Title
Leveraging New Data for Renewables Integration and Energy Efficiency

Author
Tabone, Michaelangelo

Publication Date
2016

Leveraging New Data for Renewables Integration and Energy Efficiency

by

Michaelangelo David Tabone

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Energy and Resources

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Duncan S. Callaway, Chair
Professor Pravin Varaia
Professor Kameshwar Poolla
Professor Zico Kolter

Summer 2016

Leveraging New Data for Renewables Integration and Energy Efficiency

Abstract

Leveraging New Data for Renewables Integration and Energy Efficiency

by

Michaelangelo David Tabone

Doctor of Philosophy in Energy and Resources

University of California, Berkeley

Professor Duncan S. Callaway, Chair

This dissertation develops new empirical methods that inform two questions in energy policy: How will locations of photovoltaic (PV) systems affect the the need for flexible resources in power systems? And how much energy is currently being conserved by varying set point schedules in residential households?

Chapters 2 & 3 focus on the first question regarding the siting of PV systems. Chapter 2 presents, fits, and validates a model of variability and uncertainty in PV generation that is useful for estimating the future needs of flexibility in power systems. I call this the "volatility state model", due to its reliance on latent states that I refer to as volatility states. Specifically, the model (a) accounts for spatial correlation, (b) predicts metrics of variability and uncertainty that are directly relevant to grid operation and planning, and (c) predicts boundaries on distribution tails that are consistent with observed data. I find that PV variability distributions are roughly Gaussian after conditioning on volatility states, which is helpful for finding the degree of spatial smoothing. I also propose a method for simulating volatility states that results in a very good upper bound for the probability of extreme events. Therefore the model can be used as a tool for planning additional reserve capacity requirements to balance solar variability from yet-to-be-built systems.

Chapter 3, applies the volatility state model to predict the need for reserve generation—load following and regulation—in California under different locational scenarios for PV. I find that clustering PV into small areas exacerbates the need for reserves, resulting primarily from the spatial correlation of hourly forecast errors. The benefits of dispersion diminish, and they can be saturated with a relatively small number of large utility-scale systems: 25, 500 MW systems. However, these systems need to be adequately separated, which has implications for the construction costs of transmission needed to reach them. I also identify trade-offs between locations

that minimize variability and uncertainty and locations that maximize energy or capacity value. The largest trade off in California is actually between energy and capacity value: areas of the state with the greatest energy resource tend to be cloudy on summer afternoons, when peak demand—driven by air conditioning—tends to be greatest.

In Chapter 4, I explore the ability of statistical models fit to AMI data—which I refer to as “utility meter models”—to predict the largest end use of electricity in residential homes: heating, ventilation, and air conditioning (HVAC). Specifically, I evaluate models’ abilities to predict of the timing of HVAC use, the efficiency of operation, and the amount of energy consumed. I begin by presenting a general utility meter model; our goals is to create a form that (a) directly relates to physical models of heat dynamics in buildings, and (b) encompasses utility meter models already in the literature. I then fit and validate multiple variations of this model—some similar to those in literature, and some of our own device—using data from air conditioners, thermostats, and residential electricity sub-meters. I test four specific aspects of the general model: whether to use daily or hourly data, whether to allow expected energy use to be discontinuous with respect to outdoor temperature, whether to use binary latent states to classify when HVAC is on, and which probability distribution shape to assume for model disturbances.

I find a large benefit to combining models fit to daily and hourly data; models perform best when days that are classified as without HVAC energy use cannot contain hours that are classified with HVAC. I also find that the distribution shape assumed for model disturbances greatly affects model classifications and parameters; where kernel density estimates for these distributions outperform the traditional normal distributions. Finally, I find that applying a post-fitting process that disaggregates model residuals—attributing part to HVAC and part to other end uses—increases estimations of cooling energy use.

Concluding chapter 4, I attempt to recover indoor temperature dynamics in homes metered by AMI. Though our models do not estimate these dynamics endogenously, I can infer them using out model outputs: the building is likely heating when estimated cooling energy use is less than that required to maintain a steady state, and vice versa. I find that I can infer intra-day changes in temperature well, and inter-day changes in temperature weakly.

Chapter 4 lays the groundwork for a future study, in which I hope to estimate the thermostat set point schedules of a large sample of households metered by AMI. This future work will provide an important empirical estimate of the energy currently being saved as a result of variable set point schedules. It will also provide an important behavioral baseline for the current practices of consumers; the results of energy efficiency or demand response programs should be measured against this baseline.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation could not have been conceived of or completed without the support of my advisor, Duncan Callaway—whose patience and concern I cannot claim to have deserved—nor without the help of my collaborators Peter Alstone, Sam Borgeson, Mark Dyson, Christoph Goebel, and Emre Can Kara. To each I am deeply indebted.

I also acknowledge the generous faculty—from Berkeley and elsewhere—who have reviewed and provided valuable input to this research: Dr. Sascha Von Meier, Dr. Severin Borenstein, Dr. Alan Meier, Dr. Pravin Varaiya, Dr. Kameshwar Poolla, and Dr. Zico Kolter.

Also integral to this work were those who went out of their way to collect and provide data. These include Eric Carlson and Karthik Varadarajan of SolarCity, Dr. John "Skip" Dise of Clean Power Research, Landis+Gyr, and Pacific Gas & Electric in conjunction with the Wharton School of Business.

This dissertation was financially supported by an NSF Graduate Research Fellowship, the California Public Utilities Commission, NSF grant CNS-1239567, and a Hellman Graduate Fellowship.

Finally, I would be remiss not to acknowledge the entire Energy and Resources Group (ERG). ERG is bound by a profound sense of purpose to create research that does good for society and the environment. Its community has been an irreplaceable element of my education, and I will continue to be inspired and motivated by my fellow "ERGies" wherever my career takes me.

Chapter 1

Introduction

Mitigating global climate change demands that society dramatically reduce the emission of greenhouse gases, which are intractably linked to burning fossil fuels for energy. Driven primarily—though certainly not exclusively—by these climate goals, energy systems are undergoing a great transition globally as they shift away from fossil based fuels and toward renewables.

This transition requires a re-engineering of the largest and most expensive human-made systems on Earth: power grids. Spanning continents, power grids consist of hundreds to thousands of generators, spinning in unison and responding to even slight changes in the power demanded by consumers. Assets owned by U.S. investor owned utilities are valued at 1.3 trillion dollars, and this only accounts for about half of US power generators.

Populations of industrialized countries have grown so dependent on electric power that even momentary outages cause feelings of vulnerability—and for good reason. The 2003 blackout of the Northeast U.S. and bordering Canadian provinces resulted in an estimated \$6.6 billion (USD) of lost commerce [1]. This is only a fraction of the average annual cost of outages in the U.S., estimated by the Congressional Research Service to be between \$18 and \$33 billion. But outages affect more than the economy. Without power, buildings cannot heat or cool, home medical devices are inoperable, and even quotidian tasks become foreign and risky. During the same 2003 blackout, hospitalizations for accidental injuries or respiratory illness increased two- to eight-fold in New York City, resulting in an estimated 90 additional deaths [2], [3]. Thus any changes in power systems require careful consideration, and are managed by a large number of people and institutions.

Since their inception, power grids relied on controllable generators to meet their primary objective: that power supply and demand be matched at every point in the system at all times—it is an imbalance between supply and demand that ultimately causes cascading failures leading to widespread blackouts. Shifting from fossil fuels to renewable resources challenges this paradigm. In contrast to fossil fuels, renewable resources are inherently variable and uncertain. As a result, renewable generators are less reliable than their fossil counterparts, and are often referred to as Variable Energy Resources (VERs).

Variability and uncertainty are nothing new for power systems, these have always been natural characteristics of demand. Power system operators manage variability and uncertainty by planning generation schedules over a receding horizon—one week ahead, one day ahead, one hour ahead, etc. Also, they keep a certain amount of generation on stand-by—known as reserves—to respond to changes in grid frequency, which result from imbalances between supply and demand. However renewables present a double-edged sword to this paradigm, increasing the amount of variability and uncertainty that needs to be managed, and decreasing the number of traditional, controllable generators left on-line. Maintaining sufficient controllable resources while reducing fossil fuel dependence will be costly. Options include leaving a set amount of fossil generation running in the system but using it much less often, installing batteries to store electricity, buildings excessive amounts renewable generators and curtailing them when not needed, and enabling "demand response" such that consumer demand will shift in order to match fluctuating renewable resources.

This critical juncture for the energy systems overlaps with a revolution in electronic communication and processing speed, which allows for the collection of large datasets and the application of previously infeasible statistical methods. These new data come with the promise of a great opportunity for operating power systems: to use data-driven complex models that will predict system needs with certainty, and provide additional insights into how energy is used and how it can be conserved. Furthermore, these communication tools may operate in two directions, allowing consumers to respond to the needs of the system. Some system planners hope to create a "smart grid" by installing new monitoring and communication devices, actively analyzing the data they collect, and then using this information to take actions that reduce the costs of constructing and operating the system.

But will the smart grid live up to the hype? What data should we collect? What can we realistically learn? And will any of it be helpful or necessary as we transition to sustainable sources of energy? Will data and computing revolutionize energy systems at a time when they need revolution more than ever before? Or will they be an expensive venture into monitoring people and equipment, which minimally contributes to system efficiency, reliability, or sustainability?

My research does not directly answer these questions—at least not in the broad context presented above. But they are the thread that binds my dissertation. In the following chapters, I focus on two different areas of energy policy where new data and new statistical methods allow us to learn what before we had to assume. I spend as much time proposing and evaluating statistical models as I do using them to answer questions. In doing so I try to unveil common themes: pitfalls where the needs of energy systems and the abilities statistical methods don't align as well as they seem; and general means by which we can improve models, usually by paying attention to probability distributions or underlying physical systems. My hope is that this work will be a starting point for a research career of developing empirical methods to improve energy policy.

To be more specific, my dissertation seeks to answer two questions. First, how will the locations of photovoltaic (PV) generators affect the variability and uncertainty of the power they generate? And as a result, will locations of PV systems affect how much supply—or demand—flexibility is needed in grids that rely on them? Second, how many households in California are actively changing thermostat set points throughout the day? How much energy do they save as a result? And how much could be saved if more Californians actively managed set points? To answer these questions, I develop and evaluate statistical models that can be fit to data that have recently become available with installation of new infrastructure.

In the remainder of this introduction, I will give a brief overview of the specific areas that my research addresses. The three subsequent chapters each represent a stand-alone research paper that sits in one of these two areas; Chapter 2 is published in *IEEE Transactions on Power Systems*, Chapter 3 is in peer review for publication in *Solar Energy*, and Chapter 4 will be submitted to *Energy and Buildings*. I conclude in Chapter 5 by compiling the findings from each of my individual studies, and I speculate on the future of data and statistical models in energy systems, laying out a path forward for my own data driven research in this field.

1.1 Do the locations of photovoltaics affect flexibility need?

Renewable generation poses a dual challenge to electricity system planners and operators. It adds variability and uncertainty in electricity supply, while decreasing the number of on-line controllable generators that can be used to balance supply and demand. For electric power systems to rely primarily on renewable generation, "flexible resources" will be needed. These resources could be generators, consumers or storage systems that can reliably increase or decrease supply or demand to maintain balance.

Locations of wind and solar generators affect the need for flexibility. Renewable resources exhibit different amounts of variability and uncertainty regionally, and often renewable resources are spatially correlated. Spatial correlation implies that renewable generators located closer together are more likely to experience simultaneous weather events than those far apart. Thus closely locating systems will likely exacerbate events that need to be managed by the system operator, and vice versa.

Chapters 2 & 3 of this dissertation address the question of whether the spatial dependence of cloud variability and uncertainty affect where PV generation should be sited. Specifically, what are the implications of siting large amounts of PV at the same sub-station? In the same County? Or in the same State? Is rooftop solar inherently less variable and uncertain than utility scale? And are there trade-offs to siting PV in order to minimize flexibility rather than maximize energy production?

Spatial smoothing of PV generation has been identified many times in literature [4]–[7]. Solar generation observed at one small system is much more variable than that of total generation from a set of of systems. This result holds even when systems are very close together, such as those in a utility-scale PV plant [8], [9]. However these studies do not focus on the specific needs of integrating renewable generation into power systems, for which operators are concerned with both variability and uncertainty, and are concerned with the extrema of distributions—a.k.a. the distribution tails.

Renewables integration studies account for spatial smoothing of PV generation in their analyses. However they do so in a limited way, and they do not simulate scenarios that vary the locations of PV plants beyond switching between rooftop and utility-scale systems or moving systems from one state to another. For example, both phases of the Western Wind and Solar Integration Study rely on coarse solar data—10 km spatial resolution and hourly time resolution [10], [11]. All spatial correlations in hourly variability and uncertainty is assumed to be contained in this coarse data; sub-hourly variability and uncertainty are smoothed within each system and assumed to be uncorrelated between systems.

Chapter 2 presents, fits, and validates a model of variability and uncertainty of PV generation that addresses these limitations. I call this model the "volatility state model", due to its reliance on latent states that I refer to as volatility states. Specifically, the model (a) accounts for spatial correlation, (b) predicts metrics of variability and uncertainty that are directly relevant to grid operation and planning, (c) and predicts boundaries on distribution tails that are consistent with observed data.

In Chapter 3, I use the volatility state model to estimate how locations of PV systems affect the need for reserves in future power systems. I use California's long-term planning process as a case study, modeling the need for reserves when load, wind and solar are fixed at levels that are expected to meet the state's 33%

by 2020 goals: 12 GW of photovoltaics and 9.5 GW of wind [12]. I then vary the locations of PV systems to compare centralized versus distributed arrangements of PV, centralized versus distributed balancing authorities, and regional climates.

In addition to quantifying reserve needs, I also estimate the energy produced and the capacity value of solar generation in these different scenarios. In doing so I explore trade-offs between siting PV systems to minimize reserves, or to maximize energy and capacity value.

1.2 Learning HVAC set point schedules from AMI data

Advanced metering infrastructure (AMI) measures electricity consumption at a rate of minutes as opposed to days or months and potentially holds valuable information for energy consumers, providers, and policy-makers [13], [14]. AMI is currently installed in over 40% of residential buildings in the US [15], and a utility typically installs it for every consumer within their territory. As a result, learning from AMI data can be scaled to large, representative samples or even entire populations. It may allow researchers unprecedented access to empirically study questions relating to energy efficiency and behavior without having to rely on expensive or invasive strategies. Thanks to the Green Button Initiative, AMI data are transferable—with consumer permission—to third parties in a standard format [16] and be used by companies to provide consumers with directly tailored feedback.

One of the most important end uses for energy is heating, ventilation, and air conditioning (known collectively as HVAC). HVAC accounts for more than 50% of residential energy use, and defines the timing of peak demand in almost every industrialized power system. Simulation and empirical studies have shown that set point schedules can reduce HVAC energy use by up to 50% [17], [18]. But consumers are notoriously unconscious of how to manage their thermostat set points to save energy, much less how to manage their set points to reduce system-wide costs. Surveys indicate that most consumers do not set back their thermostat at all, when they do they are conservative and only set it back a few degrees; many times, programmed thermostats are overridden to hold mode and the temperature steady [17], [19].

AMI may allow us to study HVAC schedules empirically in a way never available before, but only with the help of heroic statistics. Advanced metering infrastructure (AMI) measures electricity use at a rate of minutes as opposed to days or months. Many researchers and companies attempt to disaggregate the power consumption of specific appliances from AMI data, though this has proven to be difficult. While accuracy of algorithms is progressing it is still relatively low and can be biased. An independent analysis of EEme's algorithm from Pecan Street finds that the median estimates of weekly HVAC consumption are biased by 30% [16], [20]. In order to increase accuracy, algorithms often rely on additional or higher resolution information, such as high frequency power measurements [21], or user input from cell phone surveys [22]. While the costs of additional equipment for this sensing could be (or become) small, engaging residents to adopt it a large barrier that will greatly reduce and bias samples.

A separate set of algorithms use only utility meter data and no other input from consumers; I refer to these as "utility meter" models [23]–[26]. These algorithms trade off data needs for predictive power, seeking primarily to identify only HVAC loads and not individual plug load appliances. They are also often less focused on disaggregation and more focused on understanding thermal properties of buildings, such heat transfer coefficients.

In Chapter 4, I propose that we can use AMI data not only deduce HVAC energy use, but to roughly estimate the set point schedule of a home. I achieve this by comparing disaggregated HVAC energy use from the energy use that a thermal model predicts is needed to maintain a steady temperature. Many utility-meter models inherently estimate thermal properties of homes [25]–[28]. But there are worries over whether these estimates may be biased, as noted in a seminal paper on disaggregation of monthly data [27], and a more recent paper using hourly data from AMI [26]. Using data collected from power sub-meters and thermostats, I evaluate AMI disaggregation methods for their ability to estimate thermal models of buildings and to accurately predict cooling energy use. I conclude by using these estimates to predict a rough indoor temperature schedule for the home.

In future work, I hope to apply my model(s) from Chapter 4—or an adaptation of them—to predict set point schedules of 30,000 homes serviced by Pacific Gas & Electric in California. This study will provide a valuable baseline for the current behavior of Californians, against which energy efficiency and demand response programs may be evaluated.

Chapter references

[1] C. W. Anderson, J. R. Santos, and Y. Y. Haimes, "A Risk-based InputOutput Methodology for Measuring the Effects of the August 2003 Northeast Blackout," *Economic Systems Research*, vol. 19, no. 2, pp. 183–204, Jun. 2007, ISSN: 0953-5314. DOI: `10.1080/09535310701330233`. [Online]. Available: `http://dx.doi.org/10.1080/09535310701330233` (visited on 04/23/2016).

[2] S. Lin, B. A. Fletcher, M. Luo, R. Chinery, and S.-A. Hwang, "Health impact in New York City during the Northeastern blackout of 2003," eng, *Public Health Reports (Washington, D.C.: 1974)*, vol. 126, no. 3, pp. 384–393, Jun. 2011, ISSN: 0033-3549.

[3] G. B. Anderson and M. L. Bell, "Lights out: Impact of the August 2003 power outage on mortality in New York, NY," *Epidemiology (Cambridge, Mass.)*, vol. 23, no. 2, pp. 189–193, Mar. 2012, ISSN: 1044-3983. DOI: `10.1097/EDE.0b013e318245c61c`. [Online]. Available: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3276729/` (visited on 04/23/2016).

[4] M. Lave, J. Kleissl, and J. Stein, "A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501–509, 2013, ISSN: 1949-3029. DOI: `10.1109/TSTE.2012.2205716`.

[5] A. Mills and R. Wiser, "Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power," Lawrence Berkeley National Labratory, Tech. Rep., 2010.

[6] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. Hoff, "Parameterization of site-specific short-term irradiance variability," *Solar Energy*, vol. 85, no. 7, pp. 1343–1353, Jul. 2011, ISSN: 0038-092X. DOI: `10.1016/j.solener.2011.03.016`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X11000995` (visited on 09/12/2013).

[7] A. Murata, H. Yamaguchi, and K. Otani, "A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area," *Electrical Engineering in Japan*, vol. 166, no. 4, pp. 9–19, Mar. 2009, ISSN: 04247760. DOI: `10.1002/eej.20723`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/eej.20723/abstract` (visited on 02/09/2011).

[8] J. Marcos, L. Marroyo, E. Lorenzo, and M. Garca, "Smoothing of PV power fluctuations by geographical dispersion," en, *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 2, pp. 226–237, Mar. 2012, ISSN: 1099-159X. DOI: `10.1002/pip.1127`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.1127/abstract` (visited on 02/17/2016).

[9] A. E. Curtright and J. Apt, "The character of power output from utility-scale photovoltaic systems," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 3, pp. 241–247, May 2008, ISSN: 10627995. DOI: `10.1002/pip.786`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.786/abstract` (visited on 02/09/2011).

[10] G.E. Energy, "Western Wind and Solar Integration Study," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/SR-550-47434, May 2010.

[11] D. Lew, G. Brinkman, E. Ibanez, B. M. Hodge, and J. King, "Western wind and solar integration study phase 2," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.7781&rep=rep1&type=pdf` (visited on 04/22/2015).

[12] Energy and Environmental Economics (E3), "Investigating a Higher Renewables Portfolio Standard in California," Tech. Rep., Jan. 2014. [Online]. Available: `https://www.ethree.com/documents/E3_Final_RPS_Report_2014_01_06_with_appendices.pdf`.

[13] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, Special Section: Transition Pathways to a Low Carbon Economy, vol. 52, pp. 213–234, Jan. 2013, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2012.08.062`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421512007446` (visited on 01/28/2015).

[14] S. Darby, "Smart metering: What potential for householder engagement?" *Building Research & Information*, vol. 38, no. 5, pp. 442–457, 2010.

[15] The Institute for Electric Efficiency, “Utility-sacle Smart Meter Deployments,” Tech. Rep., Sep. 2014.

[16] Energy Information Administration, “An Assessment of Interval Data and Their Potential Application to Residential Electricity End- Use Modeling,” Tech. Rep.

[17] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, “The smart thermostat: Using occupancy sensors to save energy in homes,” in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, ACM, 2010, pp. 211–224. [Online]. Available: `http://dl.acm.org/citation.cfm?id=1870005` (visited on 04/23/2016).

[18] L. Lutzenhiser, H. Hus, M. Moezzi, A. Levenda, and A. Woods, “Lifestyles, Buildings and Technologies: What Matters Most?” *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA*, pp. 12–17, 2012. [Online]. Available: `http://www.aceee.org/files/proceedings/2012/data/papers/0193-000034.pdf` (visited on 03/20/2014).

[19] A. Meier, “How people actually use thermostats,” 2012. [Online]. Available: `http://escholarship.org/uc/item/3vd5q0cp.pdf` (visited on 03/28/2014).

[20] Pecan Street, “Setting the Benchmark for Non-Intrusive Load Monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation,” Tech. Rep., Jan. 2015.

[21] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, “Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring,” en, *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844–858, Oct. 2010, ISSN: 1530-9290. DOI: `10.1111/j.1530-9290.2010.00280.x`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9290.2010.00280.x/abstract` (visited on 03/18/2016).

[22] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, “Leveraging smart meter data to recognize home appliances,” in *2012 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2012, pp. 190–197. DOI: `10.1109/PerCom.2012.6199866`.

[23] B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, and I. H. Rowlands, “Disaggregating categories of electrical energy end-use from whole-house hourly data,” *Energy and Buildings*, vol. 50, pp. 93–102, Jul. 2012, ISSN: 0378-7788. DOI: `10.1016/j.enbuild.2012.03.025`.

[Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0378778812001739` (visited on 02/06/2015).

[24] M. Wytock and J. Z. Kolter, "Contextually Supervised Source Separation with Application to Energy Disaggregation," *ArXiv:1312.5023 [cs, math, stat]*, Dec. 2013, arXiv: 1312.5023. [Online]. Available: `http://arxiv.org/abs/1312.5023` (visited on 10/10/2014).

[25] A. Albert and R. Rajagopal, "Thermal Profiling of Residential Energy Use," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–10, 2014, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2329485`.

[26] M. E. H. Dyson, S. D. Borgeson, M. D. Tabone, and D. S. Callaway, "Using smart meter data to estimate demand response potential, with application to solar energy integration," *Energy Policy*, vol. 73, pp. 607–619, Oct. 2014, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2014.05.053`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421514003681` (visited on 09/12/2014).

[27] M. F. Fels, "Prism: An introduction," *Energy and Buildings*, vol. 9, no. 1, pp. 5–18, 1986.

[28] K. Kissock, J. S. Haberl, and D. E. Claridge, "Development of a toolkit for calculating linear, changepoint linear and multiple-linear inverse building energy analysis models. Final Report for ASHRAE Research Project, No. 1050-RP," *Energy Systems Laboratory Report No. ESLTR-02-11-02*, 2002. [Online]. Available: `http://repository.tamu.edu/handle/1969.1/2847` (visited on 12/11/2013).

Chapter 2

Modeling variability and uncertainty of photovoltaic generation: a hidden state spatial statistical approach[1]

Preface

In this chapter, I propose, fit, and validate a spatial statistical model for variability and uncertainty in PV generation. I refer to this model as the "volatility state" model because of its reliance on latent states, which we refer to as volatility states [29]. In Chapter 2, I use this model to simulate the reserve need resulting from various spatial arrangements of PV systems in California. I performed this work with the guidance of my advisor, Duncan Callaway, and it is published under the same title in *IEEE Transactions on Power Systems*, where Dr. Callaway is a co-author. I reproduce it here with the consent of Dr. Callaway and the IEEE. Within the chapter I refer to all work as both Dr. Callaway's and my own using the pronoun "we."

---

[1] ©2015 IEEE. Reprinted, with permission, from Tabone, M. and Callaway, D. "Modeling variability and uncertainty of photovoltaic generation: a hidden state spatial statistical approach," *IEEE Transactions on Power Systems*, November 2015.

Chapter Abstract

In this paper we construct, fit, and validate a hidden Markov model for predicting variability and uncertainty in generation from distributed (PV) systems. The model is unique in that it: (1) predicts metrics that are directly related to operational reserves, (2) accounts for the effects of stochastic volatility and geographic autocorrelation, and (3) conditions on latent variables referred to as "volatility states." We fit and validate the model using 1-minute resolution generation data from approximately 100 PV systems in the California Central Valley or the Los Angeles coastal area, and condition the volatility state of each system at each time on 15-minute resolution generation data from nearby PV systems (which are available from over 6000 PV systems in our data set). We find that PV variability distributions are roughly Gaussian after conditioning on hidden states. We also propose a method for simulating hidden states that results in a very good upper bound for the probability of extreme events. Therefore the model can be used as a tool for planning additional reserve capacity requirements to balance solar variability over large and small spatial areas.

2.1 Introduction

In 2013, US grid-connected solar photovoltaic (PV) capacity increased by almost 40% (4.7 GW), and solar generation accounted for 29% of all newly installed electricity generation by nameplate capacity. This growth is taking place in a diverse setting of locations and sectors [30]. Of 2,000 MW installed under California's Solar Initiative, over 99% of systems (82% of nameplate capacity) are less than 1 MW in size. [31].

PV generation (along with all solar and wind generation) is different than traditional generation in two important ways: is it **variable**, meaning that it varies uncontrollably as the sun rises and sets, and as clouds pass over PV systems, and it is **uncertain**, meaning that it cannot be perfectly predicted in advance. These properties make PV generation more like electricity demand, which has always been variable and uncertain. Power systems maintain consistent balance of supply and demand as short time-scales by employing reserves, which are readily controllable generators (or loads) placed on stand-by to quickly increase or decrease generation.

