Saharan African countries where as in India, dependence on fossil fuels is very strong and NG availability is constrained. The analytical framework developed here can be applied to these other regions to develop holistic view for electricity system decarbonization pathways by mid-century at the global scale based on considering available technology, local resources and, practically viable policy approaches relevant to each region.

In summary, this chapter explores optimization-based solutions for the exogenous uncertainty in demand growth and technology cost. However, the deep dive into the necessary modeling of the power system highlights the complexity and thus the computational burden of solving such problems. As a result, it is inefficient and seldom tractable to solve these problems using optimization methods. Therefore, this thesis transitions to the third question (see Chapter 1) in order to solve the problem of decision making under uncertainty of electricity resource design with proper consideration of exogenous uncertainties as well as endogenous ones.

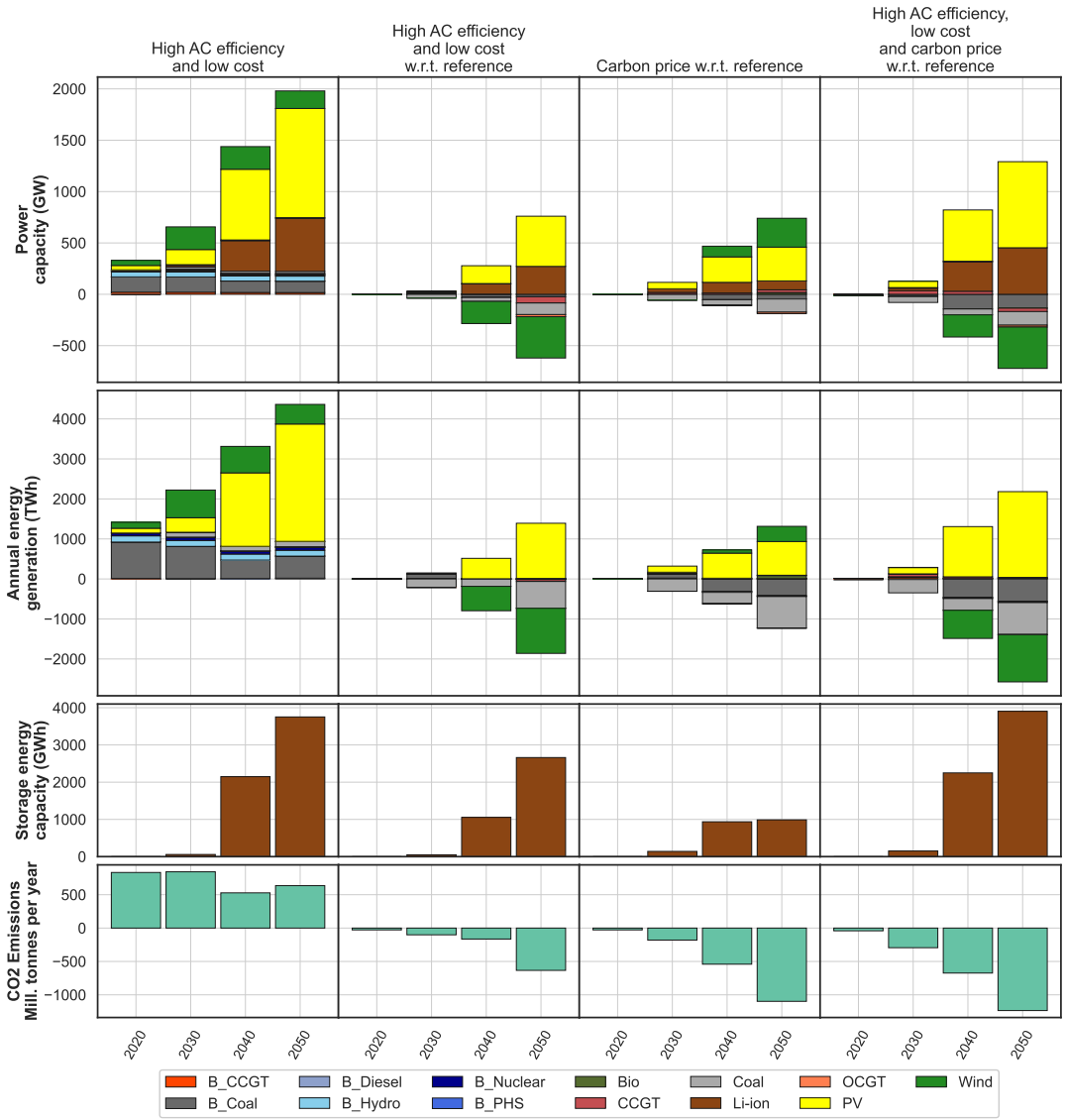

Figure 5-5: Model outcomes for high AC efficiency-low cost case, defined by low battery storage capital cost, high AC efficiency and low NG price (1^{st} column) as well as impact of carbon price with and without scenario assumptions. Model outcomes include installed capacity (1^{st} row), annual energy generation (2^{nd} row), storage energy capacity (3^{rd} row) and annual CO_2 emissions (4^th row). Columns 2-4 highlight outcomes compared to the reference case for the following cases: a) high AC efficiency-low cost case (2^{nd} column), b) low carbon price case, where CO_2 price starts at 20$/tonne in 2030 and grows by 5% each year (3^{nd} column) and high AC efficiency-low cost + low carbon price scenario case (4^{rd} column).

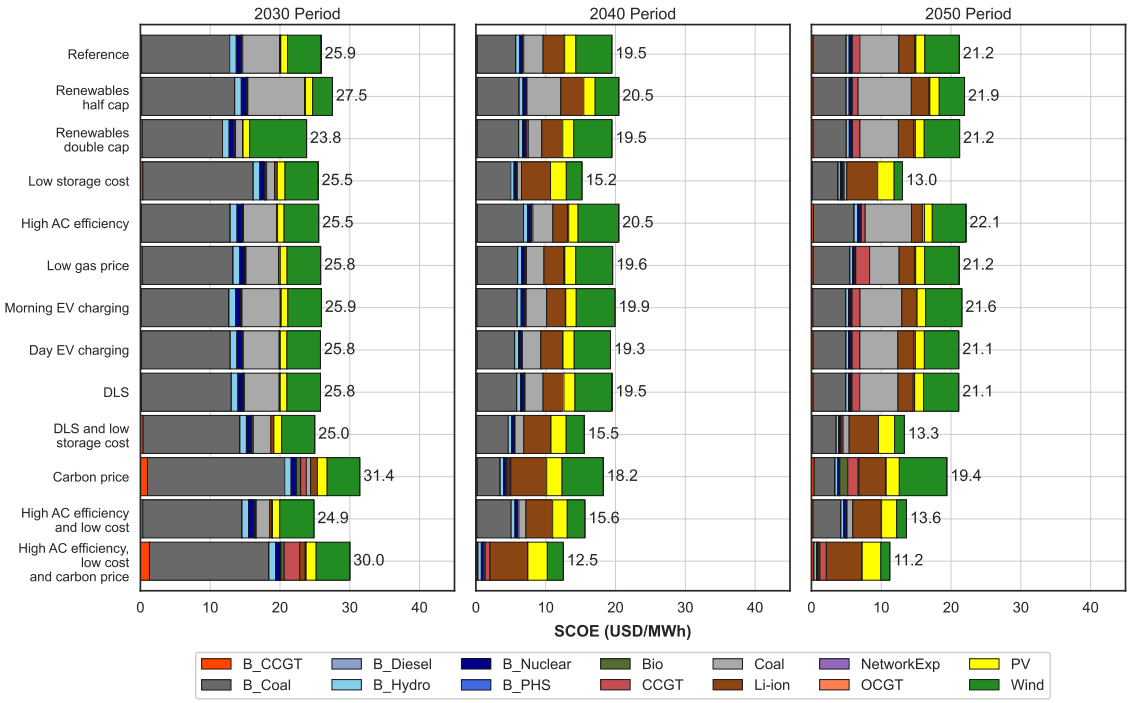

Figure 5-6: System average cost of electricity generation expansion (SCOE) per modeling period for a range of cases evaluated in the study. Note that SCOE does not include fixed costs associated with existing generation and transmission assets in 2020.

Chapter 6

Representative period selection for power system planning using autoencoder-based dimensionality-reduction

Question 3: *How to handle uncertain factors in particular investment planning of electric grid infrastructure?*

So far, we have examined three situations of capacity expansion models: distribution network investment planning, minigrids-under-the-grid, and bulk power system scenario-based generation and transmission planning. It is evident, particularly from Chapter 5, that the size of such a class of optimization and modeling problems can quickly become very large. The level of detail necessary to produce accurate results expand the size of the problem as seen in Fig. D-12. When uncertainty is considered, even a scenario-based sensitivity analysis requires significant computational resources and time. To properly address electricity resource capacity expansion planning under uncertainty, the problem must be first reduced to a size that is computationally manageable in the context of multi-stage decision-making under uncertainty. In this chapter, we present a representative period selection method for power system planning that will enable the multi-stage the later work of investment planning of electric grid infrastructure of Chapter 7.

Power sector capacity expansion models (CEMs) that are used for studying future low-carbon grid scenarios must incorporate detailed representation of grid opera-

tions to consider the capacity implications of increasing temporal variability in both electricity supply and demand. Often CEMs are formulated to model grid operations over representative periods that are sampled from the original input data using clustering algorithms. However, such representative period selection (RPS) methods are limited by the declining efficacy of the clustering algorithm with increasing dimensionality of the input data and do not consider the relative importance of input data variations on CEM outcomes. Here, we propose a RPS method that addresses these limitations by incorporating dimensionality reduction, accomplished via neural network based autoencoders, prior to clustering. Such dimensionality reduction not only improves the performance of the clustering algorithm, but also facilitates using additional features, such as estimated outputs produced from parallel solutions of simplified versions of the CEM for each disjoint period in the input data (e.g. 1 week). The impact of incorporating dimensionality reduction as part of RPS methods is quantified by quantifying the error in outcomes of the corresponding reduced-space CEM vs. the full space CEM. Extensive numerical experimentation across 1,3 and 8-bus networks and range of technology and policy scenarios establish the consistent superiority of the dimensionality-reduction based RPS methods using input and estimated output features in reproducing full-space CEM outcomes. Moreover, one of the RPS methods leads to smaller magnitude of error in reproducing full-space CEM outcomes while using half the number of representative periods as conventional RPS methods, which points to the potential solution speed up in CEM computation enabled by the method.

6.1 Introduction

Least-cost planning of deeply decarbonized bulk power systems requires contending with the unique operational attributes of the demand and supply-side resources that play a dominant role in such systems. These include: a) variable renewable energy sources (VRE) whose output varies across multiple time-scales; b) new flexible loads such as heat pumps, electric vehicles and distributed energy resources that collectively increase demand variability and c) storage resources that store energy for various time scales, ranging from intra-day to seasons, and thus couple grid operations across these periods [168, 158, 169]. Recognizing this need, a growing body of literature has focused on formulation and solution strategies for power system capacity expansion models (CEM) with improved spatial and temporal representation of grid operations [20, 53, 54, 170, 171, 172, 173, 174, 175].

A typical CEM takes the viewpoint of central planner and minimizes the sum of

investment cost and operational costs of electricity supply, network, storage and demand-side resources. The constraints of a CEM usually include operational constraints of each resource as well as the system including, supply-demand balance, reliability requirements, and environmental considerations such as carbon emissions limits [176, 177]. The number of variables in the CEM scale linearly with the number of operational periods modeled; let n be the total number of generation units, l number of load zones and t number of operational periods. Then the full-scale CEM is a problem of dimension $(n + l) \times t$. To maintain computational tractability, many CEMs evaluate grid operations over a few representative periods (e.g. days, weeks), often selected using one of the many known variations of time series clustering techniques (which we cover later in this section) [178, 20, 179]. In the context of an electricity system, time series clustering usually face common challenges such as processing multidimensional and multivariate input data and handling hidden features that are only detected in the results [58]. Therefore, along with advancements in decomposition algorithms that allow for increasing temporal resolution of CEMs while maintaining computational tractability [153, 180], improvements in representative period selection (RPS) methods to exploit unique attributes of power sector-related time series data are useful.

The problem of time series clustering for RPS finds applications in many domains outside power systems, and generally consists of four major components: 1) clustering algorithm, 2) prototype definition, 3) distance measurement, and 4) dimensionality reduction.

1. **Clustering Algorithms:** Commonly-used time series clustering algorithms can be classified into six broad classes depending on the underlying logic used for grouping the elements (e.g. hour, day, week) of the time series into clusters: hierarchical (bottom-up and top-down) [181], partition-based (e.g. k-means, k-mediod) [86], density-based (DBSCAN) [182], grid-based (e.g., finite cell division) [183], model-based (e.g., self-organizing maps) [184], and multi-step (hybrid methods) [185].

2. **Prototype Definition:** The prototype of a cluster refers to the representation of the original time series in the reduced space. There are three main prototyping methods: 1) means and medoid [186], 2) temporal averaging (e.g., Dynamic Time Warping (DTW)) [187], and 3) local search [188].

3. **Distance Metric:** Time series clustering highly depends on the choice of the distance measure used for assigning the time series elements to clusters [189]. Commonly used distance metrics include 1) time-based metrics such as

Euclidean, correlation, Fourier transform [190, 191], 2) shape-based metrics [192], and 3) change-based metrics (e.g., Hidden Markov Models) [193].

4. **Dimensionality Reduction:** Unlike the other three components, which are essential for the clustering process, dimensionality reduction is an optional component. Broadly speaking, a clustering algorithm's performance is measured by how well it can assign a cluster label to a given dataset, commonly measured by the prototype definition and distance metric. Since a clustering algorithm's performance deteriorates when the size of the data increases [86], dimensionality reduction downsizes the dataset used in the clustering process for better performance and lower computational burden. This improved performance comes at the expense of information loss in the dimensionality reduction process. Dimensionality reduction methods include 1) data-adaptive approaches [194], 2) deterministic segmentation [195], 3) model-based (e.g., Auto-Regressive Moving Average) and [196] 4) clipping [197].

The above-mentioned components of RPS have been heavily discussed in the literature in the context of electricity resource planning [52, 53, 54, 198, 199, 200, 201]. Several studies have explored the problem of RPS for CEMs applied to power and/or energy systems [55, 199, 200]. Here, we summarize a few key papers that exemplify the general approaches. Authors in[202, 170] explore modeling representative days for optimal deployment of energy storage only. Probabilistic generation snapshots method to cluster representative weeks based on mean cluster values is discussed in [203]. Operational pattern similarity in temporal data to aggregate generation dispatch models is also proposed in some studies[56, 57] but without considering the temporal variations in supply-demand balance. Several papers have resorted to using variants of k-means clustering on the original input data to select representative periods to be used in CEM evaluation [174, 20, 179, 204]. Other papers have proposed RPS methods based on hierarchical clustering [171, 198] or simulation-based methods [205, 201]. Other studies [199, 200, 58] explore optimization-based approaches at a high computational cost that limits its scalability. On the other hand, data-driven approaches [59, 60] have been proven to be effective in time series aggregation. However, the proposed methods are restricted to labeled data sets that are generally not applicable to power system planning because the time series data is not labeled and require *a priori* knowledge of the distribution function for probabilistic assignment to a cluster based on the distance metric. Additionally, the performance of such methods is often measured by the error in reproducing the input data to a CEM, whereas in power systems it is the CEM output that determines the quality of the RPS. Recent work on learning-based methods [206] has been proposed to evaluate

RPS of wind resources but without any application to CEM of power systems.

In summary, the literature on RPS for power system CEMs has generally focused on using the original multi-variate input data directly in the clustering process without considering how well the input data approximates the original output (i.e. full-space problem). The contributions of this chapter is twofold: First, we introduce a data pre-processing step involving nonlinear dimensionality reduction via neural networks. We call this process an *autoencoder* and explain in detail in section 6.2. Second, we propose a technique to incorporate CEM output information in the clustering process. The resulting approach allows us to find a balance between loss of information during the dimensionlaity reduction process and the information loss incurred during the clustering process. Our approach identifies a set of representative periods and their corresponding input data that is used to formulate a reduced-space CEM (RCEM, i.e. the CEM with representative periods) whose outcomes are shown to approximate full-space CEM (FCEM, i.e. CEM using full-dimension input data) outcomes with an acceptable accuracy. The autoencoders identify both short (intra-day variations) and long-term features (seasonal variations) in the original multi-variate time series data and reduce the dimension of the data in a way that reduces the clustering loss. We perform numerical experiments to compare the performance of the proposed RPS method with two alternative methods from the literature which rely on k-means clustering but do not use dimensionality reduction. In all four approaches, we fix the prototype definition of the clustering process to be *mediod*, use k-means as the clustering algorithm and use Euclidean distance (L2 norm) as the distance metric. Overall, the main contributions of the chapter to the field of power system planning models are as follows:

1. use of autoencoders for dimensionality reduction as a pre-processing step in the RPS method.
2. defining and incorporating features related to CEM outputs to be used in the clustering process.
3. extensive numerical experimentation spanning different number of representative periods selected (4, 8, 20) for RCEM, different network sizes (1, 3, 8 bus) and cost assumptions (90 scenario runs for each network and representative period pair) for the case study.

The remainder of the chapter is structured as follows: section 6.2 describes the methods, including the key assumptions and structure of the proposed methods and formulation of the CEM; section 6.3 details the case study input data we use for the comparative statistical analysis of the four alternative RPS methods; section 6.4

describes results of the various RPS methods from the case study, and section 6.5 discusses the results.

6.2 Autoencoder-based representative period selection

6.2.1 CEM model, input and output data

To evaluate the proposed methods, we use a standard linear programming CEM formulation [207, 41, 42] that is fully described in Appendix B. The model objective function (Eq. B.1) minimizes the total system cost which includes annualized generation expansion and, operational costs. The full-space CEM (FCEM) has $w_t = 1$ for every period t. On the other hand, the reduced-space CEM (RCEM) has time weight $w_t \geq 1$ depending on the RPS. The key operational constraints are: 1) hourly power balance constraint (Eq. B.2), 2) time-dependent capacity constraints for generation resources (Eq. B.3), 3) battery storage state of charging, energy and power capacity limits (Eq. B.4 - B.9), 4) generation unit commitment (Eq. B.10 - B.14), 5) generation minimum and maximum power output (Eq. B.15, B.16), 6) generation ramping limits (Eq. B.17, B.18), 7) direct current power flow approximation using line susceptance and bus angles (Eq. B.19), 8) network flow limits (Eq. B.20, B.21) and, 9) non-negativity constraints (Eq. B.22). This formulation is used for both the FCEM and RCEM, with the following two caveats: First, the RCEM does not include inter-temporal constraints linking resource operations across representative periods. For the inter-temporal constraints related to ramping limits and storage inventory, we look back to the last sub-period of the same representative period to define the constraint for the first sub-period (see Eq. B.5) [20, 200, 175]. This modeling approximation implies that storage inventory across two representative periods is effectively decoupled, which is generally reasonable when considering short-duration storage technologies. Second, the RCEM computes annual operating costs based on scaling up operating costs of representative periods using weights (w_t) obtained from the clustering process that correspond to number of periods approximated by each representative period.

The solution of the CEM formulation yields a set of generation, energy and charge capacities ($\Omega_{a,z}^{size}$, $\Omega_{a,z}^{energy}$ and, $\Omega_{a,z}^{charge}$). The solution also provides dispatch variables for each time period t yielding five output time series: 1) power output ($\pi_{a,t,z}$), 2) storage charge ($\Psi_{a,t,z}^{charge}$) 3) storage discharge ($\Psi_{a,t,z}^{discharge}$) 4) non-served energy ($\chi_{a,t,z}$) and 5) line power flow ($\phi_{t,z,z'}$).

To quantify the impact of input data variations on CEM outputs, we generate estimated outputs from solving the CEM for each single period in the underlying input data and use that information to inform selection of representation periods. If each representative period consists of q hours, one year input data (8760 hours) will have $p = \lfloor \frac{8760}{q} \rfloor$ periods and correspondingly p CEMs can be solved to generate output features. Note that the output produced for each period is an abstraction of the outputs obtained from solving the FCEM which considers 8760 hours of operations. The solution time required to solve the RCEM for each period can is generally much smaller than than solving the FCEM [208]. Moreover, the independence of CEM problems for each time periods allows multi-threading (or parallelization). From the outputs of the RCEM, we use the five output time series mentioned above as features to be used in the data encoding.

6.2.2 Dimensionality reduction - autoencoders

We perform dimensionality reduction using a structure of autoencoders that refer to a type of neural network which enable transforming high-dimensional input data into their lower dimensional data (encoding) and vice-versa (decoding). The encoder transforms the high dimensional input data into lower dimension, referred to as *latent representation*, while keeping the most important features. The decoder uses the latent representation of the data to reconstruct the initial input data.

The encoder and decoder are constructed using various deep learning layers [4] in Keras [209] as follows:

The encoder network architecture (Fig. 6-1) starts with a 1D convolution layer, which extracts the important short-term waveforms. The convolution layer is followed by a maximum pooling layer to reduce the dimension of the data by combining the output neurons of the convolution layer into a single neuron to the subsequent layer. Then a bidirectional long short-term memory layer is introduced to learn the temporal changes in both directions relative to each time period. The decoder consists of a fully connected upsampling layer followed by convolutional transpose layer to transform the latent representation back to the original dimension. The objective of the autoencoder is to find a code for each data sample through minimization of the mean squared error (MSE) between its output and input over all samples. The parameters of the autoencoder are updated by minimization of the reconstruction

mean squared error (MSE) as seen in Eq. 6.1:

$$L_r = \frac{1}{p}\sum_{i=1}^{p} ||D(E(x_i)) - x_i||_2^2 \tag{6.1}$$

where p is the number of periods considered (e.g., $p = 52$ if the planning horizon is one year and each period is defined as a week — $q = 168$ hours) in the data entries $x_i \in \mathbb{R}^{(n+l)\cdot q}$ is the i^{th} period. E and D represent the encoder mapping (Eq. 6.2) and decoder mapping (6.3), respectively, with k being the desired number of clusters.

$$E\colon \mathbb{R}^{(n+l)\cdot q} \mapsto \mathbb{R}^{\frac{(n+l)\cdot q}{k}} \tag{6.2}$$

$$D\colon \mathbb{R}^{\frac{(n+l)\cdot q}{k}} \mapsto \mathbb{R}^{(n+l)\cdot q} \tag{6.3}$$

Moreover, we set the maximum pooling size to the number of clusters to reduce the inputted multivariate space to latent representation. Therefore, the data is being compressed from $\mathbb{R}^{p\times(n+l)\cdot q}$ to $\mathbb{R}^{p\times\frac{(n+l)\cdot q}{k}}$ where k is the number of clusters desired. For autoencoders [4], when the dimension of the latent representation is on the order of the number of clusters of the input data then the network can be trained in an end-to-end manner without including regularization [60] such as *Dropout* and batch normalization terms [210, 211].

6.2.3 Clustering algorithm

We use the unsupervised k-means clustering algorithm that minimizes the sum of intra-cluster distances, with distance defined by the Euclidean norm in our case study. The objective function of the k-means algorithm is given in Eq. 6.4 which is unsupervised. The objective function Eq. 6.4 of k-means minimizes the mean squared error between the representative periods. In Eq. 6.4, S_j is set of periods assigned to cluster j, k is the number of clusters desired, and $\bar{x}_j$ is the centroid of cluster j [212]. For each cluster, we use the period that is nearest to the cluster centroid as the representative period [156]. The algorithm is implemented using *scikit-learn* library KMeans function with 10,000 repetitions of different centroid seeds to ensure that a global solution can be found [86]. The clustering algorithm produces period-indexed (52 in case of clustering of weeks) labels of the annual time series data. This labeling is used to generate the reduced time series data after the

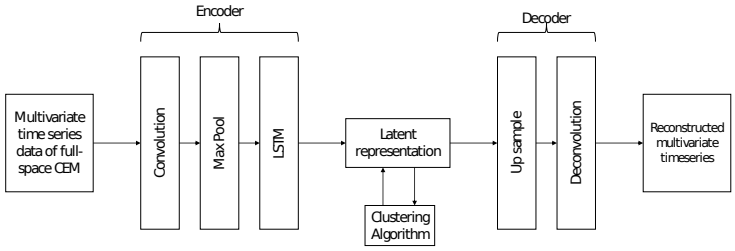

Figure 6-1: Autoencoder architecture used in the RPS method. Encoder is constructed using three main layers: Convolution, Maximum Pooling and, Long short term memory (LSTM) [4]. Decoder is constructed using two main layers: Up sampling and deconvolution [4]. Clustering is performed in the latent with the representative periods identified from clustering subsequently decoded to produce input time series for the RCEM.

latent representation is reconstructed through the decoder.

$$L_c = \sum_{j=1}^{k} \sum_{x_i \in S_j} ||x_i - \bar{x}_j||_2^2 \tag{6.4}$$

6.2.4 Overall loss-function definition

To guide the autoencoder to represent features that are important to forming clusters, the overall loss function for the RPS is defined by Eq. 6.5. Here, γ is a tunable parameter between 0 and 1 that adjusts the relative importance of the error induced in the latent representation and the error introduced via the clustering process.

$$L = L_r + \gamma L_c \tag{6.5}$$

We examine three alternative configurations for coupling autoencoders and the clustering process as illustrated in Fig. 6-2. Type 1 only uses input data, Type 2 uses both input and output data in one autoencoder and, Type 3 splits the autoencoding of the input and the output data into two separate autoencoders. Specifically, the

outputs of both encoders in Type 3 are intermediate latent representations which are fed into a second-level autoencoder that produces the latent representation on which clustering is performed. The decoding side of the Type 3 method mirrors the encoding side. Both Type 1 and 2 autoencoders use the loss function defined by Eq. 6.1. Type 3 has an expanded autoencoder loss function as detailed in Eq. 6.6, where L_{int} is the intermediate autoencoder loss function, L_I is the input data autoencoder loss function, L_O is the output data autoencoder loss function and, α and β are parameters that balance the feature importance of input relative to output data, while satisfying the condition: $\alpha + \beta = 1$. We identify the values of α and β that minimize the overall loss function by scanning through 100 combinations of values that satisfy the above-mentioned condition. Note that L_{r_i},L_I,L_O are defined based on their respective input and output data using Eq. 6.1.

$$L_r = L_{int} \times (\alpha L_I + \beta L_O) \tag{6.6}$$

By jointly reducing input and output, the Type 2 autoencoder could in principle, identify input features that have a greater impacts on CEM outputs. However, even a dimensionality reduction technique will have diminishing returns with increasing number of features i.e. the dimension of data input to the encoder. This motivated us to consider the Type 3 encoder, where we use a separate autoencoder for input and output and then combine them in a third autoencoder (Eq. 6.6).

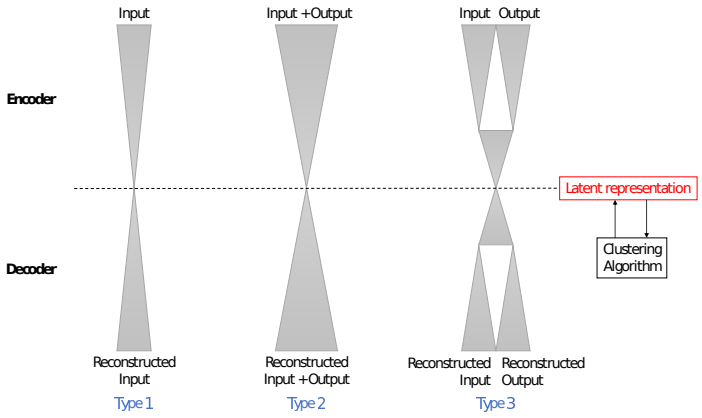

Figure 6-2: Three types autoencoder structures proposed. Type 1: input autoencoding only, Type 2: joint input and output autoencoding, Type 3: separate input and output autoencoding.

Type 1 autoencoder only uses input data therefore it is comparable to k-means on input data only (I k-means). On the other hand, Type 2 and Type 3 autoencoders use both input and output data and therefore are comparable to k-means on input and output data (I/O k-means).