These reserves will have to manage increasing amounts of variability and uncertainty as more solar and wind generators are connected. Predicting the amount of variability and uncertainty from PV generation within a balancing area (or an interconnection) is important for predicting future needs for reserves [10], [32]–[34]. Because of the geographic auto-correlation of meteorological phenomena, the locational arrangement of PV panels (i.e. centralized or distributed) will have an effect on the amount of variability and uncertainty exhibited [5], [7].

Renewable generation and operational reserves

In this paper we define variability and uncertainty to relate directly to the reserve needs of power systems. We examine two classes of operational reserves as they are defined in [35]: "load following reserves" account for the difference between a long time-scale market (typically 1 or 2 hour intervals) and a faster market (anywhere between 30 minute and 5 minute intervals); "regulation reserves" account for the difference between the scheduled generation in the faster market and actual net load.

Most renewable integration studies use the "n-sigma" method to quantify the required amount of reserves following increases in wind and solar generation – see those cited within [35], [36]. The n-sigma method plans for variability or uncertainty that is "n" standard deviations away from a mean. This method implicitly assumes the net load variability and uncertainty are Gaussian, although the true distributions often have heavier tails [5], [10], [35]–[37]. To account for this error many renewable integration studies use an artificially large "n" to compute confidence intervals [36]. A second existing approach, known as the "convolution method," computes the distribution of the sum of two random variables with any distribution shapes. However, historical data are needed to compute the original distribution shapes, making this method obsolete for studies that attempt to predict the effects of variable generators that have yet to be built [36].

Statistical models of variability in PV

A number of studies have shown that PV production is geographically autocorrelated [4], [5], which is an important factor for predicting reserve requirements. Murata *et al* [7] demonstrates that a geographic autocorrelation function relating the variability of PV generation from each pair of panels in a network[2] can be used to find the standard deviation of the aggregate generation from all panels in the network. Generation from arrangements of PV systems that are closely located are more correlated and the sum of generation from all of these systems exhibits a greater standard deviation than if the systems were more dispersed. Murata *et al* predicts the mean and standard deviation of variability or uncertainty, but these parameters are insufficient on their own to describe the full complexity of distributions of PV variability.

---

[2] We define a "network" as a spatial arrangement of PV systems, but not the electrical network connecting them.

Distributions of PV production at individual sites have high kurtosis [5]. Cloud regimes are one explanation for this distribution shape, i.e., PV generation signals will have high standard deviation in partly cloudy times with fast wind speeds (resulting in fat distribution tails) and a much lower standard deviation during sunny times or fully overcast times (resulting in a high peak).

Recent studies have attempted to identify different cloud regimes and fit separate models for each of them. Lave and Kleissl condition their geographic auto-correlation parameters on cloud size and speed predictions from numerical atmospheric models [4], [38]. Perez et al. condition their variability and geographic auto-correlation parameters on the spatial variability of satellite predicted solar insolation [39]. Hummon et al. use "variability classes" to simulate the effects of different cloud regimes on PV generation for simulation purposes [37]. Reno and Stein use a "variability index" that classifies days using the standard deviation of the cloud cover ratio [40]. Wegener et al. use a hidden Markov model on wavelet coefficients to predict the standard deviation of 1-second variability from observations of 5 minute variability [41]. However there is little to no evidence that conditioning the standard deviation and auto-correlation of PV variability on observations leads to an accurate distribution shape.

Contributions of this work

In this paper, we present, fit, and validate a model that predicts probability distributions of variability or uncertainty in distributed PV generation from networks of systems with any spatial arrangement. Our objective is to bridge a gap between statistical analysis of variability in PV generation, and power system planning models by

Using high temporal and spatial resolution data from a set of closely located distributed PV systems.

Predicting distribution shapes and geographic autocorrelation of variability and uncertainty between PV systems which allow us to estimate extreme events.

Defining variability and uncertainty in a PV generation signal to be **directly related to operational reserves**.

We note that this paper has a methodological focus. We delay application of the method to future efforts; our immediate objective is to apply the method for long term forecasts of power system reserve requirements for future high renewable penetration scenarios. However, as we will discuss later, the method could also be used for short time scale (e.g. day ahead) forecasts for reserve requirements.

The model presented here resembles hidden Markov models (HMM) for stochastic volatility, which have been used in the financial literature for some time, [42], and have already been used to downscale 15-minute resolution PV generation data to 1-second resolution data at a single system [41]. Hidden Markov models have also been used in the prior literature to forecast mean clearness index of PV insolation [43]–[45]. Our work differs from these models by including spatial dependence of variability and uncertainty from neighboring systems. This dependence exacerbates the amount of flexibility – in the form of reserves – needed by the system operator.

We note that load following and regulation reserve requirements are ultimately a function of *net load*, which also includes wind and load. Existing work on reserve requirements assumes these time series are uncorrelated at the time scales on which reserves are deployed, and can be aggregated post-simulation [36]. We will discuss strategies for dealing with any known correlation between PV generation and wind or load in the conclusions.

Model overview and structure of the paper

We define data inputs and key metrics of interest in Section 4.3, describe the model itself in Section 2.3, and fit and validate the model in Section 2.4. Though a full description of the data and model requires an extended discussion the model itself is relatively straightforward.

The volatility state model predicts metrics of variability or uncertainty, which are denoted $y(t)$ and are explained in Section 2.2. Our primary innovation is that we condition the standard deviation of $y(t)$ on an endogenously estimated latent state, referred to as a "volatility state," $v(t)$, defined in Eqs. (A.15)-(A.17) and fully explained in Section 2.3. The volatility state for each PV system allows for rapid transitions between periods of high standard deviation and low standard deviation. The transition probabilities of $v(t)$ depend on a set of input variables $x(t)$, which can be any set of discrete observations. In this paper we derive $x(t)$ from widely available 15-min resolution PV generation data, fully explained in Section 2.2.

Model parameters are fit using observations of both $y(t)$ and $x(t)$; values of $v(t)$ are latent, meaning that they are never observed and instead are endogenously estimated during the fitting process. For simulation purposes, only observations of $x(t)$ and the parameters are required; distributions of $v(t)$ and $y(t)$ are produced by the simulation; as explained in Section 2.4. The use of latent states is shown to greatly benefit the prediction of extreme events, shown in Section 2.4.

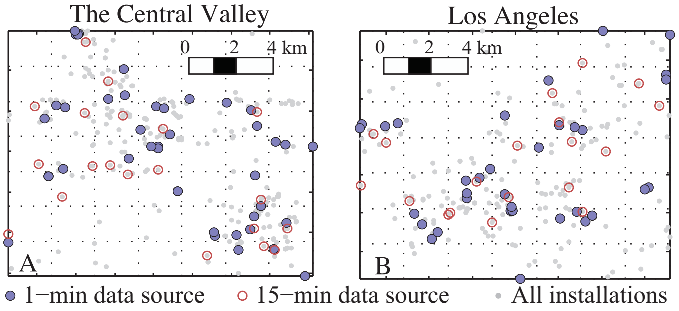

Figure 2.1: Locations of PV systems with 15 min and 1 min resolution data.

2.2 Data and Processing

The data we used for this study comprised instantaneous voltage and current measurements taken from residential and commercial PV installations provided to us by by photovoltaic integrator SolarCity. SolarCity provided 15-minute resolution data for over 6000 systems from January 2011 to late September 2012. These data also included metadata on geometry and capacity for each PV system. To study variability at faster timescales SolarCity increased the sampling rate to once per minute at a small subset of systems.

Figure 2.1 shows locations for sources of 1-minute resolution generation data in the final dataset. We chose these systems to be in one of two 256km^2 areas, each representative of different types of weather in California: the central valley (CV), and the Los Angeles coast (LA). We sampled 100 systems in each area from about 500 available using an algorithm that combined (1) quota sampling for distances between pairs of locations and (2) geographically random sampling of site pairs. Systems were monitored from mid-June to the end of August 2012. We used only systems for which there we no gaps in data over a period of 30 days, leaving us with data from 39 system inverters in LA and 55 in CV. The data in our sample cover a number of extreme events; half of the systems experience a 1 minute duration ramp rate of more than 58% of their nameplate capacity, and half of the systems experience a 5-min average ramp of more than 34% of their nameplate capacity.

The remainder of this section describes the data we used to fit and validate the model. Sections 2.2 and 2.2 describe two variables on which we condition the model. Sections 2.2 describes our metrics for variability and uncertainty.

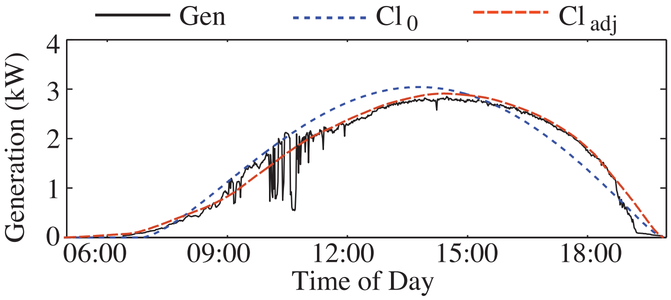

Figure 2.2: one minute resolution PV generation and clear sky signals calculated for one day in Los Angeles. Cl_0 is found using solar-earth geometry and system geometry from metadata, Cl_{adj} is the empirically corrected clear sky signal.

Empirical correction of clear sky signal

The statistical model relies on a "clear sky signal," which represents generation that would have occurred in the absence of cloud cover. Solar-earth geometry – which is predictable given time of year, and location, tilt, azimuth, and effective capacity of a PV system – determines the clear sky signal. However, using the solar-earth model described in [46] with system metadata and a derate factor of 0.77, we found that the modeled clear sky signals were poor estimates of production on sunny days. This is likely due to errors in the system metadata and periodic shading from buildings and trees.

Figure 2.2 shows 1-minute resolution generation for one day along with a clear sky signal based on metadata along with a *corrected* clear sky signal. To implement the correction we first found the difference between actual 15 minute production and the clear sky production predicted using only solar-earth geometry. Second, we identified a "clear sky deviation" as the 95th percentile those differences for each observed time of day, during a centered four-week moving window. Using this percentile excluded many low observations (which removed the effects of cloud-cover) as well as a small number of high observations (which removed the effects of occasional cloud reflection). Third, we smoothed the "clear sky deviation" signal using a 2 hour moving average. After linearly interpolating between the 15 minute intervals, we finally added the deviations back onto the clear sky signal predicted by solar-earth geometry.

Model inputs from 15-min resolution data

As we explain in Section 2.3, we condition the model's volatility states on discrete input data that are specific to the modeled times and locations. This facilitates simulation for locations and times external to this study. Figure 2.1 shows locations

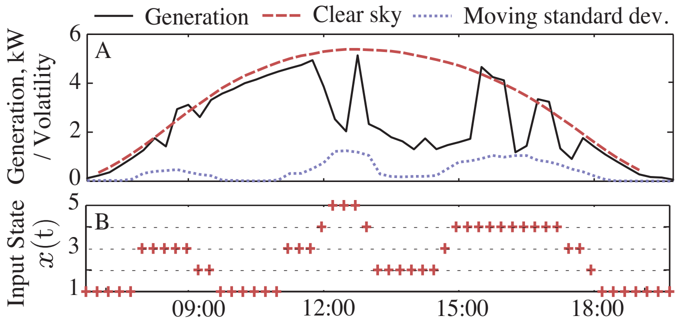

Figure 2.3: A: 15-min generation, clear sky signal, and moving standard deviation for one system. B: Volatility heuristic based on the moving standard deviation.

for our sources of input data: PV systems that continuously recorded generation at 15 minute intervals. To choose systems, we subdivide the study regions into 2km grids. For each grid cell, we chose the system that was closest to the cell's centroid and not part of the 1-min dataset.

To compute conditioning inputs, we used a heuristic to estimate slow time scale volatility. We first calculated a moving standard deviation as follows:

$$\sigma_i(t) = \frac{1}{m+1}\left[\sum_{j=(t-\frac{m}{2})}^{t+\frac{m}{2}}\left(\frac{S_i(j)}{CL_i(j)} - \mu(j)\right)^2\right]^{\frac{1}{2}} \tag{2.1}$$

where where $S_i(t)$ is the solar generation from system i at time t, $CL_i(t)$ is the clear-sky signal for system i at time t, and m is the number of intervals for the moving window ($m = 4$ to account for four 15-minute intervals in an hour). We placed each standard deviation reading into one of 5 bins, resulting in the vector of data inputs for the model:

$$x(t) \in \{1, 2, \ldots 5\}^{N_g}, \tag{2.2}$$

which contains one element for each of N_g grid cells. Binning the data was necessary because the model is conditioned on discrete (not continuous) inputs, as explained in Section 2.3. We defined the bin edges using equally spaced exponential intervals: 0, $e^{-3.5}$, $e^{-2.83}$, $e^{-2.16}$, and $e^{-1.5}$.

Panel A of Figure A.5 shows one day of generation at 15-min resolution along with the moving standard deviation of this signal; panel B shows and the resulting volatility heuristics (conditioning inputs) for this day.

Variability and uncertainty of PV generation

Operational reserves are used to manage both uncertainty and variability in net load. Uncertainty arises from forecast error on the time scale of dispatch. For example, if hour-ahead markets dispatch generators in one hour blocks, there will be error between forecasted hourly average demand and actual hourly average demand. Variability arises because dispatch instructions for an interval must be further adjusted within the interval to maintain supply-demand balance. We calculated metrics representing forecast errors (uncertainty) and deviations from perfect forecasts (variability) corresponding to the markets used by the CAISO [47], defined as:

Hourly forecast errors: the difference between an hour-ahead forecast of hourly average demand and a perfect forecast of hourly average of demand.

5 minute deviations, $\eta_{5,60}(t, S_i)$: the difference between a 5-min resolution forecast (i.e. the forecast used for the load following market) and the perfect forecast of hourly average of demand. See Fig. 2.4C.

5 minute forecast errors, $\epsilon_{5,5}(t, S_i)$: errors in a 5-min ahead persistence forecast of 5-min intervals. See Fig. 2.4D.

1 minute deviations, $\eta_{1,5}(t, S_i)$: deviation of observed generation from a perfect 5-min forecast. See Fig. 2.4E.

The total generation required from load following reserves is the sum of coincident hourly forecast errors and 5 minute deviations; the generation required from regulation reserves is the coincident 5 minute forecast error and 1 minute deviations. Because the model we present in this paper is designed to describe sub-hourly variability and uncertainty, and because numerical weather methods would likely provide a much better hour-ahead forecast than the persistence method, we will not model hourly forecast errors.

Figure 2.4 shows the decomposition process for each metric using one day of PV generation from one system. To forecast PV generation on short time scales, we used the persistence of a clearness index as suggested in [48]. Panel A of Figure 2.4 shows the hour-ahead forecast and a “perfect” hourly forecast. The hourly profiles are as described in [47] where generators provide contracted energy during the middle 40 minutes of an hour and ramp to the next hour’s contract during the following 20 minutes. Panel B of Figure 2.4 shows the perfect and persistence forecast at 5-min intervals, where generators may use the entire 5-min interval to ramp.

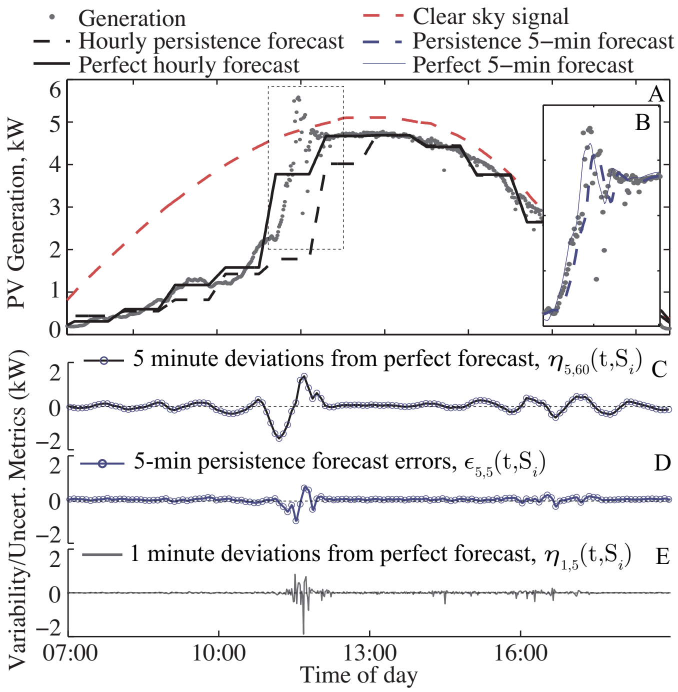

Figure 2.4: Decomposition of 1 min PV generation. Panel A: Generation, hourly persistence forecast and hourly perfect forecast. Panel B (inset): 5-min persistence forecast and 5-min perfect forecast between 11am and 12:30pm (boxed). Panels C through E show each of the metrics we use to build models.

2.3 Volatility state model

The model we develop in this section is an adaptation of a hidden Markov model (HMM) for stochastic volatility [42]. HMMs for stochastic volatility endogenously estimate the occurrence of sharp changes in the standard deviation of a signal such as those in Figures 2.2 and 2.4. For each system and each time, the HMM classifies a reading as being in one of M possible latent states (referred to as "volatility states"), where the latent state defines the standard deviation of the signal.

Figure A.3 depicts the model as a directed acyclic graph, where shaded nodes represent observed variables and unshaded nodes represent unobserved, latent variables. We will estimate a separate model for each variability or uncertainty metric. $y(t) \in \mathbb{R}^{N_s}$, represents normalized variability or uncertainty, $\eta_{1,5}$, $\epsilon_{5,5}$, or $\eta_{5,60}$, it is defined over N_s PV systems. $x(t) \in \{1, 2, ...5\}^{N_g}$ represents the inputs to the model, which in our case are volatility heuristics for each of N_g grid cells, discussed in Section 2.2. $v(t) \in \{1, 2, ...M\}^{N_s}$ is a vector of unobserved volatility states, i.e.

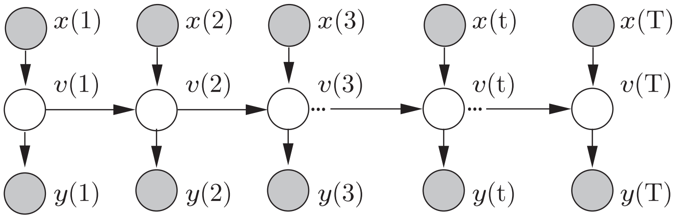

Figure 2.5: Hidden Markov model represented as a directed acyclic graph

each system at each time is in one of M volatility states. In what follows we describe these variables in further detail.

Define $y_i(t)$ as the normalization of a given variability or uncertainty metric for the i^{th} PV system. For example, normalized regulation variability is

$$y_i(t) = \frac{\eta_{1,5}(t, S_i) - \eta_{1,5}(t, CL_i)}{\max_{j \in \text{hour}(t)} CL_i(j)} \tag{2.3}$$

where S_i is the original PV generation time series and CL_i is the clear sky time series. Subtracting $\eta_{1,5}(t, CL_i)$ and dividing by the clear sky trend $CL_i(t)$ removes non-stationary variability resulting from the solar diurnal cycle. Equations for $\epsilon_{5,5}$, or $\eta_{5,60}$ follow the same form. We assume each vector of metrics is mean zero multivariate Gaussian:

$$y(t) \sim MVG\left(0, \mathbf{\Sigma}(v(t); \phi)\right) \tag{2.4}$$

where the covariance matrix $\mathbf{\Sigma}$ is dependent on the volatility states, $v(t)$, and on an exponential geographic autocorrelation function, as defined in Eq. (A.16, A.17).

$$\mathbf{\Sigma}_{ij}(v(t); \sigma^2, \phi) = \begin{cases} \sigma^2_{v_{i(t)}} & i = j \\ \sigma_{v_{i(t)}} \sigma_{v_{j(t)}} \rho(v_i(t), v_j(t); d_{i,j}, \phi) & i \neq j \end{cases} \tag{2.5}$$

$$\rho(m, n; d_{i,j}, \phi) = a_{m,n} \cdot \exp\left\{-d_{i,j}/\tau_{m,n}\right\} \tag{2.6}$$

Diagonal elements of the covariance matrix, Σ, contain the variances of each individual system such that if the i^{th} system is in the m^{th} volatility state $\Sigma_{i,i} = \sigma^2_m$. The off diagonals of Σ represent covariance between systems, defined by the exponential geographic autocorrelation functions defined in Eq. (A.17); where m and

n are the volatility states of systems i and j respectively, $d_{i,j}$ is the distance between systems i and j, and ϕ is a set of parameters $\{\mathbf{a}, \tau\}$. $\tau_{m,n}$ is a range parameter representing the distance over which correlation decreases by 63%. $a_{m,n}$ represents the correlation when $d_{i,j} = 0$. Due to heterogeneous cloud cover for adjacent systems, $a_{m,n}$ can be less than one.

We assume that the probability of being in a volatility state is conditionally dependent on the volatility state at the previous time step and on the input heuristic from the grid cell containing the system, $x_g(t)$ where g indexes the grid cell. Eq. (A.18) shows a set of Markov chain transition matrices that govern the progression of the volatility state for each system; $\mathbf{A}^{(k)} \in \mathbb{R}^{M \times M}$, k indexes the input heuristic. Eq. (A.19) describes each matrix element.

$$\mathbb{A} = \{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots \mathbf{A}^{(K)}\} \tag{2.7}$$

$$\mathbf{A}^{(k)}_{m,n} = P(v_i(t) = n | v_i(t-1) = m, x_g(t) = k) \tag{2.8}$$

Estimating the model

We tested the performance of model fits with $M = 3, 4, 5, 6, 7$ or 8 total volatility states. We cross validated the model fits by withholding 25% of the data during fitting (referred to as the "test data") and using it for validation. Data used for fitting are referred to as the "model data." We estimated parameters in two stages: the first stage estimates the entire model assuming no geographic autocorrelation, the second stage estimates autocorrelation parameters for each pair of volatility states given the output from stage 1.

Stage 1: $v(t)$, σ^2, and $\mathbb{A}$ are estimated via Expectation-maximization (EM): First, EM chooses parameters that maximize likelihood given an expected value of the volatility states. Second, it recalculates the expected value of the volatility states given the updated model parameters.This gradient ascent process repeats to convergence; it is not guaranteed to find the global maximum but will reach a local maximum.

The "expectation" step of EM provides expected values of volatility states given the model parameters, defined in Eq. (2.9), where $\gamma_{i,m}(t)$ is the expected value of an indicator for whether system i is in volatility state m at time t, i.e. it is the probability of finding a given system in a particular state.

$$\gamma_{i,m}(t) = E[\mathbb{I}\{v_i(t) = m\}] \tag{2.9}$$

Stage 2: Eq. (2.10) describes a weighted correlation coefficient $\rho_{i,j,m,n}$ for each variability or uncertainty metric when system i is in state m and system j is in state n.

$$\rho_{i,j,m,n} = \frac{\frac{\sum_{t=1}^{T} y_i(t) y_j(t) \gamma_{i,m}(t)\gamma_{j,n}(t)}{\sum_{t=1}^{T}\gamma_{i,m}(t)\gamma_{j,n}(t)}}{\sqrt{\frac{\sum_{t=1}^{T}\gamma_{i,m}(t)\gamma_{j,n}(t) y_i(t)^2}{\sum_{t=1}^{T}\gamma_{i,m}(t)\gamma_{j,n}(t)}}\sqrt{\frac{\sum_{t=1}^{T}\gamma_{i,m}(t)\gamma_{j,n}(t) y_j(t)^2}{\sum_{t=1}^{T}\gamma_{i,m}(t)\gamma_{j,n}(t)}}} \tag{2.10}$$

The autocorrelation parameters, a and τ, are fit using weighted least squares from the correlation for each pair of sites. Eq. (2.11) shows the weighted objective function for fitting the autocorrelation parameters.

$$\{a_{m,n}, \tau_{m,n}\} = \arg\min_{a,\tau} \sum_{i,j=1}^{N} \left(a \cdot e^{\frac{-d_{ij}}{\tau}} - \rho_{i,j,m,n} \right)^2 \sum_{t=1}^{T} \gamma_{i,m}(t)\gamma_{j,n},(t) \tag{2.11}$$

2.4 Results

Parameter Estimation

Figure 2.6 displays the log-likelihood of the model data (data used to fit the model) and the test data (reserved data for testing) fit to each metric. When $M > 5$ for $\eta_{1,5}$ and $\eta_{5,60}$ and $M > 7$ for $\epsilon_{5,5}$, improvements in log likelihood are small and added states are encountered less than 0.5% of the time. Therefore for the remainder of the analysis we use models with 5 states for $\eta_{1,5}$ and $\eta_{5,60}$, and 7 states for $\epsilon_{5,5}$.

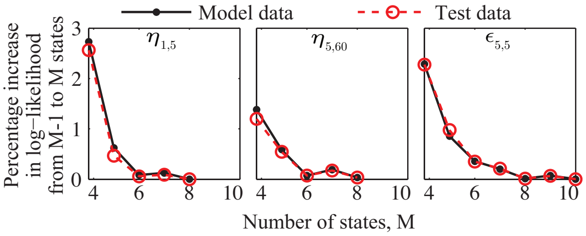

Figure 2.6: Percent log-likelihood increase from increasing the number of states.

Tables A.11 and A.10 show estimates of the autocorrelation parameters, $a_{m,n}$ and $\tau_{m,n}$, for regulation and load following variability, where higher volatility state index corresponds to higher standard deviation. For load following variability, a (correlation at a distance of 0m) generally increases with volatility state standard deviation. Trends in τ suggest a non-monotonic relationship with volatility state, where the decay range is short for high and low variance states, and long for moderate variance.

Table 2.1: Parameters for autocorrelation functions fit to $\eta_{1,5}(t)$.

	a					τ				
S#	1	2	3	4	5	1	2	3	4	5
1	0.41	0.18	0.05	0.02	0.00	0.21	0.40	0.53	0.21	0.10
2	-	0.21	0.14	0.04	0.01	-	1.54	2.05	1.76	0.10
3	-	-	0.53	0.35	0.10	-	-	1.25	1.18	0.56
4	-	-	-	0.66	0.39	-	-	-	0.97	0.68
5	-	-	-	-	0.61	-	-	-	-	0.38

Table 2.2: Parameters for autocorrelation functions fit to $\eta_{5,60}(t)$

	a					τ				
S#	1	2	3	4	5	1	2	3	4	5
1	0.20	0.17	0.07	0.06	0.00	12.76	12.94	6.94	0.61	0.10
2	-	0.23	0.17	0.06	0.00	-	8.45	5.98	2.46	0.10
3	-	-	0.49	0.38	0.18	-	-	6.51	8.73	7.58
4	-	-	-	0.61	0.51	-	-	-	7.84	5.88
5	-	-	-	-	0.74	-	-	-	-	5.42

Validation by Simulation

To validate the model we compare observations to simulated distributions, which require estimates of the volatility states. We simulate volatility states with the following methods:

$v_i(t)$ *Simulation Method 1*: Set $v_i(t)$ equal to the most likely state as estimated during model fitting. This requires that we restrict simulations to times and locations for which we have one minute data. Because we ultimately want to use the model to estimate reserves in the absence of one minute data, this method is only a baseline.

$v_i(t)$ *Simulation Method 2*: For each system at each time, simulate N_p samples of volatility states using the stationary probabilities of the transition matrices in

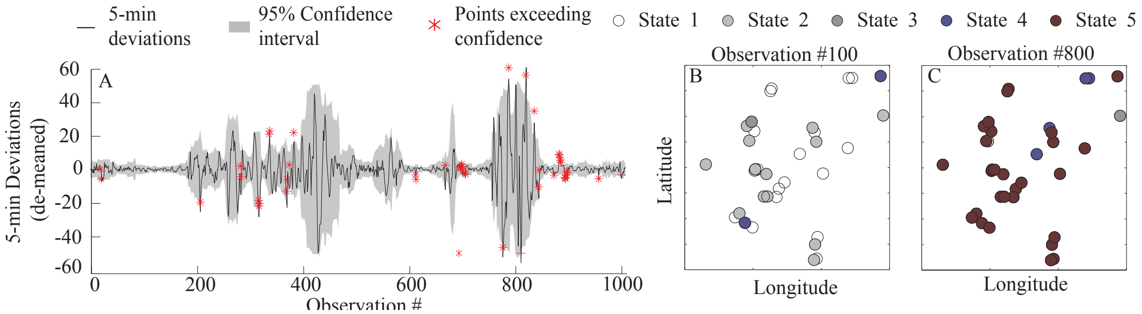

Figure 2.7: Panel A: Observed, demeaned, aggregate 5 min variability ($\eta_{5,60}$) from the entire network of observed systems. Light grey boundaries represent the 95% confidence interval from method 1; points exceeding this interval are highlighted by red stars. Panels B and C: Example volatility states of systems.

$\mathbb{A}$. In this paper we use $N_p = 40$ samples. We model the distribution for each system at each time as a Gaussian mixture with N_p equally likely components, one for each sampled covariance matrix. This method neglects correlation between volatility states.

$v_i(t)$ *Simulation Method 3*: Simulate N_p volatility states per site as in method 2, then independently sort the N_p volatility states for each system at each time from highest to lowest variance, such that $v_i^{(1)}(t)$ and $v_i^{(N_p)}(t)$ contain the highest and lowest standard deviations, respectively; this maximizes correlation of volatility states.

No latent states: We also construct, fit and test a separate benchmark model without latent states by conditioning standard deviation and the geographic autocorrelation parameters directly on the 15-min volatility heuristics, instead of on a latent state.

While it may be possible to estimate correlation between discrete volatility states, the problem is non-trivial; most geographic autocorrelation models use continuous distributions.

We use the simulated input states to calculate the standard deviation of aggregate variability or uncertainty from the entire network of monitored systems at each time. First, we calculate the normalized covariance matrix with Eq. (A.16). Second, we transform the covariance matrix to represent the de-normalization of the variability or uncertainty, i.e., the inverse of Eq. (2.3). For this transformation we multiply each element of the normalized covariance matrix by the hourly maximum of the clear sky signal for each system. Finally, we sum all elements of the covariance matrix to represent the summation of variability or uncertainty from all systems.

Figure 2.7 shows the output of the method 1 simulation for $\eta_{5,60}(t)$. The 95% confidence interval is defined as twice the simulated standard deviation, and the volatility states are ordered such that state 1 has the lowest standard deviation and state 5 has the greatest. The confidence interval is wider when simulated volatility states have a greater standard deviation.

Figure A.6 shows quantile-quantile (QQ) plots that compare standardized quantiles of the observed data to quantiles of a standard normal. We standardize the quantiles by first computing the position of the observed data within the simulated model's cumulative density function for all observations in the study period, and then taking the standard normal inverse CDF of the result. If the model and its parameterization predict the empirical distribution, the standardized quantiles should be normally distributed, and points in the QQ plot will lie along the $y = x$ line.

Row 1 of Figure A.6 shows results for the no latent state model. For 1-min and 5-min deviations, the tails of the observed data are "heavy" compared to the standard normal, meaning that the simulated distribution will under-predict extreme events. For 5-min forecast errors, the tails of the observed data are "light" compared to the standard normal, meaning that the model over-predicts extreme events. In contrast, Row 2 of Figure A.6 shows that the volatility state model more accurately estimates distributions in the baseline scenario (method 1); distributions tails are only slightly light for deviations and slightly heavy for persistence forecast errors. This result indicates that if the volatility state distribution across sites is well characterized, the model will work well in times and periods for which one minute data are unavailable.

Row 3 of Figure A.6 show that method 2 results are worse than method 1 and comparable to the model without hidden states. Because method 2 does not model volatility state correlation across systems, the probability that multiple systems will be in a high volatility state simultaneously is relatively low, and we expect this simulations to under-predict extreme events – i.e. the observed data tails will be heavy. For 1- and 5-min deviations the observed tails are heavy, as expected. For 5-min persistence forecast errors the tails are light, though trend toward crossing the $y = x$ line back to heavy.

The bottom row of Figure A.6 shows results from method 3, which maximizes volatility state correlation across systems; the effect is evident if one compares the tails of the QQ plots between methods 2 and 3. For 1-min and 5-min deviations, method 3 performs similarly in the tails to method 1, suggesting that volatility states are in fact highly correlated. For 5-min forecast errors the simulated distribution has heavy tails.

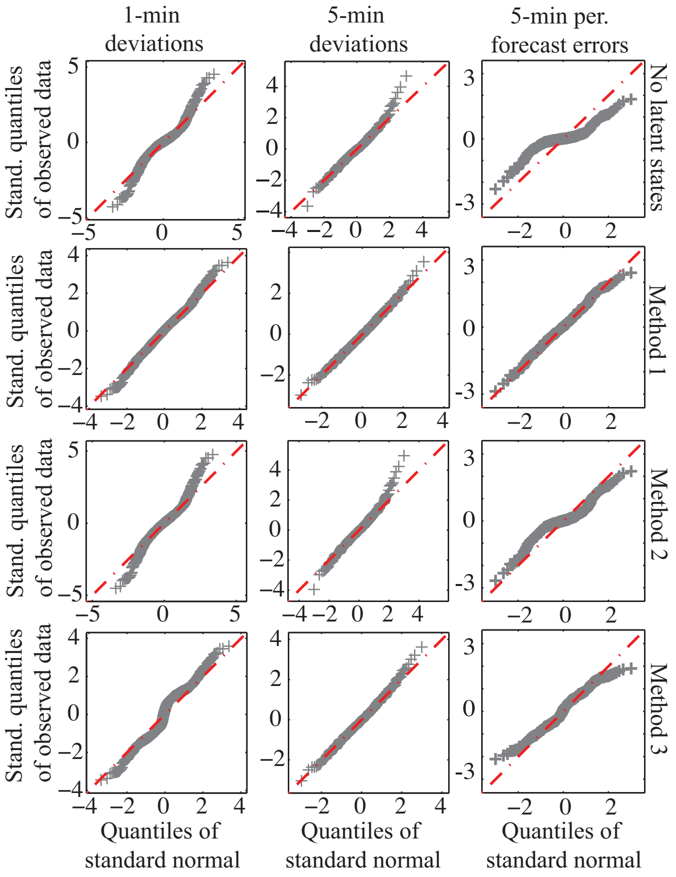

Figure 2.8: QQ plots of test data pseudo-residuals using each volatility states simulation method and for each metric.

Predicting maximum events

Power system planners are concerned with high impact, low probability events, i.e., the maximum regulation or load following reserve required within a given time period. Eq. (A.35) shows a method for finding the probability that all observations within a time period fall below some threshold, x, assuming independent observations; where T_{HOD} is a specified hour of day. The 95% confidence interval for the maximum requirement corresponds to probabilities of 0.975, and 0.025.

$$p\{\eta_{1,5}(t) \leq x : t \in T_{HOD}\} = \prod_{t \in T_{HOD}} p\left\{\eta_{1,5}(t) \leq x\right\} \tag{2.12}$$

Figure A.7 shows 95% confidence intervals for the maximum reserve requirement

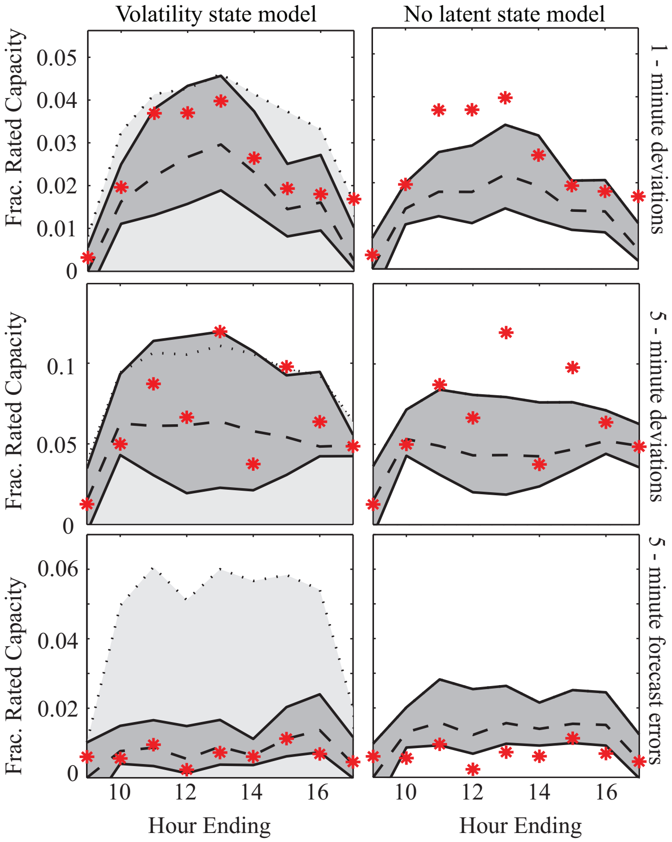

Figure 2.9: Predicted distributions of maxima observed during each hour of the day. Rows represent each metric ($\eta_{1,5}(t)$, $\epsilon_{5,5}(t)$, $\eta_{5,60}(t)$). Columns represent predictions by the volatility state model (left), and the no latent state model (right). The dark grey boundaries signify a 95% confidence interval found using the baseline scenario where the most likely volatility states predicted by EM is used, the light grey boundary signifies the 97.5^{th} percentile of the predicted distribution of maxima using the sorting based worst case assumption. Red stars signify observed maxima.

estimated for the test data, stratified by each hour of day. The left-hand column shows predictions from the volatility state model and the right-hand column shows predictions using the model without latent states. The dark grey boundaries are those calculated with known volatility states. The light grey (dotted line) boundary is 97.5% bound (i.e. the upper bound of a 95% confidence interval) of the predicted distribution using method 3, the worst-case assumption for geographic autocorrelation of volatility states. In expectation, observed maxima should fall above the 97.5% confidence bound between 0 and 1 times for 27 observations. For method 3 there are 3 observations above the bound, whereas the no latent state model has 7 above the bound. We note that the reserved test data – though randomly chosen – have slightly more extreme events than the data use to estimate the model. For 1- and 5-min deviations, the method 3 upper bound dips below the upper bound for method 1 for a few hours; this results from small differences between stationary hidden state probabilities from method 3 versus the most likely volatility states predicted in method 1.

2.5 Concluding remarks

In this paper we presented, fit, and validated a hidden Markov statistical model for variability and uncertainty in PV generation that parametrically estimates both geographic autocorrelation and stochastic volatility. The model differs from others in the literature by conditioning on latent "volatility states," which account for discontinuous changes in the standard deviation of variability or uncertainty from PV generation. We fit the model to metrics of PV generation that are useful for the planning of load following and regulation reserves: (1) 5 minute persistence forecast errors made 5 minutes ahead, (2) 1 min deviations from a perfect 5 minute interval forecast, and (3) 5 minute deviations from a perfect hour ahead forecast. These metrics relate to the use of load following and regulation reserves by power system operators.

Given knowledge of the latent states (only possible for locations or times for which we have 1 minute resolution data), the model predicts distributions well, even in the tails. For regions that lack 1 minute data we built a method to simulate latent states in a way that maximizes their correlation (see method 3 in Section 2.4). This latent state simulation method produces comparable results to those when latent states are known (though the method is overly conservative for uncertainty). We also presented results for a model similar to existing models in that it conditions on observations instead of latent variables; this model under-predicts extreme events associated with variability and over-predicts extreme events associated with uncertainty relative to the model we developed in this paper. We expect that, by increasing the number of extreme events available to fit the model, additional data would increase the performance of the latent state model relative to the model without latent states.

Though earlier research[5], [10], [35]–[37] and the results in this paper suggest that unconditioned Gaussian distributions do not characterize PV variability well, it is possible that as larger data sets become available, future research will show that a single Gaussian describes aggregate PV variability well across large spatial scales, for reasons related to the central limit theorem (CLT). We conjecture that unconditioned

Gaussian distributions will work well for an aggregation of small systems distributed across hundreds to thousands of kilometers, but not for a small number of very large utility-scale systems, even if they are relatively far apart.

Our intended application of the model is to predict the amount of reserves required to integrate PV generation into power systems. Because the model accounts for spatial auto-correlation and is conditioned on spatial inputs, it is uniquely situated to compare the additional reserve requirements from centralized versus distributed PV systems. Any model that predicts reserve requirements must also account for variability from non-PV sources, namely load variability and wind variability. For future work, we plan to identify increases in operational reserve needs in California as centralized and distributed arrangements of PV are added. To identify these requirements, we will combine this model's simulated distributions of PV variability with similar metrics for wind and load predicted by the CAISO for the same time periods (as simulated for [49]). Combining wind, load, and solar variability is achieved by summing coincident forecast errors or deviations for each; i.e., the same method used in the n-sigma method or the convolution method [36]. We note that if there is known positive or negative correlation between solar and wind or load, the model we have developed can be combined with wind and load variability models in a number of ways that preserve this correlation. For example, one could condition wind and net load models on the same inputs as the PV model, which will induce some correlation due to the common predictors. Alternately, one may include variables representing wind and load directly within the model presented and estimate correlations endogenously – using the same mathematical framework developed in this paper.

One could also use this model's estimates for reserve requirements in unit commitment optimal dispatch models for different renewable energy penetration scenarios. Recent work from the National Renewable Energy Laboratory (NREL) uses a unit commitment optimal dispatch model that accounts for time-varying reserve needs, but in a way that would be improved by this model [33], [34]. This model, or a variation of it, may also be used by system operators to predict the the required amount of reserves to procure.

2.6 Acknowledgments

We thank Eric Carlson and Karthik Varadarajan of SolarCity who provided us with the data required for this work. We also thank Alexandra Von Meier and four anonymous reviewers, whose comments greatly improved this work.

2.7 Chapter References

[4] M. Lave, J. Kleissl, and J. Stein, “A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501–509, 2013, ISSN: 1949-3029. DOI: `10.1109/TSTE.2012.2205716`.

[5] A. Mills and R. Wiser, “Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power,” Lawrence Berkeley National Labratory, Tech. Rep., 2010.

[7] A. Murata, H. Yamaguchi, and K. Otani, “A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area,” *Electrical Engineering in Japan*, vol. 166, no. 4, pp. 9–19, Mar. 2009, ISSN: 04247760. DOI: `10.1002/eej.20723`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/eej.20723/abstract` (visited on 02/09/2011).

[10] G.E. Energy, “Western Wind and Solar Integration Study,” National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/SR-550-47434, May 2010.

[29] M. Tabone and D. Callaway, “Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach,” *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–9, 2014, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2372751`.

[30] Solar Energy Industry Association, “Solar Market Insight Report 2013 Year in Review,” Solar Energy Industry Association, Tech. Rep., Mar. 2014.

[31] *The {C}alifornia {S}olar {I}nitiative*, Apr. 2014. [Online]. Available: `http://www.gosolarcalifornia.org/about/csi.php` (visited on 04/25/2014).

[32] M. Rothleder, *Track I Direct Testimony Of Mark Rothleder On Behalf Of The California Independent System Operator Corporation*, California Public Utilities Commission, 2011.

[33] M. Hummon, P. Denholm, J. Jorgenson, D. Palchak, B. Kirby, and O. Ma, “Fundamental Drivers of the Cost and Price of Operating Reserves,” National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep., 2013. [Online]. Available: `http://www.nrel.gov/docs/fy13osti/58491.pdf` (visited on 02/28/2014).

[34] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://www.smartgridinformation.info/pdf/5506_doc_1.pdf` (visited on 04/22/2015).

[35] M. Milligan, P. Donohoo, D. Lew, E. Ela, B. Kirby, H. Holttinen, E. Lannoye, D. Flynn, M. OMalley, and N. Miller, "Operating reserves and wind power integration: An international comparison," in *Proc. 9th International Workshop on large-scale integration of wind power into power systems*, 2010, pp. 18–29. (visited on 02/28/2014).

[36] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Gomez Lazaro, and N. Detlefsen, "Methodologies to determine operating reserves due to increased wind power," in *2013 IEEE Power and Energy Society General Meeting (PES)*, Jul. 2013, pp. 1–10. DOI: `10.1109/PESMG.2013.6673067`.

[37] M. R. Hummon, E. Ibanez, G. Brinkman, and D. Lew, "Sub-Hour Solar Data for Power System Modeling from Static Spatial Variability Analysis," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep. Pending, 2012.

[38] M. Lave and J. Kleissl, "Cloud speed impact on solar variability scaling Application to the wavelet variability model," *Solar Energy*, vol. 91, pp. 11–21, May 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2013.01.023`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X13000406` (visited on 09/12/2013).

[39] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. E. Hoff, "Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance," *Solar Energy*, Progress in Solar Energy 3, vol. 86, no. 8, pp. 2170–2176, Aug. 2012, ISSN: 0038-092X. DOI: `10.1016/j.solener.2012.02.027`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X12000928` (visited on 05/14/2014).

[40] M. J. Reno and J. Stein, "Using Cloud Classification to Model Solar Variability.," Sandia National Laboratories, Tech. Rep., 2013. [Online]. Available: `http://energy.sandia.gov/wp/wp-content/gallery/uploads/SAND-2013-1692C\_ASES-CloudCategoryVariability.pdf` (visited on 08/31/2013).

[41] J. Wegener, M. Lave, J. Luoma, and J. Kleissl, "Temporal downscaling of irradiance data via Hidden Markov Models on Wavelet coefficients: Application to California Solar Initiative data," 2012. [Online]. Available: `http://solar.`

ucsd.edu/datasharing/doc/UCSDReport{_1}secCSI.pdf (visited on 12/06/2012).

[42] R. Langrock, I. L. MacDonald, and W. Zucchini, "Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models," *Journal of Empirical Finance*, vol. 19, no. 1, pp. 147–161, Jan. 2012, ISSN: 0927-5398. DOI: 10.1016/j.jempfin.2011.09.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927539811000661 (visited on 04/29/2014).

[43] Y.-z. Li, L. He, and R.-Q. Nie, "Short-Term Forecast of Power Generation for Grid-Connected Photovoltaic System Based on Advanced Grey-Markov Chain," in *International Conference on Energy and Environment Technology, 2009. ICEET '09*, vol. 2, Oct. 2009, pp. 275–278. DOI: 10.1109/ICEET.2009.305.

[44] P. Poggi, G. Notton, M. Muselli, and A. Louche, "Stochastic study of hourly total solar radiation in Corsica using a Markov model," en, *International Journal of Climatology*, vol. 20, no. 14, pp. 1843–1860, Nov. 2000, ISSN: 1097-0088. DOI: 10.1002/1097-0088(20001130)20:14<1843::AID-JOC561>3.0.CO;2-O. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/1097-0088(20001130)20:14<1843::AID-JOC561>3.0.CO;2-O/abstract (visited on 05/08/2014).

[45] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, vol. 79, no. 5, pp. 469–482, Nov. 2005, ISSN: 0038-092X. DOI: 10.1016/j.solener.2004.12.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X05000204 (visited on 05/08/2014).

[46] G. M. Masters, *Renewable and Efficient Electric Power Systems*, 1st ed. Wiley-IEEE Press, Aug. 2004, ISBN: 0471280607.

[47] Y. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational Impacts of Wind Generation on California Power Systems," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039–1050, May 2009, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2009.2016364. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4808228 (visited on 06/07/2011).

[48] E. Ibanez, G. Brinkman, M. Hummon, and D. Lew, "A Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-Hourly Variability Analysis," in *2nd International Workshop on Integration of Solar Power in Power Systems Proceedings, Lisbon, Portugal*, 2012. [Online]. Available: `http://www.nrel.gov/docs/fy12osti/56169.pdf` (visited on 09/02/2013).

[49] C. Goebel and D. Callaway, "Using ICT-Controlled Plug-in Electric Vehicles to Supply Grid Regulation in California at Different Renewable Integration Levels," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 729–740, Jun. 2013, ISSN: 1949-3053. DOI: `10.1109/TSG.2012.2218629`.

Chapter 3

The effect of PV siting on power system flexibility needs

Preface

In the following chapter, Duncan Callaway, Christoph Goebel and I use the volatility state model from Chapter 2 to predict how varying the spatial arrangements of PV affects the need for reserves in California. This Chapter was submitted to *Solar Energy* and is currently in preparation for resubmission after revisions. I reproduce it here with the consent of my co-authors. I specifically acknowledge the contributions of my co-author Christoph Goebel who acquired wind and load signals from the California Independent System Operator, modeled hourly and five minute forecast errors in these signals, and provided support reviewing and editing the manuscript. The rest of the content of this chapter is my own work, performed with the guidance of Prof. Duncan Callaway. Within the chapter I acknowledge the contributions of all co-authors by using the pronoun “we.”

Chapter Abstract

Locations of photovoltaic (PV) systems affect the variability and uncertainty of their power generation, and as a result the amount of flexible resources needed to balance supply and demand. However, studies on the integration of renewable electricity into power systems focus on the total amount of renewable generation, and not their locations. This paper uses a hidden state, spatial-statistical model to simulate how locational arrangements and balancing policies affect the need for reserves—load following and regulation—in California's power system when including 12 GW of photovoltaic generators and 9.5 GW of wind.

Our results show that locations of utility-scale PV system significantly affect on the amount of reserves needed for balancing their variability and uncertainty. When PV is geographically dispersed, the additional load following and regulation reserve needs are small; on average $< 1.1\%$ and $< .05\%$ of installed PV capacity respectively. These rise to 3.8% and .2% in centralized scenarios. *Most the benefits of this dispersion can be achieved with relatively large systems, roughly 500MW, which are almost precisely the size of the largest systems in California today.* Almost all of the load following reserve need is driven by hourly forecast errors for solar generation. These forecast errors are autocorrelated, thus they can be mitigated either by better forecasts, or dispersing plants.

Siting policies for PV must weigh system flexibility against other locational concerns, such as the energy and capacity values of the solar resource in an area and the costs of developing transmission. We find a small trade off between energy and reserves; where dispersed systems require less reserves but also have lower capacity factors than more centralized systems. However, we find a much greater trade-off between energy and capacity value in California; where the regions that produce the most energy on average—in the Mojave desert—tend to be cloudy during current peak demand hours, which occur during Summer afternoons.

3.1 Introduction

Renewable generation poses a dual challenge to electricity system planners and operators. It adds variability and uncertainty in electricity supply, while decreasing the number of on-line controllable generators that can be used to balance supply and demand. For electric power systems to rely primarily on renewable generation, flexible resources will be needed. These resources could be generators, consumers or storage systems that can reliably increase or decrease supply or demand to maintain balance.

A number of renewables integration studies [10], [11], [33], [34], [50]–[52] investigate how increasing the amount of wind and solar generation will impact power system operations. Among other things, these studies estimate the amount of flexible resources needed to be held in reserve to respond to fast or unexpected changes in uncontrollable demand or generation. Notably, Wu et al. find that including the PV capacity amounting to 20% of peak demand in the Arizona Public Service territory results in increases in reserve needs by 2% to 3% of installed PV capacity [50].

Locations of wind and solar generators affect the need for flexibility. Renewable resources exhibit different amounts of variability and uncertainty regionally, and often renewable resources are spatially correlated. Spatial correlation implies that renewable generators located closer together are more likely to experience simultaneous weather events than those far apart. Thus siting PV systems in close proximity to one another may exacerbate events that need to be managed by the system operator, and the opposite may be true for systems sited far apart from one another.

Renewables integration studies often account for spatial smoothing of PV generation in their analyses; however siting of renewable generators and transmission are not the focus of these studies. They account for spatial smoothing in a limited way (we explain further below), and they do not simulate scenarios many variations in the locations of PV plants beyond switching between rooftop and utility-scale systems or moving systems from one state to another. For example, the Western Wind and Solar Integration Study (WWSIS) Phase 1 varies locations of renewable generators in only one scenario, where renewables are placed in the best location in the Western US, rather than the best locations in each state [10]. Even California's transmission planning studies perform only limited number of variations for locations of Solar—placing most if not all of it in the Tehachipi region—and do not account for the effects of spatial correlation on reserve need [53].

In this study we seek to answer questions about where PV systems should be sited within a state. What are the implications of siting large amounts of PV at the same sub-station, or in the same county? Can utility-scale systems achieve the same benefits of spatial smoothing as rooftop systems? Do regions with minimal variability and uncertainty overlap with regions of great energy resource, or with great capacity value? The need for answering such questions is demonstrated by regional transmission planning processes, such as those in California [53].

Our paper uses California as a case study to answer these questions. We model both cloud variability and uncertainty as spatially correlated processes that are dependent on regional climates, and compute the variability and uncertainty from various spatial scenarios for PV locations. We then combine these results with variability and uncertainty from wind and load. By comparing results from these scenarios, we learn how locational siting affects the need for flexibility in power systems.