6.2.5 Error metrics

We focus on error metrics that quantify the difference between the results of the FCEM and the RCEM across multiple dimensions of installed capacity ($\Omega_{a,z}^{size}$,$\Omega_{a,z}^{energy}$,$\Omega_{a,z}^{charge}$), system cost (Eq. B.1), annual non-served energy (NSE, $\chi_{t,z}$ in Eqn. B.2, summed over the entire year), generation ($\pi_{t,z}$ in Eqn. B.2, summed over the entire year), while using the same time-independent parameters inputs. For capacity, we compare capacity results retrieved directly from the outputs of RCEM and FCEM. In RCEM, each resource a in the set of resource A and each zone z in the set of zones Z of a case m ($x^{z,a,m}$) has a relative absolute error to corresponding FCEM ($y^{z,a,m}$) defined by the generalized Eq. 6.7. Where $m \in M$ corresponds to the set of cases defined by alternative technology assumptions, demand profiles, and representative periods (Appendix E Table E.1). We use absolute value based error metrics to account for both positive and negative deviations between the RCEM outputs and FCEM outputs.

$$AE_{a,z,m} = \frac{|x_{a,z,m} - y_{a,z,m}|}{y_{a,z,m}} \tag{6.7}$$

Additionally, each case m in M has multiple generation sources $a \in A$ in zones $z \in Z$. We report the weighted average absolute error per case m by the generalized equation Eq. 6.8. The weight $y^{a,z,m}$ is the result of FCEM of the corresponding resource a of zone z in case m.

$$WAE_{a,z,m} = \frac{\sum_{a\in A, z\in Z} AE_{a,z_m} y_{a,z,m}}{\sum_{a\in A, z\in Z} y_{a,z,m}} \tag{6.8}$$

In the results, we report the mean and distribution of the weighted average absolute error of all the cases $m \in M$. The weighted average absolute capacity error is defined as the weighted average installed capacity error relative to the full-space results for each considered generator type (i.e., both energy generating and dependent resources). The mean capacity error metric is therefore defined by Eq. 6.9, $AE_{a,z,m}^{\Omega^{size}}$, $AE_{a,z,m}^{\Omega^{energy}}$ and, $AE_{a,z,m}^{\Omega^{charge}}$ are the installed size, energy and charge capacity errors

(individually defined by Eq. 6.7) of generator a in zone z from case m and $\Omega_{a,z,m}^{size,FCEM}$, $\Omega_{a,m}^{energy,FCEM}$ and, $\Omega_{a,z,m}^{charge,FCEM}$ are the FCEM outcomes corresponding to installed size, energy and charge capacities for the same case m.

$$WAE^{Cap} = \frac{1}{|M|}\sum_{m\in M}\left[\begin{array}{c}\frac{\sum_a\sum_z AE_{a,z,m}^{\Omega^{size}}\cdot\Omega_{a,z,m}^{size,FCEM}}{\sum_a\sum_z\Omega_{a,z,m}^{size,FCEM}} \\ +\frac{\sum_a\sum_z AE_{a,z,m}^{\Omega^{energy}}\cdot\Omega_{a,z,m}^{energy,FCEM}}{\sum_a\sum_z\Omega_{a,z,m}^{energy,FCEM}} \\ +\frac{\sum_a AE_{a,z,i}^{\Omega^{charge}}\cdot\Omega_{a,z,m}^{charge,FCEM}}{\sum_a\sum_z\Omega_{a,z,m}^{charge,FCEM}}\end{array}\right] \tag{6.9}$$

For operational cost, generation and NSE results, direct comparison between the RCEM results and FCEM results is not plausible due to the different temporal resolutions used. Instead, we fix the capacity variable of the RCEM solution and evaluate the optimal dispatch for the fixed capacity over the entire year at an hourly resolution that can be compared against the dispatch results and cost of the FCEM. The mean system cost of electricity (SCOE) error is defined by Eq. 6.11 which resembles the optimization model's objective function Eq. B.1 without considering NSE (χ) - error in the NSE is considered separately. Eq. 6.10 is the objective function of the CEM defined in Eq. B.1 (noted as Obj) without the NSE.

$$SCOE_{a,z,m} = Obj_{a,z,m} - \sum_z\sum_t w_t\cdot\chi_{t,z}\cdot C_z^{\chi} \tag{6.10}$$

We normalize $SCOE$ by the total demand to produce the error across all the cases m with differing demand profiles as in Eq. 6.11, where $SCOE_{a,z,m}^{FCEM}$ is the full-space SCOE (as defined in Eq. 6.10) of generator a in zone z.

$$\begin{aligned}WAE^{SCOE} = \frac{1}{|M|}\sum_{m\in M}\frac{1}{\sum_t\sum_z L_{t,z,m}} \\ \times\frac{\sum_a\sum_z AE_{a,z,m}^{SCOE}\cdot SCOE_{a,z,m}^{FCEM}}{\sum_a\sum_z SCOE_{a,z,m}^{FCEM}}\end{aligned} \tag{6.11}$$

NSE is reported as a percentage of the annual demand of the corresponding load year as described in Eq. 6.12:

$$WAE^{NSE} = \frac{1}{|M|} \sum_{m \in M} \left(\frac{\sum_z \sum_t \chi_{t,z,m}}{\sum_t \sum_z L_{t,z,m}} \right) \tag{6.12}$$

The generation error is the generation weighted average error relative to the full-space for each considered generator type (i.e., energy generating resources only) as described in Eq. 6.13 where $AE^{\pi}_{a,m}$ an individual case's generation error (defined by Eq. 6.7 of generator a from case m and $\pi^{FCEM}_{a,t,z,m}$ is the full-space generation output corresponding to the same case m. Note that all the error metrics used to evvaluate the performance of the various RPS methods are dimensionless by the definitions provided above.

$$WAE^{Gen} = \frac{1}{|M|} \sum_{m \in M} \frac{\sum_a AE^{\pi}_{a,m} \cdot \sum_t \pi^{FCEM}_{a,t,z,m}}{\sum_a \sum_t \pi^{FCEM}_{a,t,z,m}} \tag{6.13}$$

6.3 Case-study description

To assess the performance of the autoencoder-based RPS method, we evaluate CEM results on three different case studies, all of which are based on the data and topology of the Electric Reliability Council of Texas (ERCOT). We consider representative period of length 168 hours with $k = 4$, 8 and, 20 . Our experiments span three networks: single bus, three bus and eight bus systems as highlighted in Fig. 6-3 and 90 scenario evaluations per representative period and network size pairing, where each scenario considers an alternative assumptions regarding time-independent parameters (e.g. VRE capital cost, see Appendix E Table E.1). The single bus system only considers the North Central region of Texas without network data. The three bus system consists of a complete linkage network of the North Central, South Central and Coastal regions of Texas. The eight bus systems considers all eight regions of Texas. Load data at the zonal level is provided by the ERCOT [213]. We aggregate historical hourly solar and wind capacity factor data at the county level [214] to the zonal level according the ERCOT load zone map [213]. Finally, we use a publicly available eight bus system [215] to model the power flow and nodal connections of the ERCOT load zones as displayed in Fig. 6-3. Each load zone has four types of generation units: combined cycle gas turbine, solar, wind and short-duration battery storage. We use the National Renewable Energy Laboratory's Annual Technology Baseline [5] for the relevant technical and cost parameters.

At each load zone, the input time series data that must be clustered are: load, solar

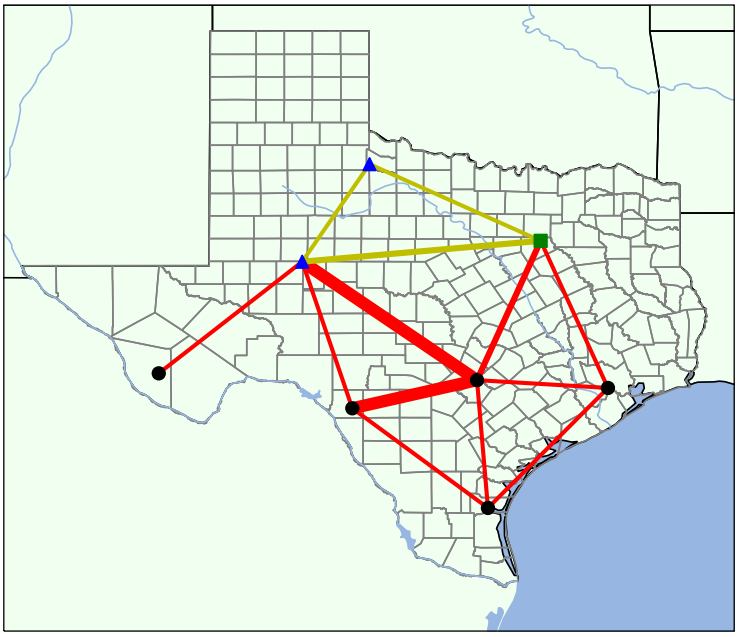

Figure 6-3: ERCOT electricity system. Single bus system (square green node); three bus network (square green node and blue triangle nodes); eight bus network is all inclusive. Line thickness highlights the average power transfer in 2018 between zones on a relative basis.

and wind profiles. Therefore, for the eight bus system $168 \times 3 \times 8$ columns of length 52 must be clustered into the desired number of representative weeks. The input time series has a shape of $52 \times 4,032$ with a latent representation of 52×504 given the autoencoder parameters (see Appendix E Table E.2) for the desired cluster size of $k = 8$. Since the pooling layer is set to the cluster size k, $504 = 4,032 \div 8$ is the latent representation size. While both learning and clustering performance can be tuned for better results on a case by case basis, we fix the hyperparameter values (see Appendix E Table E.2) of the autoencoders for all number of clusters and all cases so as to facilitate a consistent analysis. All autoencoders, clustering algorithms and optimization models were coded in the Python language and run on MIT Supercloud High Performance Computer using Intel Xeon Platinum 8260 processor with up to 48 cores and 192 GB of RAM [216].

6.4 Results

6.4.1 CEM results comparison - single case

Fig. 6-4 highlights the impact of different RPS methods on installed capacities and annual generation outputs for the 3-bus case study and 8 representative weeks (results can be seen as a difference plot in Appendix E Fig. E-1). The mean NSE of the three autoencoder-based RPS methods is 0.25 % and the mean NSE for the I and I/O k-means is 0.61%. As compared to k-means only based approaches, Type 1-3 methods also result in smaller deviations in annual generation results and capacity results as compared to the FCEM results (Fig. 6-4). Battery storage energy and charge capacities are relatively stable across all methods as seen in Fig. 6-4 and Appendix E Fig. E-1, with capacity errors below 3% relative to the full-space results. The low battery storage capacity error can be attributed to the presence of two VRE sources (solar PV and wind) which diversify the dependency of charging and discharging on the availability of VRE generation. We note that both conventional RPS methods result in lesser wind resources and more solar relative to both the FCEM results and the RCEM outcomes using autoencoder based RPS. Wind availability profiles are more variable than solar, which is not adequately captured in the conventional RPS methods and leads to over investment in natural gas which is a dispatchable resource.

6.4.2 CEM results comparison- multiple case

The general nature of the results shown in Fig. 6-4 for a single case are assessed by a broad set of numerical experiments over the three networks and four different values of representative periods. Table 6.1 summarizes the average value of different error metrics (over 90 parameter scenarios) related to system cost, NSE, annual generation and capacity (as per the definitions provided earlier see Eq. 6.9 - 6.13) across the RPS methods and network case studies. We note that Type 2 and Type 3 autoencoders, which use estimated output data, noticeably improve all four error metrics relative to both k-means and Type 1 autoencoder. In contrast, incorporating output data directly in the clustering process without dimensionality reduction, as in the I/O k-means method, does not lead to improved error metrics as compared to the I k-means method. As pointed in [217], including more dimensions to an unsupervised learning algorithm, such as k-means, may increase the complexity of identifying good quality solutions to the clustering problem.

To explore the scalability of the RPS methods, we investigated the impact of different

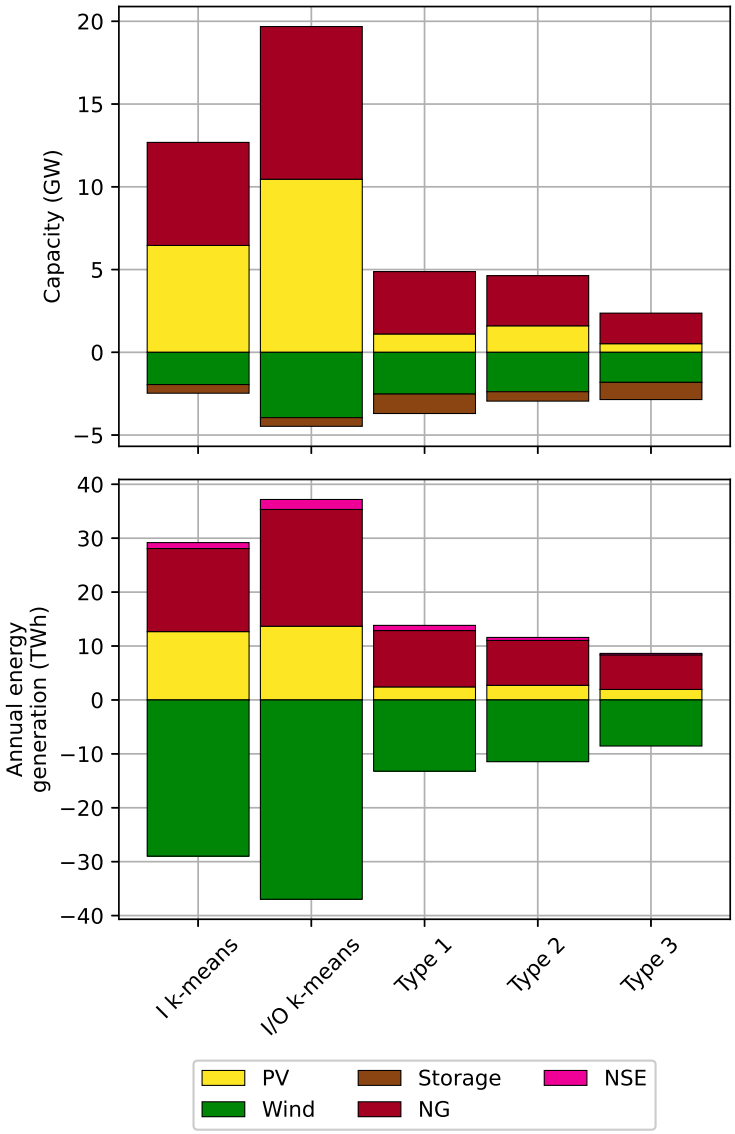

Figure 6-4: Difference in installed capacity and annual generation between reduced-space CEM (RCEM) results and full-space CEM results for three bus network. Results based on RCEM using $k = 8$ for ERCOT load year 2020, mid-range VRE technology cost, 1,000 \$/tonne CO_2 price.

representative period choices on the various error metrics using the 8-bus network as the case study (Table 6.2 and Fig. 6-5). When comparing the mean value of the

Table 6.1: Mean values of error metric defined in Eq. 6.9 - 6.13 to quantify performance of different representative period selection (RPS) methods for the 1-bus, 3-bus and 8-bus networks. Row-wise color mapping: red is highest and green is lowest value across the row.

	bus	I k-means	I/O k-means	Type 1	Type 2	Type 3
SCOE	1	12.3	11.97	11.23	9.97	7.8
	3	12.79	12.64	12.25	9.45	7.85
	8	14.24	14.05	12.07	8.57	9.09
NSE	1	1.12	0.85	1.11	0.91	0.61
	3	1.16	0.96	1.02	0.92	0.7
	8	1.33	0.99	1.02	0.85	0.77
Gen	1	23.9	19.12	21.43	18.3	15.1
	3	25.05	20.96	21.84	17.94	16.32
	8	27.13	23.35	22.24	17.47	17.77
Cap	1	29.53	27.46	17.54	15.73	19.27
	3	34.76	31.8	20.11	26.1	19.56
	8	44.22	42.61	47.19	38.61	34.00

error metrics of NSE, system cost and annual generation, Table 6.2 highlights that Type 3 autoencoder consistently outperforms the other methods across all cluster numbers. In addition to the mean values, Fig. 6-5 shows that the distribution of error metrics across the 90 experiments carried out are narrower for the Type 3 encoder as compared to the other RPS methods. In particular, for the Type 3 autoencoder, the probability density near zero is larger as compared to other RPS methods. For a total of 90 distinct cases for each cluster size k (4, 8, and 20) per method in Fig. 6-2, the Type 3 autoencoder is able to satisfy 74% of the cases' peak demand according to the dispatch-only optimization as compared to 63% for the I/O k-means method, 70% for the Type 1 autoencoder method and 62% for the I k-means method. Furthermore the standard deviation in the capacity error metric for the eight bus system (see Appendix E Table E.3) not only reflects the performance of the autoencoder-based RPS but also the consistency. Note that the violin plots in Fig. 6-5 extrapolate data points to visualize the kernel density estimator which may produce negative values, however the box plot portion of the violin plot (i.e., data points) are all greater than zero values since absolute errors are reported.

Table 6.2: Mean error metric results of different representative period selection (RPS) methods grouped by cluster number k for the 8-bus system. Row-wise color mapping: red is highest and green is lowest value across the row.

	k	**I k-means**	**I/O k-means**	**Type 1**	**Type 2**	**Type 3**
SCOE	4	14.98	14.30	14.20	7.94	10.80
	8	13.47	13.65	9.24	8.33	7.61
	20	14.27	14.21	12.77	9.44	8.85
NSE	4	1.50	0.78	0.81	0.67	0.80
	8	1.28	1.00	1.17	1.01	0.76
	20	1.22	1.19	1.08	0.88	0.75
Gen	4	28.89	23.83	24.83	16.38	20.18
	8	25.45	23.57	20.10	18.13	16.32
	20	27.04	22.66	21.79	17.89	16.80
Cap	4	52.12	46.89	56.34	46.40	40.07
	8	44.03	45.31	48.10	40.18	35.83
	20	36.50	35.62	37.13	29.26	26.10

6.4.3 Impact of input data on CEM results

As previously mentioned, the Type 3 RPS method expands the loss function to Eq. 6.6, where L_I is the input data autoencoder loss function and L_O is the output data autoencoder loss function given α and β as tunable parameters that balance the feature importance of input relative to output data. The tunable parameters satisfy $\alpha + \beta = 1$. In systems with less variable time series patterns such as a solar only VRE system without storage, the CEM outputs can be predicted quite well with input data and hence one would expect β to be relatively small compared α. However, in systems with more variability in grid operations, such as a combination of VRE and battery storage, makes the system dispatch more complicated as VRE are volatile resources and storage follows a complex charge/discharge pattern. In this case, information about estimated outputs are likely to be more important than the solar only case, hence leading to a higher value of β. This hypotheses can be evaluated by exploring the optimal value of the input and output feature importance parameters α and β for the Type 3 autoencoder in the case of CEM applied to simplified case studies.

Figure 6-6 summarizes the values of α and β for six 1-bus systems where the resources are restricted to the following options: 1) natural gas and solar, 2) natural gas, solar and battery, 3) natural gas and wind, 4) natural gas, wind and battery, 5)

natural gas, solar and wind, 6) natural gas, solar, wind and battery. We perform this sensitivity analysis on 1-bus case to eliminate power flow and network effects and highlight the impact of input and output time series on the behavior of the Type 3 autoencoder. Figure 6-6 shows that β (output data importance) is higher when more time series variability (i.e., wind) is considered in the generation design. Furthermore, β increases when battery storage is considered in the generation mix. This is due to the addition of the approximate charging and discharging output data that will improve the autoencoder's performance in identifying important features that can reproduce the FCEM results using the RCEM.

6.4.4 Solution time comparison

Evidently, the time to execute the autoencoder-based RPS methods is significantly longer than the standalone RPS methods. Appendix E Fig. 6-7 shows that run time RPS method may be larger than CEM solution time in the case of small systems such as the 1-bus cases and some of the 3-bus cases. However, for larger systems i.e., more realistic CEM use cases, the autoencoder-based RPS method run time is dwarfed by the optimization run time for any cluster number k (see Fig. 6-7).

6.5 Discussion

In this chapter, we present an autoencoder-based RPS method to enable computationally efficient solution of CEMs for large-scale power systems without sacrificing accuracy. We show that autoencoders can be better trained when considering both the input time series of an optimization model and an expected output signal that is obtained from CEM evaluation of disjoint time periods in the data set. The expected output signal need not be accurate but sufficiently directs the RPS method with the objective of translating the data to learn optimization model behavior. This is particularly important for optimization models with inter-temporal dependencies.

We report a rigorous statistical analysis of the proposed methods. The proposed Type 3 autoencoder generally outperforms all other methods on error metrics related to cost, NSE, annual generation and capacity. Furthermore, Type 3 autoencoder RPS of 4 and 8 representative weeks performs better than all other methods clustered at 8 and 20 representative weeks. Therefore the proposed architecture of the Type 3 autoencoder can significantly reduce the temporal resolution of grid operations to be modeled within a CEM without sacrificing accuracy of results or incurring significant run time burden.

In summary, we present a statistical analysis that supports the use of autoencoders with clustering algorithms to reduce the error metric in electricity resource capacity expansion planning. Further parameter tuning and loss function modification can improve error metric results depending on the objective function and the clustering algorithm of choice. Here, we tested the proposed RPS method on deterministic CEM formulations with a single investment stage. However, these methods could be more impactful in enabling computationally efficient solution of stochastic, multi-stage electricity resource CEM problems as well as those considering coordinated investment planning across multiple infrastructures.

The autoencoder-based RPS method expand the work that can be done in decision-making under uncertainty by reducing the computational burden of solution exploration without trading off performance. In the subsequent chapter, we use the proposed RPS method of this chapter to solve a peculiar problem of multi-stage decision-making under uncertainty by deriving a novel learning-based method.

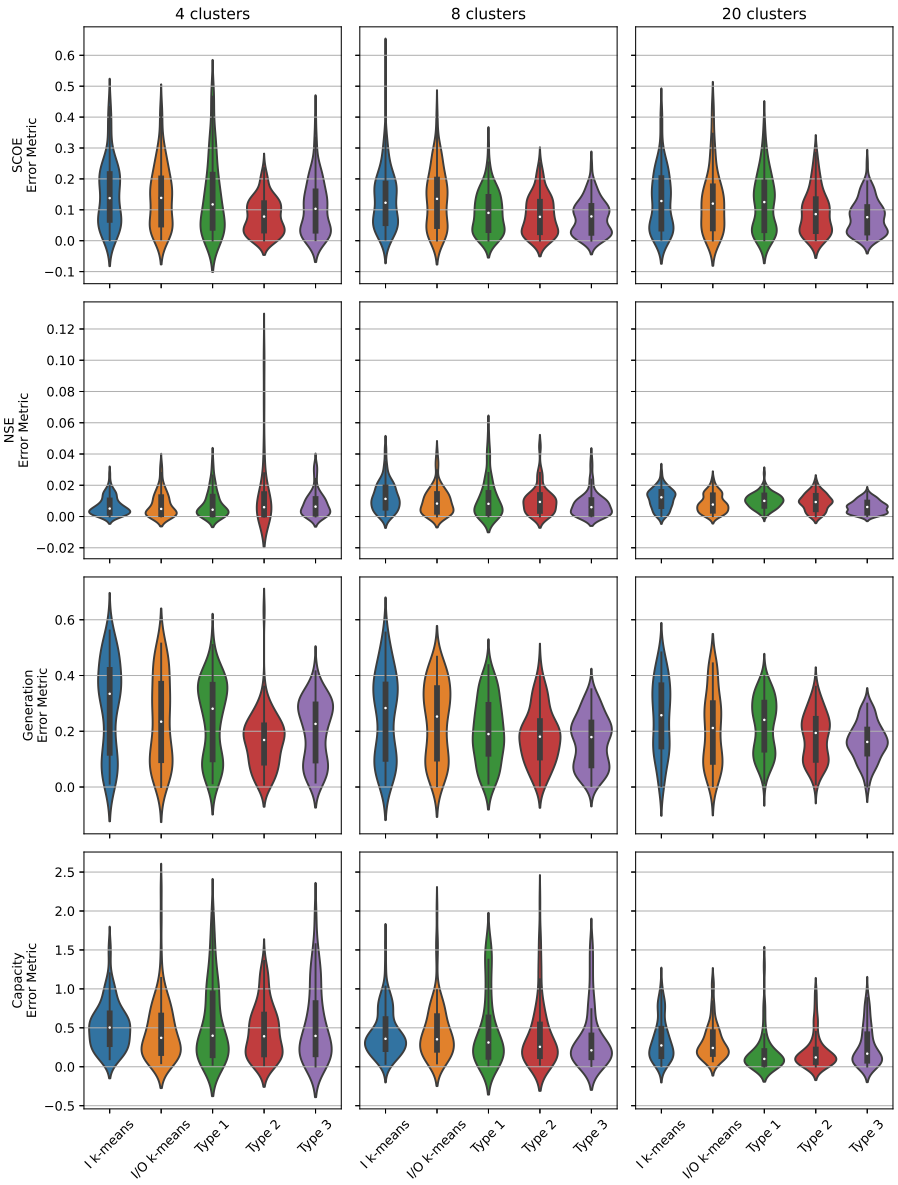

Figure 6-5: Violin plot featuring the kernel density estimator of the distribution of data points of the four weighted absolute error metrics (SCOE, NSE, generation, and capacity) grouped by RPS method for RCEM outcomes with different number of representative periods 4, 8, and 20. The data presented is based on the eight bus system discussed in section 6.3 and is consistent with the mean values presented in Table 6.2.

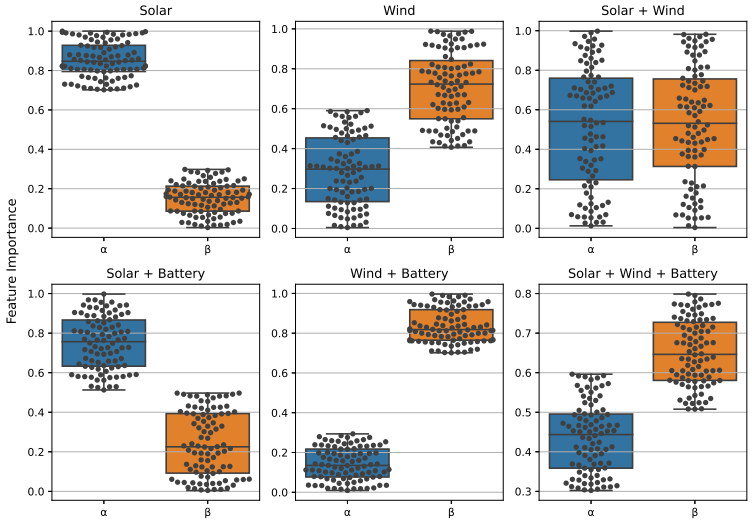

Figure 6-6: Box and swarm plot of optimized values of α and β parameters for Type 3 autoencoder across six simplified case studies described in section 6.2.3. Results based on RCEM outcomes for single bus system with 4, 8, and 20 representative periods for all 90 distinct cases as defined by parameter values in Appendix E Table E.1.

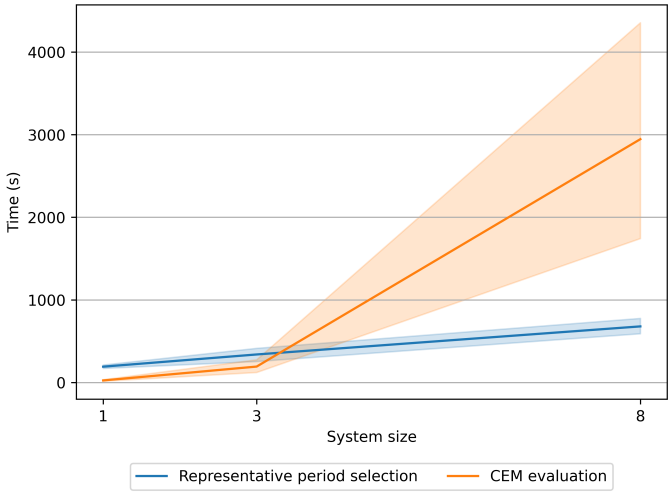

Figure 6-7: Runtime confidence band of reduced-space optimization and all autoencoder-based RPS for all sizes and scenarios across the single, three and eight bus systems.

Chapter 7

Learning-based multi-stage electricity resource capacity expansion planning under exogenous and endogenous uncertainties

Question 3: *How to handle uncertain factors in particular investment planning of electric grid infrastructure?*

The work presented in this thesis so far has been concerned with electricity resource capacity expansion planning (CEM) under (only) exogenous uncertainty. We presented an optimization-based method to solve the peculiar capacity expansion planning problems in the context of India and Nigeria as examples of Emerging Market and Developing Economy (EMDE) countries.

Real-world applications are seldom concerned with only exogenous uncertainty. Therefore, this chapter is concerned with developing a learning-based method for decision-making under uncertainty to solve a simplified real-world example of multi-stage CEM. To address this thesis's third and final question, the method presented in this chapter qualitatively and quantitatively highlights its advantages in multi-stage CEM under exogenous and endogenous uncertainties. Here, we leverage the time-domain reduction technique presented in the previous Chapter 6 to efficiently select representative weeks and reduce the temporal size of the design and dispatch of resources problem and thus enable a more efficient learning-based multi-stage decision-making under uncertainty based on decision trees.