Statistics of solar variability and uncertainty

A large body of prior work focuses on the spatial correlation of variability from wind and solar generation. See [4]–[7] for examples regarding solar with complete literature reviews, and [54], [55] for examples regarding wind. Most of these studies use spatial correlation functions to predict the standard deviation of variability or uncertainty from theoretical arrangements of PV systems [7], [39], [56], and do so accurately.

Despite accurate knowledge of the standard deviation, it remains difficult to predict variability or uncertainty from PV in a way that is useful for power system operations or planning. System operators are concerned with high impact, low probability events which are defined by distribution tails—for example, system operators may plan to balance supply and demand in $> 95\%$ of times. Mean and standard deviation completely parametrize only a few distribution shapes, most notably the Gaussian distribution. Unfortunately, the distribution of variability from PV has been shown not to be Gaussian in many studies [5], [7], [38], [57]. Two studies focus on predicting non-Gaussian distribution shapes for PV variability. Both of these models condition variability on hidden Markov processes that are analogous to cloud cover patterns [33], [41]. However these models predict variability and not uncertainty in PV generation, and do so only for individual systems, without modeling spatial correlation of neighboring systems.

Another strategy accounts for spatial diversity by filtering the generation from a small PV system—or a single irradiance point sensor—to account for the geographic smoothing that would occur in a larger plant [38], [58]. Different filters are applied at different time-scales to represent increased spatial correlation at longer timescales and vice versa. This strategy maintains the non-Gaussian shape of variability at each time-scale, but scales the width—i.e., variance—of the distribution to match what is expected from statistics. These methods provide a nice bound on the heaviness of a distribution tail, however do not account for the fact that this distribution shape will necessarily become lighter-tailed as more spatial smoothing occurs. Also, while this method accounts for smoothing in one utility-scale PV system, it does not account for spatial correlation between neighboring utility-scale systems, which is precisely what we would like to study in this paper.

In our earlier work [59], we developed a statistical model that addresses these issues by (a) accounting for spatial correlation, (b) predicting metrics of variability and uncertainty that are directly relevant to grid operation and planning, (c) and predicting boundaries on distribution tails that are consistent with observed data.

This paper makes use of this model to estimate how the spatial arrangement of PV in California may affect the need for reserves.

Predicting reserve need in renewable integration studies

Renewable integration studies treat reserve procurement in differing ways—driven both by the procedures of system operators and by the assumptions made in each study. For example, many studies define the amount of generation held in reserve to be constant throughout the year, or to be a fixed percentage of total load, wind, and/or solar in the system [10], [32]. In contrast, more recent studies define the amount of reserves procured each hour to be dynamic, reflecting varying expectations of load and renewable generation in each specific hour [33], [34], [50].

Methods to calculate the amount of reserves to be procured—dynamic or constant—also vary among studies. Most early studies used the n-sigma method, which defines reserve need to be n standard deviations of expected variability and/or uncertainty. If distribution shape of variability or uncertainty is Gaussian, the n-sigma method defines a confidence interval for reserve needs. However the distribution shape has been shown not to be Gaussian. It has heavier tails, signifying that extreme events are more likely than a Gaussian would predict [36], [60]. Integration studies will often use an arbitrarily large n as a conservative approximation for boundaries on these wide tails.

More recent studies overcome this limitation by attempting to directly compute the bounds within which reserves will be sufficient for 95% of times. For example, the 2nd Phase of the Western Wind and Solar Integration Study define load following reserve procurement by the the square root of the sum of squares of the 70th percentile hourly forecast errors for wind and solar [11]. Wu et al. actually simulate 200 possibilities for reserve need at each interval, and procure reserves that will cover 95% of simulated need. We adopt Wu's strategy in this paper [50].

Overview

In this study, we build on our earlier modeling work and upon previous renewable integration studies to estimate how locations of PV systems affect the need for reserves in future power systems. We use California's long-term planning process as a case study, modeling the need for reserves when load, wind and solar are fixed at levels that are expected to meet the state's 33% by 2020 goals: 12 GW of photovoltaics and 9.5 GW of wind [12]. We then vary the locations of PV systems to compare centralized versus distributed arrangements of PV, centralized versus distributed balancing authorities, and regional climates. In doing so, we answer the question of how geographic dispersion of PV systems can mitigate flexibility need in future power systems.

In addition to quantifying reserve needs, we also estimate of the energy production and capacity value of solar generation in these different scenarios. In doing so we explore trade-offs between siting PV systems to minimize reserves, or to maximize energy or capacity value.

3.2 Method

Locational scenarios

Table 3.1 lists each of the locational scenarios we examine in this paper; each includes 12 GW of rated PV capacity.[1] Six scenarios represent arrangements of **utility-scale** PV plants ranging from highly concentrated to highly dispersed.[2] We limited candidate locations to 130 "ecologically preferred" regions, which were randomly selected from all possible locations outside of the environmentally excluded zones identified in [61].[3]

In the first five utility scale scenarios we site plants where they are expected to have the maximal capacity factor (see Figure 3.1), i.e. where they have the greatest expected energy production. For the sixth utility-scale scenario we site a single 12 GW system where it is expected to have the greatest capacity value, i.e. the site with the maximal expected capacity factor of PV during the peak 50 hours of demand minus wind production. In Section 3.3 we find the actual capacity value of each scenario after simulation.

We ran eight additional scenarios, referred to as **dispersion tests**, to more specifically study the effect of spacing between utility-scale PV systems and climate conditions. These scenarios consist of 100, 120 MW PV systems randomly located in areas of different sizes: 50km, 100km, 200km, and 300km squares. Areas used for the dispersion test scenario are shown on the rightmost maps in Figure 3.1. We created two scenarios for each of these area sizes, one in the "best" available climate and one in the "worst". All areas used for the dispersion test scenarios are shown on the

---

[1] We define rated capacity by the AC power produced under 1000 W/m^2 of direct solar irradiance, after the balance of systems and other efficiency losses.

[2] For all non-rooftop scenarios, we assume that utility-scale PV systems have ground cover ratio of 0.15 and are aligned to face due south with a tilt equal to latitude.

[3] Wu et al. identify multiple different criteria for exclusion. At the recommendation of the authors, we use the second most stringent criteria—their own work finds the most stringent criteria to be infeasible. These exclude all land already excluded by law, land as high risk for biodiversity, and land that contains unique natural characteristics, historical value, or cultural value.

Table 3.1: Locational scenarios for 12GW of rated PV capacity in California.

Scenario	# of plants	Plant cap.	Location
Utility-scale scenarios			
100-sys	100	120 MW	(E) energy value
50-sys	50	240 MW	
25-sys	25	480 MW	
4-sys	4	3 GW	
1-sys (E)	1	12 GW	
1-sys (CV)	1	12 GW	(CV) capacity value
Dispersion tests, best climate			
50 km	100	120 MW	Best climate
100 km			
200 km			
300 km			
Dispersion tests, worst climate			
50 km	100	120 MW	Worst climate
100 km			
200 km			
300 km			
Rooftop	2.4M	5 kW	Urban areas
No Solar	-	-	-

rightmost maps in Figure 3.1. We define the favorability of climate using a weighted average of three metrics: (1) frequency of sub-hourly cloud volatility, (2) frequency of hourly cloud volatility, and (3) expected annual capacity factor. These scenarios are meant to be illustrative but not realistic; we removed the ecological constraints applied in the utility-scale scenarios to achieve adequate dispersion. Section A.1 of the supporting information further explains these placements.

The **rooftop scenario** distributes 2.4 million 5 kilowatt PV systems throughout California based on the density of residential structures within each census tract.[4] Figure 3.2 displays locations for these rooftop systems on a coarse scale. A more

---

[4] The census does not directly report the number of structures in each census tract, instead

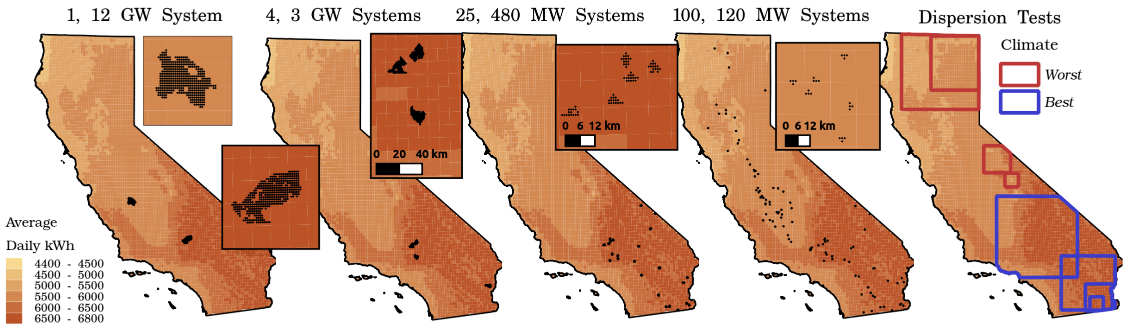

Figure 3.1: Utility-scale PV scenarios. Black dots represent 1km grid cells, each of which contain 30 MW of PV. The left panel contains scenarios where 12 GW are contained in one contiguous system, (E) located for the greatest energy value and (CV) located for the greatest capacity value. Locations are increasingly dispersed in each subsequent scenario, ending with 100, 120 MW. The rightmost panel contains regions used for the dispersion test scenarios.

detailed account of PV placement in the rooftop scenario is included in Section A.1 of the supporting information.

For the utility-scale scenarios, we also investigated two **balancing policies**: (1) all solar, wind, and load outputs are balanced by the system operator, and (2) wind and load are balanced by the system operator, but solar plants balance themselves by independently procuring load following and regulation. The later scenario reflects policies where solar PV plants "smooth" their own generation using co-located controllable resources.

Defining system balancing

Wind and solar generators have nearly zero marginal cost and can be viewed as a load modifier, meaning that system operators will dispatch controllable resources to balance **net-load** $= load - solar - wind$.[5] Controllable resources include generators, energy storage and communicating demand-side resources. Many controllable resources have long start-up times or cannot change output quickly, e.g., large coal, nuclear, and to some extent combined cycle gas generators. To accommodate these resources the system operator schedules them to change their output at most once an hour. System

---

this value is estimated using the number of housing units per tract and the percentage of units in multi-unit buildings see Section A.1 for more information

[5] System operators may curtail renewable generators when controllable generators reach minimum output levels or net load becomes negative. Although these circumstances are rare today, they could become more prevalent as more renewable generators are connected. As we will explain, this curtailment can be seen a flexibility product itself that is separate from net-load.

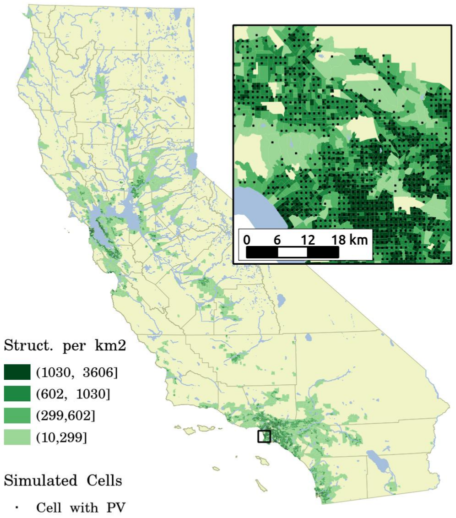

Figure 3.2: Map of the rooftop scenario of PV locations. Shades of green indicate the density of residential structures within census tracts, and black dots indicate 1 km^2 grid cells in which PV systems are located.

operators then hold in reserve enough flexible capacity to respond, with high reliability, to the variability and uncertainty of net-load around this schedule. Deviations between supply and demand manifest as variation in power system frequency. The fastest reserve capacity will respond to these changes in frequency and are thus referred to as frequency **regulation** reserves. To limit regulating capacity requirements, system operators may run an intermediate scheduling market that coordinates additional resources that are not providing frequency regulation but are available to respond faster than the hourly signal. This intermediate responding capacity is often referred to as **load following** reserves. Figure 3.3 shows both load following and regulation reserves as the difference between schedules: load following is the difference between hourly and 5-min schedules, and, regulation is the difference between actual net-load and the 5-min schedule.

Reserves compensate for uncertainty and variability of net-load. Uncertainty is the inability to perfectly predict net-load, resulting in erroneous schedules. Variability is the result of each subsequent schedule changing more quickly than its predecessor, even when both are based on perfect forecasts. Please see Makarov 2009 for a more complete explanation [47]. Our analysis rests on simulating four metrics of uncertainty or variability of solar, wind, and load:

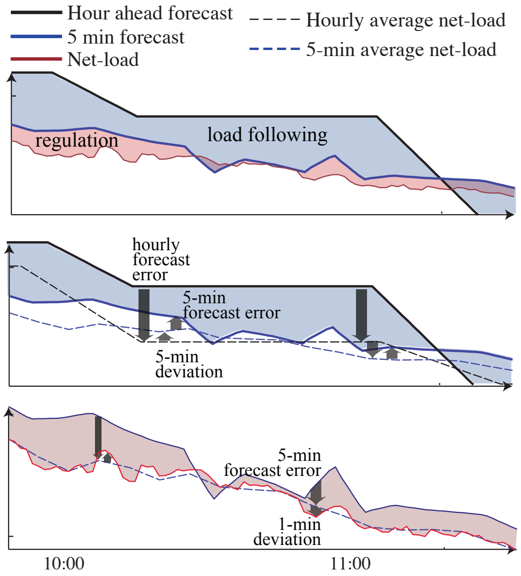

Figure 3.3: Quickly-responding controllable resources, set aside as reserves, balance net-load as it deviates from hourly scheduled generation. (Top) load following reserves follow a 5-min schedule as it deviates from the hourly schedule, and regulation reserves respond to grid frequency to balance real-time load as it deviates from the 5-min schedule. (Middle) load following reserves effectively compensate for hourly forecast errors, 5-min forecast errors, and 5-min deviations. (Bottom) regulation reserves compensate for 5-min forecast errors and 1-min deviations.

Hourly forecast errors, $\epsilon_{60,60,t}$, errors of an hour-ahead forecast of hourly average demand or generation.

5 minute deviations, $\eta_{5,60,t}$, the differences between a 5 minute and an hourly schedule that are based on perfect forecasts. This is constructed to contain only variability and not uncertainty.

5 minute forecast errors, $\epsilon_{5,5,t}$, errors in a 5 minute ahead forecast of 5 minute demand or generation. Forecasts assume persistence of a the clearness index.

1 minute deviations, $\eta_{1,5,t}$, deviation of 1-min generation/demand from a 5-min schedule based on a perfect forecast.[6]

The middle and bottom panels of Figure 3.3 show how reserves are composed of uncertainty and variability. Section A.2 of the supporting information shows mathematically what we show graphically in Figure 3.3: the use of load following reserves is the sum of $\epsilon_{60,60}(t)$, $\eta_{5,60}(t)$ and $\epsilon_{5,5}(t)$, and the use of regulation reserves is the sum of $\epsilon_{5,5}(t)$ and $\eta_{1,5}(t)$.

It is not intuitive that 5 minute forecast errors should affect both regulation and load following reserves. As a descriptive example of this double counting, consider what will happen if a 5-minute forecast predicts that demand will increase, but instead it remains constant: load following reserves will increase generation in order to meet the erroneous forecast, frequency will speed up as generation exceeds demand, and then regulation reserves will respond to frequency by decreasing generation. System operators must account for these forecast errors when planning both the load following market and frequency regulation.

Our method is general and can be applied to any set of timescales. For this study, we chose the time-scales previously used by the California system operator to run energy and reserve markets, as defined in [47]. In response recent regulations,[7] CAISO and many other utilities in the Western United States now reschedule generation on a 15 minute basis rather than hourly. Forecasts for these new markets are made 40 minutes in advance of the interval, which is similar to the hour-ahead forecasts that we use. In contrast to our analysis, 15 minute schedules ramp continuously over each interval rather than plateauing in the middle; thus new schedules will better match predictable changes of net-load over time, such as the effect of sunrise and sunset on solar generation. In Section 3.3, we remove the effect of sunrise and sunset on reserves to better model the effect of new scheduling.

Simulating PV

We modified our model from [59] to simulate variability and uncertainty of PV generation in each locational scenario. In this section we provide a brief explanation of the model framework, including updates. Appendix A.3 of the supporting information contains a detailed description.

Figure 3.4 (left) shows a graphical depiction of dependencies in our model, using variables that we explain in this section. This structure is applied to generate models

---

[6] Note that our data do not support investigations of deviations on time scales faster than one minute, but if such data were available this metric could constructed on the relevant time scale – e.g. if 4 second data were available one would compute 4 second deviations from the 5 minute schedule.

[7] Order 746 of the Federal Energy Regulatory Commission.

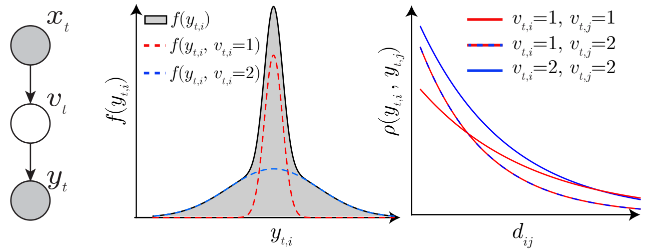

Figure 3.4: **Left,** Graphical model representing dependence between variables. **Center,** a pdf of variability or uncertainty at one PV system, $y_{t,i}$, represented as a mixture of Gaussians. Volatility states, $v_{i,t}$, define which mixture component is exhibited. **Right.** Correlation of y between two systems, $\rho(y_{t,i}, y_{t,j})$. Correlation decays exponentially with distance, the parameters of this exponential decay are dependent on the volatility states of both systems.

for each of the four metrics variability or uncertainty defined in Section 3.2. To explain the structure, we generally refer to these metrics as $y_{i,t}$; subscripts signify the metric from PV system i at time t.

Our primary innovation is that we condition the standard deviation of $y_{i,t}$ on an endogenously estimated latent state, $v_{i,t}$, referred to as a volatility state – defined and fully explained in Section **??**. The volatility state for each PV system allows for rapid transitions between periods of high standard deviation and low standard deviation. Figure 3.4 (center) shows how the distribution of $y_{i,t}$ is a mixture of Gaussian distributions. The mean of $y_{i,t}$ is 0 for forecast errors, signifying that forecasts are unbiased; for deviations the mean is defined as the deviations of a deterministic clear sky signal.

$y_{i,t}$ from neighboring systems are spatially correlation, with correlation coefficients dependent on volatility states of both systems. Figure 3.4 (right) shows multiple autocorrelation functions representing the pearson correlation coefficient between $y_{i,t}$ from two systems as a function of the distance separating those systems, $d_{i,j}$. One function is estimated for each possible pairing of volatility states.

The distribution of $v_{i,t}$ depends on a set of climate variables, x_t, which can be any set of discrete observations from the geographic areas and time periods modeled. In this study, x_t represent cloud variability in 1-km^2 grid cells, and are derived from measurements of global horizontal irradiance provided to us by Clean Power Research.[8]

---

[8] Full derivations of the climate data are included in Sections A.4 and A.4 of the supporting

Table 3.2: Summary of parameters for modeling variability and uncertainty from PV.

Variability / Uncertainty metric	Standard deviation (% of clear sky)		Correlation decay rate (km)	
	range	weighted mean	range	weighted mean
Hourly forecast errors, $\epsilon_{60,60}$	[6.3, 21.9]	11.8	[46.2, 62.5]	53.34
5-min forecast errors, $\epsilon_{5,5}$	[0.08, 22.1]	1.6	[0.10, 60.7]	2.48
5-min deviations, $\eta_{5,60}$	[0.37, 16.4]	3.2	[0.10, 20.2]	9.6
1-min deviations, $\eta_{1,5}$	[0.18, 16.1]	2.0	[0.10, 2.21]	0.87

Model parameters are fitted using observations of both y_t and x_t; values of v_t are latent, meaning that they are never observed but are estimated during the fitting process. For the simulations in this study, only observations of x_t, the locations of PV systems, and the parameters are required; distributions of y_t and v_t are produced by the simulation; explained in Section A.3.

Table 3.2 presents ranges and time weighted averages for standard deviations and correlation decay rates for each metric. Noticeably, hourly forecast errors exhibit relatively large standard deviations and long correlation decay rates, indicating that they will have the greatest effect on reserve needs.

Utlity-scale PV plants in our scenarios range from 120 MW to 12 GW. However we fit our models using systems that only range from 2 kW to 12 kW. We expect that there will be significant effects of geographic diversity within the utility-scale systems beyond what we observed in the distributed systems. We account for this intra-plant geographic diversity by modeling utility-scale PV systems as collections of 5kW systems arranged in grids that are spaced such that the ground cover ratio is 0.15 at a panel efficiency of 0.20.[9] In all cases, the utility-scale systems also span more than 1km ^{2}, and thus are affected by multiple observations of climate at each time. Thus our simulations account for geographic smoothing and varying climate within each utility-scale plant.

Reducing the computational costs of the simulation was necessary for this study. Each of our locational scenarios requires simulating thousands of $M \times M$ covariance matrices, where $M = 2.4$ million 5 kW PV systems. This simulation is theoretically possible however is computationally infeasible on common computers. To reduce computational costs, we make two approximations, which allow us to run simulations

---

information.

[9] This arrangement results in 30MW of PV capacity per km^2, which is similar to land use efficiencies found in a recent survey of utility-scale solar installations[62].

in two, tractable stages. Our approximations and our two-stage method are fully explained in Section A.3 of the supporting information.

Predicting hourly forecast errors with our model is a new innovation in this paper that we fully explain in Section A.4 of the supporting information. In brief, hourly forecasts are generated using a linear model conditioned on lagged values of solar generation, the expected solar irradiance given a clear sky, and time of day and seasonal fixed effects. From our model, we simulate a root mean square error (RMSE) of between 9.6 and 11.4 MW for the 120 MW plants in our scenario with 100 utility-scale systems. This is consistent with or better than commercial forecasts evaluated in [63]; this is unsurprising for hour-ahead predictions, for which persistence forecasts are difficult to out-perform [64].

The contribution of PV to reserves is calculated by adding forecast errors to deviations, again resulting in Gaussian mixture distributions. We assume no correlation between forecast errors and deviations given the volatility states, however we do assume correlation of the volatility states. Thus forecast errors and deviations are more likely to be volatile at the same time than if they were independent, but are not more likely to fluctuate in the same direction. Section A.3 of the supporting information explains this method for summing the forecast errors and deviations, and validates our simulation method using data from model fitting. Overall, our method accurately predicts the distributions of load following reserve use, and it provides a conservative estimate of the distribution tails of regulation reserve use.

Variability and Uncertainty from Wind and Load

The CAISO provides one-minute resolution wind generation and load data for the year 2005. We obtained these data and then linearly scaled them to represent growth between 2005 and 2020, when California's independently owned utilities are required to procure 33% of energy from renewable sources. For load, we used an annualized growth rate of 1.5%, resulting in a scaling factor of 1.25 over 15 years. For wind, we scaled it to match current projections for wind energy in California as presented in a recent long term planning paper, [12]. These scaled data are used to create one year of variability metrics – $\eta_{5,60}$, and $\eta_{1,5}$ – for wind and load in the year 2020. We simulate forecast errors – $\epsilon_{60,60}$, and $\epsilon_{5,5}$ – for both wind and load using an autoregressive processes defined in [47] and used in CAISO's recent renewables projections.[10]

We adapt a method put forward by Wu et al. in [50] to simulate reserve needs from wind and load during our study period. The method bins all observations of reserve need using hourly observations. Wind and load's contributions to reserves are

---

[10] We describe this method in more detail in [49].

kept separate and binned based on different data. We bin wind generation by the decile of hourly average wind generation, implying that wind may be more or less variable depending on the average generation. We bin load's contribution to reserve use using only hour of day and month of year, capturing diurnal and seasonal changes in load variability.

For each bin, we create a mean profile within the hour; this effectively captures steady increases or decreases in load during the morning or evening. We then store all observed deviations from this mean profile. A sample of wind or load's contributions then includes the mean profile plus randomly sampled deviations. I.e. the load following reserve need from load at 10:35 am on Jan 2 is simulated as the average load at 10:35 in January, plus a random selection of any of the deviations observed between 10:00 and 10:55 in January.

We seek to identify the nuanced effects of PV system siting in the context of a more general larger system, thus we are less careful when modeling of wind and load than we are when modeling solar. In interest of clarity, we list here some limitations of our wind and load models:

- Additional wind generations will be in new locations, increasing geographic diversity and likely smoothing wind's variability and uncertainty. Thus our model will over-predict the variability and uncertainty from wind.
- Load growth results from a growing and changing stock of appliances and consumers, which will consume energy at different times than the current stock. I.e. ownership of air conditioners, electric vehicles, and electric heaters are projected to increase over the next decades. Thus peak load and common ramps in load may occur at different times in the future than they do today; as our linear scaling assumes.
- Finally, coincident patterns in the solar, wind and load data are reflected in our analysis only insofar as they are contained in the input data: binning data for wind and load, and the climate data for solar. Beyond this, variability and uncertainty from solar, wind, and load are simulated independently.

Procuring reserves

Figure 3.5 shows the process of simulating reserve procurement from the solar, wind, and load simulations. We use the models described above to create probability distribution shapes for wind, load, and PV's contributions to reserves. From these distributions, we sample 200 realizations at each appropriate interval: 1-min for regulation and 5-min for load following. We calculate the reserves needed to follow net-load by subtracting solar and wind's contributions from load's contribution. This creates 200 equally likely values for reserve need at each interval in our study year. Thus for each hour of the year we have 2,400 equally likely values for load following reserves, and 12,000 equally likely values for regulations reserves.

We procure reserves such that they are sufficient in 95% of hours. This is achieved by defining reserve "up" and "down" procurement as the 97.5^{th} and 2.5^{th} percentiles of reserve needs realized in that hour.

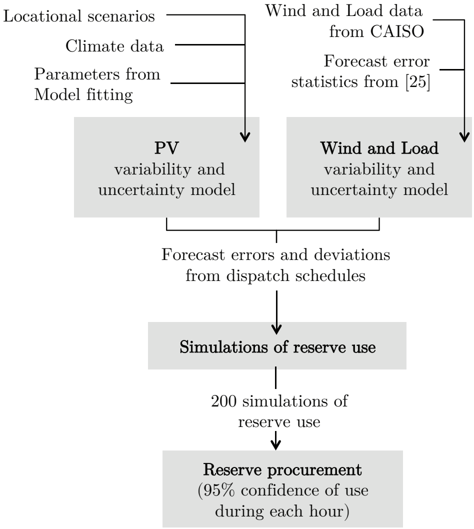

Figure 3.5: Process diagram for the construction of load following and regulation reserve signals

We condition the distributions of wind and solar reserves on information that will only be available during the period reserves are used, and not during the earlier time at which reserves are procured. A system operator will rely on less certain forecasts, and will be less able to determine when more or less reserves are needed. As a result, we expect our estimates for hourly reserve procurement to be more variable than those of an actual system operator. However, conditioning variables are coarse and potentially easy to predict: there are only 10 possible deciles of hourly average wind generation, and only 5 possible climate states for solar. Thus we do not expect this assumption to greatly affect our results.

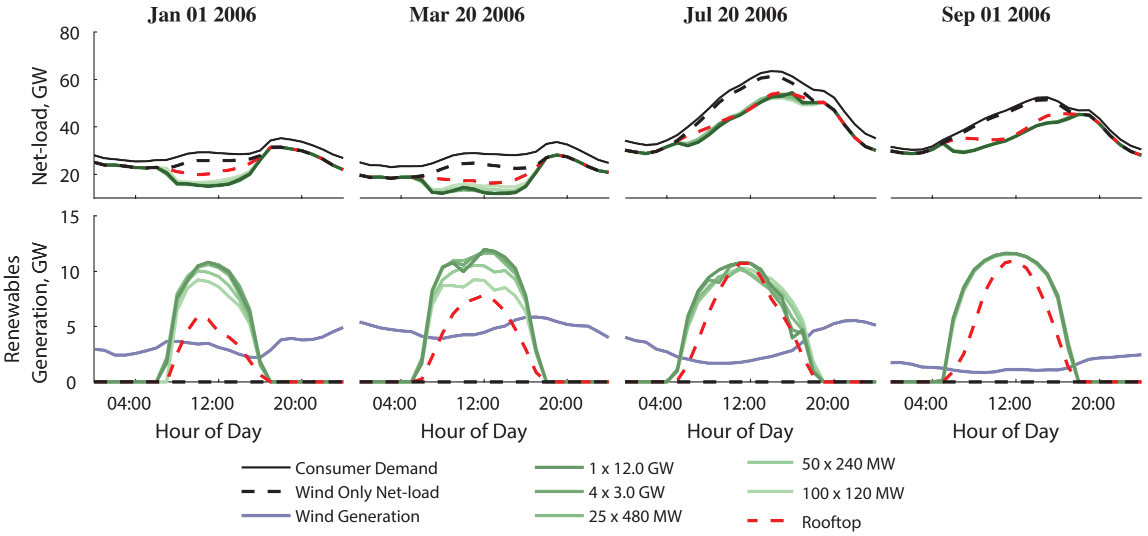

Figure 3.6: A few daily net-load profiles under different scenarios.

3.3 Results

Net-load

Figure 3.6 shows four example days of net-load and renewable generation from the utility-scale and rooftop scenarios. The most centralized scenario shows large amounts of variability on the example days in January and October. The rooftop scenario generally produces less overall generation than the centralized scenarios, because rooftop systems don't track the sun, have less ideal geometries, and are not sited in areas with the greatest energy resource.

Notably, morning and afternoon ramps in PV generation—those ramps resulting from sunrise and sunset—have opposite effects on generation ramping requirements. On some days, the morning rise in PV generation mitigates a simultaneous rise in demand reducing the amount other generation needs to ramp, while the afternoon drop in PV generation exacerbates the power ramp already needed to meet an afternoon peak in demand—this is most pronounced in colder months (January and April). This effect—known as the duck curve—is well known to system planners.

Hourly reserve procurement

Figure 3.7 shows the minimum, mean, and maximum of hourly reserve procurement for each scenario. Hourly reserve procurements depend strongly on spatial configuration:

the load following reserve increase in the worst utility-scale scenario (single 12 GW system sited to maximize capacity value) is 4 times larger than in the best utility-scale scenario. Though the maximum percentage increase in frequency regulation reserve is smaller than for load following, the range of percentage increases is more pronounced because the strongly dispersed cases result in essentially no increase in reserve requirements.

Reserve requirements are greater when we site a single 12 GW system to maximize capacity value rather than energy production (see the “CV” case versus the “E” case). This is because the CV site has a higher total fraction of partly cloudy hours. We will show later, however, that though on average need for reserves is greater, at peak hours the need for reserves at the CV site is lower than at the E sites.

As one might expect, the dispersion tests show that the size of the area over which systems are distributed clearly matters. However reserve reductions diminish as the size increases. For squares with side lengths larger than 100 km, reserve reductions for regulation are negligible. Constraining systems to be in areas with different climates has little effect on reserve requirements for highly dispersed systems, but a large effect on concentrated systems.

The right-most scenarios in Figure 3.7 show that requiring individual plants to balance their own output locally—for example with energy storage—significantly increases reserve capacity requirements. Balancing locally removes two benefits of aggregation: First, solar is balanced separately from wind and load, and this is the driver for the increase in reserve requirements in the single system cases. Second, the benefits of decaying geographic autocorrelation are lost, which explains why the reserve requirement increase is greatest for the 100 system cases. Taken together these results indicate that that local balancing is likely to be very expensive if balancing costs scale linearly with required capacity.

Sensitivity analyses

We ran four sensitivity analyses on the utility-scale scenarios to better qualify our results. Black dots in Figure 3.8 show our original estimates for mean and maximum procurement of load following reserves for the baseline and utility-scale scenarios; these match those values presented in Figure 3.7.

One major driver for load following reserves is predictable variability driven by the solar diurnal cycle. PV generation ramps steadily throughout morning and evening while generation schedules ramp only for the 20 minutes between hours, then plateau for 40 minutes the middle. In our model, the solar diurnal cycle greatly affects the mean of the 5-min deviation distributions, $\eta_{5,60}^{solar}$, which is based on a clear sky signal. Removing this predicable variation models more dynamic generation schedules, such

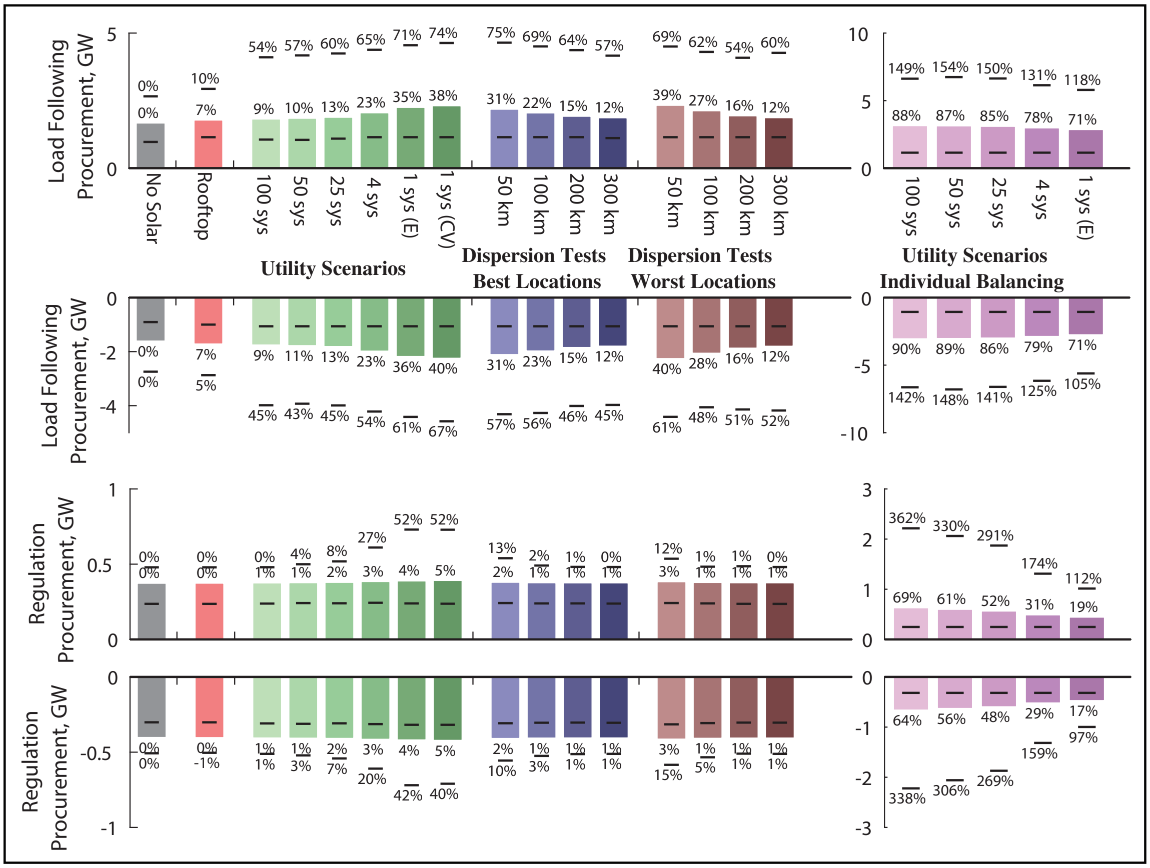

Figure 3.7: Average hourly reserve procurement in each scenario. Lines above and below each bar represent the maximum and minimum reserve procurement. Labels above each bar are the percent change in average procurement from the baseline scenario; labels above each maximum line are the percent change in the maximum procurement of reserves.

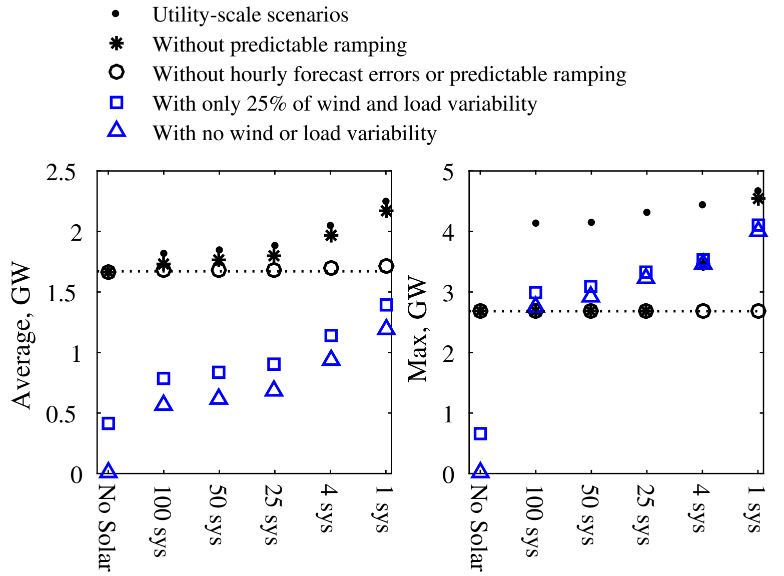

Figure 3.8: Average and maximum hourly load following procurement for the baseline and utility-scale scenarios, original results are presented along with four sensitivity cases.

as CAISO's updated system which asks generators to ramp continuously within 15-minute intervals, rather to plateau within each hour.

Black stars in Figure 3.8 present results from a scenario when these predictable ramps are removed. Average procurement reduced by about 80 MW in each scenario (≈5% of baseline). Maximum procurement reduced to that in the baseline case for all except the two most centralized scenarios[11], signifying that the large increases in maximum procurement were driven entirely by the steep morning and evening ramps of utility-scale solar.

Black circles in Figure 3.8 display results from our analysis when forecast errors are also set to zero. After removing the effects of predictable PV generation ramps, we find that forecast errors account for between 90% and 95% of the reserve increases from solar in all scenarios.

Figure 3.8 also contains two scenarios that test the effect of variability from wind and load: one where we include only 25% of this variability (blue squares) and one where we include none (blue triangles). These scenarios represent systems with much lower loads but similar geographies, or systems with a much higher penetration of

---

[11] This maximum occurs at 6pm in the winter, during the evening peak in consumption depicted in Figure 3.6

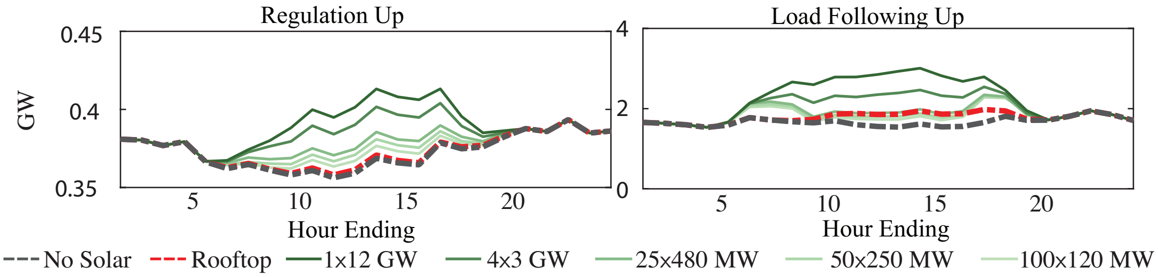

Figure 3.9: Average regulation and load following reserves procured in each hour of day for the utility-scale, rooftop, and baseline scenarios.

solar. We find that the inclusion of solar in general has a greater relative impact in these cases, particularly for maximum procurement—compare the baseline case to all other cases. However the general findings regarding spatial diversity remain: there are diminishing returns to dispersion, particularly after the 25 system scenario.

Diurnal variations in reserve need

The solar diurnal cycle drives not only the generation but also magnitude of variability and uncertainty in PV. Averaging the effect of PV on reserve need throughout the year belies the consistently larger effect PV will have at midday, compared to the non-existent effect at night.

Figure 3.9 shows the average procurement of load following and regulation reserves stratified by hour of day. As expected, reserve increases in the solar scenarios are greatest at midday, although these increases are rapidly mitigated by geographic diversity. There is a consistent increase in load following reserve requirements in the afternoon that is not mitigated by geographic diversity. This is due to the correlation of sunset with increased demand, and the effect of solar ramping on load following reserve use – explained in the previous section.

Curtailing PV to provide reserves

Renewable generators may provide reserves by dynamically curtailing generation, as suggested in [65], [66] and currently practiced by the Public Service Company of Colorado [67]. In this section we estimate the amount of PV energy that must be curtailed to provide the reserve procurement that exceeds that in the baseline scenario. While there will often be less expensive reserve capacity than curtailing

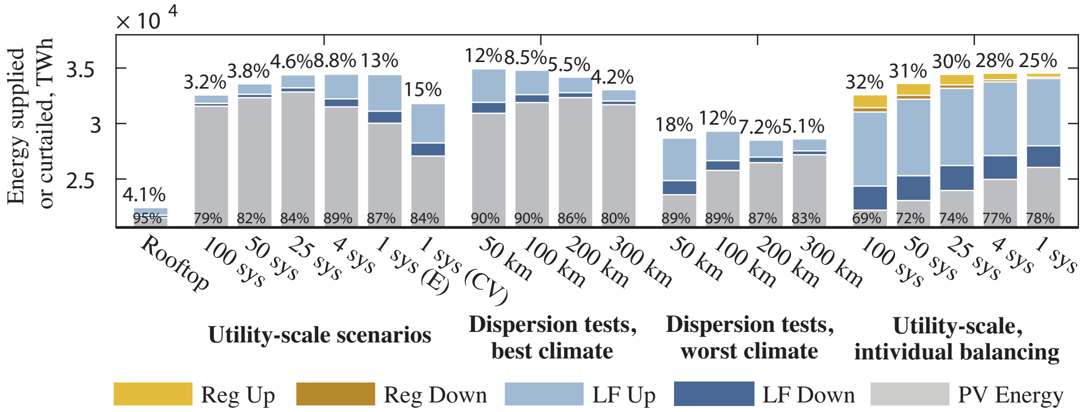

Figure 3.10: Total energy produced in each scenario, and energy curtailed to provide additional reserves resulting from the inclusion of solar. The percentage of PV energy curtailed in order to provide reserves is labeled *above* each bar; The percentage of additional reserves able to be supplied by curtailing PV is labeled *below* each bar.

solar, this estimate provides a conservative boundary for the costs of variability and uncertainty that PV adds to supply.

To provide *up reserves*, a PV system must pre-curtail power output by the amount of up reserves contracted and reduce this curtailment to reliably increase generation when needed. To provide down reserves, a PV system could simply curtail power output to reduce its own generation when needed. As in [66], we conservatively assume that 25% of the power committed to up reserves is used on average.[12] Thus, PV will curtail only 75% of the MW-h committed to up reserves in a given hour. Analogously, we assume that 25% of down reserves committed will be curtailed on average.[13] PV systems can only commit as much capacity as they reasonable expect to produce consistently throughout an hour. We do not allow PV to commit more than 70% of its hourly average production to reserves; as a result, PV is insufficient to provide additional reserves in some hours.

---

[12] Assuming that up and down reserves are mutually exclusive and equally likely, each have a 50% likelihood of being called. Given that a set of reserves are called, we conservatively assume that 50% of committed capacity is used on average. The result is $.5 \times .5 = 25\%$ of committed capacity being used on average.

[13] If up and down reserve procurement are symmetric, then total curtailment will always be equal to total amount of reserves procured in one direction. This is true regardless of the percentage of committed reserves employed on average. I.e., if $x\%$ of committed up and down reserves are used on average, then up reserves curtail (100 - x)% of committed power on average, and down reserves will curtail x%; thus the total reserve procurement adds to 100% of up or down procurement.

Figure 3.10 shows the total PV energy supplied annually in each scenario, and highlights the amount of this energy that would be curtailed to provide all reserves in excess of the baseline scenario. The total height of each stacked bar is the annual PV energy that could be supplied in each scenario. Colored bars at the top of each stack represent PV energy that is curtailed to provide each type of reserves. Thus, the bottom gray bar represents PV energy supplied after curtailment. Percentages above each stack signify the percent of PV energy curtailed, while percentages below each stack are the amount of hours that enough PV energy is available to provide the additional reserves. Typically, PV is insufficient to meet additional reserves in the early morning or afternoon hours, when both PV generation and the additional reserve need are low.[14]

More concentrated arrangements are able to supply more energy, but they also increase the need for reserves and thus increase curtailment. This trade-off is most apparent for the dispersion tests in the best climates: though potential PV generation is greatest when concentrating systems in a 50km square area, 12% of this energy is curtailed to provide reserves. Conversely, in the most dispersed 300km scenario there is less total potential for PV generation, but also less is curtailed for reserves, resulting in more net generation being added to the system. The 200km scenario provides the optimal net contribution of energy among the dispersion tests in the best climates.

Capacity value of PV

The capacity value of a resource reflects its ability to reduce the peak need for controllable resources on the system, e.g., by producing energy or reducing demand during the hours when the need for controllable resources is greatest. We define the capacity value of solar in each of our scenarios as the reduction in this peak as compared to the baseline scenario. We define the need for controllable resources as the sum of hourly net-load, load-following up procurement, and regulation up procurement; down reserves are excluded because controllable resources are conventionally generators, and additional generation capacity is not needed in order to provide down reserves. Solar generation may affect peak need for controllable resources in counteracting ways: first it may reduce net-load during peak hours, and second it may increases the need for reserves during peak hours.

Figure 3.11 displays peak need for controllable resources in each scenario. Percentages above each bar represent the reduction from the baseline, no-solar, scenario.

---

[14] During night-time hours PV doesn't generate energy but, as a result, doesn't contribute to reserves need. Thus there is no increase in reserve need during these hours.

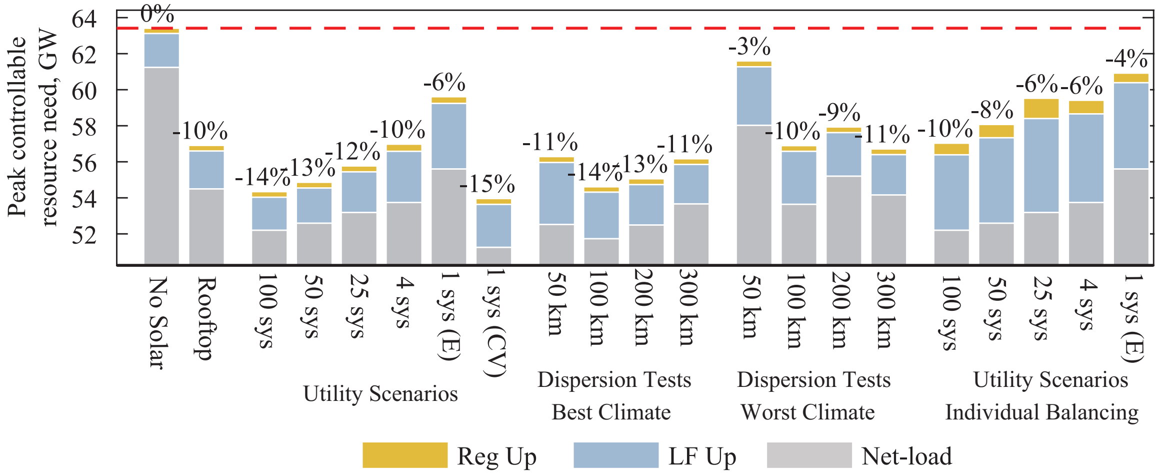

Figure 3.11: Maximum hourly need for dispatchable generation during the 2020 study year. Need for generation is broken by net-load (grey), load following up need (blue), and regulation up need (red).

Surprisingly, the scenarios in which PV is sited to achieve the greatest annual energy production have relatively small capacity values. This is most stark by comparing the dispersion tests in the "best" climates to those in the "worst". While capacity value increases with dispersion in the utility-scale scenarios, dispersion is certainly not necessary: by siting one system intentionally for capacity value, it achieves a higher capacity value than the 100 dispersed systems.

The low capacity value for systems with high capacity factors is due to the weather patterns in California during our study year, where locations that on average were least cloudy—particularly those in the Mojave desert—happened to be cloudy on summer afternoons when peak demand is greatest. While our study can only show this effect for the 2005 study year, solar resource data from the National Renewable Energy Laboratory (NREL) suggest that it is a general trend. The NREL data contain annual and monthly average PV resource, which is derived from satellite observations collected from 1998 to 2009. Figure 3.12 presents the annual averages along with averages for three summer months. On an annual basis, the Mojave Desert contains a far greater solar resource. But in the months of July and August—which most often contain peak demand—the Central California valley and even some of the mountainous Northern California regions have a greater resource. Some areas in the Northern Mojave offer optimal mixes: large resources over the year and during summer months. Together, these data suggest a limited trade off between maximizing total energy, and energy during peak hours, where a few regions are optimal for both.

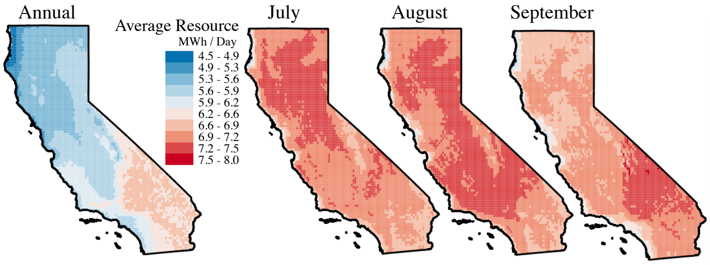

Figure 3.12: Average daily PV generation by region and month.

Our capacity value results are highly dependent on our assumptions of how much renewable generation is included in the system and the timing of peak demand (we assume that load shapes are linearly scaled from 2005). California's future renewable energy goals are far more ambitious than their 33% goals modeled here, thus we expect there to be more wind and solar included eventually. The timing of peak demand may also change in the future, particularly if end-uses change significantly, for example due to heating or transportation electrification.

3.4 Discussion

The need for reserves is greatly reduced by dispersing PV systems over large areas. However, almost all of the benefits of this dispersion can be achieved with relatively large, roughly 500 MW, systems. This is almost precisely the size of the largest utility-scale PV systems in California today, implying that current utility-scale PV systems do not generate significantly additional need for reserves than their more dispersed counterparts. Climate also plays a role in the variability and uncertainty of PV generation, although the effect of climate is greater for highly centralized systems than for dispersed ones.

This study's results should be considered in policy discussions over where to site PV, in particular where to locate competitive renewable energy zones (CREZ). Limiting solar to only one or two CREZ could greatly increase the need for system reserves. However the number of zones need not be excessively large, as most of the benefits of geographic dispersion are achieved with even 25 dispersed, large PV plants. Currently utility-scale PV systems are clustered into about 3 regions; and transmission planning studies simulate them in only one region: Tehachipi [53].

When dispersed, PV contributes a small amount to load following reserves, and a *negligible amount to regulation reserves.* When PV is contained in plants smaller than 480 MW average load following procurement increased by less than 1.1% of installed solar capacity (12 GW), and regulation reserves increased by less than 0.05%.

Errors in hourly forecasts are the greatest contributor to load following reserves, particularly in the centralized scenarios. Our study uses a simple linear model to forecast average solar generation that relies on only historical generation, system location, and system geometry. More informed models may reduce these errors. However, we note that the performance of our simple model is similar to those found in literature, and the performance of models informed by numerical weather predictions or satellite imagery only marginally exceed persistence forecasts at intervals of one hour-ahead [64], [68].

PV forecasting algorithms could reduce the need for reserves by reducing autocorrelation in forecast errors, if not by reducing the forecast errors themselves. I.e., even if the predictions of generation at individual PV plants are poor, low autocorrelation between plants will increase the statistical smoothing resulting from summing their many forecasts. The obvious way to reduce this autocorrelation is to spatially disperse PV plants, making them less likely to be affected by the same weather patterns. However we imagine some prediction methods may be better at reducing autocorrelation than others, such as the cloud tracking methods described in [69]. An incoming cloud could be predicted with high certainty to cover some fraction of PV plants, though which subset is uncertain. The forecast for all plants may be poor, but the errors in these predictions will be anti-correlated: the subset of panels not covered by the cloud will have been under-predicted, and the subset covered by the cloud will have been over-predicted.

The second largest contributors to load following reserves were predictable, diurnal ramps in PV generation. I.e. the effects of sunrise and sunset. Load following reserves are expected to follow these ramps, which continuously increase throughout hours while controllable generation remains at a constant plateau. This effect accounts for roughly 50% of load following reserve increases in the most dispersed scenarios. Load shifting could greatly reduce this need by altering ramps in load to counteract ramps in solar and wind. However, system operators—in partial response to FERC order 764—have already taken strides to reduce this effect by scheduling generation to ramp continuously within 15 minute intervals rather than plateau within 60 minute intervals.

PV systems can provide reserves by curtailing a small amount of generation in a controlled manner, as noted in recent work [65], [66]. We find that in the most distributed scenarios, PV systems can provide 97% of the additional reserves resulting from their inclusion while curtailing less than 5% of the energy they provide. For spatially concentrated scenarios, only 88% of additional reserves could be provided by PV, and it would require 15% curtailment to do so. We reiterate here that, in most hours, if will likely be uneconomical for PV to provide reserves because cheaper sources will often be available.