7.1 Introduction

Long-term least-cost planning of power systems requires considering supply resources such as variable renewable energy (VRE), energy storage, and the evolution of the regulatory, economic and social landscape. A large body of literature has focused on the formulation of multi-stage power system capacity expansion models with an improved spatial, and temporal representation of the supply of the grid and its evolution [168, 158, 169, 156, 162]. A typical capacity expansion model (CEM) takes the viewpoint of the utility and minimizes the sum of investment cost and operational costs of the portfolio of electricity supply, network, storage resources and non-served energy, while accounting for operational constraints of individual resources as well as the system's reliability, emissions, and cost. Solving such a problem for multiple planning periods is a significant mathematical programming effort that requires capacity-carry-over constraints for the deterministic case and non-anticipativity constraints for the stochastic case [42, 41].

To properly plan a power system, multi-stage least-cost capacity expansion models must consider the uncertainty in demand growth and any external event that can have a significant impact on the system during each planning stage. Two types of uncertainty can affect such planning: exogenous and endogenous (further explained below). Incorporating demand uncertainty in capacity expansion models is a challenging task [218]. When uncertainty is introduced, the model quickly becomes intractable and requires several constraints to solve it. Such relaxations yield results that are seldom useful to a utility in practice. Therefore, there is a need to develop multi-stage least-cost planning tools that consider exogenous and endogenous uncertainties to inform the capacity expansion decisions that a utility must make in practice.

Generally, there are two types of uncertainties in multi-stage stochastic problems (MSSP): exogenous and endogenous. Exogenous is defined as decision-independent, and endogenous is defined as decision-dependent. Additionally, there are two types of endogenous uncertainties [61]: Type 1 and Type 2. In Type 1, decisions alter the probability of uncertain parameters. In Type 2, decisions change the temporal evolution of the uncertain parameter in the multi-stage problem. There are various MSSP formulation for exogenous uncertainties [62, 63, 64, 65, 66, 67]. Some multi-stage stochastic optimization models have used decomposition methods such as Lagrangian and Benders [62, 63, 64]. Stochastic Dual Dynamic Programming (SDDP) [66] improves nested Benders Decomposition [65] by avoiding the requirement to solve an exponential number of scenarios. In SDDP, the stochastic process has to be

stage-wise independent. A major difference between Lagrangian decomposition and Benders decomposition is that Benders decomposes the full space problem by nodes, whereas Lagrangian decomposes the full space problem by scenarios. Augmented Lagrangian decomposition called Progressive Hedging [67] is also effective in solving multi-stage stochastic problems with only exogenous uncertainty. In summary, problems typically addressed by MSSP do not account for endogenous uncertainty. Some work has been done on Type 1 endogenous uncertainty for multi-stage stochastic problems [219, 69] but without consideration for operational constraints of the system being analyzed. When the operation is considered, it is either solved through domain-specific decomposition [70] or is a reduced problem (dispatch of 1 day or a few hours) [68]. Solving design and dispatch problems of electricity resources necessitates a prolonged temporal resolution to account for time (e.g., solar PV, demand profile) and energy-dependent resources (e.g., battery storage). Accounting for uncertainty in large-scale problems is practically challenging to apply due to the computational burden [220, 221]. Learning-based models such as Markov Decision Processes can be applied to Type 2 endogenous uncertainty multi-stage problems [222, 223] by intelligently searching for feasible solutions without complete knowledge of the full domain of solutions. To effectively apply simulation-based approaches to stochastic problems, meaningful formulation and search mechanisms that are often problem-dependent are required to ensure satisfactory results [224].

In the context of EMDE countries, demand growth, VRE integration, and technology costs are more volatile, hindering progress of electricity access and power supply reliability improvements. Effective multi-stage power system capacity expansion modeling can empower local utilities to inform immediate decisions that optimize future return given external event realization (e.g., demand growth, cost decline, grid reinforcement) that would alter the multi-stage plan. We focus on developing a multi-stage power system planning model under uncertainty with application to a case study that is a simplified version of a real problem in the rural parts of Odisha in India. Furthermore, we turn to simulation-based approaches to address the multi-stage problem under uncertainty due to their ability to consider endogenous uncertainty. We carefully examine the parameters of the case study and propose a novel simulation-based approach that aims to improve classical Monte Carlo search methods [221, 158] in the context of the simplified version of the real problem in Odisha. The application of the proposed method highlights the potential value of effective multi-stage decision-making under uncertainty in revealing ample cost savings opportunities to local utilities in EMDE countries.

7.1.1 Odisha design problem

In the rural parts of the state of Odisha in India, certain electricity demand clusters of residential, commercial, and small industrial loads are faced with unique electricity design challenges. Presently, most clusters are connected to the main electric grid via power lines with poor reliability (R_0) at the medium voltage that are maintained by the distribution company (disco). The disco cannot let the supply reliability deteriorate further, as the Regulatory Agency of the State of Odisha (OERC) would have to step in and impose penalties. We simulate this in our case study by assuming that a reliability level worse than R_0 would result in a penalty imposed by OERC. Notably, the lack of target reliability enforcement is probably due to OERC's reluctance to increase the end-customer tariff incurred by the additional investments needed to operate and maintain the network by the disco. On the other hand, when R_0 deteriorates, then OERC would penalize the disco. Due to the high operation and maintenance cost of the existing long medium voltage lines, the disco is probably losing money since the demand L_0 is low — and is not expected to grow significantly at the present reliability R_0 — and the present tariff is not sufficient to cover the costs of power delivery. Problematically, the feeder lines are long and cross hazardous areas; therefore, the operational costs of maintaining these lines are high.

The analysis that is presented is from the disco's perspective. We consider multiple planning stages in which the disco can decide what to do. We discretize the decisions of the disco into just three choices, for the sake of simplicity and interpretability:

1. **Business-as-usual**: in this decision the situation remains as described above.
2. **Grid upgrade**: in this decision, the disco does what is needed to improve the line's reliability for the proper operation and maintenance costs. The new reliability R_N at the cluster connection point may remain lower than the target reliability. However, electricity demand would grow at a higher rate. We assume that OERC does not acknowledge the costs incurred by the disco from such a decision.
3. **Minigrid**: this decision disconnects the cluster from the main grid, salvages network equipment, and deploys local generation. This decision would satisfy the target reliability, resulting in an even more substantial demand growth than the grid upgrade decision.

Furthermore, there are uncertain external events which are fully decided by external agents. For the sake of simplicity, the three external events considered are:

1. **Nothing**: no external events happen and the existing environment remains

the same.

2. **Grid reinforcement**: there is a possibility that the federal government may extend the medium voltage network and install a substation that is close to the considered clusters. The reliability of the cluster will significantly improve through the medium voltage connection. This grid reinforcement would be funded by the Indian Central of State Government. Operation and maintenance costs remain the responsibility of the disco.

3. **Independent minigrid**: as long as there is a potential to improve the cluster's reliability, then there is a chance that an independent minigrid developer comes in and installs the necessary system to service the electricity demand of some or all of the consumers. From the perspective of the disco this is perceived as a partial or complete demand (and revenue) reduction.

In the context of the Odisha design problem, multiple cross-dependencies complicate the development of meaningful decision-making under uncertainty framework. We characterize the uncertainty in demand growth as exogenous and endogenous due to various parameters that can influence future demand growth (e.g., time, regulation, and reliability). From the perspective of the disco, the independent minigrid event is also a decision-dependent uncertainty since a developer would install a minigrid only when they can improve the cluster's reliability and thus bring value to the cluster. Therefore, the independent minigrid event also affects the level of demand from the perspective of the disco. On the other hand, the grid reinforcement external event is an exogenous random event (decision-independent). Therefore, demand has a three-dimensional dependency: time, decision, external events. Given the Odisha design problem, designing a multi-stage decision-making model must consider both the decision-dependent and independent uncertainties and their respective probabilities to inform the disco on which initial decision is best suited for their long-term objective.

7.1.2 Tree-based modeling

Due to the nature of the Odisha design problem, this thesis focuses on both exogenous and endogenous uncertainties where the decisions influence the realizations of uncertain events or parameters by altering their underlying probability. Simulation-based models are an effective yet simple tool to do so. Generally, simulation-based (also known as learning-based) methods involve an agent interacting with an environment. The agent is the mathematical model we want to use to solve the multi-stage stochastic problem [225]. The environment is the solution space that cannot be fully

explored due to computational limitations. In a multi-stage problem, the agent aims to maximize some defined objective value. The basic definition of such a class of methods is broken down into three elements: decision, state, and return [226]. A decision is a possible move that the agent can make given various uncertainties. A state is a current situation where the agent initially finds itself or is given previous decisions. The return is the feedback value that the environment yields to evaluate an agent's decision and, therefore, impact the objective value.

Various adaptations of this learning-based methods have been proposed [227] to maximize the objective value that the agent yields [228, 229]. Monte Carlo Tree Search (MCTS) stands out as a versatile and smart search algorithm that can navigate through a large state-space [230] and identify a sequence of solutions given an objective function [231, 232]. MCTS methods provide means of planning in complex sequential decision problems. MCTS uses traditional tree search with node evaluations that are based on stochastic simulations. MCTS led to spectacular results in numerous combinatorial game theory problems with imperfect information of sequential games [233]. At its core, MCTS is a time and resource-constrained repetition of iterating over the decision, state, and return learning-based process. However, MCTS uses decision trees, and therefore that process is expanded into four steps: 1) selection, 2) expansion, 3) simulation, and 4) backpropagation. Decisions are made by selecting a node in the defined tree. Therefore, a state is synonymous with a node in the decision tree. The return is collected based on the tree expansion, simulation, and backpropagation.

Selection

In the selection process, a tree is used to construct paths from the root to various leaf-nodes. A leaf-node is a node that has unexplored child node(s) before the maximum number of iterations has been reached. In a leaf-node, a decision needs to be made in order to advance in the stages of the tree. Each decision has a return that is associated with the results of the decision that impact the multi-stage problem. One important consideration for the selection step is the problem of exploration versus exploitation. The Upper Confidence Bounds (UCB) is a fundamental algorithm of MCTS that provides a strong worst-case distribution-free convergence guarantee and a numerical bound on the cumulative return of an initial decision at low computational complexity. UCB algorithm elegantly solves the exploration-exploitation problem using Eq. 7.1, where $N = \sum n_D$ and B_D is the upper confidence bound of each decision D that is calculated at each stage T and r_D is the average return. The node with the lower bound B_D is selected during the selection process.

$$B_{D_T} = r_{D_T} + \sqrt{2\ln\frac{N}{n_{D_T}}} \tag{7.1}$$

Expansion

At each selection, the MCTS reaches a leaf-node. In the expansion step, the model picks a subsequent unexplored leaf-node based on probability distributions of event realization and available decisions given the leaf-node.

Simulation

The simulation step ensures that the MCTS is not only seeing the future through leaf-node selection and expansion but many versions and combinations of it.

Backpropagation

After selecting a leaf-node, the total score of its parent nodes must be updated by rolling back up the tree. The updated score changes the state of the tree and can change future node selection.

The selection step must be of leaf-nodes with the best possible signal regarding uncertainty realization to effectively explore a scenario tree. As mentioned earlier, MCTS is heavily employed in multi-stage decision problems in a broad range of problems such as strict tree structures and Markov Decision Processes (MDP) [234]. Various tree policy enhancement strategies have been proposed to accelerate the convergence and search of the algorithm [235]. UCB is the most widely used, commonly available, and modular one [236]. The $1/\sqrt{n_i}$ factor in UCB may be viewed as the underestimation of uncertainty. Bayesian derivation of UCB has proven to be helpful in improving accuracy over a limited number of trials. However, Bayesian MCTS assumes a single random variable at a leaf-node in the tree, and all nodes share the same distribution function [237]. However, practical applications of MCTS to MDP to solve a multi-stage stochastic problem must consider domain knowledge to enhance the convergence of the heuristic algorithm. Algorithms that incorporate the exogenous and endogenous uncertainties of the problem could potentially outperform the distribution-free methods or generic Bayesian posterior distribution ones.

Domain knowledge can be heavily leveraged in the numerical definition of the return value that the model observes. Therefore, formulating the return value collected after a decision and interaction with the environment is a significant component of effective

MCTS. Two standard formulations of return exist; reward and regret. In reward-based models, MCTS aims to maximize the return while minimizing it in regret-based models. The more informative the return value is, the better, and MCTS can search the environment and solve the multi-stage stochastic problem. Therefore, return value formulation is a domain-dependent numerical value that serves as a consistent feedback signal from the environment to determine whether a decision (or sequence of decisions) improves the model's objective (maximization or minimization). In the context of the Odisha problem, we aim to empower local utilities to io inform immediate decisions that mitigate future regret given external event realization that would alter the multi-stage plan. Thus, we use a regret-based formulation where the objective of our proposed MCTS formulation is to minimize the return value. Therefore we refer to return value as regret.

We propose a novel tree policy enhancement strategy based on classical MCTS with a domain-dependent formulation of the regret value that aims to inform the model on the environment of the Odisha problem given exogenous and endogenous uncertainties. First, regret values at leaf-nodes are sampled in a selection-expansion-simulation search. We then use the moments of the distribution of regrets at the leaf-nodes to inform the confidence bound B_i, which is used for further selection in the complete MCTS iterative process. We hold the backpropagation phase for a fixed number of iterations defined as the burn-in rate. The burn-in rate is a tunable parameter [238] to reduce simulation bias. Trivially, increasing the burn-in rate improves the accuracy of a simulation-based model. We empirically fix the burn-in rate to 25% of the total number of iterations. Backpropagation after a burn-in phase enables the better formulation of distributions of regrets at various leaf-nodes that are informative of the tree's state-dependencies. Furthermore, We introduce a tolerance bound to allow for higher regret selection and reduce the exploration bias. The bound formulation is based on domain knowledge of electricity resource design optimizations. The multi-stage decision problem is solved through an MCTS of single-stage reduced-order multi-period optimizations using representative period selection [239] that was further detailed in Chapter 6 to enable fast tree exploration.

The local distribution company in Odisha is faced with several decisions, uncertainties, and external events. We simplify the multi-stage problem to three stages (initial and two uncertain stages). Additionally, we explore a limited list of the possible decisions and external events and discretize the probability of space of external event realization over each stage through transition matrices. Finally, we apply our proposed tree search model to the Odisha Design problem to highlight the reliability improvement that effective decision-making under uncertainty can provide and the

performance of the model relative to other tree searches.

The remainder of the chapter is organized as follows: section 7.2 presents our proposed Regret MCTS and its various components, section 7.3 reports the results of the Regret MCTS from a simulation and planning perspectives and section 7.4 discusses the findings.

7.2 Regret MCTS

The Odisha design problem includes multiple sources of exogenous and endogenous uncertainties and therefore the tree structure of MCTS rapidly expands with state-dependent branches. Our proposed methodology of Regret MCTS is divided into four steps highlighted in Fig. 7-1:

1. Tree reduction: the entire tree that describes decision-event combinations under endogenous uncertainty quickly becomes very large. We reduce the number of leaf-nodes by aggregating the various states based on common decisions.
2. Selection-expansion-simulation: in this step, the model does not build any decision-making beliefs, thus not selecting any leaf-nodes based on confidence bound values. The model only collects the defined leaf-node regret values and stores them in the corresponding leaf-node.
3. Complete MCTS: The complete MCTS logic is deployed (i.e., selection, expansion, simulation, and backpropagation). Selection is made using the observed confidence bounds at a given leaf-node.
4. Result extraction from simulated tree search.

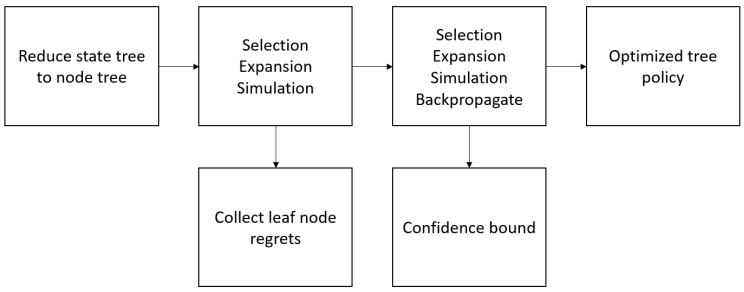

Figure 7-1: Proposed Regret MCTS methodology flowchart.

7.2.1 Tree reduction

Fig. 7-2 describes the structure of the decision tree of the Odisha design problem from the perspective of the disco. The left-hand side tree shows all potential transitions from the initial state. Since we are considering three decisions and three external events, then there are nine possible states to transition to, where a state (circle) is defined as a decision and external-event combination. The right-hand side tree aggregates the states based on decisions where the arrows represent a decision. We define a leaf-node as a group of states of the initial tree (left-hand side of Fig. 7-2) based on the same decision, i.e., a leaf-node groups various states which are realized through the same decision but witnessed different external events. There are three leaf-nodes to transition to from the initial decision given various realized external events.

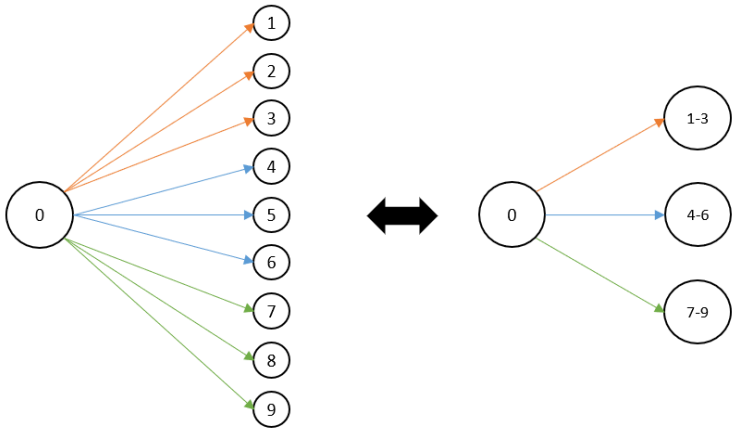

Figure 7-2: Example of single-stage decision tree with one parent node and nine state nodes that are grouped together into three leaf-nodes. Color map refer to the same decision but different external event realizations.

The objective of the Odisha design situation is to optimize the initial decision that will yield a long-term least-cost solution to the multi-stage design and dispatch of electricity resources. Given the tree's rapid expansion, exploring all sequences is computationally intractable. Therefore, we solve for an approximate solution of the initial decision that will minimize the regret of subsequent decisions given various state-dependent event realizations and demand growth. A state starts at time step T when a decision is made. An event is realized at time step $T + \tau$ (where τ is a defined time interval). Therefore a state depends on the previous state's decision

and event. As mentioned earlier, a leaf-node groups states by decisions in order to produce the reduced tree (right-hand side of Fig. 7-2). A leaf-node is associated with one decision at time step T but contains multiple regrets due to the multitude of external events that may realize within the leaf-node and in subsequent leaf-nodes.

7.2.2 Leaf-node regret

We define the problem of the Odisha design as a multi-stage decision under uncertainty problem with a multi-step design and dispatch optimization of grid investment and off-grid resources. The balance of system equation of the multi-step optimization (Eq. 7.2) is:

$$\text{Demand}_t = \text{Solar}_t - \text{Storage}_t^{in} + \text{Storage}_t^{out} + \text{Diesel}_t + \text{Grid}_t + \text{Ind}_t + \text{NSE}_t \quad (7.2)$$

where t is the time period (hour) considered in the dispatch optimization, Demand is the electrical load of the cluster, Storage^{in} and Storage^{out} are the charging and discharging of short duration battery storage, respectively, Diesel is the back-up diesel generator, Grid is the supply from the main grid, Ind is the supply from the independent minigrid developer (which is modeled as an external supply profile that offsets the total demand of the cluster) and, NSE is the non-served energy. With reliability defined by Eq. 7.3:

$$R = \sum_T \gamma^T \frac{\sum_t (L_t - NSE_t)}{\sum_t L_t} \quad (7.3)$$

To solve the Odisha design multi-stage problem, we search for the initial decision with minimum expected regret using the MCTS of sequences of single-stage optimization results. Each single-stage design optimization has a regret value as part of a sequence of decisions in the multi-stage problem. We define a generation asset's regret as its cost contribution to the revenue stream and its generation contribution to the total demand. When the total cost C_g of generation g increases, the regret of g will increase. On the other hand, when the total generation of asset g increases, then the regret of g decreases as defined in Eq. 7.4:

$$\rho_g = \frac{\frac{C_g}{\sum_{a \in A} C}}{1 + \frac{G_g}{\sum_t L_t}} \tag{7.4}$$

where C_g is the annualized total cost of generator g in USD, $\sum_{a \in A} C$ is the annualized total cost of all considered generators $a \in A$ (note that also $g \in A$), G_g is the annual generation of g in MWh, and, $\sum_t L_t$ is the annual load in MWh. The regret (ρ) of generation g is the ratio of cost to generation. When the cost contribution of generator g increases with respect to the total cost $\sum_{a \in A} C$, the regret ρ_g increases. On the other hand, when generator g increases with respect to the total demand $\sum_t L_t$, ρ_g decreases. ρ is dimensionless.

Therefore, solving the multi-stage problem becomes a tree search of the initial decision that will yield minimum regret on expectation. To solve the multi-stage problem using single-stage optimizations only, we compare regret values of different designs at a given leaf-node for the selection process of the MCTS in the direction of lower regret. At each leaf-node, there are two single-stage optimization designs based on the cost objective function: low and high cost, as detailed in Table 7.1.

Table 7.1: Design options

	Low Cost	**High Cost**
Definition	Absolute least-cost design	Least-cost design given existing capacities
Explanation	No capacity carry-over from previous state	Carries over installed capacities from previous state
Sequence	Disjoint	Compounded
Analogy	Lease lowest cost equipment at time step T and for time interval τ	No second-hand market to re-purpose installed capacities
Cost	$C_T = \gamma^T C_T(\sum_t L_t^T)$	$C_T = \gamma^{T-1} C_{T-1}(\sum_t L_t^{T-1}) + \gamma^{T-1}[C_T(\sum_t L_t^T) - C_{T-1}(\sum_t L_t^{T-1})]$

The high-cost design will have the highest regret at each leaf-node in the decision tree. The difference between the high-cost design and the low-cost one (Eq. 7.5) provides insight into which asset has the highest regret and therefore should be minimized

in a subsequent trial. Therefore, we collected leaf-node regret vector Υ_g^S indexed by generation g in each leaf-node during the second step of Regret MCTS. Thus, Υ_g^S is a random variable, and each asset will have distribution-dependent confidence bound.

$$\Upsilon_g^S = \rho_g^{high} - \rho_g^{low} \tag{7.5}$$

7.2.3 Confidence bound

Given the numerical vector formulation of leaf-node regret, the confidence bound is also a numerical vector indexed by generation g for time T indexed decision D_T. We propose the formulation of Eq. 7.6:

$$B_{D_T}^g = \mu_{D_T}^g + \kappa_{D_T}^g \sqrt{2\ln\frac{N}{n_{D_T}}} \tag{7.6}$$

where μ and κ are the expected value and kurtosis of the distribution of regrets of the random variable Υ_g^S at decision i for a given asset. A much more informed uncertainty estimate can be obtained from Bayesian probability by the number of child nodes and their associated values. Therefore r_i of Eq. 7.1 is replaced by μ_i which is the expected value of the probability distribution of regrets $P(\Upsilon)$ at each leaf-node node. When new trials are sampled, results are combined with priors in the standard way to compute the posterior. Given the number of possible sequences that can be built, the distribution of regrets at a particular leaf-node will be fat-tailed. From the perspective of the MCTS, the expected value alone is not descriptive of the endogenous uncertainties that govern the tree. Variance is the second moment of a random variable, and with a higher selection uncertainty, the variance tends to explode [240]. On the other hand, kurtosis is the fourth moment of a random variable that describes the distribution's spread in a scale-independent manner. Therefore kurtosis is a clear function of uncertainty and can improve the confidence in the bound B_D^g.

7.2.4 Tree search

We decouple the uncertainty dependence from the decisions via sampling by iterating through the selection, expansion, and simulation steps only without any backpropagation. We do so since backpropagation can bias subsequent node selection, and we wish to model the impact of decisions on uncertainty before making any decisions.

In the complete MCTS iterative process, we use the approximated probability distributions of regrets (identified through backpropagation) to alter the decision's design by trading off short-term cost optimality for cumulative long-term savings. The goal is the minimize the cumulative regret of the initial decision made by the disco in the Odisha design problem. The second step of Regret MCTS explores sequences based on the selection confidence bound (Eq. 7.6) given the collected samples. In the third step of Regret MCTS, the decision tree is selected using Eq. 7.7, which will explore sequences that minimize the cumulatively discounted regret of decisions with bounded confidence. A decision D_T evolves with the simulation due to the endogenous uncertainty considered. Therefore, a decision D_T^g will minimize the highest regret generation and improve the overall confidence bound accordingly.

$$D_T = \arg\min_{D} \Big[\sum_{T+1} \gamma^T \sum_{D} \sum_{g} \big(B^g_{D_T} + \epsilon \big) \Big] \tag{7.7}$$

In the second step of Regret MCTS, regret is formulated as the dimensionless difference between a *low* design and an *high* design. In reality and for a sequence of decisions, the cumulative cost of *low* designs does not account for the cost of equipment replacement and sunk cost of discarded ones. On the other hand, *high* is a greedy approach to the electricity resource design problem. The Regret MCTS starts with the greedy *high* optimization that is solved in single-stage to construct a multi-stage design, then identifies the decision's regret from the comparison to the *low* lower cost.

In the third step of Regret MCTS, backpropagation is now included in the tree search. Eq. 7.7 consolidates the generation decisions D^g to minimize the regret of a decision by reducing the highest regret asset of the decision that was selected based on the second step and replacing it with the lowest one. A sequence of decisions is always an *high* optimization that takes over the electricity resource design of the previous decision unless the confidence bound satisfies the constraint of Eq. 7.8, where at time T, $C^{D_{T-1}}$ the total cost of the design based on decision D_{T-1} for the previous period $T-1$ and C^{D_T} is the cost of the design based on decision D_T.

$$C_{T-1}^{D_{T-1}}(\sum_t L_t^{T-1}) + \gamma C_T^{high}(\sum_t L_t^T) \geq C_{T-1}^{D_{T-1}}(\sum_t L_t^{T-1}) + \gamma C_T^{D_T}(\sum_t L_t^T) \tag{7.8}$$

In summary, the Regret MCTS algorithm starts with a *high* optimal solution at

time period T. Based on inputted probabilities of realizations, an external event is observed at time $T + \tau$ and a *low* optimal solution is identified for the same time period. Regret at the leaf-node regret is calculated as per Eq. 7.5. This process is repeated for every leaf-node explored node in the tree throughout the second step of Regret MCTS, and no decisions are made in this step. In the third step, decisions are selected using backpropagation that includes leaf-node regret values from the second step. Eq. 7.7 explores minimum regret designs for the initial decision without violating Eq. 7.8. To avoid exploration bias, ϵ of Eq. 7.7 introduces an error to the confidence bound, which allows for tree search that may result in low regret but was not sufficiently explored in the second step of Regret MCTS due to computational limitations that restrict the number of iterations. After N iterations, the Regret MCTS produces an initial decision that will minimize the probability of regret in leaf-nodes of the tree based on expected values of regrets in subsequent time steps and a measure of risk associated with the branching of the tree.

7.2.5 Design and dispatch optimization

An optimization model evaluates the cost-optimal design and dispatch of the generation mix subject to operational constraints. This is achieved by formulating and solving a linear program to solve for capacity expansion and dispatch of a power system network as described in Appendix B. The model objective is to minimize the total system cost which includes annualized resource expansion (generation, storage) and, operational costs as described in Eq. B.1. The operational constraints are: 1) hourly power balance of system (Eq. B.2), 2) time-dependent capacity constraints for generation resources (Eq. B.3), 3) battery storage state of charge, energy and power capacity limits (Eq. B.4 - B.9), 3) generation unit commitment (Eq. B.10 - B.14), 4) generation minimum and maximum power (Eq. B.15, B.16), 5) generation ramping limits (Eq. B.17, B.18) and 6) non-negativity constraints (Eq. B.22).

7.2.6 Transition-matrices formulation

Transition matrices define the space from which MCTS samples. Each stage of the decision tree has its decision-event probability matrix and decision-event load matrix. Both matrices are described by $d \cdot e$, where d is the number of possible decisions, and e is the number of possible events. The decision-event matrix informs the probability of realization of an event given a decision. The decision-event load matrix informs the system's peak load given a decision-event combination. Decision and event transitions are modeled in Fig. 7-3 and Fig. 7-4, respectively.

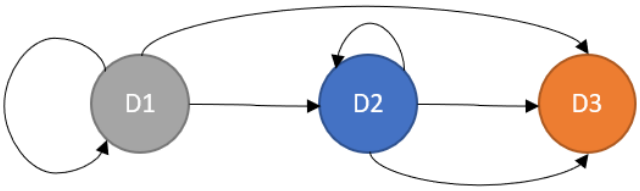

Figure 7-3: Decision transition graph; D1: business-as-usual, D2: grid upgrades, D3: minigrid.

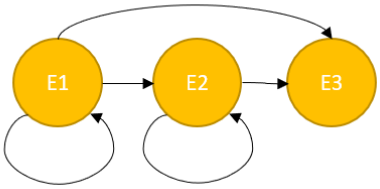

Figure 7-4: Events transition graph; E1: nothing, E2: grid reinforcement, E3: independent minigrid.

7.3 Results

We present a three-node planning case study of the Odisha design problem. Three decisions can be made (nothing, grid upgrades, and minigrid), and three external events can realize (nothing, grid reinforcement, and independent minigrid) as presented in the context section. For the three-node planning, the tree is a 9x9x9 states with 3x3x3 leaf-nodes, which incorporates the three external events within a decision node as highlighted in Fig. 7-2. We set the time interval τ to 5 years. Therefore the multi-stage problem has 15 years planning horizon with a parent node ($S1$) and two uncertain leaf-nodes ($S2$ and $S3$). We use cost projections from NREL's Annual Technology Baseline report [5] as presented in Appendix F Tables F.1 and F.2. For the uncertain leaf-nodes, decision-dependent external event realization probabilities are defined in Appendix F Table F.3, and the decision and external event dependent load growth projections are defined in Appendix F Table F.4. In the three-node problem, there are three leaf-nodes in $S1$ and nine leaf-nodes in $S2$ in addition to the parent node. We present detailed results on the calculation of the regret values for three sequences of decisions in Table 7.2 and Fig. 7-5.

Table 7.2: Decisions and events sequences.

	Sequence 1	**Sequence 2**	**Sequence 3**
Initial Decision	Install minigrid	Business-as-usual	Maintain grid
External Event 1	Nothing	Nothing	Nothing
Decision 1	Maintain minigrid	Business-as-usual	Maintain grid
External Event 2	Grid Reinforcement	Independent minigrid	Grid Reinforcement
Decision 2	Done	Install minigrid	Done

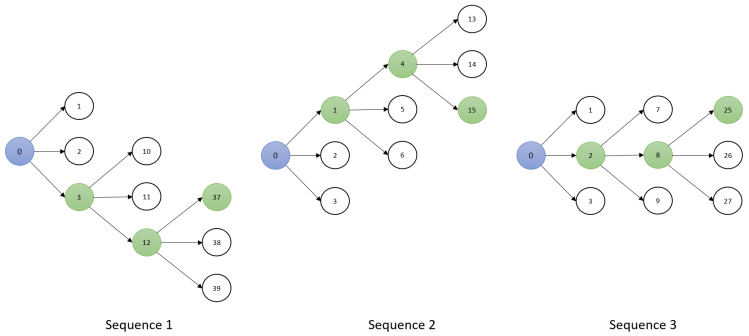

Figure 7-5: Three sequences example in decision tree.

Regret values calculated using Eq. 7.4 using the results of the optimization model and time domain reduction method of chapter 6. The optimization solves the design and dispatch problem on a reduced time series as seen in Fig. 7-6.

By using the various resulting regrets, the margin of regret (ρ_g) and asset regret (Υ_g^S) values for each leaf-node in the tree is solved as per the sequence of decisions and events. The sequencing is randomly generated using the decision and event dependent transition probabilities of Appendix F Table F.3.

Regret values are stored in each leaf-node of the tree as per the second step of Regret MCTS (see Fig. 7-7). Normal, Gamma, and Beta distributions are used to fit various leaf-node regret values. The selection is based on the backpropagation of the fitted distributions parameters and backward induction of a Monte Carlo Tree Search algorithm in the third step. Decision optimization is then carried out with Eq. 7.7 to properly size the system with an error parameter ϵ to eliminate exploration bias.

We apply the Regret MCTS on 9x9x9 (three nodes) and a 9x9x9x9 (four nodes) of the

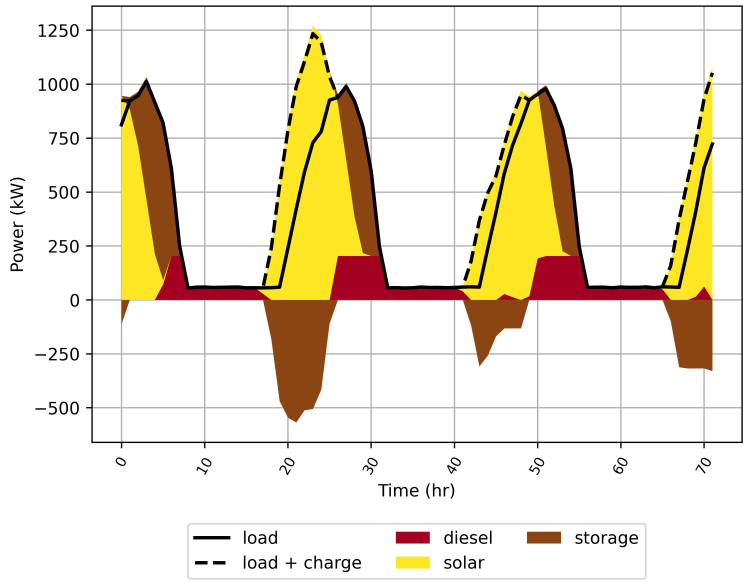

Figure 7-6: Three days dispatch result of minigrid design solution of optimization.

Table 7.3: Margin of regret (ρ_g) and asset regret (Υ_g^S) values for leaf-nodes $S1$ and $S2$ for the three defined sequences.

	Sequence 1				**Sequence 2**				**Sequence 3**			
	S1		S2		S1		S2		S1		S2	
	high	low	high	low	high	low	high	low	high	low	high	low
PV + Battery	0.1	0.08	0.1	0	0	0	0.1	0	0	0	0	0
Diesel	0.08	0.08	0.01	0	0	0	0.01	0	0	0	0	0
Grid	0.09	0.09	0.14	0.23	0.01	0.23	0.14	0.23	0.4	0.4	0.23	0.23
NSE	0	0	0	0	0.5	0	0	0	0	0	0	0
Regret	0.27	0.25	0.25	0.23	0.51	0.23	0.25	0.23	0.4	0.4	0.23	0.23

Odisha design problem. We apply each case to both endogenous and exogenous uncertainties. To highlight the algorithm's performance, we eliminate the endogenous uncertainty by converting the probability of demand growth and event realization decision-independent. Finally, we compare the Regret MCT to the standard UCB and the Bayesian UCB as seen in Fig. 7-8. Regret MCTS has similar performance to the other two models in the exogenous uncertainty-only cases. However, it outperforms the other two on an average error decision error basis, i.e., the true loss of the initial decision with the highest estimated mean value. In the case of exogenous uncertainty alone, the model performs similarly to the Bayesian UCB since the as-

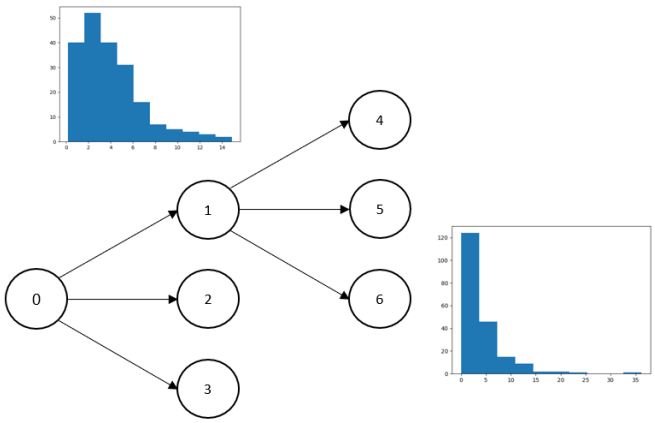

Figure 7-7: Resulting tree from selection-expansion-simulation only.

sumptions of Bayesian UCB regarding probability distributions hold. However, in the case of endogenous uncertainty, restricting the backpropagation of the tree to further collect regret values to fit for distribution functions parameterized by mean and kurtosis provides better insight on the endogenous uncertainty's impact on decision-making. Therefore, Regret MCTS outperforms Bayesian UCB and standard UCB. All simulations were run on MIT Supercloud High Performance Computer [216].

We compare the performance of Regret MCTS against UCB and Bayes on a reliability performance basis. We previously defined reliability by Eq. 7.3 as the discounted sum of demand served over multiple periods T. We use the four nodes case (9x9x9x9) to assess the models' performance in recommending solutions that maximize the expected reliability of the system by interrupting the simulations at 400, 600, 800, and 1,000 iterations for every tree search method. Additionally, we report a deterministic case that is based on random selection and the expected value of regret backpropagation only. We sample 100 trials at each interruption phase to present statistically rigorous results of the models' performances. We report these results in Fig 7-9. At 400 iterations, Regret MCTS can recommend decisions that will result in a reliable performance that UCB and Bayes need 1,000 and 800 iterations.

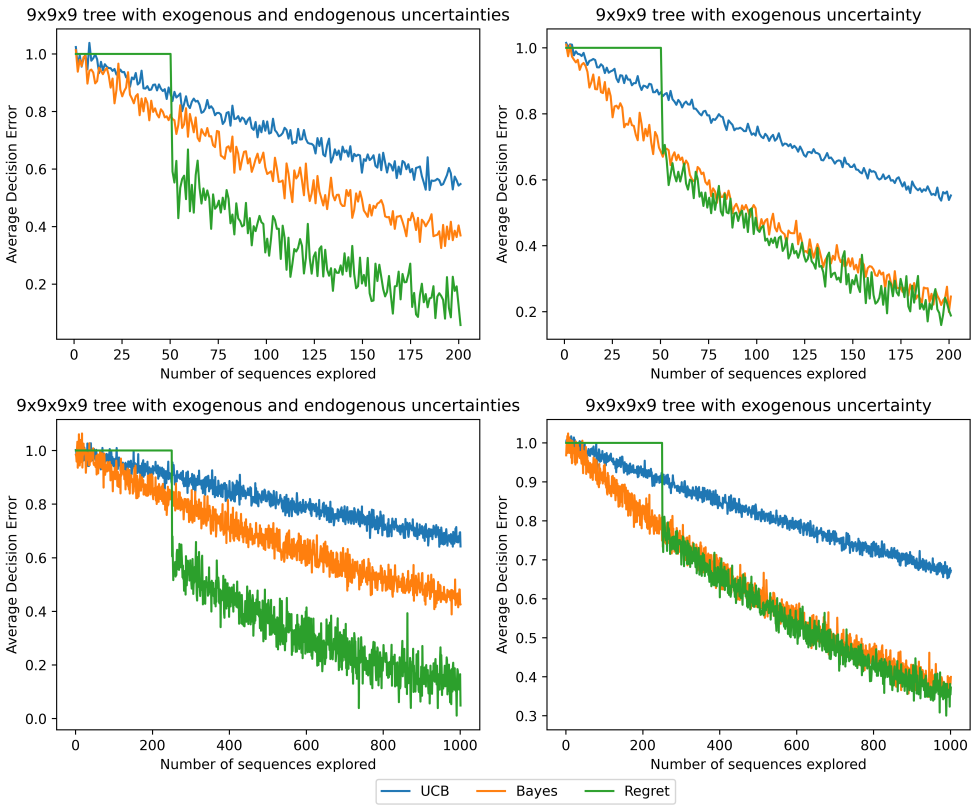

Figure 7-8: Average Decision error relative to ground truth results of standard UCB, Bayesian UCB and Regret MCTS.

7.4 Discussion

Due to the similarity in confidence bound formulation, when Regret MCTS is applied to exogenous uncertainty trees only, the performance is very close to Bayesian-UCB. Both perform better than UCB. On the other hand, when Type 2 endogenous uncertainty is considered in the scenario tree, Regret MCTS overperforms both UCB and Bayesian-UCB when the number of iterations possible is the same for all three algorithms. Therefore, splitting the MCTS four-step process (selection, expansion, simulation, and backpropagation) into two steps, as proposed by Regret MCTS, reduced the exploration bias that backpropagation has on the selection process. This bias may not be problematic with exogenous uncertainties where the regret value distributions are fixed. However, it is critical with endogenous uncertainties where more collected regret values may shift the distribution of regret. When enough iter-

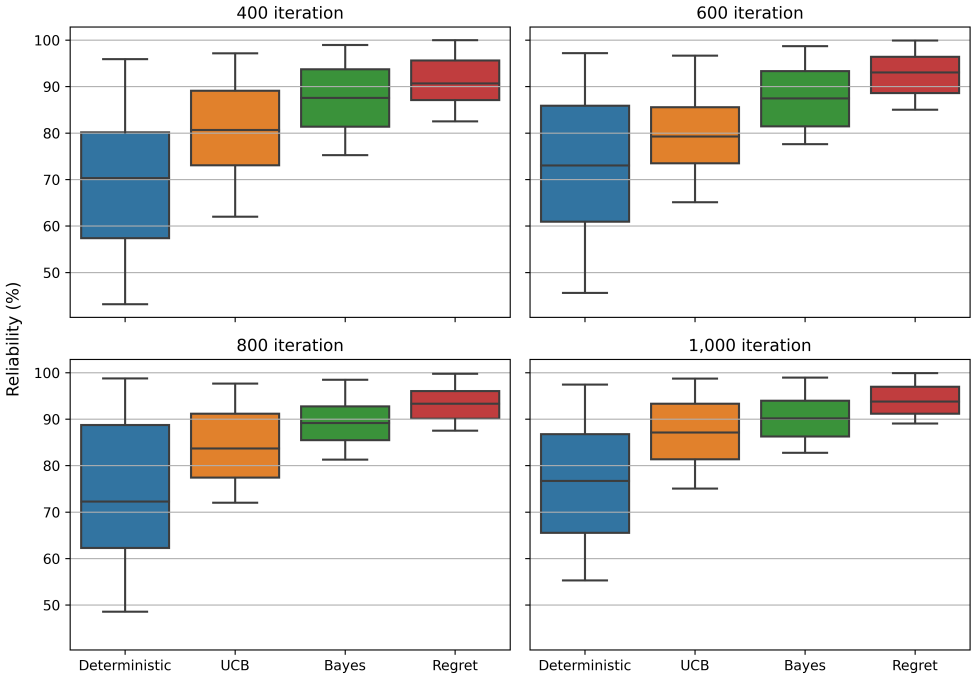

Figure 7-9: Box plots of sampled results from interrupted tree search at 400, 600, 800 and 1,000 iterations of four decision-making methods: Deterministic UCB, Bayes and Regret for the four nodes (9x9x9x9) case.

ations are dedicated to selection, expansion, and simulation, only Regret MCTS can decouple the uncertainty of the scenario tree from the decision-making. Hence, when backpropagation is enabled, Regret MCTS is better at selecting leaf-nodes, leading to lower overall regret of the multi-stage Odisha design problem. In summary, the reliability performance of Regret MCTS is higher than UCB and Bayes, which not only shows that Regret MCTS is a better heuristic tree search algorithm for the Odisha Design Problem, but it is also faster at doing so.

Our formulation of regret given low and high cost single-stage scenarios is leverages domain knowledge to reformulate the confidence bound of the MCTS. Additionally, we introduced the concept of burn-in to the MCTS and highlighted the performance improvement when endogenous uncertainty is considered in the scenario tree. We report statistically rigorous results to validate the Regret MCTS performance. Through Regret MCTS, we present a novel, efficient, and useful method to address the problem of decision-making under both exogenous and endogenous uncertainties.

Chapter 8

Conclusion

8.1 Contributions of the thesis

This thesis demonstrates the value of decision-making under various types of uncertainties through application to particular situations in the electric power system of Emerging Market and Developing Economy (EMDE) countries. However, all models and methodologies presented in this thesis can be adapted to other contexts (e.g., Developed Economies) and sectors (e.g., cooking, gas, heat, and water). The takeaways and discussions presented in each thesis chapter are particular to their described situations, nonetheless, generalizable key findings and contributions are reported in this chapter.

This thesis describes a way to determine future electricity demand from organic growth (or decay) of existing and new loads. We considered that the existing loads consist of a bundle of household appliances, while space-cooling and electric vehicles are new ones. The methodology of Chapter 2 enables the creation of a wide range of scenarios. Such scenarios make possible the examination of the uncertainty that stems from electricity demand growth. In the context of India, a large gap between the high AC efficiency and the baseline scenario creates various planning problems at both distribution and transmission levels, as further examined in Chapters 3 and 5.

Given the uncertainty in future electricity demand growth, Chapter 3 of this thesis presents a flexible framework that evaluates and compares two planning strategies of electricity distribution networks. The flexible valuation framework combines system

design optimization with multi-stage decision-making under uncertainty for distribution network planning. Given the unique situation of distribution networks in major metropolitan areas in India, the strategy of Chapter 3 analyzes the feasibility of short-duration battery storage as an alternative to network upgrades given the uncertainty in demand growth. We highlight that in the case of Delhi, non-wire alternative (NWA) battery storage is driven by capital savings in network investment planning, where we assume that the battery storage and the network lines are owned by the same entity (the utility). However, this may not always be the case and could result in different outcomes than presented in this thesis.

Chapter 4 presents another assessment of the impact of uncertainty in power systems distribution infrastructure planning. We apply an optimization-based approach to address the problem of decision-making under uncertainty in the context of a commercial load (the Wuse Market) in Abuja, Nigeria. This thesis demonstrates that the system's design can drastically change when uncertainty is accounted for. In the case of the Wuse Market, energy-dependent resources such as battery storage play a minimal role in the backup generation when grid supply is uncertain.

Chapter 5 presents the magnitude of the challenge of detailed electricity resource capacity expansion planning given a wide range of demand projections (from Chapter 2) and cost structures (e.g., technology cost and carbon price policy). To present meaningful results on decision-making, the optimization-based approach that is employed in the chapters 3, 4, and 5 do not scale, and the computational burden becomes large. Chapter 5 highlights the need for detailed capacity expansion planning in order to properly quantify the contribution of clean energy resources and the role of battery storage.

With the limitations and challenges presented in Chapter 5, Chapter 6 of this thesis demonstrates that reducing the size of the optimization does not mean a significant loss in the accuracy of the results. We develop an autoencoder architecture for better representative period selection to reduce the size of the capacity expansion planning linear programming problem. It is shown that reducing the optimization size without increasing the error is potentially helpful for the multi-stage stochastic problem of electricity resource capacity expansion planning.

Finally, Chapter 7 explores a learning-based algorithm that uses tree search methods to inform multi-stage decision-making in power system design. We leverage the representative period selection method presented in Chapter 6 to rapidly search the decision tree and arrive at a robust heuristic solution faster than classical methods. This work describes a practical approach to dealing with exogenous and endogenous

uncertainties in capacity expansion planning in the simplified real-world case study of Odisha in India.

In summary, this thesis answers the three questions presented in Chapter 1 through case studies and applications to situations that are unique to EMDE countries. Chapter 3 address the first question; *what are the uncertain factors that can impact the design of the different segments of the power supply?* Chapters 3, 4, and 5 are concerned with the second question; *How do these uncertain factors impact electric grid planning at the various stages — generation, transmission, and distribution?* Finally, Chapters 6 and 7 present rigorous methods to answer the third question; *How to handle uncertain factors in particular investment planning of electric grid infrastructure?*

8.2 Limitations and future work

The research presented in this thesis is limited in certain ways, and some associated investigations could grow out of these limitations. One explicit limitation is the amount of data available for this work. This thesis aims to address specific problems evident in EMDE countries regarding decision-making under uncertainty. Unfortunately, rigorous amounts of data on electricity demand and technology costs are challenging to find. Hence, the demand forecasting model of Chapter 2 is limited to new load consideration of only space-cooling and electric vehicles. Further work can be done to expand the electricity demand projection model proposed in Chapter 2 by including more electrification drivers such as electric cooking and consideration of technology phase-out such as incandescent light bulbs.

In Chapter 3, the flexible valuation framework focuses only on the use of battery storage for network deferrals without considering the added value of storage in ancillary services and arbitrage. On the other hand, the impact of peak-shifting battery storage on the overall system is also not considered. Investigating the deployment of storage at the distribution level can only be looked at in isolation when the displacement of demand is a small fraction of the total demand. However, as shown in the megacities aggregate analysis of NWA battery storage potential, adopting the flexible valuation framework strategy can significantly impact transmission level. Further work to couple the distribution-level study with a transmission level is necessary to portray NWA battery storage value fully. Additionally, the system design optimization does not consider generation expansion on the distribution network to allow for the proliferation of distributed energy resources (DER), such as rooftop solar PV. Future work may also include investigating NWA battery storage with

DERs on the distribution network. Finally, when modified, several input values, such as discount rate and demand growth, can impact the result (and potentially eliminate NWA battery storage). While Delhi shares similar traits with large EMDE cities (i.e., Cairo, Abuja), in terms of demand growth from space cooling and cost of capital, each environment is inherently different and must carefully be studied. The results demonstrate the value of storage as NWA under certain circumstances in the Indian context and call for further investigation of such planning strategies in EMDE. Furthermore, in the context of distribution-level planning, Chapter 4 leaves many challenges open to further discussion and can present alternative pathways, namely overnight improvement of grid supply. For example, the DER assets are stranded if grid supply is improved due to a dramatic external factor. However, in the case of solar PV and, to a lesser extent, storage, these assets can be replaced at transmission-level contrary to diesel generators. To identify the optimal ownership structure for an improved overall outcome, further analysis encompassing the large uncertainties of an interconnected minigrid is necessary.

We note several limitations of the work presented in Chapter 5 which addresses the impact of uncertainty at the bulk power system level. On the technology aspect, the cases evaluated do not consider deploying certain low-carbon resources like hydro, nuclear, or carbon capture and sequestration (CCS) equipped fossil-fuel power plants. Since investments in hydro and nuclear are not driven solely by economics, they are not considered in the model. Additionally, while hydro has been deployed in India recently, the expected increase in capacity is 12 GW [27, 19] which is minimal compared to the projected peak demand. Similarly, India's nuclear generation goals set by the Central Electricity Authority are also low and are therefore not likely to drastically change the modeling outcomes presented here [19]. CCS has not yet been considered in the Indian national electricity plan [19] or is considered a post-2050 technology which is out of scope of the presented results [241]. We also restrict short-duration battery storage technology to lithium-ion due to its rising popularity and declining costs. While lead-acid batteries are presently prevalent in India, we assume that lithium-ion dominates the market by 2040- where we note the mass deployment of grid-scale short duration storage. On the modeling aspect, the capacity expansion model [113] simulates grid economic dispatch, which does not reflect the current structure of the electricity system dispatch in India. Additionally, we do not take into consideration administrative transmission losses due to theft and other exogenous events when modeling simplified regional transmission flows (see Fig. 5-1). Finally, the resource availability maps used for variable renewable energy characterization were processed using satellite capacity factor data [242] which includes 14% system losses, with 1.5% corresponding to light-induced degradation [243]. However, ground

truth data may differ due to smog and poor air quality, particularly in the case of PV. This might lower the value of PV compared to our modeling outcomes.

The autoencoder-based representative period selection methodology presents a statistical analysis that supports autoencoders with clustering algorithms to reduce the error metric in electricity resource capacity expansion planning. Further parameter tuning and loss function modification can improve error in the results depending on the objective function and the clustering algorithm of choice. Additionally, the work presented in Chapter 6 is limited to the well documented and heavily used clustering algorithm k-means. Further work can be done to apply the proposed methodology to other clustering algorithms and compare performance. Additionally, we use data from the Electric Reliability Council of Texas as a source of ground truth with real noise in the time series. Application of a data-driven method on synthetic data such as the one presented in Chapter 2 may have sub-par performance relative to the standalone clustering algorithm.

Another limitation of the work presented in this thesis is the simplification of the real-world case study of Chapter 7. Data scarcity, especially in EMDE countries, makes it challenging to present a complete picture of the situation being addressed in this thesis. A set of assumptions regarding the inter-dependencies of decisions and external events force the discretization of the evolutionary probability space employed in the learning-based model. Future work may include continuous random variables that describe the uncertainties of the situation being analyzed.

8.3 Discussion

This thesis aims to shed light on the need for better modeling of decision-making in system design problems to pinpoint the challenges in policy design and the need for technological breakthroughs, both at a reasonable cost to overcome the challenges of uncertainty. This thesis demonstrates some of the advantages of contextualizing the problem of decision-making under uncertainty to achieve results and performance improvements. Appropriate modeling of the situation is being addressed leads to higher confidence in the design decisions and conclusions. Furthermore, when electricity resources are modeled in detail, as in this thesis, regulatory and execution challenges of electric power system operation and expansion planning are better appreciated in achieving clean and reliable power.

Appendix A

Supplementary information — Chapter 2

Additional information pertaining to Chapter 2 which includes a description of the data record that is published online at [3], and additional reported results of electricity demand projections for India.

A.1 Data Records

Data is published in [3]. The path leading to a CSV file indicate the scenario corresponding to the results of that file. Breakdown of the folder hierarchy listed as:

1. GDP Growth: slow, stable, rapid
2. EV charging: home, work, public
3. Cooling: baseline, efficient
4. Type: detailed, summary

The *detailed* results are tables of the itemized hourly demand profile of each considered scenario; all files will produce 8760 rows (number of hours in a year). The *summary* are tables of the itemized annual energy consumption for the considered years; all files will produce seven rows (number of considered future years). Both file types are itemized the same way as per Table A.1. The path of each file is the reference to the specific scenario the data in the tables represents. For example

SR.csv file under *slow/home/efficient/summary* is the summary file of the case of slow economic growth, home EV and energy efficient air conditioning consumption.