Requiring PV systems to smooth their own output before it is connected to the system results in massive increases in the need for reserves, 50%*, and diminishes the capacity value of solar by as much as 33%.* Operators should be additionally concerned because these results are only for utility-scale scenarios; if rooftop systems or micro-grids were to balance themselves these number would be far worse. Conversely, if only one small system or micro-grid balances itself it will have only a small effect on the balancing requirements of the system. Thus policies to promote the use of storage or distributed generation to balance PV should consider their objective, responding to a reserve signal from the grid will provide much more value than smoothing on site generation.

As modeled here, utility-scale scenarios have the added benefit of optimal system geometry and being located in areas with greater solar resource. However, what isn't modeled is the effect of co-locating solar near consumer demand on distribution system needs. Predicting this effect is an active area of research. Cohen and Callaway [70] find that PV can have net costs or benefits to distribution systems depending on the system topology, the climate, and the timing of loads. Future work for these models and methods is to apply them to much smaller spatial scales and assess the implications of PV spatial arrangements on distribution systems.

3.5 Acknowledgements

This work was supported in part by the NSF Graduate Research Fellowship, the California Public Utilities Commission, and NSF grant CNS-1239567. The authors would like to thank E. Carlson and K. Varadarajan of SolarCity, and J. Dise of Clean Power Research, who provided data required for this work.

3.6 Chapter References

[4] M. Lave, J. Kleissl, and J. Stein, "A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501–509, 2013, ISSN: 1949-3029. DOI: `10.1109/TSTE.2012.2205716`.

[5] A. Mills and R. Wiser, "Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power," Lawrence Berkeley National Labratory, Tech. Rep., 2010.

[6] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. Hoff, "Parameterization of site-specific short-term irradiance variability," *Solar Energy*, vol. 85, no. 7, pp. 1343–1353, Jul. 2011, ISSN: 0038-092X. DOI: `10.1016/j.solener.2011.03.016`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X11000995` (visited on 09/12/2013).

[7] A. Murata, H. Yamaguchi, and K. Otani, "A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area," *Electrical Engineering in Japan*, vol. 166, no. 4, pp. 9–19, Mar. 2009, ISSN: 04247760. DOI: `10.1002/eej.20723`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/eej.20723/abstract` (visited on 02/09/2011).

[10] G.E. Energy, "Western Wind and Solar Integration Study," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/SR-550-47434, May 2010.

[11] D. Lew, G. Brinkman, E. Ibanez, B. M. Hodge, and J. King, "Western wind and solar integration study phase 2," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.7781&rep=rep1&type=pdf` (visited on 04/22/2015).

[12] Energy and Environmental Economics (E3), "Investigating a Higher Renewables Portfolio Standard in California," Tech. Rep., Jan. 2014. [Online]. Available: `https://www.ethree.com/documents/E3_Final_RPS_Report_2014_01_06_with_appendices.pdf`.

[32] M. Rothleder, *Track I Direct Testimony Of Mark Rothleder On Behalf Of The California Independent System Operator Corporation*, California Public Utilities Commission, 2011.

[33] M. Hummon, P. Denholm, J. Jorgenson, D. Palchak, B. Kirby, and O. Ma, "Fundamental Drivers of the Cost and Price of Operating Reserves," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep., 2013. [Online]. Available: `http://www.nrel.gov/docs/fy13osti/58491.pdf` (visited on 02/28/2014).

[34] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://www.smartgridinformation.info/pdf/5506_doc_1.pdf` (visited on 04/22/2015).

[36] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Gomez Lazaro, and N. Detlefsen, "Methodologies to determine operating reserves due to increased wind power," in *2013 IEEE Power and Energy Society General Meeting (PES)*, Jul. 2013, pp. 1–10. DOI: `10.1109/PESMG.2013.6673067`.

[38] M. Lave and J. Kleissl, "Cloud speed impact on solar variability scaling Application to the wavelet variability model," *Solar Energy*, vol. 91, pp. 11–21, May 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2013.01.023`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X13000406` (visited on 09/12/2013).

[39] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. E. Hoff, "Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance," *Solar Energy*, Progress in Solar Energy 3, vol. 86, no. 8, pp. 2170–2176, Aug. 2012, ISSN: 0038-092X. DOI: `10.1016/j.solener.2012.02.027`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X12000928` (visited on 05/14/2014).

[41] J. Wegener, M. Lave, J. Luoma, and J. Kleissl, "Temporal downscaling of irradiance data via Hidden Markov Models on Wavelet coefficients: Application to California Solar Initiative data," 2012. [Online]. Available: `http://solar.ucsd.edu/datasharing/doc/UCSDReport{\_1}secCSI.pdf` (visited on 12/06/2012).

[47] Y. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational Impacts of Wind Generation on California Power Systems," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039–1050, May 2009, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2009.2016364`. [Online]. Available: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4808228` (visited on 06/07/2011).

[49] C. Goebel and D. Callaway, "Using ICT-Controlled Plug-in Electric Vehicles to Supply Grid Regulation in California at Different Renewable Integration Levels," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 729–740, Jun. 2013, ISSN: 1949-3053. DOI: `10.1109/TSG.2012.2218629`.

[50] J. Wu, A. Botterud, A. Mills, Z. Zhou, B.-M. Hodge, and M. Heaney, "Integrating solar PV (photovoltaics) in utility system operations: Analytical framework and Arizona case study," *Energy*, vol. 85, pp. 1–9, Jun. 2015, ISSN: 0360-5442. DOI: `10.1016/j.energy.2015.02.043`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S036054421500198X` (visited on 12/19/2015).

[51] D. Corbus, J. King, T. Mousseau, *et al.*, "Eastern wind integration and transmission study," *NREL (http://www. nrel. gov/docs/fy09osti/46505. pd f), CP-550-46505*, 2010. [Online]. Available: `http://www.nyiso.com/public/webdocs/markets_operations/committees/bic_miwg/meeting_materials/2010-02-17/EWITSNYISO.pdf` (visited on 04/22/2015).

[52] X. Bai, K. Clark, G. A. Jordan, N. W. Miller, and R. J. Piwko, "Intermittency Analysis Project: Appendix B Impact of Intermittent Generation on Operation of California Power Grid," California Energy Commission, Tech. Rep., 2007.

[53] California Public Utilities Commission, "Planning Assumptions & Scenarios For The 2016 Long Term Procurement Plan Proceeding and The CAISO 201617 Transmission Planning Process," California Public Utilities Commission, Tech. Rep. R.13-12-010 MP6/avs, May 2016.

[54] L. M. Hinkelman, "Differences between along-wind and cross-wind solar irradiance variability on small spatial scales," *Solar Energy*, vol. 88, pp. 192–203, Feb. 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2012.11.011`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X12004021` (visited on 05/14/2014).

[55] C. M. S. Martin, J. K. Lundquist, and M. A. Handschy, "Variability of interconnected wind plants: Correlation length and its dependence on variability time scale," en, *Environmental Research Letters*, vol. 10, no. 4, p. 044 004, Apr. 2015, ISSN: 1748-9326. DOI: `10.1088/1748-9326/10/4/044004`. [Online]. Available: `http://iopscience.iop.org/1748-9326/10/4/044004` (visited on 04/08/2015).

[56] M. Lave, J. Stein, J. Kleissl, A. Ellis, C. Hansen, and Y. Miyamoto, *Simulating Solar Power Plant Variability for Grid Studies: A Wavelet-based Variability Model (WVM).* Maui, HI, Oct. 2011.

[57] M. D. Tabone and D. S. Callaway, "Parameterizing Fluctuations in Solar Photovoltaic Generation Using Hidden Markov Models," in *Proceedings of the 2013 IEEE Power & Energy Society General Meeting*, Vancouver, BC, Jul. 2013.

[58] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "Power output fluctuations in large scale pv plants: One year observations with one second resolution and a derived analytic model," en, *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 2, pp. 218–227, Mar. 2011, ISSN: 1099-159X. DOI: `10.1002/pip.1016`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.1016/abstract` (visited on 02/17/2016).

[59] M. Tabone and D. Callaway, "Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2965–2973, Nov. 2015, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2372751`.

[60] M. Milligan, B. Kirby, J. King, and S. Beuning, "Operating reserve implication of alternative implementations of an energy imbalance service on wind integration in the Western Interconnection," in *2011 IEEE Power and Energy Society General Meeting*, Jul. 2011, pp. 1–8. DOI: `10.1109/PES.2011.6039622`.

[61] G. C. Wu, "Incorporating land-use requirements and environmental constraints in low-carbon electricity planning for California," *Environmental science & technology*, 2014. [Online]. Available: `http://pubs.acs.org/doi/abs/10.1021/es502979v` (visited on 01/29/2015).

[62] R. R. Hernandez, M. K. Hoffacker, and C. B. Field, "Land-Use Efficiency of Big Solar," *Environmental Science & Technology*, vol. 48, no. 2, pp. 1315–1323, Jan. 2014, ISSN: 0013-936X. DOI: `10.1021/es4043726`. [Online]. Available: `http://dx.doi.org/10.1021/es4043726` (visited on 10/28/2015).

[63] O. Bartholomy, T. Vargas, M. Simone, C. Hansen, S. Fitchett, and A. Pohl, "Benchmarking solar power and irradiance forecasting accuracy at Sacramento Municipal Utility District," in *Photovoltaic Specialist Conference (PVSC), 2014 IEEE 40th*, Jun. 2014, pp. 0063–0068. DOI: `10.1109/PVSC.2014.6925196`.

[64] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renn, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, vol. 84, no. 12, pp. 2161–2172, Dec. 2010, ISSN: 0038-092X. DOI: `16/j.solener.2010.08.014`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X10002823` (visited on 08/29/2011).

[65] A. Olson, R. A. Jones, E. Hart, and J. Hargreaves, "Renewable Curtailment as a Power System Flexibility Resource," *The Electricity Journal*, vol. 27, no. 9, pp. 49–61, 2014. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1040619014002322` (visited on 11/23/2015).

[66] J. H. Nelson and L. M. Wisland, "Achieving 50 Percent Renewable Electricity in California," Union of Concerned Scientists, Oakland, CA, Tech. Rep., Aug. 2015. [Online]. Available: `http://www.ucsusa.org/clean-energy/california-and-western-states/achieving-50-percent-renewable-energy-in-california` (visited on 09/03/2015).

[67] L. Bird, J. Cochran, and X. Wang, "Wind and solar energy curtailment: Experience and practices in the United States," *National Renewable Energy Laboratory2014*, 2014. [Online]. Available: `http://www.nrel.gov/docs/fy14osti/60983.pdf` (visited on 11/23/2015).

[68] J. Zhang, B.-M. Hodge, A. Florita, S. Lu, H. F. Hamann, and V. Banunarayanan, "Metrics for evaluating the accuracy of solar power forecasting," in *3rd International workshop on integration of solar power into power systems, London, England*, 2013. [Online]. Available: `http://www.nrel.gov/docs/fy14osti/60142.pdf` (visited on 09/09/2015).

[69] V. P. A. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Solar Energy*, vol. 97, pp. 58–66, Nov. 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2013.08.002`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X13003125` (visited on 05/14/2014).

[70] M. A. Cohen and D. S. Callaway, "Physical Effects of Distributed PV Generation on California's Distribution System," *ArXiv:1506.06643 [physics]*, Jun. 2015, arXiv: 1506.06643. [Online]. Available: `http://arxiv.org/abs/1506.06643` (visited on 09/17/2015).

Chapter 4

Learning HVAC Energy Use from Smart Meters

Preface

In the following chapter, I propose a general statistical model for utility meter data, which estimates thermal properties of residential buildings and disaggregates energy used for heating, ventilation and air conditioning (HVAC). The model's definition is flexible, and it encompasses many of those already in literature. I fit and validate multiple variations of it to test which forms perform best. I then use its output to estimate indoor temperature dynamics in monitored homes by identifying where HVAC energy use deviates from that needed to maintain the building at a steady temperature. I performed this work with the guidance of my advisor, Duncan Callaway, with whom I plan to submit this chapter to *Energy and Buildings* during the summer of 2016. I reproduce our work here with the consent of Dr. Callaway. Within the chapter I refer to all work as both Dr. Callaway's and my own using the pronoun "we."

Chapter Abstract

Advanced metering infrastructure (AMI) measures electricity use at a rate of minutes as opposed to days or months and may provide valuable information to energy consumers, providers, and policy-makers. In this paper, we explore statistical models fit to AMI data—which we refer to as "utility meter models"—that predict the largest end use of electricity in residential homes: heating, ventilation, and air conditioning (HVAC). Specifically, we evaluate each model's ability to estimate the frequency of HVAC use, the efficiency of operation, and the amount of energy consumed.

Models fit to daily data—versus those fit to hourly data—are better at classifying whether or not HVAC is active. Hourly classifications improve when requiring that a day must be classified as containing HVAC before an hour within it can be. The distribution shape assumed for model disturbances affects both classifications and parameter estimates. And a post-fitting disaggregation process improves estimates of cooling energy for all models studied.

Finally, we attempt to recover indoor temperature dynamics in homes metered by AMI. Though our models do not estimate these dynamics endogenously, we can infer intra-day changes in temperature well.

4.1 Introduction

Advanced metering infrastructure (AMI) measures electricity consumption at a rate of minutes as opposed to days or months and could hold valuable information for energy consumers, providers, and policy-makers [13], [14].

AMI is currently installed in over 40% of residential buildings in the US [15], and a utility typically installs it for every consumer within their territory. As a result, learning from AMI data can be scaled to representative samples of households. It may allow researchers unprecedented access to empirically study questions relating to energy efficiency and behavior without having to rely on expensive or invasive strategies. Thanks to the Green Button Initiative, AMI data are transferable—with consumer permission—to third parties in a standard format [16] and be used by companies to provide consumers with services based on their energy use.

AMI data have been used to segment customers [71], [72], to estimate the effects of electricity pricing programs [71], [73], and to predict appliance ownership [74]. But by far the most interest in AMI data has focused on energy disaggregation, which apportions total energy use to different devices within a building. The driving force behind this interest is to create information-based products that enable consumers better manage their consumption habits to reduce energy use [13]. We envision an additional use for this information in energy efficiency and demand response research.

While accuracy of disaggregation algorithms is progressing it is still relatively low and can be biased. An independent analysis of EEme's algorithm from Pecan Street finds that the median estimates of weekly HVAC consumption are biased by 30% [16], [20]. To increase the accuracy of these methods, companies or researchers collect additional information from sensors on appliances, or from surveys of building occupants, or they increase the sampling rate of the smart meters [21], [22], [75]–[79]. While the costs of additional equipment for this sensing could be (or become) small, one still must convince consumers to provide this information. Engaging residents to install additional appliances or to fill out surveys is a large barrier that will greatly reduce and bias samples, making them less useful for research purposes.

Utility meter models in literature

A separate set of statistical models use only AMI and climate data; we refer to these as "utility meter models". These models trade off data needs for predictive power, seeking primarily to identify only HVAC loads and not specific plug load appliances. They also focus less on disaggregation and more on estimating the thermal properties of buildings, such as heat transfer coefficients. While only one end-use for energy in the home, HVAC accounts for more than 50% of residential energy use, and defines the timing of peak electricity demand in almost every industrialized power system. Also, the amount of HVAC energy used is determined as much by behavior as by appliance efficiency [17], [18]. But consumers are notoriously unconscious of how to manage their thermostat set points to save energy, much less how to manage their set points to reduce system-wide costs [17], [19].

This motivates us to build tools for AMI data that both disaggregate HVAC energy use as well as estimate HVAC set point schedules. In addition to consumer information programs, such methods could find baselines for the efficiency of HVAC in homes as well as the thermostat scheduling practices of consumers; it is against these baselines that energy efficiency and demand response programs should be evaluated. They could also be used in impact evaluations for energy efficiency or demand response programs, providing additional insight into why some consumers saved while others didn't. They could be used in research studies that identify drivers of energy efficiency or behavior, such as geography or landlord/tenant effects. They could even help us explain the large variance in energy use by otherwise similar consumers, or explain why some consumers' energy use has increased or decreased over time.

Utility meter models began with monthly data, and were standardized with the Princeton Score-keeping Method (PRISM), the purpose of which is to transform observations in any year to the energy that would have been used in an average weather year. Recent studies have applied PRISM or PRISM-like models to predict how the existing building stock will be affected by warming temperatures, or how consumers react to anomalous climate events. See [80] for a recent review of modeling and results.

More recently, researchers have adapted the PRISM model to leverage the additional information in hourly data. Birt et al. apply a piecewise linear analysis to identify a relationship between energy use and outdoor temperature [23]. Dyson et al. use latent variables to endogenously find times when air conditioning is being used, and fit a relationship between energy use and outdoor temperature during only these times [26]. Albert and Rajagopal similarly identify thermal responsiveness conditioned on a latent variables, however they allow the building to be in as many as 7 different thermal regimes; they also use a hidden semi-Markov model to induce correlation between latent variables in time [25]. Finally Wytock and Kolter are an exception who focus on disaggregation rather than on thermal modeling: they use an optimization method to disaggregate cooling load, heating load, base load, and variable load by relating each to exogenous climate observations and then constraining their sum to equal the energy measured at the utility meter [24].

Overview

This study seeks to build on utility meter models in multiple ways: relating a general form for these models to physical models of heat dynamics in buildings; tracking their assumptions; adapting their form in an attempt to make them more consistent; rigorously testing and validating their predictions; and extending their abilities to predict internal temperature dynamics.

4.2 Utility meter models

We use the term “utility meter model” to refer to statistical models that are fit to utility meter data. Their premise is simple: first identify times when heating or cooling is taking place, and second identify a relationship between energy use and outdoor temperature during these times. Energy use that is predicted by the outdoor temperature relationship is attributed to heating or cooling. We show four examples of this process in the panels of Figure 4.1; where blue points represent readings when cooling is active, and the blue line represents an estimated linear relationship during these times.

Eqs. (4.1a)-(4.1d) present a general form for utility meter models; where three linear component models add to predict the total energy consumption of a home: (4.1a) is a component model for cooling energy, (4.1b) is for heating, (4.1c) is for

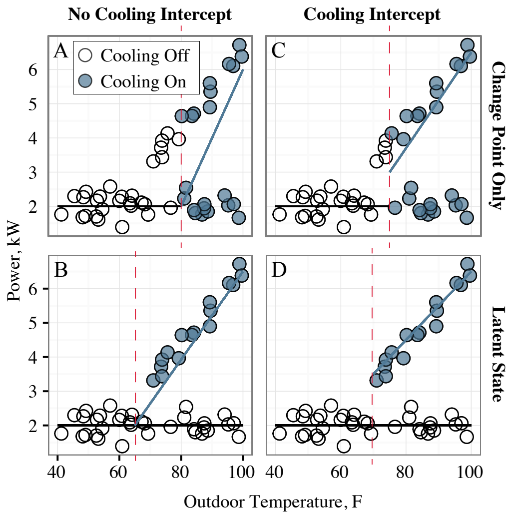

Figure 4.1: Descriptive examples of models with and without endogenous classifications (rows), and with and without cooling specific intercepts (columns).

non-HVAC, and (4.1d) is the model for total for energy use—the sum of all three. Each component model contains observations in the vector $X_{\cdot,t}$, coefficients in the vector $\beta_{\cdot}$, and a disturbance, $\epsilon_{\cdot,t}$; $c_{\cdot,t} \in \{0,1\}$, are binary variables that signify heating or cooling being on or off. This form can easily contain more component models, as presented in [23]–[25], which could represent different modes of heating or cooling, or different non-HVAC appliances.

$$\textbf{cooling} \quad P_{c,t} = c_{c,t}\left[X_{c,t}\beta_c + \epsilon_{c,t}\right] \tag{4.1a}$$

$$\textbf{heating} \quad P_{h,t} = c_{h,t}\left[X_{h,t}\beta_h + \epsilon_{h,t}\right] \tag{4.1b}$$

$$\textbf{non-HVAC} \quad P_{o,t} = X_{o,t}\beta_o + \epsilon_{o,t} \tag{4.1c}$$

$$\textbf{total} \quad P_t = \underbrace{X_{o,t}\beta_o}_{\text{non-HVAC}} + \underbrace{\sum_{m\in c,h} c_{m,t}X_{m,t}\beta_m}_{\text{HVAC}} + \underbrace{\epsilon_{o,t} + \sum_{m\in c,h} c_{m,t}\epsilon_{m,t}}_{\text{Lumped errors}} \tag{4.1d}$$

In the remainder of this section we further explain each aspect of the model, and we identify variations of the model that we test.

Switching HVAC on and off

We evaluate two means for switching HVAC on/off. Eq. (4.2a) defines a traditional change point model (shown in Panels A and C of Figure 4.1) where cooling is assumed to be on above an outdoor temperature change point and off below—$\mathbb{I}(\cdot)$ is an indicator function, which equals one when the condition is true and zero otherwise. This model is similar to—and in some cases identical to—using heating and cooling degree days as regressors, which are standard for utility meter models fit to monthly data. Additional heating or cooling models could be defined with additional change points, as they are in Birt et al[23]. Each additional change point will create another region that exhibits its own temperature responsiveness.

$$c_{c,t} = \mathbb{I}(T_{a,t} > \bar{T}_{cp}) \tag{4.2a}$$

$$p\{c_{c,t} = 1\} = \begin{cases} w_c & T_{a,t} > \bar{T}_{cp} \\ 0 & otherwise \end{cases} \tag{4.2b}$$

We also evaluate a method that treats $c_{c,t}$ as a latent—i.e. unobserved—random variable. Latent variables allow HVAC to follow more variable human behavior, such as turning HVAC off when not at home or changing the thermostat set point. Eq. (4.2b) defines the probability distribution of this latent variable, which is still dependent on a change point: when outdoor temperature is below the change point cooling is off; when above the change point cooling is on with probability $w_c \in [0, 1]$. This is identical to the formulation in Dyson et al. [26].

Heating and cooling models
In (4.1a), we write a linear model for cooling with general vectors for observations, $X_{c,t}$, coefficients, β_c, and disturbances $\epsilon_{c,t}$. The model for heating is directly analogous, and for brevity we will only refer to the cooling model in this section. Eq. (4.3a) shows the common model where only outdoor temperature minus the change point are included in $X_{c,t}$. This allows cooling energy to linearly increase with outdoor temperature and forces it to be zero (in expectation) at the change point, shown in panels A and B of Figure 4.1.

$$X_{c,t} = [T_{a,t} - \bar{T}_{cp}] \tag{4.3a}$$

$$X_{c,t} = [1 \quad T_{a,t} - \bar{T}_{cp}] \tag{4.3b}$$

$$X_{c,t} = [1 \quad T_{a,t} - \bar{T}_{cp} \quad (T_{a,t} - \bar{T}_{cp})^2] \tag{4.3c}$$

In Eq. (4.3b) we include an intercept in $X_{c,t}$ (analogous to methods from Dyson et al. and Albert et al. [25], [26]) The intercept allows cooling energy and non-HVAC energy prediction lines to be discontinuous; i.e. at the change point, there may be some non-zero amount of energy consumed for cooling. Panels C and D of Figure 4.1 show discontinuous models fit with these intercepts. The coefficient on the intercept could represent many things, including a constant amount of fan energy needed for the HVAC system.

Finally, Eq. (4.3c) includes a quadratic term in X_c. This term could account for diminishing efficiency of air conditioners or heat pumps at extreme temperatures, as we explore later in this section.

non-HVAC models

We predict non-HAVC use using simple daily and weekly schedules. The set of regressors for non-HVAC use, $X_{o,t}$, contains only indicator variables that identify the hour of day and weekdays versus weekends.

Distribution of disturbances

The probability distribution of disturbances is a significant assumption for models with latent states. Models in the literature almost always use a normal distribution for these disturbances; this assumption is sometimes implicit in the choice to minimize square errors, which is identical to maximizing the logged likelihood of a normal distribution. Model parameters and latent states are fit such that prediction errors best match the assumed distribution. Prior work has shown that poorly defined distributions can greatly bias both the estimates of latent states and parameters [81].

We test two distribution shapes for disturbances: a normal distribution, and a kernel density estimate (KDE). A KDE is a non-parametric estimate of the distribution shape, similar to a smoothed histogram. We use a Gaussian kernel function for this smoothing, and define the bandwidth variance of the kernel using Scott's rule [82]. Figure 4.2 shows an example of a normal and a KDE distribution fit to the same utility meter data; notice the high peak and wide tail of the KDE distribution versus the normal. The KDE is also bimodal, with an additional small peak around 2.5.

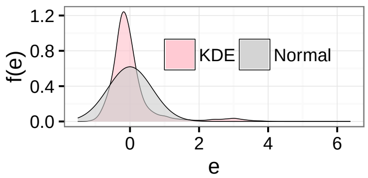

Figure 4.2: Example of a kernel density estimate for the distribution of disturbances versus a normal distribution.

Without latent states, normal distributions are less likely to bias estimates. With the exception of the choice of change point, these are linear models; and maximizing likelihood given normally distributed disturbances is equivalent to ordinary least squares (OLS). OLS is an unbiased estimator for linear models, regardless of the distribution of disturbances. For models with only change points, we only test normally distributed disturbances.

Disaggregation of residuals

Most studies use the expected value of the cooling model, $X_{c,t}\beta_c$, to estimate the amount of cooling energy at each time [23], [25], [26]. However, given an observation total home energy use, P_t, we also obtain the total model's residual, which is the sum of the residuals from each component model. If heating and cooling are off, this total residual can be attributed to non-HVAC consumption. If either is on, we can attribute part of this residual HVAC and the remainder to non-HVAC.

We define a maximum likelihood method for disaggregating a total residual into component model residuals in Section B.2 of the supporting information, for which we assume the disturbances to be normally distributed and independent. We also evaluate how correlation between HVAC and non-HVAC energy use affects the disaggregation problem, and find that the effect will be small under expected conditions.[1] We also show that the maximum likelihood method is equivalent to the disaggregation method from Wytock and Kolter: “contextually supervised source separation” (CSSS) [24].

We make the normal assumption even in the case where we use a KDE of the distribution for classification purposes; this is because the maximum likelihood problem for KDE disturbances must be performed numerically and can be inconsistent. We consider the normal assumption to be a good estimate that can be improved upon in the future.