Column Header	Description
DateTime	Hourly or yearly time resolution
Base	Business-as-usual model resulting demand
Com AC	Commercial Air Conditioning demand
Res AC	Residential Air Conditioning demand
E2W	Electric Two-Wheelers demand
E3W	Electric Three-Wheelers demand
E4W	Electric Four-Wheelers demand

Table A.1: Output data headers descriptor

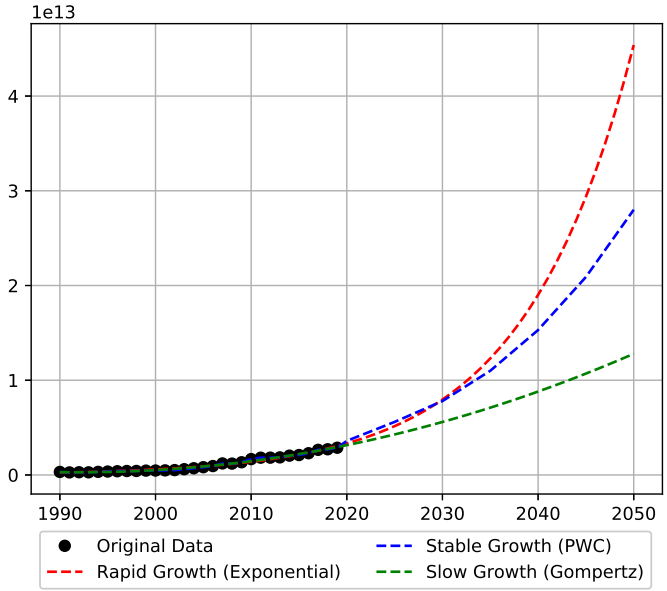

Figure A-1: India's GDP curve-fit and forecasting to 2050

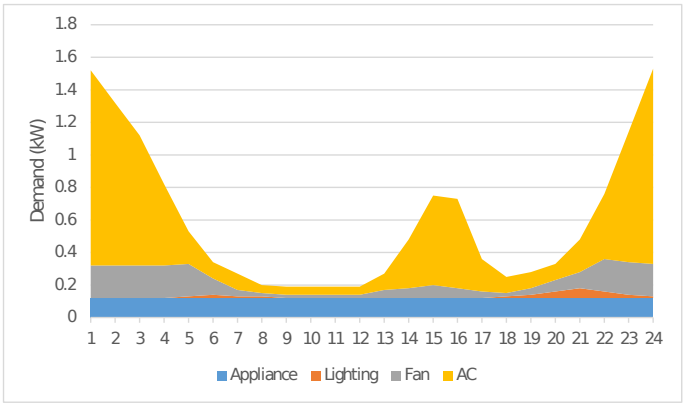

Figure A-2: Residential survey categorized hourly demand profile

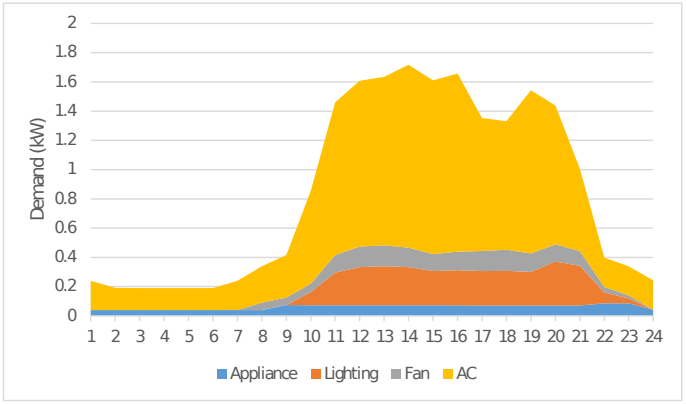

Figure A-3: Commercial survey categorized hourly demand profile

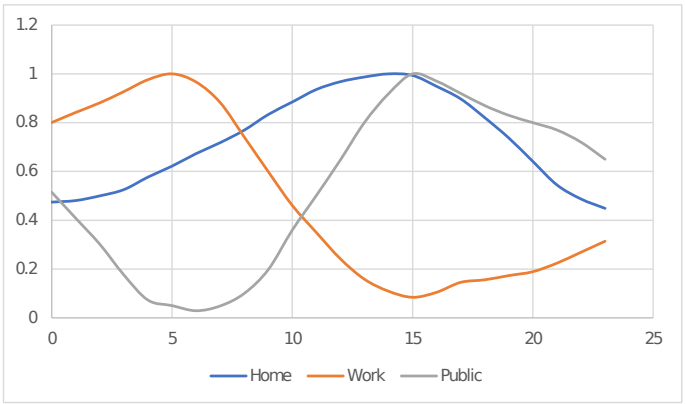

Figure A-4: Normalized sample charging profile schemes

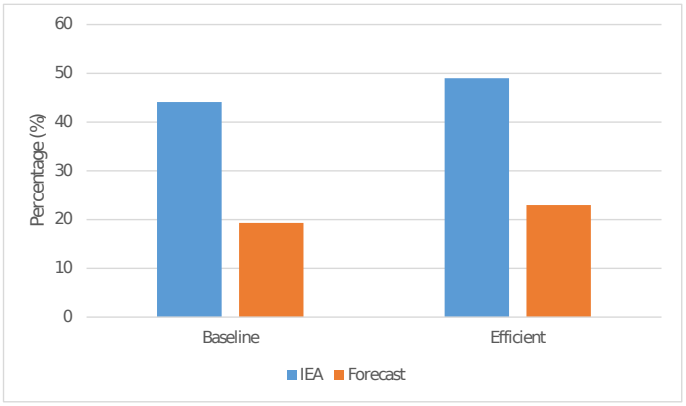

Figure A-5: 2050 cooling demand contribution to peak results comparison with IEA's Future of Cooling

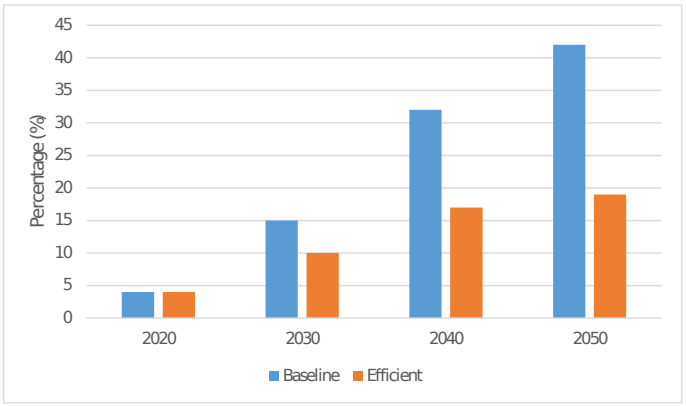

Figure A-6: Cooling demand contribution to peak demand

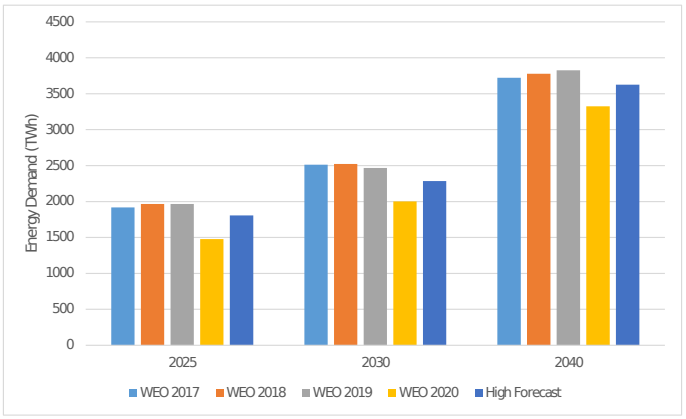

Figure A-7: Results comparison with stated policy World Energy Outlook projections

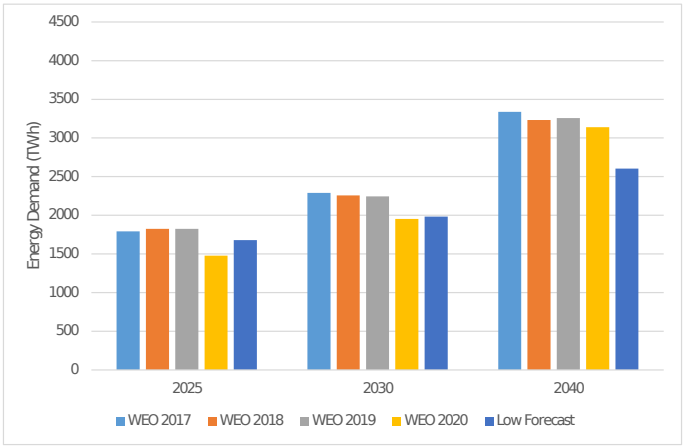

Figure A-8: Results comparison with sustainable policy World Energy Outlook projections

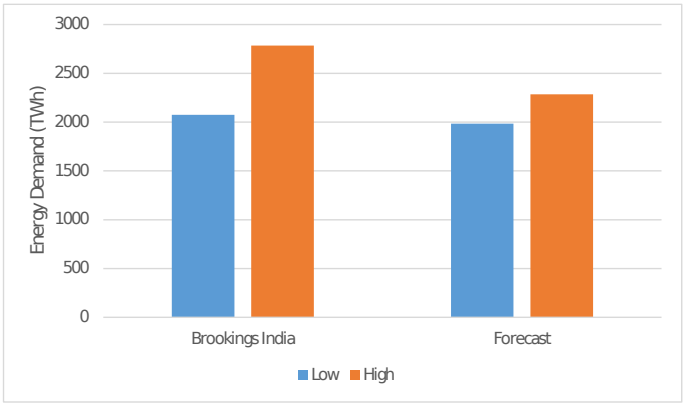

Figure A-9: Results comparison with Brookings India 2030 projections

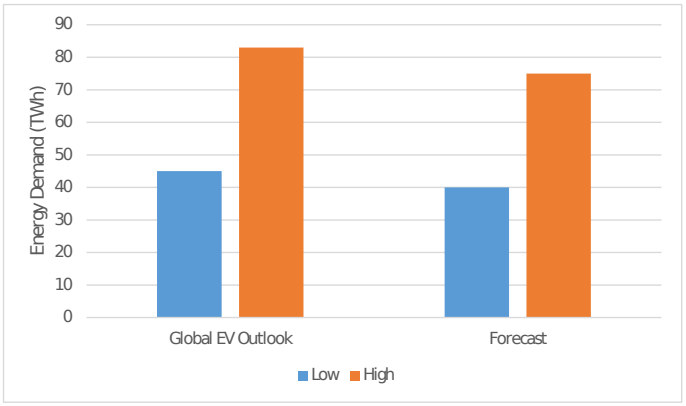

Figure A-10: Electric Vehicle demand results comparison with IEA's Global EV Outloook

Appendix B

Optimization model

Linear programming formulation of the capacity expansion model used in this thesis with supporting nomenclature in section B.8 Tables B.1, B.2, B.3, and B.4.

B.1 Objective function

Capital and operational cost object of all generation types, battery storage degradation, network expansion, and value of lost load.

$$
\begin{aligned}
\min \sum_{a} \sum_{z} \big(& \Omega^{size}_{a,z} \cdot (C^{I}_{a,z} + C^{F}_{a,z}) \\
& + \Omega^{energy}_{a,z} \cdot C^{e}_{a,z} \cdot (1 + C^{d}_{a,z}) \\
& + \Omega^{charge}_{a,z} \cdot (C^{c}_{a,z} + C^{Fc}_{a,z}) \\
& + \sum_{t} w_t \cdot \pi_{a,t,z} \cdot (C^{V}_{a,z} + C^{Vf}_{a,z}) \\
& + \sum_{t} w_t \cdot \Psi^{charge}_{a,t,z} \cdot C^{Vc}_{a,z} \\
& + \sum_{t} w_t \cdot n_{a,t,z} \cdot C^{start}_{a,z} \big) \\
+ \sum_{z} \sum_{t} & w_t \cdot \chi_{t,z} \cdot C^{\chi}_{z}
\end{aligned}
\tag{B.1}
$$

B.2 Demand balance constraint with non-served energy

$$\begin{aligned} \text{s.t.} \quad L_{t,z} = \sum_{a} (\pi_{a,t,z} + \Psi^{discharge}_{a,t,z} \\ - \Psi^{charge}_{a,t,z}) + \chi_{t,z} + \phi_{t,z,z'} \\ \forall z, z' \in Z \quad \forall t \in T \end{aligned} \tag{B.2}$$

B.3 Variable renewable energy availability constraint

$$\begin{aligned} \pi_{a,t,z} \leq \Omega^{size}_{a,z} \cdot A_{a,t,z} \\ \forall a \in R, \forall t \in T, \forall z \in Z \end{aligned} \tag{B.3}$$

B.4 Storage constraints

Battery storage state of charging, energy and power capacity limits.

$$\begin{aligned} \Gamma_{a,t,z} = \Gamma_{a,t-1,z} - \frac{\psi^{discharge}_{a,t,z}}{\eta^{discharge}_{a,z}} + \eta^{charge}_{a,z} \cdot \psi^{charge}_{a,t,z} \\ \forall a \in S, \forall t \in T^{interior} \in T, \forall z \in Z \end{aligned} \tag{B.4}$$

$$\begin{aligned} \Gamma_{a,t,z} = \Gamma_{a,t^{period},z} - \frac{\psi^{discharge}_{a,t,z}}{\eta^{discharge}_{a,z}} + \eta^{charge}_{a,z} \cdot \Psi^{charge}_{a,t,z} \\ \forall a \in S, \forall t \in T^{start}, \forall z \in Z \end{aligned} \tag{B.5}$$

$$\begin{aligned} \Gamma_{a,t,z} \leq \delta_{a,z} \cdot \Omega^{energy}_{a,z} \\ \forall a \in S, \forall t \in T, \forall z \in Z \end{aligned} \tag{B.6}$$

$$\Psi_{a,t,z}^{charge} \leq \Omega_{a,z}^{charge} \quad \forall a \in S, \forall t \in T, \forall z \in Z \tag{B.7}$$

$$\Psi_{a,t,z}^{charge} + \Psi_{a,t,z}^{discharge} \leq \Omega_{a,z}^{charge} \quad \forall a \in S, \forall t \in T, \forall z \in Z \tag{B.8}$$

$$\Psi_{a,t,z} \leq \Gamma_{a,t-1,z} \quad \forall a \in S, \forall t \in T, \forall z \in Z \tag{B.9}$$

B.5 Thermal generation constraints

B.5.1 Generation unit commitment

$$v_{a,t,z} \leq \frac{\Omega_{a,z}^{size}}{\Omega_{a,z}^{unit}} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.10}$$

$$u_{a,t,z} \leq \frac{\Omega_{a,z}^{size}}{\Omega_{a,z}^{unit}} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.11}$$

$$n_{a,t,z} \leq \frac{\Omega_{a,z}^{size}}{\Omega_{a,z}^{unit}} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.12}$$

$$v_{a,t,z} = v_{a,t-1,z} + u_{a,t,z} - n_{a,t,z} \quad \forall a \in M, \forall t \in T^{interior}, \forall z \in Z \tag{B.13}$$

$$v_{a,t,z} = v_{a,t+t^{period}-1,z} + u_{a,t,z} - n_{a,t,z} \quad \forall a \in M, \forall t \in T^{start}, \forall z \in Z \tag{B.14}$$

B.5.2 Generation minimum and maximum power output

$$\pi_{a,t,z} \geq \rho_{a,z}^{min} \cdot \Omega_{a,z}^{size} \cdot v_{a,t,z} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.15}$$

$$\pi_{a,t,z} \leq \rho_{a,z}^{max} \cdot \Omega_{a,z}^{size} \cdot v_{a,t,z} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.16}$$

B.5.3 Generation ramping limits

$$\pi_{a,t,z} - \pi_{a,t-1,z} \leq \Omega_{a,z}^{size} \cdot \kappa_{a,z}^{up} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.17}$$

$$\pi_{a,t-1,z} - \pi_{a,t,z} \leq \Omega_{a,z}^{size} \cdot \kappa_{a,z}^{down} \quad \forall a \in M, \forall t \in T, \forall z \in Z \tag{B.18}$$

B.6 Network expansion

B.6.1 Direct-current power-flow approximation through line susceptance and voltage deviation

$$\phi_{t,z,z'} = B_{z,z'} \cdot (\theta_{t,z} - \theta_{t,z'}) \quad \forall z, z' \in Z, \forall t \in T \tag{B.19}$$

$$\begin{aligned} \phi_{t,z,z'} &\leq \Phi_{z,z'}^{max} \\ \forall i, j \in Z, \forall t \in T \end{aligned} \tag{B.20}$$

$$\begin{aligned} \phi_{t,z,z'} &\geq -\Phi_{z,z'}^{max} \\ \forall i, j \in Z, \forall t \in T \end{aligned} \tag{B.21}$$

B.7 Non-negativity constraints

$$\begin{aligned} \Omega_{a,z}^{charge}, \Omega_{a,z}^{discharge}, \Omega_{a,z}^{size} &\geq 0 \\ \pi_{a,t,z}, \chi_{t,z}, \Gamma_{a,t,z} &\geq 0 \\ \Psi_{a,t,z}^{charge}, \Psi_{a,t,z}^{discharge} &\geq 0 \\ \theta_{t,z,z'}^{min} \leq \theta_{t,z} - \theta_{t,z'} &\leq \theta_{t,z,z'}^{max} \\ \theta_{t,1} &= 0 \\ \phi_{t,z,z'} &\in \mathbb{R} \\ \forall z, z' \in Z, \forall t \in T, \forall a \in M \cup R \cup S & \end{aligned} \tag{B.22}$$

B.8 Optimization model nomenclature

Table B.1: Set nomenclature of the electricity resource capacity expansion model.

Set	Description
R	Variable renewable energy resources
S	Battery storage resources
M	Thermal generation resources
Z	Power system zones
T	Hours of operation in a model period
$T^{interior}$	Interior time steps inside T
T^{start}	First time step of T

Table B.2: Index nomenclature of the electricity resource capacity expansion model.

Index	Description
a	Generation resource
t	Time step
z, z'	Load zone

Table B.3: Parameter nomenclature of the electricity resource capacity expansion model.

Parameter	Description
t^{period}	Total number of time steps
w_t	Time step weight
C^I	Investment cost (USD/MW)
C^e	Energy capacity investment cost (USD/MWh)
C^c	Charge investment cost (USD/MWh)
C^d	Battery energy capacity degradation per annum (%)
C^F	Fixed operational cost (USD/MW-yr)
C^{Fc}	Fixed operational charge cost (USD/MWh-yr)
C^V	Variable cost (USD/MWh)
C^{Vf}	Fuel cost (USD/MWh)
C^{Vc}	Variable charge cost (USD/MWh)
C^χ	Value of lost load (USD/MWh)
C^{start}	Startup cost (USD/MW)
μ	Storage round-trip efficiency
A	Generation availability profile
Ω^{max}	Maximum generation capacity
ρ^{min}	Minimum generation power
ρ^{max}	Maximum generation power
κ^{up}	Ramp up limit
κ^{down}	Ramp down limit
B	Line susceptance
Φ^{max}	Maximum line power capacity
δ	Storage depth of discharge
η	Storage efficiency
Ω^{unit}	Generation unit capacity
$\theta^{min}, \theta^{max}$	Minimum and maximum voltage angle

Table B.4: Variable nomenclature of the electricity resource capacity expansion model.

Variable	**Description**
$\Omega_{a,z}^{size}$	Generation capacity
$\Omega_{a,z}^{energy}$	Energy capacity
$\Omega_{a,z}^{charge}$	Charge capacity
$\pi_{a,t,z}$	Power output
$v_{a,t,z}$	Number of units committed
$u_{a,t,z}$	Number of startup decisions
$n_{a,t,z}$	Number of shutdown decisions
$\Psi_{a,t,z}^{charge}$	Storage charge
$\Psi_{a,t,z}^{discharge}$	Storage discharge
$\chi_{t,z}$	Non-served energy
$\Gamma_{a,t,z}$	Storage state of charge
$\theta_{t,z}$	Line voltage angle
$\phi_{t,z,z'}$	Line power flow

Appendix C

Supplementary information — Chapter 4

Supporting capital, fuel, and operational cost results of the minigrid-under-the-grid designs for the three time periods 2020, 2025, and 2030 of Chapter 4.

Table C.1: Capital expenditure results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030.

	2020	**2025**	**2030**
ES	$ 272,461	$ 0	$ 6,490
GF	$ 311,101	$ 154,912	$ 252,499
GF O1	$ 282,682	$ 179,753	$ 252,893
GF O2	$ 297,326	$ 176,699	$ 246,653
GF LS	$ 328,771	$ 174,984	$ 213,025
GF LS O1	$ 322,573	$ 182,490	$ 212,630
GF LS O2	$ 334,659	$ 178,566	$ 209,674

Table C.2: Annual fuel cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030.

	2020	**2025**	**2030**
ES	$ 179,569	$ 315,043	$ 494,694
GF	$ 120,821	$ 49,271	$ 718
GF O1	$ 169,263	$ 51,607	$ 718
GF O2	$ 213,112	$ 62,972	$ 718
GF LS	$ 74,520	$ 11,575	$ -
GF LS O1	$ 96,991	$ 12,161	$ -
GF LS O2	$ 138,163	$ 22,858	$ -

Table C.3: Annual fixed operation and maintenance cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030.

	2020	**2025**	**2030**
ES	$ 21,000	$ 16,250	$ 14,516
GF	$ 24,317	$ 32,065	$ 46,044
GF O1	$ 20,265	$ 31,986	$ 46,044
GF O2	$ 22,250	$ 32,763	$ 46,044
GF LS	$ 26,677	$ 32,989	$ 44,301
GF LS O1	$ 25,219	$ 33,060	$ 44,301
GF LS O2	$ 27,045	$ 33,532	$ 44,301

Table C.4: Annual variable operation and maintenance cost results from optimization of the seven scenarios for modeled periods 2020, 2025, 2030.

	2020	**2025**	**2030**
ES	$ 192,372	$ 287,061	$ 416,344
GF	$ 171,765	$ 248,960	$ 201,865
GF O1	$ 161,313	$ 247,665	$ 201,865
GF O2	$ 122,372	$ 232,112	$ 201,865
GF LS	$ 182,505	$ 212,479	$ 115,292
GF LS O1	$ 173,994	$ 211,487	$ 115,292
GF LS O2	$ 139,332	$ 199,243	$ 115,292

Appendix D

Supplementary information — Chapter 5

Supplementary information regarding the supply-side study of the Indian bulk power system in India of Chapter 5. Result figures and supporting analyses are reported below. Additionally, input information such as cost structure, scenario definition, and abbreviations used in the capacity expansion model are available here.

Table D.1: DLS system cost of electricity comparison

Reference case			
	Reference case	DLS case	DLS SCOE
2030	25.9	25.8	0.4
2040	29.5	19.5	0.5
Low cost storage case			
	DLS SCOE	DLS SCOE	Low storage SCOE
2030	26.5	0.4	26.4
2040	17.7	0.5	18

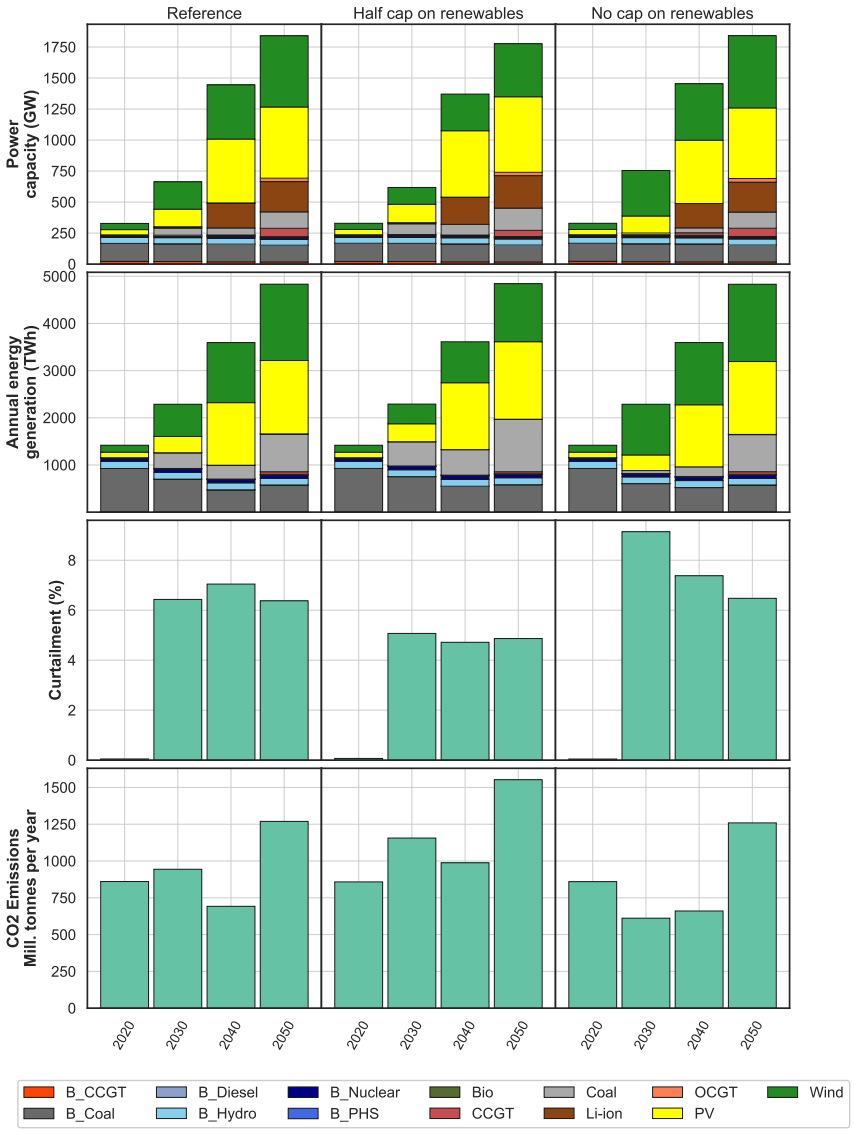

Figure D-1: Reference case model outcomes with alternative assumptions about decadal renewables installation limits. Reference = decadal installation limits as shown in Table D.13. Half cap = decadal installation limits are 0.5 of the reference values. No cap = no decadal installation limits.

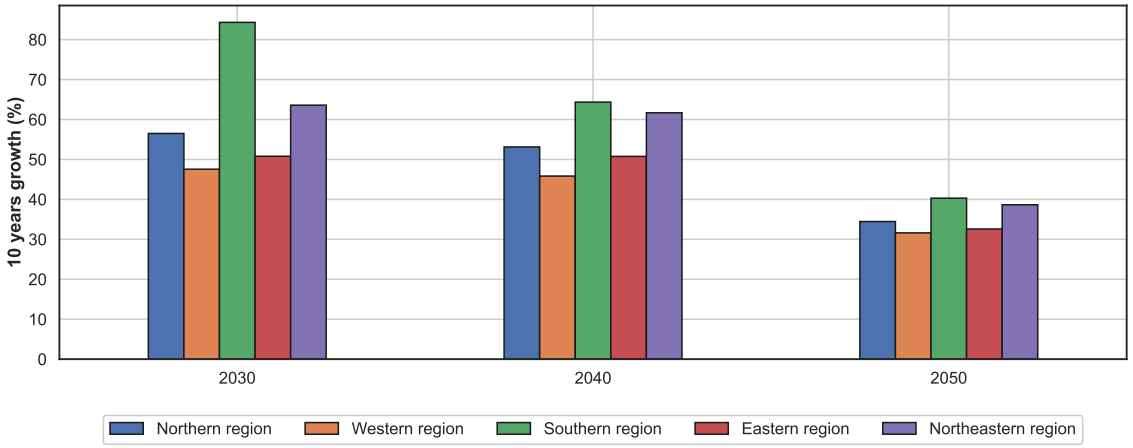

Figure D-2: Regional demand growth projections for India as per demand forecasting model results [3]

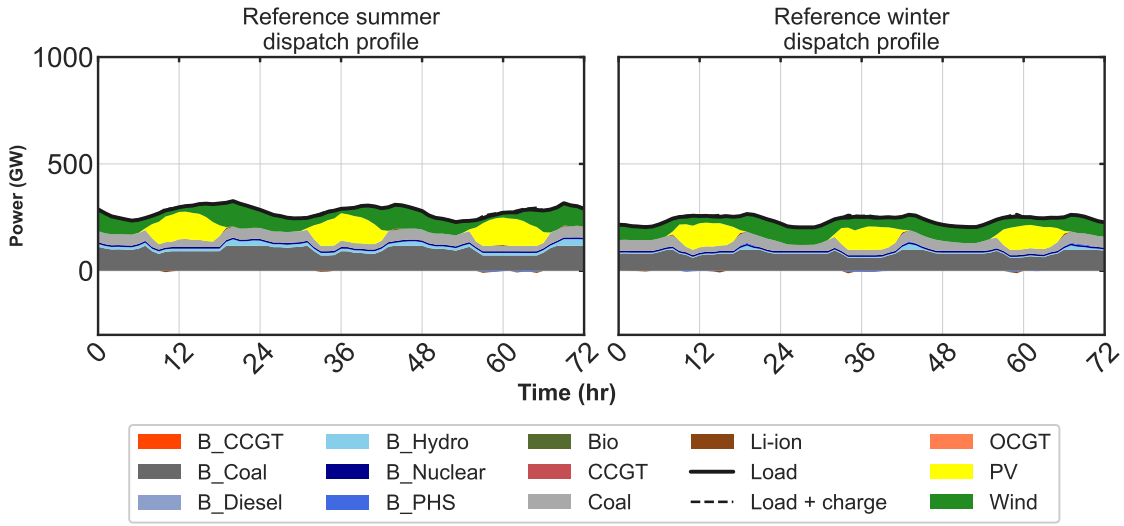

Figure D-3: Hourly generation dispatch for reference winter and summer load profiles for 2030. Technology names and their respective abbreviations in Supplementary Table D.4

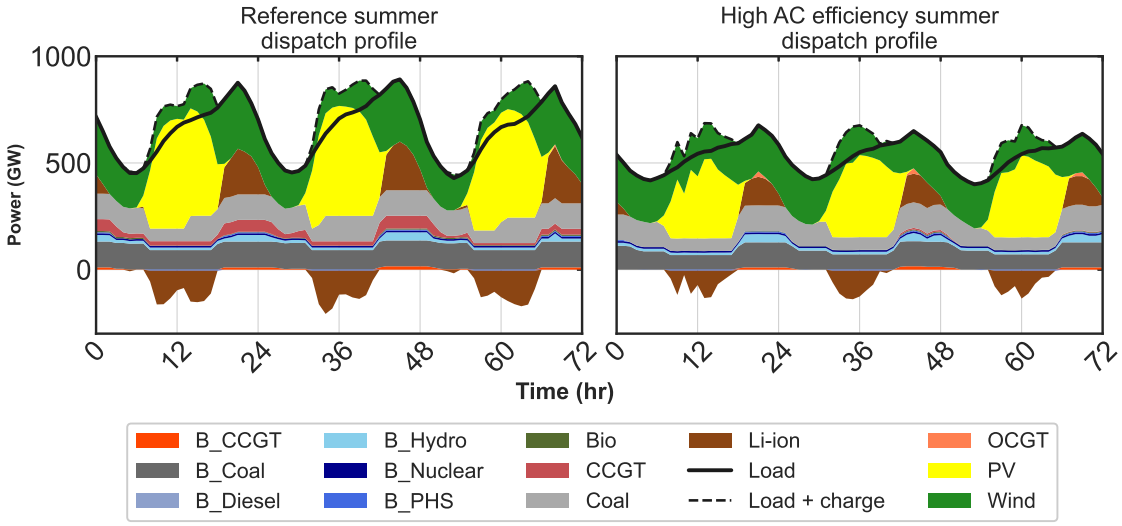

Figure D-4: Hourly generation dispatch for summer reference and high AC efficiency load profiles for 2050. Technology names and their respective abbreviations in Supplementary Table D.4

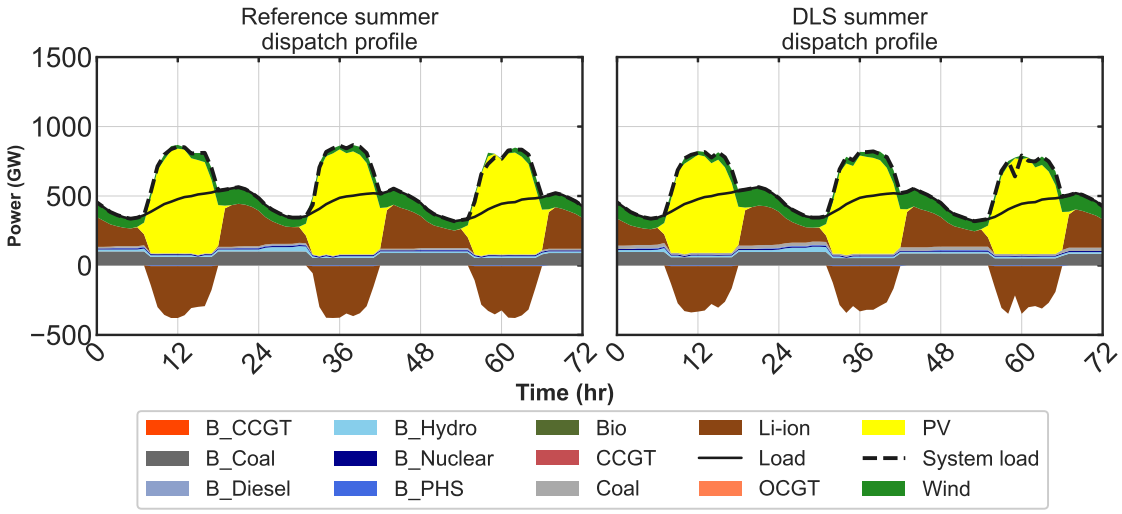

Figure D-5: Hourly generation dispatch for summer reference and DLS load profiles for 2040. Technology names and their respective abbreviations in Supplementary Table D.4

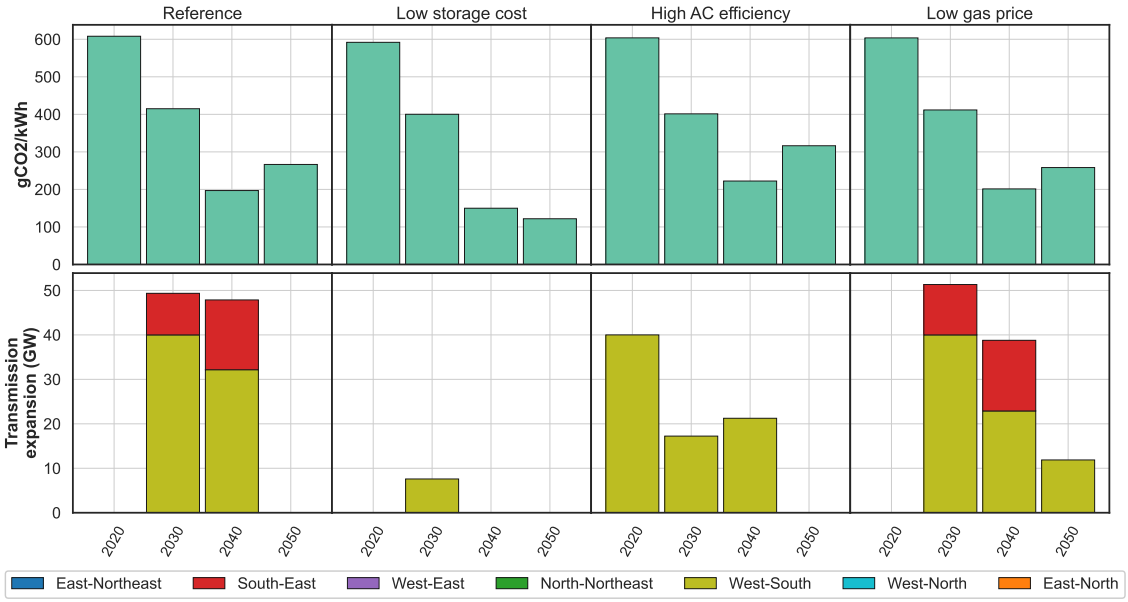

Figure D-6: Emissions intensity and Transmission expansion outcomes for modeled cases considered in the main text

.

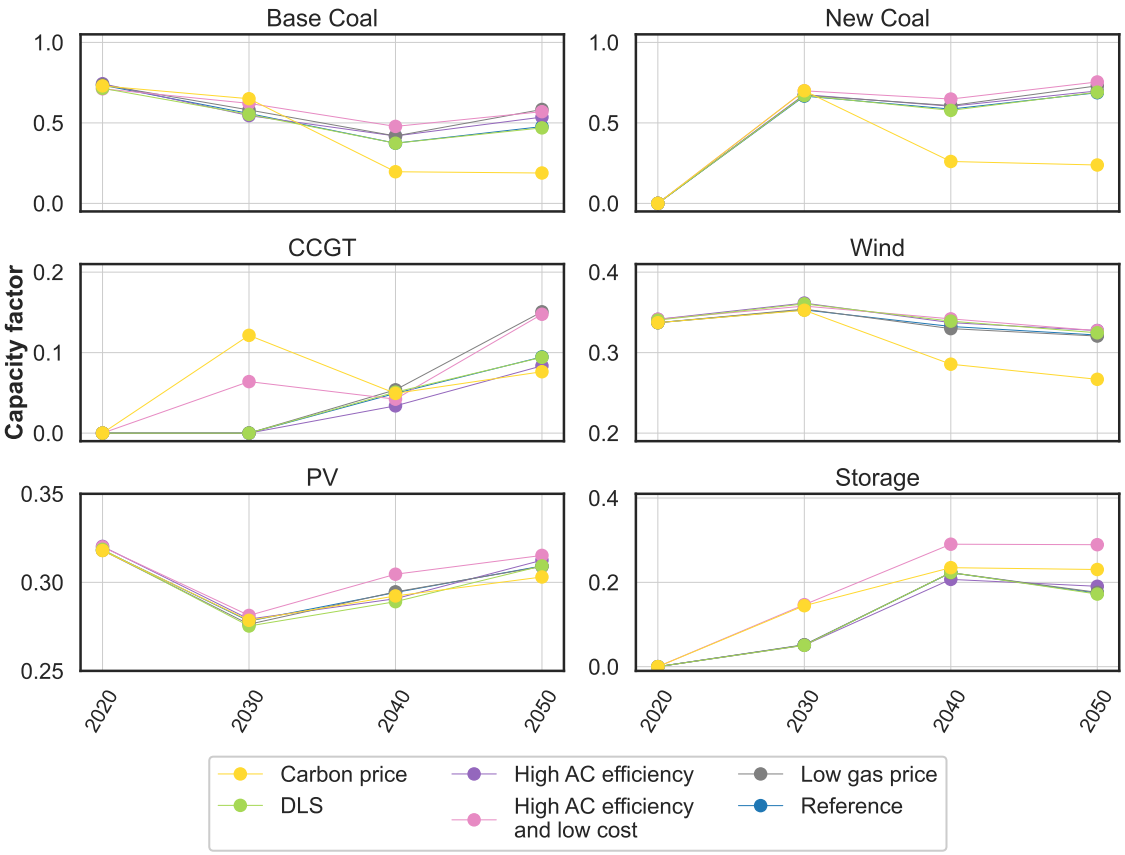

Figure D-7: Technology capacity factors across the scenarios over different modeling periods

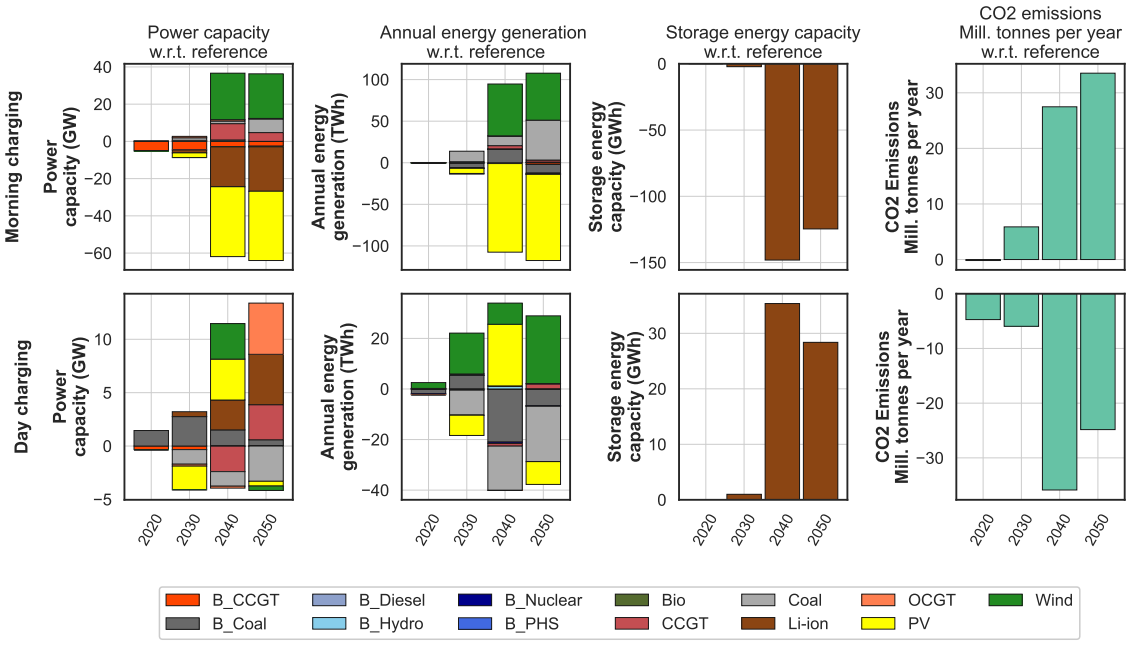

Figure D-8: Impact of Morning (top) and day (bottom) electric vehicle (EV) charging schemes relative to evening EV charging scheme under the reference case

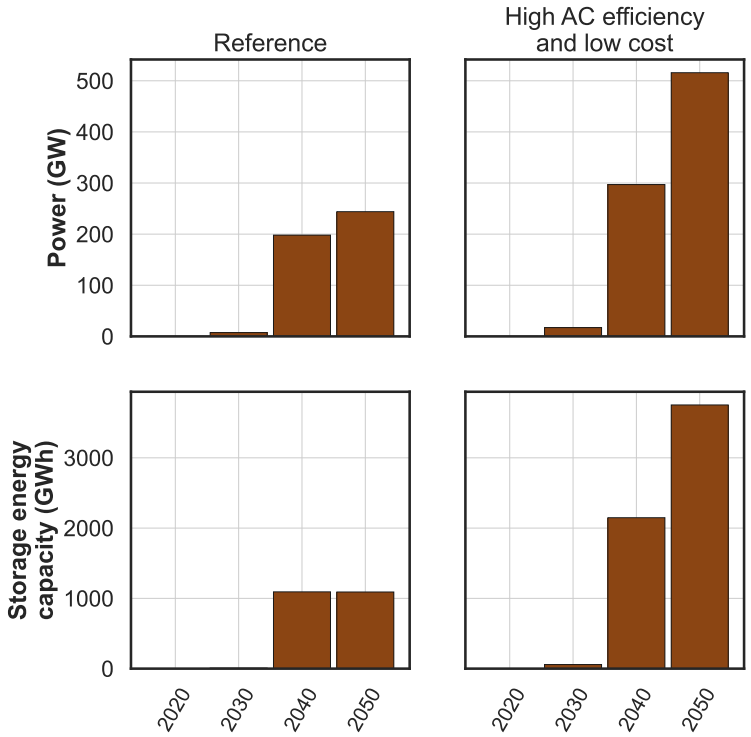

Figure D-9: Storage power and energy capacity deployment trends in the reference and sensitivity cases

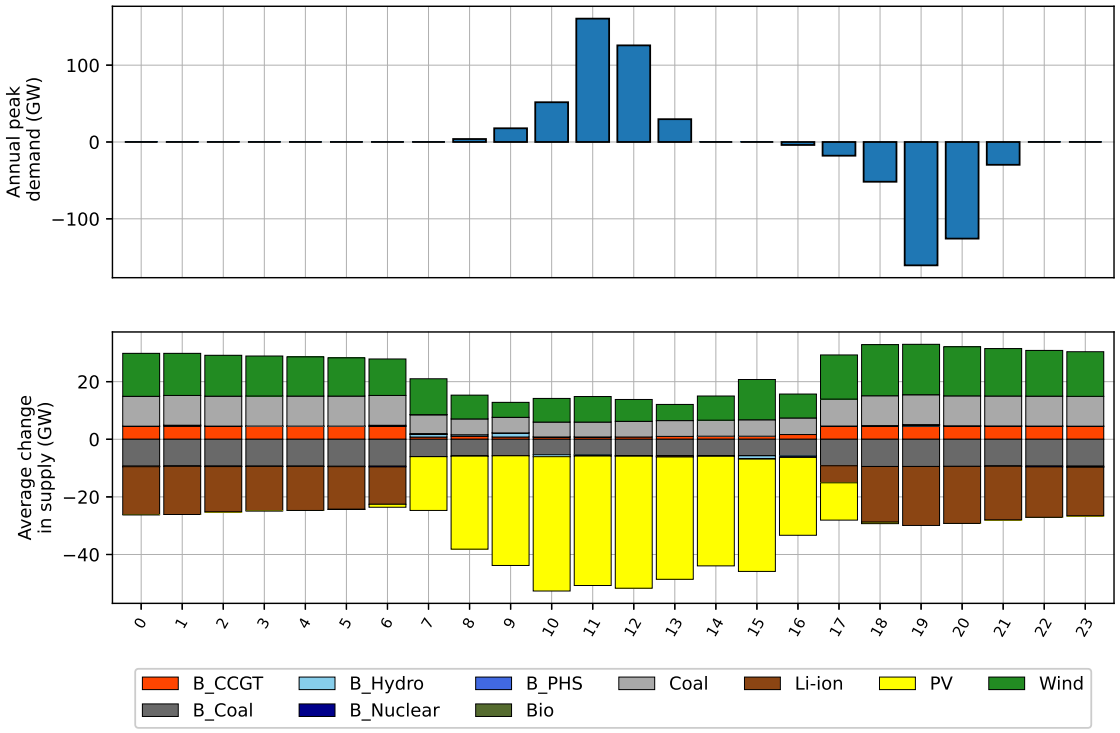

Figure D-10: Yearly average DLS impact on supply and demand per hour for 2030

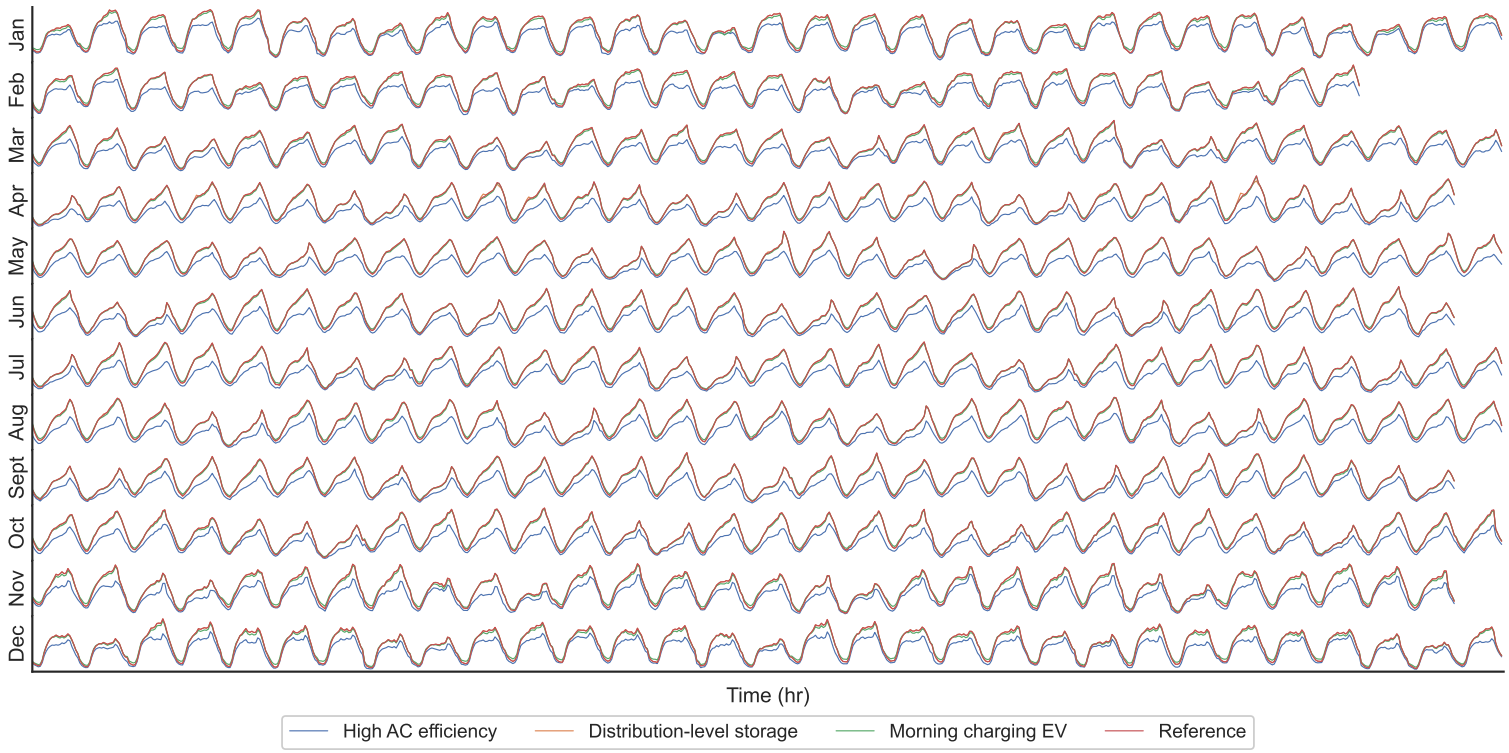

Figure D-11: Hourly load profile by month in 2040 across various demand scenarios considered here

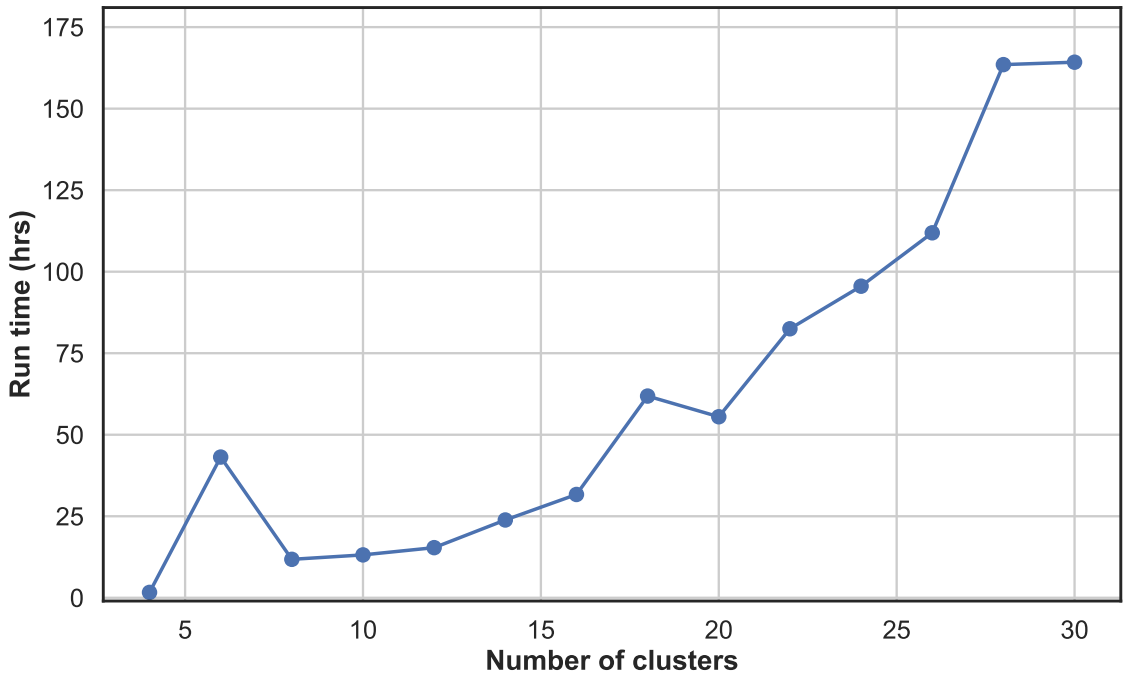

Figure D-12: Indicative GenX capacity expansion optimization model run time with respect to number of clustered weeks. Outputs based on reference case.

Table D.2: Carbon price scenarios

Year	Carbon price
2020	0
2030	20
2040	33
2050	53

$$SCOE_y = \frac{AIC_y + \sum_{p \in M} AIC_{y,p} + FOM_y + VAR_y + U_y + S_y}{D_y} \quad \text{(D.1)}$$

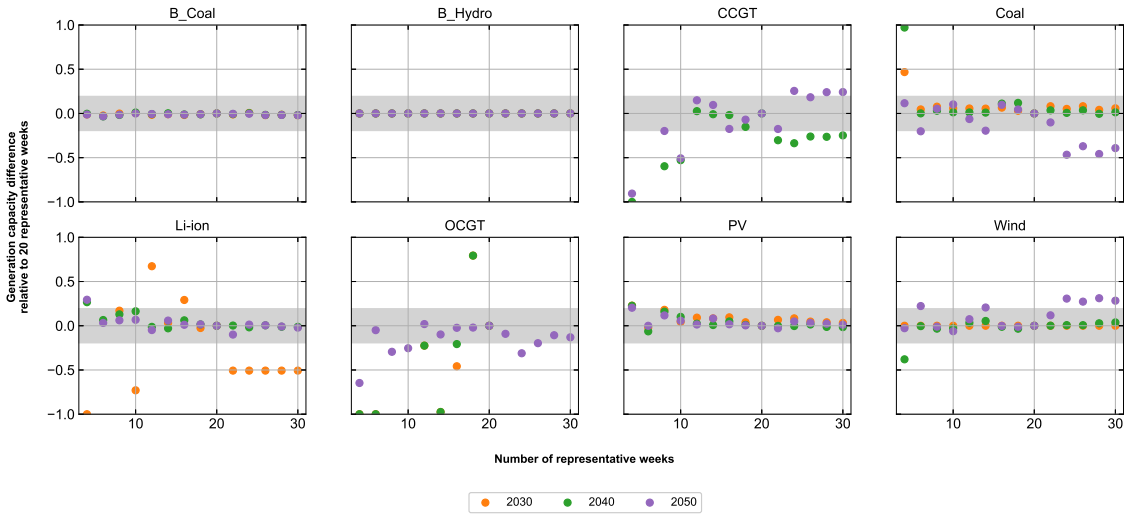

Figure D-13: Generation capacity difference relative to 20 representative weeks for the reference case.

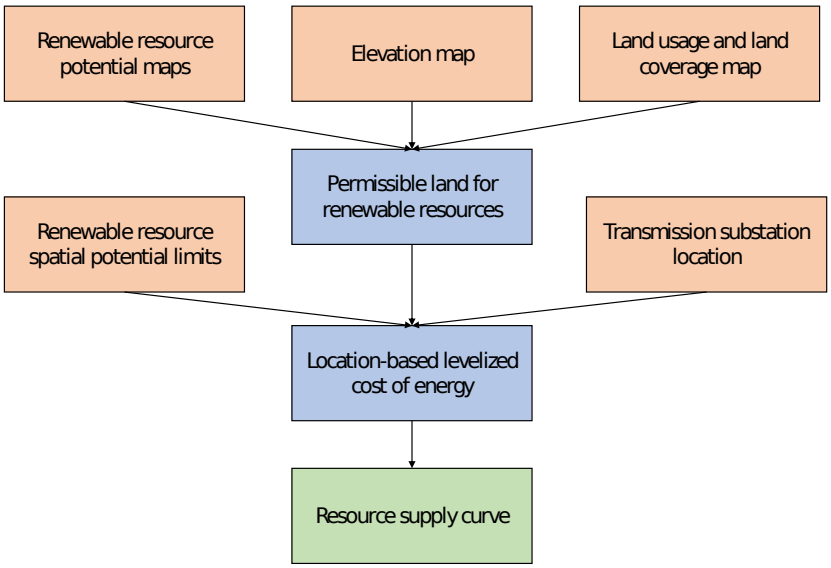

Figure D-14: Renewable resources supply curve calculation flowchart

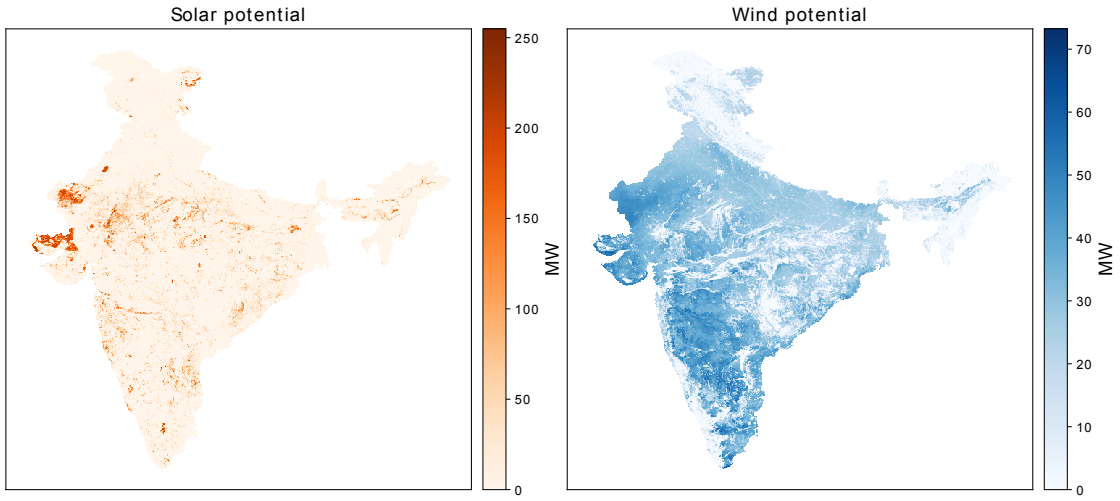

Figure D-15: Deployable solar and wind resources potential maps

Table D.3: Abbreviations for Eqn. D.1

y	Model period	
t	Technology	Table D.4
p	Previous model period	
M	Set of model periods	Table 5.2
$SCOE$	System cost of electricity	USD/MWh
AIC	Annualized investment cost	USD
FOM	Total fixed operation and maintenance cost	USD
VAR	Variable cost	USD
U	Total fuel cost	USD
S	Total startup cost	USD
D	Total electricity demand	MWh

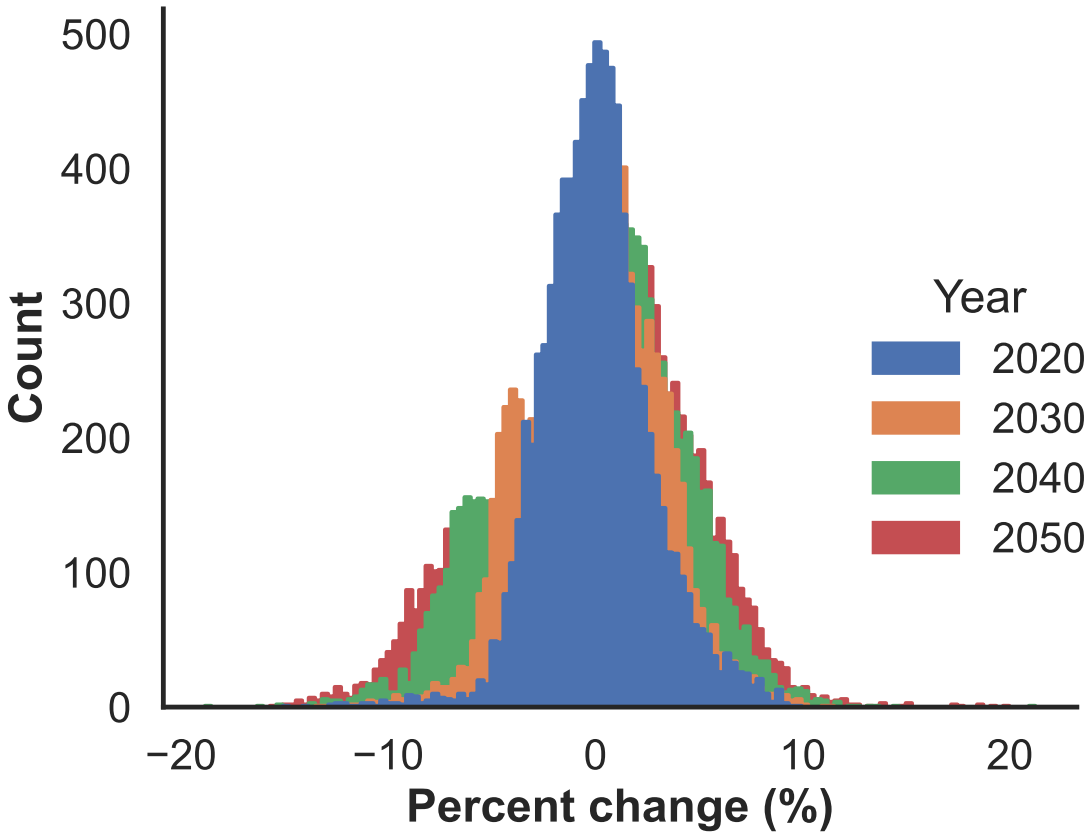

Figure D-16: Reference case frequency of hourly load variation

Table D.4: Modeled technology abbreviations

Technology	**Abbreviation**
Existing combined cycle gas turbine generation	B_CCGT
Existing coal generation	B_Coal
Existing diesel generation	B_Diesel
Existing hydro generation	B_Hydro
Existing nuclear generation	B_Nuclear
Existing pumped hydro storage generation	B_PHS
New biomass generation	Bio
New combined cycle gas turbine generation	CCGT
New coal generation	Coal
New Lithium ion battery storage power	Li-ion
New open cycle gas turbine generation	OCGT
New solar generation	PV
New wind generation	Wind
Network expansion	NetworkExp

Table D.5: System assumptions

Parameter	**Value**	
Discount rate	9%	
Value of lost load ($/MWh)	20,000	
Fuels assumptions	Cost ($/MMBtu)	Emissions intensity (tonnes CO_2/MMBtu)
Uranium	1	0.000
Coal	3	0.096
Natural gas (Reference / low)	11 / 8	0.052
Diesel	18	0.073
Biomass	3.7	0.000

Table D.6: Storage duration comparison across scenarios and modeled periods

	Reference	**Low storage cost**	**High AC efficiency**	**Low gas price**	**Sensitivity**
2040	4.5	5.4	4.4	4.6	6.8
2050	3.7	5.5	3.7	3.6	6.2

Table D.7: Zonal power transfer limits (Annex 7.1 of [11]). Zonal definitions as shown in Chapter 2 Fig. 2-1.