Assumptions in heating and cooling models

We explore the assumptions inherent in the simplified cooling model by comparing it to a first order model for heat dynamics in a building. Eq 4.4 presents this first order model, which already assumes that all heat in the building is stored in one lumped thermal mass. In Eq 4.4, $\eta_{c,t}$ is the efficiency of the cooling device at time t, $T_{i,t}$ is the temperature of the lumped thermal mass, UA, is the thermal resistance of the building envelope, C_i is the thermal capacity of the mass, P_s is the solar insolation at time t, A is the buildings aperture, and $\phi_{occ,t}$ is the heat from internal sources at

---

[1] correlation being positive and the variances of HVAC and non-HVAC energy use within a factor of four. Conversely, negative correlation can greatly affect the results of disaggregation.

time t (people, electronic devices). Three additional assumptions reduce this model to the simple form in Eq. 4.1a.

1. Indoor temperature is constant: $T_{i,t} = \bar{T}_i$ and $\Delta T_{i,t} = 0$
2. Internal heat gains are constant: $\phi_{occ,t} = \bar{\phi}_{occ}$
3. Solar heat gains are negligible: $AP_s \approx 0$

$$\begin{aligned} \eta_{c,t} P_{c,t} &= UA(T_{a,t} - T_{i,t}) - C_i \Delta T_{i,t+1} + AP_s + \phi_{occ,t} + \epsilon_{c,t} && (4.4) \\ &= UAT_{a,t} - \underbrace{UA\bar{T}_i + \bar{\phi}_{occ}}_{UA \cdot \bar{T}_{cp}} + \epsilon_{c,t} && (4.5) \end{aligned}$$

Because both indoor temperature and occupant heat-gains are constant, they become indistinguishable—a.k.a., unidentifiable—in a statistical model. Instead a model will estimate a combination of the two, which becomes the change point T_{cp}.

If we assume the efficiency of HVAC to be constant, $\eta_{c,t} = \bar{\eta}_c$, we recover cooling model without an intercept; where the coefficient on outdoor temperature is an estimation of $UA/\bar{\eta}_c$, shown in Eq. (4.6)

$$P_{c,t} = -\frac{UA}{\bar{\eta}_c}(T_{a,t} - \bar{T}_{cp}) + \epsilon_{c,t} \tag{4.6}$$

For air conditioners and other vapor liquid compression devices (VLC), this efficiency is called the coefficient of performance (COP) and is dependent on outdoor temperature. For ideal VLC devices, this relationship is inverse linear. We mimic this inverse linear relationship using $\eta_{c,t}^{-1} = \eta_{cp}^{-1} + \beta_\eta(T_{a,t} - \bar{T}_{cp})$; where η_{cp} is the COP at the change point and β_η is an inverse linear relationship between outdoor temperature and COP. Eq. 4.7 applies this assumption, which resembles the quadratic model presented in (4.3c).

$$P_{c,t} = -\frac{UA}{\eta_{cp}}(T_{a,t} - \bar{T}_{cp}) - UA\beta_\eta(T_{a,t} - \bar{T}_{cp})^2 + \epsilon_{c,t} \tag{4.7}$$

These assumptions only apply when HVAC is on, $c_{c,t} = 1$, and when the disturbance term is zero, $\epsilon_{c,t} = 0$. Thus each model deviates from these assumptions otherwise. In Section 4.5 we use these deviations to infer indoor temperature dynamics.

Model Fitting

We used the expectation-maximization (EM) algorithm to find the maximum likelihood estimate for all parameters for each model except the change-point, which must be specified exogenously.[2] To fit the change point, we performed a grid search over values from 50F to 90F in increments of 5F, choosing the model with the maximum likelihood as fit by EM. Section B.1 of the appendix defines all equations used to update parameters and latent states.

Eq. 4.8 defines an important output of EM, $\gamma_{c,t}$: the expected value of each latent state, $c_{c,t}$, given the model parameters and all observations.

$$\gamma_{c,t} = \mathbb{E}[c_{c,t}|\Theta, \sigma, \bar{T}_{cp}] \tag{4.8}$$

4.3 Data

Table 4.1 shows each of our data sources alongside its time resolutions and coverage. We used interval averages—starting at the top of each hour—to scale all data to a 60 minute time resolution, which matches the time resolution of many AMI meters.

Table 4.1: Data sources

Source	Coverage	Res.	Variables	Location
Landis+Gyr	June 2012 - Aug 2012	5 min	Indoor temperature Set point temperature Relative humidity HVAC energy use	Texas
Pecan St.	Dec. 2012 - Nov. 2013	1 min	Sub-metered energy use	Texas
Forecast.io		60 min	Outdoor temperature Relative cloudiness	U.S.

Landis+Gyr provided readings from over 5000 thermostats and air conditioners installed in over 3000 residential buildings in the San Antonio Texas area. Ideally, we would fit and validate models using data from thermostats, air conditioners, and

---

[2] The discontinuity approach to the change point definition makes updating it intractable in EM, this could be corrected in future models by using a logistic or probit link function to estimate the change point.

smart meters all located at the same home. Unfortunately, we did not have smart meter data from the homes monitored by Landis+Gyr.

To overcome this limitation, we created "artificial AMI" data by combining the air conditioning data form Landis+Gyr with non-air conditioning data from Pecan St. Pecan Street sub-meter electricity use at the circuit level in residential buildings near Austin Texas. We obtained 1 year of sub-metered energy use from 10 homes from December 2012 to November 2013. Because this time coverage did not overlap with the coverage of Landis+Gyr's set, we subtracted 364 days from each time-stamp. This shift maintains the original day of week of observations and roughly maintains season; though it does not preserve variations from holidays.

We generated 1000 sets of artificial AMI data by combining the non-cooling consumption of the 10 sub-metered homes with the AC consumption from a sample of 100 homes metered by Landis+Gyr. Combining the data in this way allowed us to test how properties of AC and non-HVAC consumption affect model fitting. Restricting ourselves to only 100 air conditioners limited the computation time when testing each model.

Tables 4.2 and 4.3 show summary statistics from the air conditioners and homes used to create the artificial AMI Data. The 100 air conditioners from the Landis+Gyr dataset were intentionally chosen to contain homes that often change their temperature set-point as well as homes that rarely do. They also contain a wide range AC power capacities (reflected in maximum power consumption); and contain a few systems which appear to have undersized AC units, and spend a large amount of time operating in excess of 90% of AC capacity.

Table 4.2: Descriptive statistics of validation data from Landis+Gyr. These data contain readings from thermostats and sub-metered air conditioning energy use.

	Quantiles				
	min	25	med	75	max
Ave. AC Power (kW)	0.044	1.1	1.4	1.8	2.9
Max. AC Power (kW)	2.2	3.5	4.2	5.1	8.8
Hours >90% max AC (frac)	7.5e-05	0.011	0.032	0.055	0.21
Days <2F T_i Change (frac)	0	0.011	0.1	0.73	0.95
Days >4F T_i Change (frac)	0	0.11	0.53	0.75	0.99
Days >2F SP Change (frac)	0.12	0.16	0.61	0.78	1
Days >4F SP Change (frac)	0.011	0.051	0.13	0.57	1

Table 4.3: Descriptive statistics of validation data from Pecan Street. These data measure all non-cooling power consumption, sub-metered by circuit.

	Quantiles				
	min	25	med	75	max
Ave. power (kW)	1.4	2.5	3.5	3.8	5.1
% Ave. from one circuit	13	26	28	39	79
Peak power (kW)	2.9	4	6.9	7.7	9.3
Ratio of home to circuit peak	.22	.46	.58	.74	.78
Kurtosis	4.6	7.7	12	18	28

The Pecan Street data also represent a variety of consumption patterns, shown in Table 4.3. We compare the peak consumption of each circuit to the peak consumption of whole-home energy use, which ranges from 22% to 78%. High values typically result from dryers. The kurtosis of non-cooling energy use in our validation sets ranges from 4.6 to 28. Kurtosis measures of how heavy a distribution's tails are: where greater values represent heavier tails and normal distributions have a kurtosis of 3.

Though the artificial AMI data are each compiled from separate households, they are not unrelated. In about 40% of artificial sets, AC energy use from Landis+Gyr and non-AC energy use from pecan street are correlated with a Pearson coefficient of greater than 0.1, and this correlation can be as great as 0.4. Though the pecan street homes are in different locations than the Landis+Gyr households, non-AC energy use is sometimes strongly anti-correlated with outdoor temperature with Pearson coefficients between -0.1 and -0.3. This is likely due to coincidence of daily schedules with diurnal temperature patterns, such that residents are more likely to be home and using energy at night.

4.4 Evaluation of utility meter models

The choice of model affects estimates of physical parameters: the change point and the overall efficiency, UA/η. To offer some intuition for how this occurs, Figure 4.3 presents results for one set of artificial AMI data fit to all models; this particular set has a high kurtosis of non-HVAC use. Each panel plots power consumption versus average outdoor temperature, similarly to those in Figure 4.1. The efficiency of cooling, UA/η, is represented by the slope of expected power consumption given that HVAC is on—the red line. For quadratic models, this relationship changes with outdoor temperature. The change-point, $\bar{T}_{cp}$, is the threshold below which air conditioning is assumed to be off.

Models fit to daily data are greatly affected by which variables are included in the cooling model, i.e. whether an intercept or quadratic term is included. In continuous models—without intercept or quadratic term—with latent states, there is a region near this intersection where classifications of HVAC state are uncertain. Uncertainty like this reduces the likelihood of the data, thus is isn't surprising that discontinuous models skip this region by choosing a positive intercept and a higher change point. The model with latent states, KDE disturbances, and a quadratic term chooses a very low change point—at which the cooling efficiency is negative. This could be the model capturing some heating at these low temperatures[3].

Models fit to hourly data were more varied in general, but were less affected by extra terms included in the cooling model. Intercepts were estimated to be around zero, and change points were unaffected by intercept or quadratic terms. The change point does vary with other modeling choices. The change point also affects the UA/η estimate: observe that models with lower change points have lower slopes and vice versa.

Models with latent states and normally distributed disturbances chose the lowest possible change point, and as a result chose a low slope (UA/η) compared to others. By visual inspection, this slope seems far too low for the data: it over-predicts at low temperatures and under-predicts at high temperatures. Including a quadratic term somewhat corrects for the error in slope, but it does not correct for the low change point. This poor fit is likely due to high non-HVAC consumption at low temperatures, which receive a high penalty given a normal distribution; thus an additional model with a higher mean (even slightly) may greatly reduce this penalty.

Applying a KDE estimate for the distribution of disturbances appears to correct the low change point error; possibly because its ability to account for patterns in these outliers. In fact, in this case KDEs for disturbance distributions are adept at classifying these high non-HVAC readings. In the actual data, there is a high cloud of readings around the change point where air conditioning is actually off. Models that assume normal distributions misclassify these. Models with KDE distributions correctly classify these readings due to a bimodal distribution shape estimated by the KDE, which recognizes that these readings are common for the non-HVAC model, even though they are far from the mean.

Figure 4.4 presents results for all artificial AMI data, and shows that some of the

---

[3] The heating data and climate data are not from the same location, but they are within the same state. Seasonal trends in temperature will still be the same

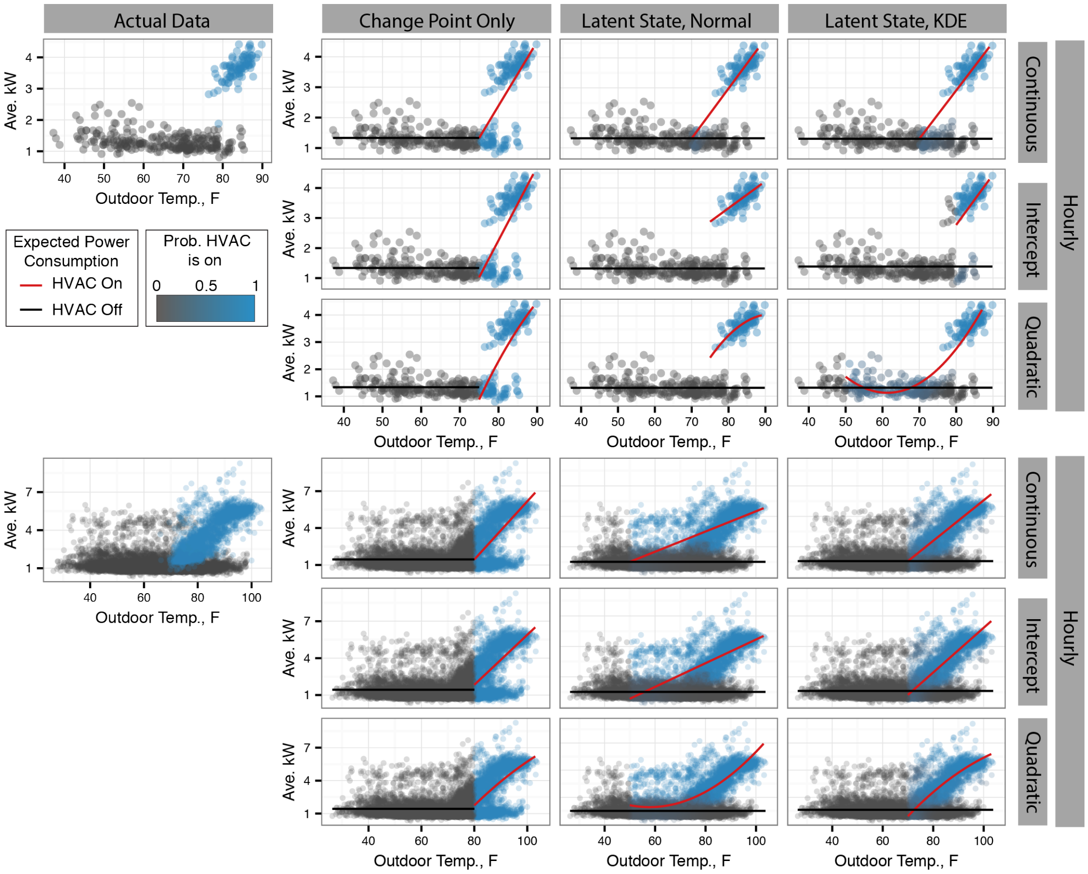

Figure 4.3: One set of artificial AMI data fit with each utility meter model.

trends from Figure 4.3 are general[4]. For daily data, using a KDE for disturbances greatly affected results. Including an intercept increased the change point, likely to skip uncertain regions. Using a quadratic relationship affected the KDE models even more, causing them to choose very low change points in some cases.

When fitting hourly data, latent state models with normally distributed disturbances choose the lowest possible change point in almost 50% of cases; they are also more likely to choose low UA/η values. However, in the other half of cases they chose change points with roughly the same distribution as models with KDE distributions. As we suspected, low change points are fit when the kurtosis of non-HVAC con-

---

[4] Models with quadratic temperature responses estimate a separate UA/η value at each temperature. We present an average UA/eta value here. UA/η values from each reading are weighted by the expected value of cooling being on. Our intention is to offer the average efficiency of cooling during all hours of use.

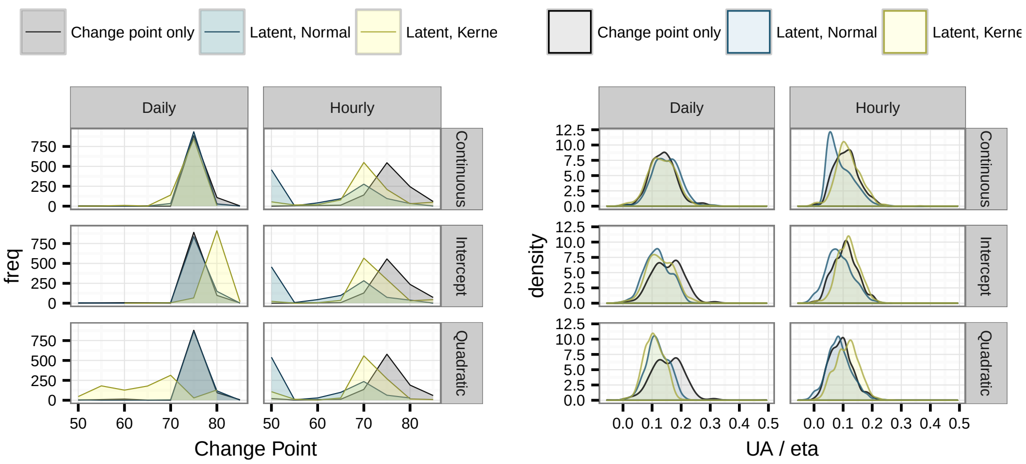

Figure 4.4: Change points and slopes (UA/η) fit by utility meters models to all artificial AMI data.

sumption is high; indicating that the high penalty on outliers defined by the normal distribution affects these parameters. Change-point-only models and model with kernel density estimates were unaffected by the kurtosis of non-HVAC consumption.

Classifying the cooling state

Classifying whether or not HVAC is on helps to exclude dates and times when there is no HVAC use. This is useful for home energy reports, which would not want to report HVAC use during a month when in fact there was none at all. As we will see later, it can also help identify thermostat schedules.

We use precision and recall to evaluate each model's ability to identify hours when cooling is active. Precision measures the fraction of "HVAC on" classifications that are correct, while recall measures the fraction of all cooling hours (or days) that are correctly classified. Our definitions of precision and recall differ from the traditional ones, because we predict an expected value of latent states, $\gamma_{c,t}$, not a binary "on" or "off". Eqs. (4.9) and (4.10) explicitly define our metrics for precision and recall. For the change-point-only models (without latent states) we do use a binary classification, where HVAC is on when outdoor temperature is above the change point, and off otherwise.

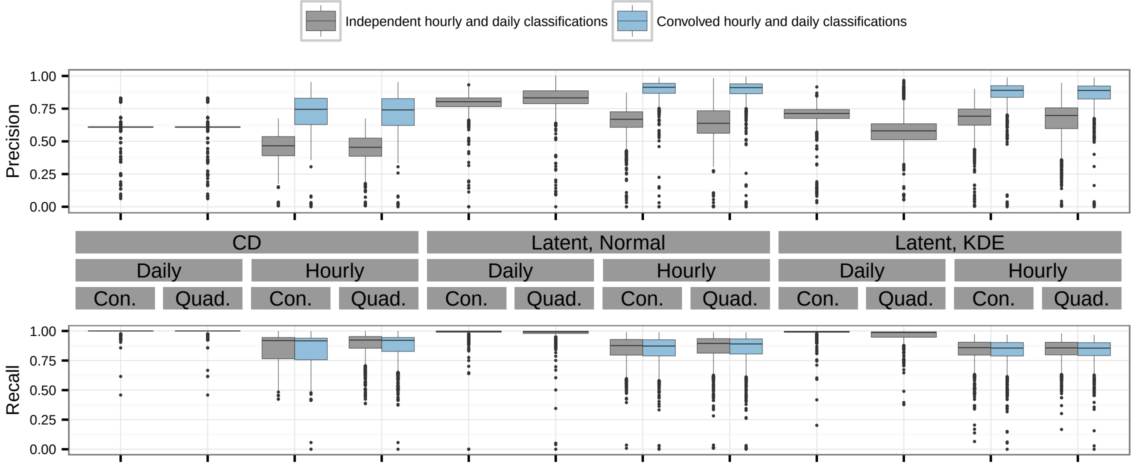

Figure 4.5: Precision and recall of HVAC on/off classifications from each utility meter model.

$$Prec = \frac{\sum_{t=1}^{T} \gamma_{c,t}\mathbb{I}(P_{c,t} > 0.02P_{c,max})}{\sum_{t=1}^{T} \gamma_{c,t}} \tag{4.9}$$

$$Rec = \frac{\sum_{t=1}^{T} \gamma_{c,t}\mathbb{I}(P_{c,t} > 0.02P_{c,max})}{\sum_{t=1}^{T} \mathbb{I}(P_{c,t} > 0.02P_{c,max})} \tag{4.10}$$

Grey box plots in Figure 4.5 display precision and recall metrics for each model—we address the blue box plots later in this section. Unsurprisingly, change-point-only models were poorer classifiers than the models with latent states.

In general, models fit to daily data perform much better than models fit to hourly data, particularly with respect to recall. This is expected because cooling use is clustered on a daily basis, compounding the difference between a cooling or non-cooling reading. The best performing daily classifier is latent state model with normal disturbances and an intercept.

Improving classifications by convolving hourly and daily estimates

While daily classifications are more reliable, hourly classifications give us more information. We improve hourly classifications by convolving the two: requiring that a day be classified as containing cooling (by the latent state model with normal distributions) before an hour within it could be. Eq. 4.11 shows this convolution process, where $c_{c,t}^{(conv)}$ is the combined daily and hourly classification. This relationship takes advantage of the high recall of daily classifiers. I.e., because the the daily model is accurate at identifying negatives, we can use these negatives to remove false positives from the hourly classifications.

$$c_{c,t}^{(conv)} = c_{c,t}^{(hourly)} * c_{c,t}^{(daily)} \tag{4.11}$$

$$\gamma_{c,t}^{(conv)} = \gamma_{c,t}^{(hourly)} \gamma_{c,t}^{(daily)} \tag{4.12}$$

We test the ability of this convolution process by applying it after fitting, using the formula in Eq. (4.12).[5] Because the change point serves as a threshold for classifications, convolving the hourly and daily classification effectively uses the higher of the two thresholds as the change point. Thus, this convolution corrects the very low change points chosen by models with latent states and normally distributed disturbances; however it does not change the UA/η value.

Blue box-plots in Figure 4.5 show the precision and recall of hourly classifications when convolved with daily classifications. Hourly precision increased greatly for all models while recall was only slightly reduced. Notably, this convolution improved the precision of the normal classification models to be greater than the kernel classification models.

Though useful for understanding our models, precision and recall are imperfect metrics for evaluating them. Our definitions of precision and recall penalize uncertainty: the expected value of classifications may be less than one but not zero, which negatively affects both precision or recall regardless of the true classification. However, the same result is preferred when estimating cooling energy because a value of $\gamma_{c,t} < 1$ may scale expected energy use to be positive but still less than expected under our our assumptions. We explore the ability of each model to predict hourly and daily cooling energy in the following section.

Disaggregating cooling energy consumption

Disaggregating the cooling energy is crux of these models. This is the result that makes up home energy reports, defines load predictions, creates a baseline for evaluating the demand response or energy efficiency of a consumer. As we will show in the next section, these estimates may also tell us about thermostat schedules.

Figure 4.6 summarizes the effectiveness of hourly cooling energy predictions—only hourly models are included because daily models don't make this prediction. All values are normalized to the power capacity of the air conditioner, as identified by

---

[5] Such a relationship could be endogenously included in the model, in which case it would change estimates of all parameters.

the maximum power consumption of AC during the study period. The top panel displays RMSE: gray box-plots show results without disaggregating residuals, yellow box-plots show results with disaggregated residuals, and blue box plots represent results with disaggregated residuals and convolved daily and hourly classifications.

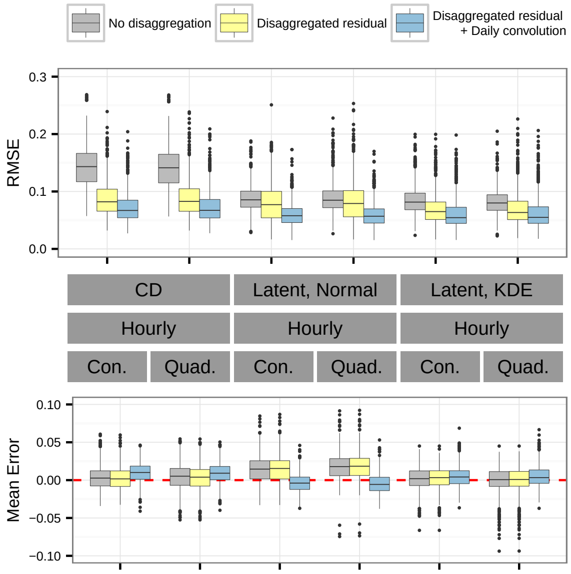

Figure 4.6: Errors in estimates of hourly AC energy use. Box plots represent distributions of errors from fits to each of the 1000 sets of artificial AMI data. (Top) Root mean square error (RMSE), (Bottom) mean error (ME).

Without disaggregation the change-point-only models performed much worse than the latent state models, but after disaggregation they became comparable; signifying that, for some purposes, disaggregation methods like CSSS can replace the need for latent states. Convolving hourly and daily classifications improved all models; after which the KDE and normal models exhibited comparable RMSE.

The bottom panel of Figure 4.6 displays each model's mean prediction errors, which are a measure of bias. Without disaggregation or convolution, the cooling degree (CD) and the KDE models exhibited a range of biases, with the median at about zero, meaning they were equally likely to over predict or under-predict. Convolving hourly and daily classifications induced a slight bias on the KDE models and a larger bias on the cooling degree models.

Models with latent states and normally distributed disturbances exhibited a positive bias (over-prediction), which changed to a negative bias after convolving daily and hourly classifications. This is likely a result of the low UA/η estimates from these models, which caused them to generally over predict cooling energy during low temperature times. Convolving hourly and daily classifications removed many classifications at low temperatures but did not change the slope of the fit line; as a result the remaining predictions at high temperatures under-predicted cooling.

4.5 Recovering indoor temperature dynamics

We infer indoor temperature dynamics in each building by evaluating how disaggregated cooling energy deviates from expected cooling energy given our assumption of constant indoor temperature.

To start, we define the cooling power needed to maintain a steady building temperature, $p_{need,t}$, as that predicted by Eq. (4.6). Panel A (top) of Figure 4.7 shows $p_{need,t}$, along with cooling energy estimated by latent state model with KDE disturbances and an intercept, $p_{est,t}$, and cooling energy measured at the air conditioner, $p_{meas,t}$. Data are shown for four days of recordings in one building. Utility meter models allow disaggregated cooling energy to deviate from $p_{need,t}$ by two means: HVAC may turn off, and a portion of the overall model's residual will be attributed to cooling.