Zonal links	**2020 zonal capacity limits (MW)**	**Distance (km)**	**Line loss**
East-North	22,530	1,140	7.1%
West-North	36,720	851	5.3%
West-South	23,920	812	5.0%
North-Northeast	3,000	1,684	10.5%
West-East	21,190	937	5.8%
South-East	7,830	1,241	7.7%
East-Northeast	2,860	863	5.4%

Table D.8: Regional thermal power existing capacity and parameters [9, 5, 12, 13, 14, 15, 16]

Resource	Fuel	Region	2020 Capacity (MW)	Minimum Retirement (MW) 2030	2040	2050	VOM ($/MWh)	FOM ($/kW/y)	Heat rate (MMBtu per MWh)	Average plant size (MW)	Maximum Capacity (MW)
Coal	Coal	North	41,220	3,919	11,755	13,588	1.0	55.1	10.0	400	
		West	87,431	5,812	28,890	37,398	0.9	55.1	9.1	400	
		South	40,965	5,150	10,868	16,368	1.1	55.1	10.6	450	
		East	39,080	4,010	5,295	5,895	0.9	55.0	10.5	440	
		Northeast	750	0	0	250	1.0	55.0	9.8	230	
CCGT	Natural gas	North	5,752	179	685	910	1.2	9.4	7.8	480	
		West	10,239	870	2,686	3,022	1.5	12.0	6.9	410	
		South	6,505	1,147	1,948	2,115	1.4	11.0	6.2	470	
		Northeast	1,306	19	351	390	1.9	10.8	7.7	140	
Nuclear	Uranium	North	1,720								
		West	3,240								
		South	3,820								
Biomass	Biomass	North	2,431								9,721
		West	678.75								6,835
		South	2,934								5,336
		East	463								1,906
		Northeast	0								274
Backup	Diesel	South	761.58								
		Northeast	36								

Table D.9: National thermal power parameters [9, 17, 5, 10, 12, 18, 19, 20]

Resource	Fuel	VOM ($/MWh)	FOM ($/kW/y)	Start cost ($/MW)	Start fuel (MMBtu/ MW)	Heat rate (MMBtu/ MWh)	Min up time (hours)	Min down time (hours)	Ramp up	Ramp down	Min stable power	Max power	Average plant size (MW)	Lifetime
Coal	Coal			236.8			24	24	60%	60%	55%	90%		30
CCGT	Natural gas			106.5			8	8	100%	100%	50%	90%		30
Nuclear	Uranium	0.6	75	1,000		10.1	36	36			90%	90%	1,000	40
Biomass	Biomass	0	37.88			16.7	24	24	60%	60%	55%	90%	1	20
New Coal	Coal	0.9	30	214	0	9.5	24	24	60%	60%	45%	90%	620	30
New CCGT	Natural gas	1.5	10	106.5	0	6.6	8	8	100%	100%	33%	90%	573	30
New OCGT	Natural gas	7	11	96	0	9.1	2	2	100%	100%	26%	90%	384	30
Backup	Diesel		0	0	0	10.9	0	0	100%	100%		90%		

Table D.10: Hydro power existing capacity and parameters [9, 17]

Resource	2020 Capacity (MW) North	West	South	East	Northeast	VOM ($/MWh)	FOM ($/kW/y)	Lifetime	Efficiency Up/down	Power to energy ratio	Initial hydro level (% of reservoir)
Hydro reservoir	7,103	5,494	7,429	4,217	2,061	0	34.85	50			
North	7,103									8.89×10^{-4}	0
West	5,494									6.19×10^{-4}	0
South	7,429									4.18×10^{-4}	0
East	4,217									1.1×10^{-3}	0
Northeast	2,061									8.24×10^{-3}	0
Hydro run-of-river	16,235	1693	2,430	1,417	839	0	34.85	50			
North	16,235										
West	1,693										
South	2,430										
East	1,417										
Northeast	839										
Pumped hydro storage		1,840	2,005	940			34.85	50	89.4%	0.083333	

Table D.11: Existing Variable Renewable Energy [9, 5, 12, 13, 14, 15, 16]

		2020 capacity (MW)	Minimum retirement in 2050 (MW)	Lifetime	Interconnection cost ($/MW) Bin 1	Bin 2	Bin 3	Maximum capacity (MW) Bin 1	Bin 2	Bin 3
Wind	North	7,267	7,267	25	5,323	11,910	8,357	1,381,894	572,033	779,177
	West	19,659	19,659	25	5,253	6,468	4,593	1,078,581	1,093,093	458,577
	South	22,979	22,979	25	6,251	5,645	5,860	758,149	168,913	745,844
Solar	North	8,393	8,393	30	7,724	24,726	6,198	898,332	114,045	1,106,127
	West	10,889	10,889	30	5,251	7,997	5,159	774,465	898,509	402,062
	South	21,522	21,522	30	5,259	7,041	5,660	376,212	234,610	265,542
	East	1,100	1,102	30	4,139	4,656	4,980	183,627	301,833	129,668
	Northeast	323	323	30	6,033	5,347	56,541	152,712	45,735	40,998

Variable renewable energy resource characterization

Supplementary information pertaining to the renewable resource supply curve formulation of Chapter 5 section 5.2.4.

Fig. D-14 describes our approach to generate parameters used to parameterize VRE resource availability in the GenX model. First, we translate wind and solar resource data available for each location from the Renewable Energy Potential Model (reV)[12] into hourly normalized power output (or capacity factor (CF)) profiles. For wind, this is done using the power curve of the Siemens Gamesa 126/2500 [244] wind turbine with hub height 84 meters, while for solar, we model a single-axis tracking, horizontally oriented PV system using the NREL System Advisor Model[245]. Second, we estimate the land area available for wind and solar generation based on 2005 land use and land coverage classifications across India [16] and considering constraints on elevation [15]. Similar to [13] we restrict land usage for solar development to: shrubland, wasteland, salt pan, grassland, while land eligible for wind development also cropland, barren and fallow land. We further exclude parcels of land with a slope greater than 5% for solar development and 20% for wind development. We use 32 MW/km^2 and 4 MW/km^2 for spatial density of solar and wind resources respectively [9] to convert available land area within each grid cell ($25km^2$ resolution) into nameplate capacity that can be deployed. Third, we identify the closest point sampled from reV to associate a CF profile to each pixel. Fourth, we identify the cost of interconnection of each pixel by extending a straight line from the centroid of the pixel to the minimum cost substation by factoring voltage dependent cost of interconnection with distance. Here substation data is sourced from from OpenStreetMaps [14]. Fifth, we aggregate different resource sites into a small number of resource bins (3 for wind and 3 for solar for each zone) that can be represented in the GenX model based on clustering the sites using the levelized cost of energy (LCOE) metric. The LCOE for each site is computed using site-specific CF and interconnection costs as well as capital costs and Fixed O&M costs from NREL Annual Technology Baseline (ATB) 2020 [5] (utility-scale PV and class 6 wind data). Parameter inputs developed to characterize each resource bin in the GenX model [157] include : a) hourly CF, computed as the weighted average CF for sites within each bin, where the weights correspond to the developable area associated with that site., b) total developable capacity and c) weighted average annualized interconnection cost associated with each bin.

India's growth in the twenty-first century can most closely compared to China's with China ahead of India with respect to VRE capacity installation [10]. Given how China has been deploying VRE for a longer time, there are more data points

to fit a Gompertz growth curve [80]. We use the Gompertz sigmoid function to simulate slow initial adoption, rapid ramp-up followed by slow progress which is a good representation of new technology deployment. Results of the curve fitting are shown in Supplementary Table D.12. In 2019, India had 37.5 GW of wind and 33.7 GW of solar capacity installed nationally [10]. These capacity values are inserted in their corresponding fitted Gompertz curve to map to the year China was at that capacity. The mapped years are 2014 and 2010 for solar and wind respectively. Projecting the decadal installation limits are then extracted from the curves with the starting points being the identified mapped years. Results are shown in Supplementary Table D.13.

Table D.12: Gompertz curve fitting results

Parameter	**Solar**	**Wind**
A	2,952.55	-1,119.48
μ	89.53	39.40
d	2,018.93	2,042.84
y_0	-5.69	1,042.94
R-squared	0.999434	0.996302

Table D.13: Decadal VRE installation limit (MW) for the reference case

Resource	**2020**	**2030**	**2040**	**2050**
Wind	0	171,000	320,000	364,000
Solar	0	443,000	854,000	746,000

Appendix E

Supplementary information — Chapter 6

Supporting information relevant to the autoencoder-based dimensionality reduction of time series clustering algorithms for electricity resource capacity expansion planning optimization problems of Chapter 6.

Table E.1: Input data considered in statistical analysis. Complete description of each parametric scenario, including description of low, medium and high technology cost assumptions.

Solar Tracking	Load Year	Technology Cost	CO_2 \$/tonne
Single	2018	Low	50
Dual	2019	Mid	100
	2020	High	500
			1,000
			1,500

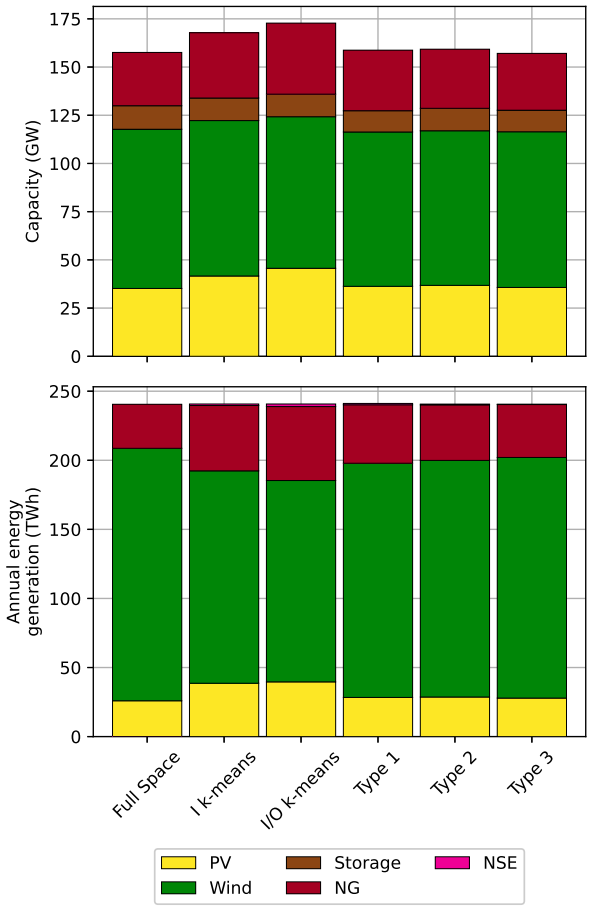

Figure E-1: Installed capacity (top) and annual generation (bottom) for reduced-space CEM (RCEM) results and full-space CEM results for three bus network. Results based on RCEM using $k = 8$ for ERCOT load year 2020, mid-range VRE technology cost, 1,000 \$/tonne CO_2 price.

Table E.2: Autoencoder parameters. Encoder maximum pooling layer size is divisible by flattened time series of one period to satisfy dimensionality reduction to cluster latent representation by k.

Parameter	Value
Filter size	50
Kernel size	10
Stride	1
Pooling	Cluster size (k)
Long-short term memory (LSTM) Units	50
Deconvolution size	10

Table E.3: Standard deviation of capacity error metric grouped by cluster number k for the eight bus system. Row-wise color mapping: red is highest and green is lowest.

	I k-means	**I/O k-means**	**Type 1**	**Type 2**	**Type 3**
4	30.24	39.35	46.33	34.27	37.87
8	29.3	35.83	29.39	31.25	26.69
20	27.86	26.1	29.64	26.8	24.72

Appendix F

Supplementary information — Chapter 7

Input information relevant to the Regret Monte Carlos Tree Search case study of the Odisha problem of Chapter 7.

Table F.1: Capital expenditure costs for each node for the three node case study.

	S0	**S1**	**S2**
PV ($/kW-AC)	1,354	1,095	836
Li-ion ($/kWh)	299	206	168
Li-ion ($/kW)	260	179	146
Diesel ($/kW)	400	400	400

Table F.2: Operational expenditure costs for each node for the three node case study

	S0	**S1**	**S2**
PV ($/kW-yr)	16	13	10
Li-ion ($/kW-yr)	36	25	20
Fuel cost ($/gal)	2.2	2.2	2.2
Diesel ($/kW)	100	100	100
Grid ($/kWh)	500	500	500
VoLL ($/kWh)	1000	1000	1000

Table F.3: Decision-dependent external event realization probabilities for leaf nodes 1 and 2.

	S1			**S2**		
	E1	E2	E3	E1	E2	E3
D1	0.7	0.05	0.25	0.5	0.15	0.35
D2	0.8	0.05	0.15	0.6	0.15	0.25
D3	0.95	0.05	0	0.85	0.15	0

Table F.4: Decision and external event dependent load projections for leaf nodes 1 and 2.

	S1			**S2**		
	E1	E2	E3	E1	E2	E3
D1	1,200	1,200	600	1,400	1,400	700
D2	1,600	1,600	800	1,800	1,800	900
D3	1,800	1,800	900	2,000	2,000	1,000

Appendix G

Code availability

All the code developed for this thesis is published on a university Github repository: `https://github.mit.edu/mbarbar/phd-thesis`.

G.1 Demand forecasting model

`https://github.mit.edu/mbarbar/phd-thesis/tree/master/demand-forecast`

G.2 Real options model

`https://github.mit.edu/mbarbar/phd-thesis/tree/master/roma`

G.3 Design and dispatch optimization model

`https://github.mit.edu/mbarbar/phd-thesis/tree/master/optmodel`

G.4 Autoencoder model

`https://github.mit.edu/mbarbar/phd-thesis/tree/master/autoencoder`

G.5 Regret MCTS model

https://github.mit.edu/mbarbar/phd-thesis/tree/master/mcts

Bibliography

[1] World Bank Group, "Electric power consumption (kwh per capita)," 2020.

[2] Tata Power Delhi Distribution Limited, "Technical specification cover sheets: 33KV and 11KV Grid," 2012.

[3] M. Barbar, D. S. Mallapragada, M. Alsup, and R. Stoner, "Scenarios of future Indian electricity demand accounting for space cooling and electric vehicle adoption," *Scientific Data*, vol. 8, July 2021.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[5] NREL, "Annual Technology Baseline: Electricity," 2020.

[6] K. Horowitz, "2019 distribution system upgrade unit cost database current version," 2019.

[7] P. Ciller, D. Ellman, C. Vergara, A. Gonzalez-Garcia, S. J. Lee, C. Drouin, M. Brusnahan, Y. Borofsky, C. Mateo, R. Amatya, R. Palacios, R. Stoner, F. de Cuadra, and I. Perez-Arriaga, "Optimal electrification planning incorporating on- and off-grid technologies: The reference electrification model (REM)," *Proceedings of the IEEE*, vol. 107, pp. 1872–1905, Sept. 2019.

[8] AEDC, "Wuse market energy audit report," tech. rep., 2019.

[9] A. Rose, l. Chernyakhovskiy, D. Palchak, S. Koebrich, and M. Joshi, "Least-Cost Pathways for India's Electric Power Sector," tech. rep., National Renewable Energy Laboratory, Golden, CO, 2020.

[10] International Energy Agency, *World Energy Outlook 2020*. IEA, 2020.

[11] C. E. Authority, "Revised Draft National Electricity Plan Volume II: Transmission," tech. rep., Ministry of Power, Government of India, Delhi, India, 2017.

[12] G. J. Maclaurin, N. W. Grue, A. J. Lopez, and D. M. Heimiller, "The renewable energy potential (rev) model: A geospatial platform for technical potential and supply curve modeling," tech. rep., national Renewable Energy Laboratory, 2019.

[13] R. Deshmukh, G. C. Wu, D. S. Callaway, and A. Phadke, "Geospatial and techno-economic analysis of wind and solar resources in India," *Renewable Energy*, vol. 134, pp. 947–960, 2019.

[14] OpenStreetMap contributors, "India Power Substations," 2020.

[15] World Bank Group, "Terrain elevation above sea level ele gis data."

[16] P. Roy, A. Roy, P. Joshi, M. Kale, V. Srivastava, S. Srivastava, R. Dwevidi, C. Joshi, M. Behera, P. Meiyappan, Y. Sharma, A. Jain, J. Singh, Y. Palchowdhuri, R. Ramachandran, B. Pinjarla, V. Chakravarthi, N. Babu, M. Gowsalya, P. Thiruvengadam, M. Kotteeswaran, V. Priya, K. Yelishetty, S. Maithani, G. Talukdar, I. Mondal, K. Rajan, P. Narendra, S. Biswal, A. Chakraborty, H. Padalia, M. Chavan, S. Pardeshi, S. Chaudhari, A. Anand, A. Vyas, M. Reddy, M. Ramalingam, R. Manonmani, P. Behera, P. Das, P. Tripathi, S. Matin, M. Khan, O. Tripathi, J. Deka, P. Kumar, and D. Kushwaha, "Development of decadal (1985–1995–2005) land use and land cover database for india," Feb. 2015.

[17] I. Rudnick, "Decarbonizing the Indian Power Sector by 2037: Evaluating Different Pathways that Meet Long-Term Emissions Targets," Master's thesis, Massachusetts Institute of Technology, 2019.

[18] J. Cochran, *Greening the Drid: Pathways to Integrate 175 Gigawatts of Renewable Energy into India's Electric Grid*. National Renewable Energy Laboratory, 2017.

[19] Central Electricity Authority, "National electricity plan," 2018.

[20] D. S. Mallapragada, N. A. Sepulveda, and J. D. Jenkins, "Long-run system value of battery energy storage in future grids with increasing wind and solar generation," *Applied Energy*, vol. 275, p. 115390, 2020.

[21] V. Foster and A. Rana, *Rethinking Power Sector Reform in the Developing World*. World Bank, 2020.

[22] M. J. Kochenderfer, *Decision making under uncertainty theory and application*. The MIT Press, 2015.

[23] International Energy Agency, *World Energy Outlook 2020*. IEA, 2020.

[24] International Energy Agency, "IEA Atlas of Energy - India," 2020.

[25] International Energy Agency, *The Future of Cooling*. IEA, 2018.

[26] International Energy Agency, *Global EV Outlook 2020*. IEA, 2020.

[27] International Energy Agency, *India Energy Outlook 2021*. IEA, 2021.

[28] S. Ali, The future of Indian electricity demand: how much, by whom and under what conditions? The Brookings Institute, 2018.

[29] T. Spencer and A. Awasthy, Analysing and projecting Indian electricity demand to 2030. The Energy and Resources Institute, 2019.

[30] Stockholm Environment Institute, "Long range energy alternatives planning (LEAP) system."

[31] T. Fleiter, M. Rehfeldt, A. Herbst, R. Elsland, A.-L. Klingler, P. Manz, and S. Eidelloth, "A methodology for bottom-up modelling of energy transitions in the industry sector: The forecast model," *Energy Strategy Reviews*, vol. 22, pp. 237–254, 2018.

[32] M. A. McNeil, V. E. Letschert, S. de la Rue du Can, and J. Ke, "Bottom–up energy analysis system (BUENAS)—an international appliance efficiency policy tool," *Energy Efficiency*, vol. 6, pp. 191–217, May 2013.

[33] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1966–1979, 2019.

[34] A. Garg, J. Maheshwari, and J. Upadhyay, "Load research for residential and commercial establishments in gujarat," tech. rep., Indian Institute of Management Ahmedabad, 2010.

[35] S. Nhalur and A. Josey, "Electricity in megacities," tech. rep., Prayas Energy Group, 2012.

[36] M. E. Samper and A. Vargas, "Investment decisions in distribution networks under uncertainty with distributed generation—part i: Model formulation," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2331–2340, 2013.

[37] P. Das, P. Mathuria, R. Bhakar, J. Mathur, A. Kanudia, and A. Singh, "Flexibility requirement for large-scale renewable energy integration in Indian power system: Technology, policy and modeling options," *Energy Strategy Reviews*, vol. 29, p. 100482, 2020.

[38] K. Kim, H. Park, and H. Kim, "Real options analysis for renewable energy investment decisions in developing countries," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 918–926, 2017.

[39] S. Pilpola and P. D. Lund, "Different flexibility options for better system integration of wind power," *Energy Strategy Reviews*, vol. 26, p. 100368, 2019.

[40] S. Sgouridis, S. Griffiths, S. Kennedy, A. Khalid, and N. Zurita, "A sustainable energy transition strategy for the United Arab Emirates: Evaluation of options using an integrated energy model," *Energy Strategy Reviews*, vol. 2, no. 1, pp. 8–18, 2013. Strategy Options and Models for the Middle East and North Africa (MENA) Energy Transition.

[41] J. Jenkins and N. Sepulveda, "Enhanced Decision Support for a Changing Electricity Landscape: the GenX Configurable Electricity Resource Capacity Expansion Model," tech. rep., MIT Energy Initiative, 2017.

[42] B. Knueven, J. Ostrowski, and J.-P. Watson, "On mixed-integer programming formulations for the unit commitment problem," *INFORMS Journal on Computing*, June 2020.

[43] A. Sahlberg, B. Khavari, A. Korkovelos, F. Fuso Nerini, and M. Howells, "A scenario discovery approach to least-cost electrification modelling in Burkina Faso," *Energy Strategy Reviews*, vol. 38, p. 100714, 2021.

[44] J. M. Olsson and F. Gardumi, "Modelling least cost electricity system scenarios for Bangladesh using osemosys," *Energy Strategy Reviews*, vol. 38, p. 100705, 2021.

[45] I. Perez-Arriaga, R. Stoner, D. Nagpal, and G. Jacquot, "Global commission to end energy poverty: Inception report," tech. rep., 2019.

[46] C. de Abajo, S. Díaz-Pastor, A. González, and I. Pérez-Arriaga, "A business plan to achieve full electrification in Rwanda under the integrated distribution framework (IDF)," tech. rep., 2020.

[47] B. Tenenbaum, C. Greacen, T. Siyambalapitiya, and J. Knuckle, "From the bottom up: How small power producers and mini-grids can deliver electrification and renewable energy in Africa. directions in development, energy and mining," tech. rep., 2014.

[48] B. W. Tenenbaum, C. Greacen, and A. Shrestha, "Undergrid mini grids in Nigeria and India: Interconnected and non-interconnected," tech. rep., 2021.

[49] S. Graber, P. Mong, and J. Sherwood, "Reliability of rural electricity service with undergrid minigrids," tech. rep., 2018.

[50] S. Graber, O. Adesua, C. Agbaegbu, I. Malo, and J. Sherwood, "Electrifying the underserved: Collaborative business models for developing minigrids under the grid," tech. rep., 2019.

[51] A. Evans, "The value of flexibility : application of real options analysis to electricity network investments," Master's thesis, Massachusetts Institute of Technology, 2020.

[52] H. Teichgraeber and A. R. Brandt, "Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison," *Applied Energy*, vol. 239, pp. 1283–1293, Apr. 2019.

[53] E. Kuepper, H. Teichgraeber, N. Baumgaertner, A. Bardow, and A. Brandt, "Wind data introduce error in time series reduction for capacity expansion modeling," *Preparation*, vol. 123, p. 124, 2020.

[54] H. Teichgraeber, C. P. Lindenmeyer, N. Baumgärtner, L. Kotzur, D. Stolten, M. Robinius, A. Bardow, and A. R. Brandt, "Extreme events in time series aggregation: A case study for optimal residential energy supply systems," *Applied Energy*, vol. 275, p. 115223, 2020.

[55] M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius, "A review on time series aggregation methods for energy system models," *Energies*, vol. 13, no. 3, p. 641, 2020.

[56] C. E. Lythcke-Jørgensen, M. Münster, A. V. Ensinas, and F. Haglind, "A method for aggregating external operating conditions in multi-generation system optimization models," *Applied Energy*, vol. 166, pp. 59–75, 2016.

[57] M. Sun, F. Teng, X. Zhang, G. Strbac, and D. Pudjianto, "Data-driven representative day selection for investment decisions: A cost-oriented approach," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2925–2936, 2019.

[58] H. Teichgraeber and A. R. Brandt, "Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 157, p. 111984, 2022.

[59] N. Tavakoli, S. Siami-Namini, M. A. Khanghah, F. M. Soltani, and A. S. Namin, "An autoencoder-based deep learning approach for clustering time series data," *SN Applied Sciences*, vol. 2, Apr. 2020.

[60] X. Guo, X. Liu, E. Zhu, and J. Yin, "Deep clustering with convolutional autoencoders," in *Neural Information Processing*, pp. 373–382, Springer International Publishing, 2017.

[61] V. Goel and I. E. Grossmann, "A class of stochastic programs with decision dependent uncertainty," *Mathematical programming*, vol. 108, no. 2, pp. 355–394, 2006.

[62] M. Guignard, "Lagrangean relaxation," *Top*, vol. 11, pp. 151–200, Dec. 2003.

[63] R. M. V. Slyke and R. Wets, "L-shaped linear programs with applications to optimal control and stochastic programming," *SIAM Journal on Applied Mathematics*, vol. 17, pp. 638–663, July 1969.

[64] G. Laporte and F. V. Louveaux, "The integer l-shaped method for stochastic integer programs with complete recourse," *Operations Research Letters*, vol. 13, pp. 133–142, Apr. 1993.

[65] J. R. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs," *Operations Research*, vol. 33, pp. 989–1007, Oct. 1985.

[66] M. V. F. Pereira and L. M. V. G. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Mathematical Programming*, vol. 52, pp. 359–375, May 1991.

[67] R. T. Rockafellar and R. J.-B. Wets, "Scenarios and policy aggregation in optimization under uncertainty," *Mathematics of Operations Research*, vol. 16, pp. 119–147, Feb. 1991.

[68] K. Tong, Y. Feng, and G. Rong, "Planning under demand and yield uncertainties in an oil supply chain," *Industrial & engineering chemistry research*, vol. 51, no. 2, pp. 814–834, 2012.

[69] S. Ahmed, *Strategic planning under uncertainty: Stochastic integer programming approaches*. PhD thesis, University of Illinois at Urbana-Champaign, 2000.

[70] L. Hellemo, *Managing uncertainty in design and operation of natural gas infrastructure*. PhD thesis, Norwegian University of Science and Technology, 2016.

[71] Ministry of Statistics and Programme Implementation, "*National accounts statistics 2019*," 2019.

[72] Shakti Sustainable Energy Foundation, GE India Exports Pvt. Ltd, "*Modelling with power sector planning for India*," 2018.

[73] Government of India National Institution for Transforming India, "Zero emission vehicles (zevs): Towards a policy framework," 2018.

[74] Society of India Automobile Manufacturers, "*Automobile domestic sales trends 2015-2020*," 2020.

[75] A. Garg, J. Maheshwari, and J. Upadhyay, *Load Research for Residential and Commercial Establishments in Gujarat*, vol. ECO-III-1024 of *Energy Conservation and Commercialization*. Indian Institute of Management Ahmedabad, 2010.

[76] Ministry of Power, "*Power system operation corporation: load dispatch center monthly reports*," 2020.

[77] Ministry of Statistics and Programme Implementation, "*Energy statistics 2019*," 2019.

[78] S. Sankhe, K. Chockalingam, A. Madgavkar, J. Woetzel, S. Smit, and G. Kumra, "*India's turning point: an economic agenda to spur growth and jobs*," 2020.

[79] PricewaterhouseCoopers, "*The long view how will the global economic order change by 2050?*," 2017.

[80] C. P. Winsor, "The gompertz curve as a growth curve," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, no. 1, pp. 1–8, 1932.

[81] P. Datta, "Population projection of EAG states of India: vision until 2051," 2011. submitted.

[82] United States Energy Information Administration, "*India's economic growth is driving its energy consumption*," 2013.

[83] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao, "The modern-era retrospective analysis for research and applications, version 2 (MERRA-2)," *Journal of Climate*, vol. 30, pp. 5419–5454, July 2017.

[84] M. Alsup, "Forecasting electricity demand in the data-poor Indian context," Master's thesis, Massachusetts Institute of Technology, 2020.

[85] United Nations, Department of Economic and Social Affairs, Population Division, *World Population Prospects 2019: Highlights*. No. Report No. ST/ESA/SER.A/423, United Nations, 2019.

[86] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[87] I. R. García, "Decarbonizing the Indian power sector by 2037 : evaluating different pathways that meet long-term emissions targets," Master's thesis, Massachusetts Institute of Technology, 2019.

[88] International Energy Agency, *World Energy Outlook 2019*. IEA, 2019.

[89] Datanet India Development Team, "*Electrical energy consumption by industry*," 2020.

[90] C. Mateo Domingo, T. Gomez San Roman, A. Sanchez-Miralles, J. P. Peco Gonzalez, and A. Candela Martinez, "A reference network model for large-scale distribution planning with automatic street map generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 190–197, 2011.

[91] R. Batra, S. Kumar, S. Bose, R. Grant, K. Garg, S. Gupta, R. J. Miller, A. Valsan, S. Kapoor, G. Batra, A. Chandna, and T. Madaan, *Standing up India's EV ecosystem - who will drive the charge?* Ernest and Young, 2018.

[92] Government of India National Institution for Transforming India, "*Zero emission vehicles: towards a policy framework*," 2018.

[93] M. Patil, B. B. Majumdar, P. K. Sahu, and L. T. Truong, "Evaluation of prospective users' choice decision toward electric two-wheelers using a stated preference survey: An indian perspective," *Sustainability*, vol. 13, no. 6, p. 3035, 2021.

[94] A. von Meier, "Ev infrastructure planning and grid impact assessment: a case for Mexico." submitted, 2018.

[95] International Energy Agency, *World Energy Outlook 2018*. IEA, 2018.

[96] International Energy Agency, *World Energy Outlook 2017*. IEA, 2017.

[97] I. Pérez-Arriaga, "Utility of the future: An MIT Energy Initiative response to an industry in transition," tech. rep., Massachusetts Institute of Technology, 2016.

[98] M. Barbar, "Resiliency and reliability planning of the electric grid in natural disaster affected areas," Master's thesis, Massachusetts Institute of Technology, 2019.

[99] N. D. Rao, J. Min, and A. Mastrucci, "Energy requirements for decent living in India, Brazil and South Africa," *Nature Energy*, vol. 4, pp. 1025–1032, Dec 2019.

[100] K. Olaniyan, B. C. McLellan, S. Ogata, and T. Tezuka, "Estimating residential electricity consumption in Nigeria to support energy transitions," *Sustainability*, vol. 10, no. 5, 2018.

[101] Economic and Social Council, "Short-Term Financing, Creation of Repo Market Crucial to Assist Poor Countries Facing Escalating Debt, Economic Contraction, Speakers Tell Financing for Development Forum," 2021.

[102] World Energy Council, "Transmission and distribution in india," 2009.

[103] S. Pargal and S. G. Banerjee, *More Power to India: The Challenge of Electricity Distribution*. Washington, DC: World Bank, 2014. License: Creative Commons Attribution CC BY 3.0 IGO.

[104] Central Electricity Authority, "Growth of Electricity Sector in India from 1947-2019," 2018.

[105] International Finance Council Corporation, "The dirty footprint of the broken grid: The impacts of fossil fuel back-up generators in developing countries," tech. rep., World Bank Group, 2019.

[106] A. Q. Jakhrani, A. R. H. Rigit, A.-K. Othman, S. R. Samo, and S. A. Kamboh, "Estimation of carbon footprints from diesel generator emissions," in *2012 International Conference on Green and Ubiquitous Technology*, pp. 78–81, 2012.

[107] V. Grimm, J. Grübel, B. Rückel, C. Sölch, and G. Zöttl, "Storage investment and network expansion in distribution networks: The impact of regulatory frameworks," *Applied Energy*, vol. 262, p. 114017, 2020.

[108] P. Shi, Y. Li, Y. Du, Q. Cai, and H. Chen, "Distribution network planning considering uncertainty of incremental distribution network access," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 2244–2247, 2020.

[109] H. Wang, Z. Yan, M. Shahidehpour, Q. Zhou, and X. Xu, "Optimal energy storage allocation for mitigating the unbalance in active distribution network via uncertainty quantification," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 303–313, 2021.

[110] P. Boonluk, A. Siritaratiwat, P. Fuangfoo, and S. Khunkitti, "Optimal siting and sizing of battery energy storage systems for distribution network of distribution system operators," *Batteries*, vol. 6, no. 4, 2020.

[111] M. Jooshaki, S. Fattaheian-Dehkordi, M. Fotuhi-Firuzabad, and M. Lehtonen, "Planning a flexible distribution network with energy storage systems considering the uncertainty of renewable sources and demand," in *CIRED 2020 Berlin Workshop (CIRED 2020)*, vol. 2020, pp. 132–135, 2020.

[112] G. Muñoz-Delgado, J. Contreras, and J. M. Arroyo, "Multistage generation and network expansion planning in distribution systems considering uncertainty and reliability," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3715–3728, 2016.

[113] N. Sepulveda, J. Jenkins, D. Mallapragada, A. Schwartz, N. Patankar, Q. Xu, J. Morri, and S. Chakrabarti, "Source code for: Genx." *Github*. `https://github.com/GenXProject/GenX`, 2021.

[114] B. Knueven, J. Ostrowski, and J.-P. Watson, "Source code for: Egret," 2021.

[115] Tata Power Delhi Distribution Limited, "Customer solutions tariff details and net-metering," 2020.

[116] W. E. Hart, C. D. Laird, J.-P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson, and J. D. Siirola, *Pyomo–optimization modeling in python*, vol. 67. Springer Science & Business Media, second ed., 2017.

[117] I. I. Cplex, "V12. 1: User's manual for cplex," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.

[118] C. P. Winsor, "The gompertz curve as a growth curve," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, no. 1, pp. 1–8, 1932.

[119] K. Rudion, A. Orths, Z. A. Styczynski, and K. Strunz, "Design of benchmark of medium voltage distribution network for investigation of dg integration," in *2006 IEEE Power Engineering Society General Meeting*, pp. 6 pp.–, 2006.

[120] Office of the economic adviser department for promotion of industry and internal trade. Ministry of Commerce and Industry. Government of India., "Index files for wpi series," 2019.

[121] Lazard Ltd., "Levelized cost of energy and levelized cost of storage – 2020," 2020.

[122] RBSA Advisors, "Cost of capital in India," 2020.

[123] Central Electricity Authority, "Guidelines for distribution utilities for development of distribution infrastructure," 2018.

[124] Tata Power Delhi Distribution Limited, "Excellence journey progress report," 2016.

[125] M. Kojima and C. Trimble, "Making power affordable for africa and viable for its utilities," tech. rep., 2016.

[126] M. El-Hendawi, H. A.Gabbar, G. El-Saady, and E.-N. A. Ibrahim, "Optimal operation and battery management in a grid-connected microgrid," *Journal of International Council on Electrical Engineering*, vol. 8, no. 1, pp. 195–206, 2018.

[127] J. Zhou, J. Zhang, X. Cai, G. Shi, J. Wang, and J. Zang, "Design and analysis of flexible multi-microgrid interconnection scheme for mitigating power fluctuation and optimizing storage capacity," *Energies*, vol. 12, no. 11, 2019.

[128] A. Khodaei, S. Bahramirad, and M. Shahidehpour, "Microgrid planning under uncertainty," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2417–2425, 2015.

[129] H. Khorramdel, J. Aghaei, B. Khorramdel, and P. Siano, "Optimal battery sizing in microgrids using probabilistic unit commitment," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 834–843, 2016.

[130] M. Husein and I.-Y. Chung, "Optimal design and financial feasibility of a university campus microgrid considering renewable energy incentives," *Applied Energy*, vol. 225, pp. 273–289, 2018.

[131] M. V. Kirthiga, S. A. Daniel, and S. Gurunathan, "A methodology for transforming an existing distribution network into a sustainable autonomous microgrid," *IEEE Transactions on Sustainable Energy*, vol. 4, pp. 31–41, Jan 2013.

[132] I. Perez-Arriaga and R. Stoner, "Abuja electric's proposed franchising model: Dessa," tech. rep., 2020.

[133] Nigerian Electricity Regulatory Commission, "Regulation for mini-grids," 2016.

[134] National Bureau of Statistics, Nigeria, "Diesel price watch," 2021.

[135] M. Barbar, J. Nsengiyaremye, R. Stoner, and I. P. Arriaga, "Minigrid design under grid supply uncertainty." 2021.

[136] M. Waite, E. Cohen, H. Torbey, M. Piccirilli, Y. Tian, and V. Modi, "Global trends in urban electricity demands for cooling and heating," *Energy*, vol. 127, pp. 786–802, 5 2017.

[137] F. P. Colelli and E. D. Cian, "Cooling demand in integrated assessment models: a methodological review," *Environmental Research Letters*, vol. 15, p. 113005, nov 2020.

[138] N. Zhou, N. Khanna, W. Feng, J. Ke, and M. Levine, "Scenarios of energy efficiency and co2 emissions reduction potential in the buildings sector in China to year 2050," *Nature Energy*, vol. 3, pp. 978–984, Nov 2018.

[139] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy*, vol. 3, pp. 193–201, Mar 2018.

[140] K. B. Debnath and M. Mourshed, "Challenges and gaps for energy planning models in the developing-world context," *Nature Energy*, vol. 3, pp. 172–184, Mar 2018.

[141] T. Mai, P. Jadun, J. Logan, C. McMillan, M. Muratori, D. Steinberg, and L. Vimmerstedtt, "Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States," tech. rep., National Renewable Energy Laboratory, Golden, CO, 2018.

[142] D. Palchak, l. Chernyakhovskiy, T. Bowen, and V. Narwade, "India 2030 Wind and Solar Integration Study: Interim Report," tech. rep., National Renewable Energy Laboratory, Golden, CO, 2019.

[143] D. S. Mallapragada, I. Naik, K. Ganesan, R. Banerjee, and I. J. Laurenzi, "Life Cycle Greenhouse Gas Impacts of Coal and Imported Gas-Based Power Generation in the Indian Context," *Environmental Science & Technology*, vol. 53, pp. 539–549, 1 2018.

[144] R. Deshmukh, A. Phadke, and D. S. Callaway, "Least-cost targets and avoided fossil fuel capacity in India's pursuit of renewable energy," *Proceedings of the National Academy of Sciences*, vol. 118, no. 13, 2021.

[145] T. Lu, P. Sherman, X. Chen, S. Chen, X. Lu, and M. McElroy, "India's potential for integrating solar and on- and offshore wind power into its energy system," *Nature Communications*, vol. 11, pp. 1–10, 12 2020.

[146] T. Spencer, N. Rodrigues, R. Pachouri, S. Thakre, and G. Renjith, "Renewable Power Pathways: Modelling the Integration of Wind and Solar in India by 2030," tech. rep., The Energy and Resources Institute, Delhi, India, 2020.

[147] I. Rudnick, P. Duenas-Martinez, A. Botterud, D. J. Papageorgiou, B. K. Mignone, S. Rajagopalan, M. R. Harper, and K. Ganesan, "Decarbonization of the indian electricity sector: Technology choices and policy trade-offs," *iScience*, vol. 25, no. 4, p. 104017, 2022.

[148] G. He, A.-P. Avrin, J. H. Nelson, J. Johnston, A. Mileva, J. Tian, and D. M. Kammen, "SWITCH-china: A systems approach to decarbonizing china's power system," *Environmental Science & Technology*, vol. 50, pp. 5467–5473, May 2016.

[149] N. Kittner, F. Lill, and D. M. Kammen, "Energy storage deployment and innovation for the clean energy transition," *Nature Energy*, vol. 2, July 2017.

[150] M. T. Craig, P. Jaramillo, and B.-M. Hodge, "Carbon dioxide emissions effects of grid-scale electricity storage in a decarbonizing power system," *Environmental Research Letters*, vol. 13, p. 014004, jan 2018.

[151] A. Malik, C. Bertram, J. Despres, J. Emmerling, S. Fujimori, A. Garg, E. Kriegler, G. Luderer, R. Mathur, M. Roelfsema, S. Shekhar, S. Vishwanathan, and Z. Vrontisi, "Reducing stranded assets through early action in the Indian power sector," *Environmental Research Letters*, vol. 15, p. 094091, 9 2020.

[152] J.-P. Carvallo, B. J. Shaw, N. I. Avila, and D. M. Kammen, "Sustainable low-carbon expansion for the power sector of an emerging economy: The case of kenya," *Environmental Science & Technology*, vol. 51, pp. 10232–10242, Aug. 2017.

[153] C. L. Lara, D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, and I. E. Grossmann, "Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm," *European Journal of Operational Research*, vol. 271, pp. 1037–1054, 12 2018.

[154] B. S. Palmintier, *Incorporating operational flexibility into electric generation planning : impacts and methods for system design and policy analysis*. PhD thesis, Massachusetts Institute of Technology, 2013.

[155] K. Poncelet, E. Delarue, and W. D'haeseleer, "Unit commitment constraints in long-term planning models: Relevance, pitfalls and the role of assumptions on flexibility," *Applied Energy*, vol. 258, p. 113843, 1 2020.

[156] D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, C. L. Lara, and I. E. Grossmann, "Impact of model resolution on scenario outcomes for electricity sector system expansion," *Energy*, vol. 163, pp. 1231–1244, 2018.

[157] M. Barbar, D. S. Mallapragada, and R. Stoner, "Impact of demand growth on decarbonizing india's electricity sector and the role for energy storage," 2022.

[158] P. R. Brown and A. Botterud, "The Value of Inter-Regional Coordination and Transmission in Decarbonizing the US Electricity System," *Joule*, 12 2020.

[159] G. He and D. M. Kammen, "Where, when and how much solar is available? a provincial-scale solar resource assessment for china," *Renewable Energy*, vol. 85, pp. 74–82, 2016.

[160] P. Denholm, M. O'Connell, G. Brinkman, and J. Jorgenson, "Overgeneration from solar energy in california: A field guide to the duck chart," 2015.

[161] Federal Reserve Bank of St. Louis, "Global price of natural gas, eu, u.s. dollars per million metric british thermal unit, monthly between 2010 and 2021."

[162] M. Barbar, D. S. Mallapragada, and R. Stoner, "Decision making under uncertainty for deploying battery storage as a non-wire alternative in distribution networks," 2021.

[163] H. Gil and G. Joos, "On the quantification of the network capacity deferral value of distributed generation," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1592–1599, 2006.

[164] S. P. Burger, J. D. Jenkins, C. Batlle, and I. J. Perez-Arriaga, "Restructuring revisited part 2: Coordination in electricity distribution systems," *The Energy Journal*, vol. 40, July 2019.

[165] B. Yarlagadda, S. J. Smith, B. K. Mignone, D. Mallapragada, C. A. Randles, and J. Sampedro, "Climate and air pollution implications of potential energy infrastructure and policy measures in india," *Energy and Climate Change*, vol. 3, p. 100067, 2022.

[166] International Institute for Sustainable Development, "*The evolution of the clean energy cess on coal production in India*," 2018.

[167] CDP India, "Putting a price on carbon: A handbook for Indian companies," tech. rep., CDP, 2017.

[168] N. A. Sepulveda, J. D. Jenkins, A. Edington, S. Mallapragada, Dharik, and R. K. Lester, "The Design Space for Long-duration Energy Storage in Decarbonized Power Systems," *Nature Energy*, p. Accepted, 2021.

[169] O. J. Guerra, J. Zhang, J. Eichman, P. Denholm, J. Kurtz, and B.-M. Hodge, "The value of seasonal energy storage technologies for the integration of wind and solar power," *Energy & Environmental Science*, vol. 13, no. 7, pp. 1909–1922, 2020.

[170] D. A. Tejada-Arango, M. Domeshek, S. Wogrin, and E. Centeno, "Enhanced representative days and system states modeling for energy storage investment analysis," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6534–6544, 2018.

[171] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3029–3039, 2018.

[172] Y. Yin, C. He, T. Liu, and L. Wu, "Risk-averse stochastic midterm schedule of thermal-hydro-wind system: A network-constrained clustered unit commitment approach," *IEEE Transactions on Sustainable Energy*, pp. 1–1, 2022.

[173] C. Feng, M. Cui, B.-M. Hodge, S. Lu, H. F. Hamann, and J. Zhang, "Unsupervised clustering-based short-term solar forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2174–2185, 2019.

[174] M. A. Z. Alvarez, K. Agbossou, A. Cardenas, S. Kelouwani, and L. Boulon, "Demand response strategy applied to residential electric water heaters using dynamic programming and k-means clustering," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 524–533, 2020.

[175] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1936–1948, 2016.

[176] N. E. Koltsaklis and A. S. Dagoumas, "State-of-the-art generation expansion planning: A review," *Applied Energy*, vol. 230, pp. 563–589, 2018.

[177] K. Poncelet, E. Delarue, D. Six, J. Duerinck, and W. D'haeseleer, "Impact of the level of temporal and operational detail in energy-system planning models," *Applied Energy*, vol. 162, pp. 631–643, 2016.

[178] S. Wogrin, P. Dueñas, A. Delgadillo, and J. Reneses, "A new approach to model load levels in electric power systems with high renewable penetration," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2210–2218, 2014.

[179] N. Baumgärtner, B. Bahl, M. Hennen, and A. Bardow, "Rises3: Rigorous synthesis of energy supply and storage systems via time-series relaxation and aggregation," *Computers & Chemical Engineering*, vol. 127, pp. 127–139, 2019.

[180] K. Kim, A. Botterud, and F. Qiu, "Temporal decomposition for improved unit commitment in power system production cost modeling," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5276–5287, 2018.

[181] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[182] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," KDD'96, p. 226–231, AAAI Press, 1996.

[183] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "Wavecluster: A multi-resolution clustering approach for very large spatial databases," in *VLDB*, 1998.

[184] T.-C. Fu, F. lai Chung, R. W. P. Luk, and Ng, "Pattern discovery from stock time series using self-organizing maps," 2016.

[185] S. Aghabozorgi, T. Y. Wah, T. Herawan, H. A. Jalab, M. A. Shaygan, and A. Jalali, "A hybrid algorithm for clustering of time series data based on affinity search technique," *The Scientific World Journal*, vol. 2014, pp. 1–12, 2014.

[186] L. Kaufman and P. J. Rousseeuw, eds., *Finding Groups in Data*. John Wiley & Sons, Inc., Mar. 1990.

[187] E. J. Keogh and M. J. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback," in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, KDD'98, p. 239–243, AAAI Press, 1998.

[188] V. Hautamaki, P. Nykanen, and P. Franti, "Time-series clustering by approximate prototypes," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, 2008.

[189] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, p. 264–323, sep 1999.

[190] E. Keogh and S. Kasetty *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349–371, 2003.

[191] A. Bagnall and G. Janacek, "Clustering time series with clipped data," *Machine Learning*, vol. 58, pp. 151–178, Feb. 2005.

[192] S. Chu, E. Keogh, D. Hart, and M. Pazzani, *Iterative Deepening Dynamic Time Warping for Time Series*, pp. 195–212.

[193] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, vol. 13, pp. 335–364, May 2006.

[194] C. Ratanamahatana, E. Keogh, A. J. Bagnall, and S. Lonardi, "A novel bit level time series representation with implication of similarity search and clustering," in *Advances in Knowledge Discovery and Data Mining*, pp. 771–777, Springer Berlin Heidelberg, 2005.

[195] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery - DMKD '03*, ACM Press, 2003.

[196] A. Bagnall, C. ". Ratanamahatana, E. Keogh, S. Lonardi, and G. Janacek, "A bit level representation for time series data mining with shape based similarity," *Data Mining and Knowledge Discovery*, vol. 13, pp. 11–40, May 2006.

[197] J. Shieh and E. J. Keogh, "isax: indexing and mining terabyte sized time series," in *KDD*, 2008.

[198] W. W. Tso, C. D. Demirhan, C. F. Heuberger, J. B. Powell, and E. N. Pistikopoulos, "A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage," *Applied Energy*, vol. 270, p. 115190, 2020.

[199] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1936–1948, 2017.

[200] A. Almaimouni, A. Ademola-Idowu, J. Nathan Kutz, A. Negash, and D. Kirschen, "Selecting and evaluating representative days for generation expansion planning," in *2018 Power Systems Computation Conference (PSCC)*, pp. 1–7, 2018.

[201] S. Pineda and J. M. Morales, "Chronological time-period clustering for optimal capacity expansion planning with storage," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7162–7170, 2018.

[202] D. A. Tejada-Arango, M. Domeshek, S. Wogrin, and E. Centeno, "Enhanced representative days and system states modeling for energy storage investment analysis," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6534–6544, 2018.

[203] D. Z. Fitiwi, F. de Cuadra, L. Olmos, and M. Rivier, "A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty," *Energy*, vol. 90, no. P2, pp. 1360–1376, 2015.

[204] C. Li, A. J. Conejo, J. D. Siirola, and I. E. Grossmann, "On representative day selection for capacity expansion planning of power systems under extreme operating conditions," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107697, 2022.

[205] M. Sun, F. Teng, X. Zhang, G. Strbac, and D. Pudjianto, "Data-driven representative day selection for investment decisions: A cost-oriented approach," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2925–2936, 2019.

[206] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571–583, 2020.

[207] J. Taylor, *Convex Optimization of Power Systems*. Convex Optimization of Power Systems, Cambridge University Press, 2015.

[208] M. Barbar, D. S. Mallapragada, and R. Stoner, "Impact of demand growth on decarbonizing india's electricity sector and the role for energy storage," 2022.

[209] F. Chollet *et al.*, "Keras." `https://github.com/fchollet/keras`, 2015.

[210] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[211] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 448–456, PMLR, 07–09 Jul 2015.

[212] D. MacKay, *Information Theory, Inference and Learning Algorithms*, ch. Chapter 20. An Example Inference Task: Clustering. Cambridge University Press, 2003.

[213] Electric Reliability Council of Texas, "Hourly load data archive." `https://www.ercot.com/gridinfo/load/load_hist`.

[214] Rojowsky, K., Gothandaraman, A. & Beaucage, P., "Hourly wind and solar generation profiles (1980-2019)," tech. rep., 2020.

[215] Pacific Northwest National Laboratory, "8-Bus ERCOT Bulk System Models." `https://github.com/pnnl/tesp`.

[216] A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor, B. Bergeron, V. Gadepally, M. Houle, M. Hubbell, M. Jones, A. Klein, L. Milechin, J. Mullen, A. Prout, A. Rosa, C. Yee, and P. Michaleas, "Interactive supercomputing on 40,000 cores for machine learning and data analysis," in *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pp. 1–6, IEEE, 2018.

[217] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 5, pp. 363–387, 2012.

[218] G.-C. Lee, M. Höhenrieder, J.-P. Watson, and D. L. Woodruff, "Chance and service level constraints for stochastic generation expansion planning," *Netnomics*, vol. 16, p. 169–191, dec 2015.

[219] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff, "A class of stochastic programs withdecision dependent random elements," *Annals of operations research*, vol. 82, pp. 83–106, 1998.

[220] S. Sen, "Stochastic programming: Computational issues and challenges."

[221] J. R. Birge, "State-of-the-art-survey—stochastic programming: Computation and applications," *INFORMS Journal on Computing*, vol. 9, pp. 111–133, May 1997.

[222] L. A. Hannah, "Stochastic optimization," April 2014.

[223] B. Chen, P. L. Donti, K. Baker, J. Z. Kolter, and M. Bergés, "Enforcing policy feasibility constraints through differentiable projection for energy optimization," in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, e-Energy '21, (New York, NY, USA), p. 199–210, Association for Computing Machinery, 2021.

[224] A. Shapiro and T. Homem-De-Mello, "A simulation-based approach to two-stage stochastic programming with recourse," *Math. Program.*, vol. 81, p. 301–325, may 1998.

[225] M. Wiering and M. van Otterlo, eds., *Reinforcement learning.* Adaptation, learning, and optimization, Berlin, Germany: Springer, 2012 ed., Mar. 2012.

[226] R. Sutton and A. Barto, *Reinforcement Learning, second edition: An Introduction.* Adaptive Computation and Machine Learning series, MIT Press, 2018.

[227] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[228] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu, "Deep blue," *Artificial intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

[229] J. Schaeffer, "The games computers (and people) play," in *Advances in computers*, vol. 52, pp. 189–266, Elsevier, 2000.

[230] G. Tesauro and G. Galperin, "On-line policy improvement using monte-carlo search," in *Advances in Neural Information Processing Systems* (M. C. Mozer, M. Jordan, and T. Petsche, eds.), vol. 9, MIT Press, 1996.

[231] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, Jan 2016.

[232] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[233] G. Chaslot, S. Bakkes, I. Szita, and P. Spronck, "Monte-carlo tree search: A new framework for game ai," in *Proceedings of the Fourth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE'08, p. 216–217, AAAI Press, 2008.

[234] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2/3, pp. 235–256, 2002.

[235] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, 2012.

[236] L. Weng, "A (long) peek into reinforcement learning," *lilianweng.github.io*, 2018.

[237] G. Tesauro, V. T. Rajan, and R. Segal, "Bayesian inference in monte-carlo tree search," 2012.

[238] C. P. Robert and G. Casella, "The metropolis—hastings algorithm," in *Springer Texts in Statistics*, pp. 267–320, Springer New York, 2004.

[239] M. Barbar and D. S. Mallapragada, "Representative period selection for power system planning: autoencoder based dimensionality-reduction coupled with output features." 2022.

[240] N. N. Taleb and P. Cirillo, "Branching epistemic uncertainty and thickness of tails," 2019.

[241] J. Morris, J. Farrell, H. Kheshgi, H. Thomann, H. Chen, S. Paltsev, and H. Herzog, "Representing the costs of low-carbon power generation in multi-region multi-sector energy-economic models," *International Journal of Greenhouse Gas Control*, vol. 87, pp. 170–187, 2019.

[242] M. Rossol, G. Buster, and M. Bannister, "Nrel/rev: lat_lon_cols cleanup," Feb. 2021.

[243] A. Dobos, "Pvwatts version 5 manual," 2014.

[244] Gamesa, "G126-2.5 MW," tech. rep., 2017.

[245] National Renewable Energy Laboratory, "System advisor model version 2020.11.29 (sam 2020.11.29)."