We refer to deviations of disaggregated cooling from cooling need as the "cooling deficit." Such a deficit is indicative that at least one of our assumptions is being violated. Equation (4.13) identifies how violations may create the cooling deficit by subtracting (4.6) from the cooling energy predicted by the first order model in (4.4). The drivers of the cooling deficit are (1) deviations of indoor temperature from average temperature, which affects the heat transfered across the envelope; (2) the indoor thermal mass absorbing or emitting energy by changing indoor temperature; and (3) varying occupant or solar heat gains.

$$P_{def,t} = P_{need,t} - P_{est,t} \approx \underbrace{\frac{UA}{\eta_c}(\bar{T}_i - T_{i,t})}_{\substack{\text{Change in}\\ \text{outdoor-}\\ \text{indoor}\\ \text{temperature}}} + \underbrace{\frac{C_i}{\eta_c}\Delta T_i}_{\substack{\text{Heat}\\ \text{absorbed by}\\ \text{the thermal}\\ \text{mass}}} + \underbrace{\frac{\bar{\phi}_{occ} - \phi_{occ,t} + AP_s}{\eta_c}}_{\substack{\text{Variable}\\ \text{heat gains}}} \tag{4.13}$$

Panel B (middle) of Figure 4.7 displays indoor temperature and set point for the same home. Points along the indoor temperature line are colored to represent the cooling deficit, where red is positive and blue is negative. Following a set point increase, cooling energy ceases until room temperature rises to reach the new set-point—if it ever does. The latent state models may capture this effect by identifying the AC as off during this time. As a result, the model will predict a large cooling deficit as the building heats following a set point increase.

Following a set-point decrease, cooling energy consumption increases above P_{need} in order to cool the home. We see this effect in the actual measured value of cooling, $P_{meas,t}$, but it is not as strong in the estimated need for cooling, $P_{est,t}$. This is a result of how we've structured the model: there is a latent state that turns cooling "off" but not a latent state to turn in "on." Thus this surge may only be estimated by disaggregating residuals, and not by latent states. As a result, negative cooling deficits are less common and of lower magnitude; though they do correlate with decreases in room temperature.

Panel C (bottom left) of Figure 4.7 compares indoor temperature changes to cooling deficit (truncated by kW) for the same home during the entire study period. For this home, a cooling deficit above 1kW is indicative that the home is heating, during these hours temperature change is above 0 more than 75% of the time. A cooling deficit below -1kW is indicative that the home is cooling, however negative cooling deficits are less likely than positive deficits.

General predictions of indoor temperature

We posit that cooling deficit is driven by thermal mass effects during a day, while cooling deficit is driven by average indoor temperature on an inter-day basis. Our reasoning is that a cooling deficit caused by the building's thermal mass heating and cooling will be relatively large; but these temperature changes are only likely to take place during a day. Average temperature may change from one day to the next, but not as drastically as it does during a day. Thus the average cooling deficit over an entire day will driven by the average indoor/outdoor temperature difference, and not by an overall change in the temperature of the thermal mass.

Under these conditions, we attempt to predict two things, (A) the average daily room temperature of a home (as referenced to the average temperature overall) and (B) the range of room temperatures a home exhibits during a day. Together they will create a rough estimate of whether and how much an occupant has set back the thermostat during a day.

To predict average temperature, we assume that the second two terms in (4.13) are negligible when averaged over a day. As a result, the deviation of daily average room temperature from the overall mean temperature should equal UA/η_c times the average cooling deficit. Panel D (bottom middle) of Figure 4.7 plots average indoor

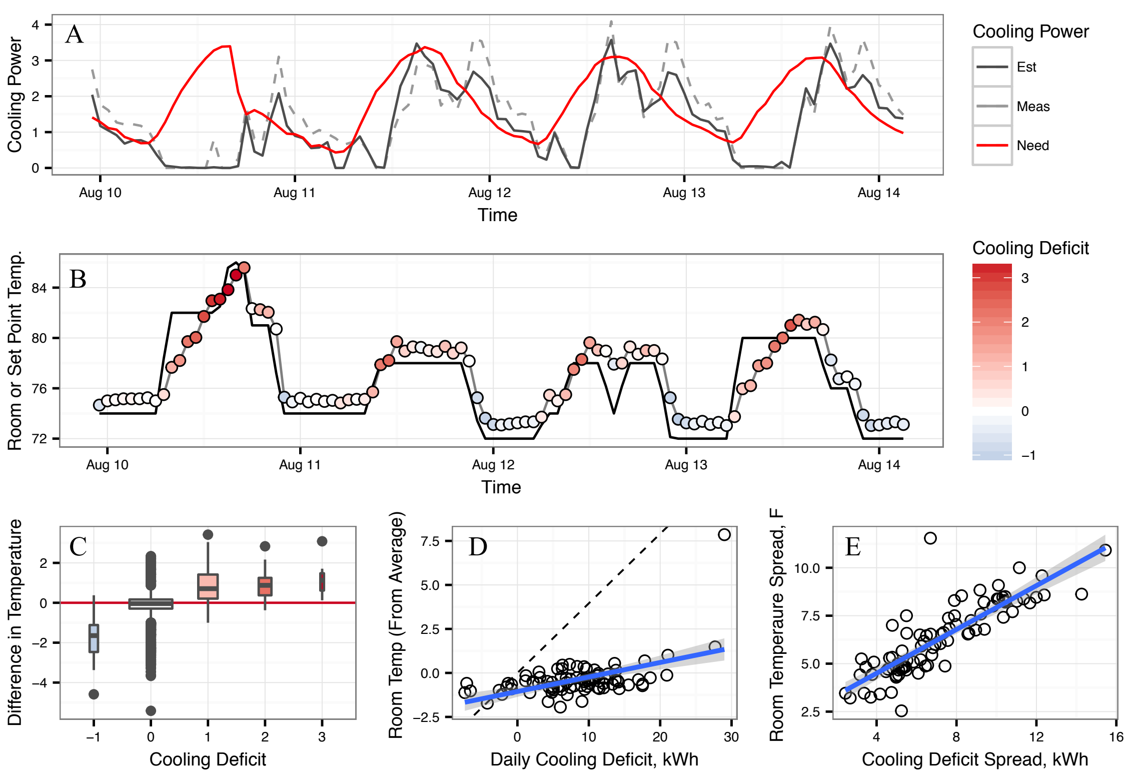

Figure 4.7: Prediction of indoor temperature dynamics using utility meter models. An example from one home and the LHK1 model.

temperature versus the daily cooling deficit for our sampled building. While there is a relationship, it is not strong, and much lower than that predicted by UA/η_c—which is shown as a black dashed line. This poor performance could be because of the bias toward negative cooling deficits. Summing over a day exacerbates this bias, leading to much larger than actual cooling deficits.

We predict the total range of indoor temperatures over a day by assuming that the first and third terms of Eq. (4.13) are negligible; thus any cooling deficit is absorbed by the thermal mass. These deficits accumulate in the indoor thermal mass, representing cumulative changes in temperature over many hours. Eq. (4.14) explains how accumulated cooling deficits are an approximation of room temperature (referenced to some initial temperature). To correct for bias in the cooling deficit—which we identified earlier—we de-mean it before finding how it accumulates.

$$P_{def}^{accum}(t) = \sum_{k=0}^{t} P_{def,k} - \bar{P}_{def} \approx C_i(T_{i,t} - T_{i,0}) \tag{4.14}$$

$$\bar{P}_{def} = \frac{1}{T}\sum_{t=0}^{T} P_{def,t}$$

To test this method we identify the range of accumulated cooling deficits over the course of a day, and compare it to the range of indoor temperatures that day. Panel E of Figure 4.7 shows this test for each day of data in our example building. For this building, the accumulated cooling deficit is a good predictor for how much indoor temperature varies on a given day.

However, if we are to use this relationship to predict for buildings without indoor temperature sensors, we need a means of estimating the slope of this relationship. The slope should be the thermal capacity of the indoor thermal mass, C_i, which we do not estimate. In its place, we use the maximum estimated cooling energy use during the year. Our reasoning is that this maximum energy is an estimate of the capacity of the air conditioner, and that this capacity should be sized appropriately for the thermal mass it cools. Thus the lager the thermal mass in a building, the larger an air conditioner will be installed to cool it.

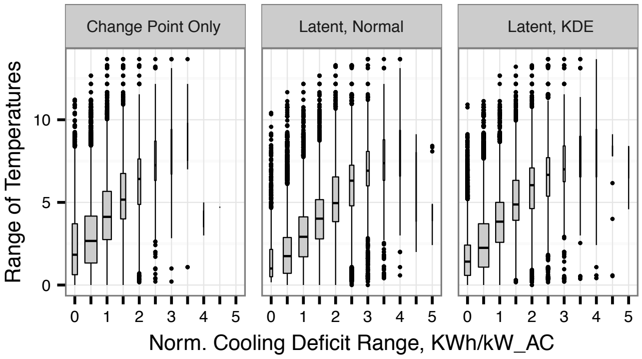

Figure 4.8: Range of indoor temperature versus the range of accumulated cooling deficits (daily) for every day and thermostat in the dataset. Accumulated cooling deficits are normalized to an estimate of the capacity of the cooling device, in kW. Results are binned and presented as box plots to better display trends and ranges.

Figure 4.8 plots the daily range of indoor temperatures versus the daily range of accumulated cooling deficits for each day in each artificial AMI dataset, resulting in roughly 80,000 thermostat days. We bin the cooling deficits and present data in box plots to make trends more visible. We also show the results from only three of the hourly models, excluding models without intercepts here only because they are similar to models with intercepts.

All of the models exhibit a strong positive relationship between the accumulated cooling deficit and the range of room temperature. Overall, the models with latent states isolate this effect the best, as shown by the overlap in interquartile ranges. For the latent state model with kernel emissions and an intercept, a range of accumulated cooling deficits equivalent to the capacity of the air conditioner—I.e. a normalized deficit range of 1—is indicative that the indoor temperature range is between 2.5 and 5 degrees Fahrenheit; a value of zero indicates a range of 0 to 2.5 F.

4.6 Discussion

In this work we evaluate statistical models fit to AMI data—a.k.a "utility meter models"—for their ability to estimate thermal properties of buildings, to disaggregate cooling energy consumption, to classify when HVAC is being used, and to infer indoor temperature dynamics. We test 4 different properties of these models: whether or not to include latent states that endogenously classify when HVAC is active; whether to use normal distributions for disturbances or to estimate kernel density estimates; whether or not to include a constant amount of cooling energy; and whether to fit to daily or hourly resolution data.

Utility meter models estimate both a change point temperature above which cooling becomes active, and an effective cooling efficiency. Without a ground truth, we cannot confidently validate predictions of change points or cooling efficiency, however we can evaluate trends that are or are not consistent with our data. We believe that models without latent states tend to choose higher than actual change points and efficiencies, though only slightly. Latent state models that assume normally distributed disturbances choose very low change points and high efficiencies; though only when non-HVAC use has a high kurtosis—often resulting from electric heating devices like dryers. Using KDE distributions for disturbances models greatly reduces this propensity to choose low change points. For daily data, we find models with latent states and normally distributed disturbances to be the best performing. For hourly data, we believe the latent state models with a KDE of disturbance distributions to be the best performing.

We find that daily classifications are much more consistent than hourly classifications, particularly at identifying days without any AC use. We improve hourly classifications by convolving them with the daily ones; such that if a day is classified as without HVAC, then an hour is too. After this convolution, latent state models achieve about 90% precision and 80% recall; these results are consistent across all latent state models regardless of the distribution of disturbances assumed.

We propose a means of disaggregating model residuals using maximum likelihood. This process is a special case of contextually supervised source separation, as presented by Wytock and Kolter in [24]. Such a method improves the cooling degree model to be almost comparable with the latent state models. Overall, we find that all models are highly effective at disaggregating cooling energy consumption, though assuming normally distributed disturbances caused a bias in predictions, resulting from their bias toward choosing lower change points.

Finally, we infer indoor temperature dynamics from each model's output. We calculate a "cooling deficit," which is the difference between estimated cooling energy consumed and the predicted cooling needed to maintain the building at a steady temperature. We find that the cooling deficit can predict how much indoor temperature ranges in a given day. Such estimates could be helpful for predicting which consumers vary their thermostat set points to conserve energy. In the future we would also potentially use this deficit to predict conserved energy from thermostat settings.

Future Work

For future work, we propose to test an additional HVAC state that would represent when AC is at maximum capacity. Such an event would occur when the set point is decreased, and the air conditioner is actively cooling the house. We believe including such a state could increase our ability to predict indoor temperature dynamics.

We also note that our method for separating lumped residuals using maximum likelihood is equivalent to a very limited case of CSSS. In their original paper, Wytock and Kolter use an additional penalty in their objective function that induces autocorrelation in source signals [24]. Including such a penalty could improve all of our models' abilities to disaggregate cooling energy. Wytock and Kolter also include an extra source for variable non-HVAC loads as opposed to baseload. Separating these sources may not only provide more information, but also may account for the non-Gaussian shape of disturbances without having to rely on a KDE. In the future, we may integrate more aspects of CSSS into our modeling—possibly even creating a form that will endogenously estimate the effects of cooling on indoor temperature.

While each model's ability to recover indoor temperature dynamics is impressive, different information would be more useful for energy analysis. In particular, we may like to know the amount of energy saved as a result of set point changes. This information would represent the current amount of energy savings from thermostat set-backs, and would provide a baseline from which the effects smart thermostats or demand response programs should be evaluated. We also may be identify whether set point scheduling explains why some households are able to consume much less than others, or why the total consumption of some households has dramatically increased or decreased over time.

Future work should also focus on issues that will arise if we apply this method to a wider set of smart meter data. These include identifying whether or not heaters or coolers are present in a home, and testing whether these methods work on less conventional sources of cooling—such as window ACs or swamp coolers.

4.7 Chapter References

[13] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, Special Section: Transition Pathways to a Low Carbon Economy, vol. 52, pp. 213–234, Jan. 2013, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2012.08.062`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421512007446` (visited on 01/28/2015).

[14] S. Darby, "Smart metering: What potential for householder engagement?" *Building Research & Information*, vol. 38, no. 5, pp. 442–457, 2010.

[15] The Institute for Electric Efficiency, "Utility-sacle Smart Meter Deployments," Tech. Rep., Sep. 2014.

[16] Energy Information Administration, "An Assessment of Interval Data and Their Potential Application to Residential Electricity End- Use Modeling," Tech. Rep.

[17] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat: Using occupancy sensors to save energy in homes," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, ACM, 2010, pp. 211–224. [Online]. Available: `http://dl.acm.org/citation.cfm?id=1870005` (visited on 04/23/2016).

[18] L. Lutzenhiser, H. Hus, M. Moezzi, A. Levenda, and A. Woods, "Lifestyles, Buildings and Technologies: What Matters Most?" *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA*, pp. 12–17, 2012. [Online]. Available: `http://www.aceee.org/files/proceedings/2012/data/papers/0193-000034.pdf` (visited on 03/20/2014).

[19] A. Meier, “How people actually use thermostats,” 2012. [Online]. Available: `http://escholarship.org/uc/item/3vd5q0cp.pdf` (visited on 03/28/2014).

[20] Pecan Street, “Setting the Benchmark for Non-Intrusive Load Monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation,” Tech. Rep., Jan. 2015.

[21] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, “Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring,” en, *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844–858, Oct. 2010, ISSN: 1530-9290. DOI: `10.1111/j.1530-9290.2010.00280.x`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9290.2010.00280.x/abstract` (visited on 03/18/2016).

[22] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, “Leveraging smart meter data to recognize home appliances,” in *2012 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2012, pp. 190–197. DOI: `10.1109/PerCom.2012.6199866`.

[23] B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, and I. H. Rowlands, “Disaggregating categories of electrical energy end-use from whole-house hourly data,” *Energy and Buildings*, vol. 50, pp. 93–102, Jul. 2012, ISSN: 0378-7788. DOI: `10.1016/j.enbuild.2012.03.025`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0378778812001739` (visited on 02/06/2015).

[24] M. Wytock and J. Z. Kolter, “Contextually Supervised Source Separation with Application to Energy Disaggregation,” *ArXiv:1312.5023 [cs, math, stat]*, Dec. 2013, arXiv: 1312.5023. [Online]. Available: `http://arxiv.org/abs/1312.5023` (visited on 10/10/2014).

[25] A. Albert and R. Rajagopal, “Thermal Profiling of Residential Energy Use,” *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–10, 2014, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2329485`.

[26] M. E. H. Dyson, S. D. Borgeson, M. D. Tabone, and D. S. Callaway, “Using smart meter data to estimate demand response potential, with application to solar energy integration,” *Energy Policy*, vol. 73, pp. 607–619, Oct. 2014, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2014.05.053`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421514003681` (visited on 09/12/2014).

[71] A. Todd, *Insights from Smart Meters: Identifying Specific Actions, Behaviors and Characteristics that drive savings in Behavior-Based Programs*, 2014.

[72] A. Albert, T. Gebru, J. Ku, J. Leskovec, and R. Rajagopal, "Drivers of variability in energy consumption," in *Submitted for review to European Conference on Machine Learning*, 2013. [Online]. Available: `http://www.stanford.edu/~adalbert/papers/drivers_variability.pdf` (visited on 03/20/2014).

[73] A. Todd, *Insights from Smart Meters: The Potential for Peak-Hour Savings from Behavior-Based Programs*, 2014.

[74] K. Basu, V. Debusschere, and S. Bacha, "Residential appliance identification and future usage prediction from smart meter," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov. 2013, pp. 4994–4999. DOI: `10.1109/IECON.2013.6699944`.

[75] J. St. John, *Putting Energy Disaggregation Tech to the Test : Greentech Media*, Nov. 2013. [Online]. Available: `http://www.greentechmedia.com/articles/read/putting-energy-disaggregation-tech-to-the-test` (visited on 07/22/2015).

[76] ——, *Perfecting Energy Disaggregation in the Home : Greentech Media*, Nov. 2013. [Online]. Available: `http://www.greentechmedia.com/articles/read/perfecting-energy-disaggregation-in-the-home` (visited on 07/22/2015).

[77] K. X. Perez, W. J. Cole, J. D. Rhodes, A. Ondeck, M. Webber, M. Baldea, and T. F. Edgar, "Nonintrusive disaggregation of residential air-conditioning loads from sub-hourly smart meter data," *Energy and Buildings*, vol. 81, pp. 316–325, Oct. 2014, ISSN: 0378-7788. DOI: `10.1016/j.enbuild.2014.06.031`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0378778814005131` (visited on 02/18/2016).

[78] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "Nilmtk: An open source toolkit for non-intrusive load monitoring," in *Proceedings of the 5th international conference on Future energy systems*, ACM, 2014, pp. 265–276. [Online]. Available: `http://dl.acm.org/citation.cfm?id=2602051` (visited on 03/08/2016).

[79] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, "Unsupervised Disaggregation of Low Frequency Power Measurements.," in *SDM*, vol. 11, SIAM, 2011, pp. 747–758. [Online]. Available: `http://epubs.siam.org/doi/pdf/10.1137/1.9781611972818.64` (visited on 06/22/2015).

[80] R. Fazeli, M. Ruth, and B. Davidsdottir, "Temperature response functions for residential energy demand A review of models," *Urban Climate*, vol. 15, pp. 45–59, Mar. 2016, ISSN: 2212-0955. DOI: `10.1016/j.uclim.2016.01.001`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S2212095516300013` (visited on 01/25/2016).

[81] R. Langrock, T. Kneib, A. Sohn, and S. L. DeRuiter, "Nonparametric inference in hidden Markov models using P-splines," *Biometrics*, 2015. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1111/biom.12282/full` (visited on 01/18/2015).

[82] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996. [Online]. Available: `http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1996.10476701` (visited on 04/29/2016).

Chapter 5

Conclusion

This dissertation develops new empirical methods that inform two questions in energy policy. How will locations of photovoltaic (PV) systems affect the the need for flexible resources in power systems? And what are the current HVAC set point schedules in residential households?

5.1 Summary of findings

In Chapters 2 and 3, I answer the questions of whether the locations of photovoltaic (PV) generators will affect the variability and uncertainty of the power they generate, and as a result, whether locations of PV systems affect how much supply—or demand—flexibility is needed in grids that rely on them. Chapter 2 presents, fits, and validates a model of variability and uncertainty of PV generation. I call this model the "volatility state model", due to its reliance on latent states that I refer to as volatility states. Specifically, the model (a) accounts for spatial correlation, (b) predicts metrics of variability and uncertainty that are directly relevant to grid operation and planning, (c) and predicts boundaries on distribution tails that are consistent with observed data. Latent variables prove to be very useful in this model. Without these variables, models predict tails that are light compared to observed data, and thus under-predict extreme events.

Chapter 3, applies the volatility state model to predict the need for reserve generation—load following and regulation—in California under different locational scenarios for PV. I find that clustering PV into small areas exacerbates the need for reserves, resulting primarily from the spatial correlation of hourly forecast errors. The benefits of dispersion can be saturated with a relatively small number of large utility-scale systems: 25, 500 MW systems. However, these systems need to be adequately dispersed, which may have implications for the costs of transmission needed to reach them. I also identify trade-offs between siting systems to minimize variability and uncertainty or to maximize energy or capacity value. In California, there exists a trade-off between energy and capacity values, where areas of the state with the greatest energy resource tend to be cloudy on summer afternoons, when peak demand—driven by air conditioning—tends to be greatest.

In Chapter 4, I explore the ability of statistical models fit to AMI data to predict the largest end use of electricity in residential homes: heating, ventilation, and air conditioning (HVAC). Specifically, I evaluate models' abilities to predict of the timing of HVAC use, the efficiency of operation, and the amount of energy consumed. I begin by presenting utility meter model, that directly relates to both physical models of heat dynamics in buildings, and to other utility meter models that are already present in literature. I then fit and validate multiple variations of this model—some similar to those in literature, and some of our own device—using data from air conditioners, thermostats, and residential sub-meters.

I find a great benefit to combining models fit to daily and hourly data; where days that are classified as without HVAC energy use cannot contain hours that are classified with HVAC. I also find that the distribution shape assumed for model disturbances greatly affects model classifications and parameters; where kernel density estimates for these distributions outperform the traditional normal distributions. Finally, I find that applying a post-fitting process that disaggregates model residuals—attributing part to HVAC and part to other end uses—increases estimations of cooling energy use, and almost obviates any need for hourly classifications of HVAC activity.

Concluding chapter 4, I attempt to recover indoor temperature dynamics in homes metered by AMI. Though our models do not estimate these dynamics endogenously, I can infer intra-day changes in temperature well, and inter-day changes in temperature weakly.

In future work, I hope to use insights form Chapter 4 to better construct utility meter models such that they may uncover set point schedules in residential buildings endogenously. Additionally, I hope to also estimate the savings that results from variable set point schedules compared to a flat schedule. This work will provide an important empirical verification of the ability of set-point schedules to reduce energy use. It will also provide an important behavioral baseline which will be useful when evaluating the effects of energy efficiency or demand response programs.

5.2 Future work

The need for flexibility that I study in Chapters 2 & 3 is defined by the timing of the presiding energy market. This market can change to better manage uncertainty and variability, and has already done so in California and the rest of the Western United States with the onset of the Energy Imbalance Market. However, controllable resources that participate in these markets have inherent flexibility limits. These inherent limits are a better benchmark by which to estimate the effect of variability and uncertainty on future power systems. Recent work attempts to inventory the total supply of and demand for flexibility by evaluating the inherent constraints for controllable generators, and the variability and uncertainty of variable generators [83].

In future work, I hope to survey the ability of data-driven tools to either reduce the variability and uncertainty of uncontrollable resources—by making better forecasts—or to increase the flexibility of controllable resources—by improving forecasting or control. I then plan to compare these abilities to the existing (and projected) supply and demand for flexibility in future power systems. By this process, I plan to to identify which applications for data-driven modeling and control will have the greatest effect on power systems abilities to integrate large amounts of renewable generation.

Chapter 3 also leaves many open questions regarding where optimally to site PV. Optimal dispersion to reduce flexibility can occur with large utility-scale systems: in California amounting to 500 MW. These utility-scale systems have many benefits over rooftop systems: their geometric arrangement is optimal for capturing solar energy, they can employ single axis tracking, their central location allows for better forecasting via sky-imaging, and they will exhibit lower costs for maintenance and cleaning—not to mention they are simply more likely to be cleaned. However many of these same benefits could be achieved with community scale PV systems, with the additional benefits of being located nearer to load centers, and creating opportunities for financing that is only available to consumers—i.e. bill reductions. In future work I also hope to explore the benefits of community scale versus utility-scale solar options for the future of California's power system.

Chapter 4 presents a model but not an application. I have already suggested to answer two questions as an application of this work, which I hope to complete as soon as possible: What are the current set point schedules in California's households? And, using these schedules as a baseline, how much savings can I expect from energy efficiency or demand response programs that will more optimally set these schedules?

However, there are many more questions that can be asked with utility meter data given effective statistical models. Such as what are the drivers for thermal inefficiency in homes? Is the average income of occupants, the age of the building, construction materials, or whether the property is rented or owned among the contributing factors? How do these results relate to conventional wisdom for the drivers of energy efficiency? What would these results imply for energy policy and building standards? I.e., should California enact laws for energy efficiency testing or retrofitting in rental properties? Are low-income individuals paying excessive amounts to heat and cool homes due to inefficiency? What is the potential (in energy savings and increased comfort) for energy efficiency retrofit programs?

Answering these questions is the driving force behind my research in utility meter models, though complete answers will require additional data on the buildings and occupants behind each meter. In the future I hope to work more closely with utilities—who have these data—to pursue these questions in detail.

Bibliography

[1] C. W. Anderson, J. R. Santos, and Y. Y. Haimes, "A Risk-based InputOutput Methodology for Measuring the Effects of the August 2003 Northeast Blackout," *Economic Systems Research*, vol. 19, no. 2, pp. 183–204, Jun. 2007, ISSN: 0953-5314. DOI: `10.1080/09535310701330233`. [Online]. Available: `http://dx.doi.org/10.1080/09535310701330233` (visited on 04/23/2016).

[2] S. Lin, B. A. Fletcher, M. Luo, R. Chinery, and S.-A. Hwang, "Health impact in New York City during the Northeastern blackout of 2003," eng, *Public Health Reports (Washington, D.C.: 1974)*, vol. 126, no. 3, pp. 384–393, Jun. 2011, ISSN: 0033-3549.

[3] G. B. Anderson and M. L. Bell, "Lights out: Impact of the August 2003 power outage on mortality in New York, NY," *Epidemiology (Cambridge, Mass.)*, vol. 23, no. 2, pp. 189–193, Mar. 2012, ISSN: 1044-3983. DOI: `10.1097/EDE.0b013e318245c61c`. [Online]. Available: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3276729/` (visited on 04/23/2016).

[4] M. Lave, J. Kleissl, and J. Stein, "A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501–509, 2013, ISSN: 1949-3029. DOI: `10.1109/TSTE.2012.2205716`.

[5] A. Mills and R. Wiser, "Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power," Lawrence Berkeley National Labratory, Tech. Rep., 2010.

[6] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. Hoff, "Parameterization of site-specific short-term irradiance variability," *Solar Energy*, vol. 85, no. 7, pp. 1343–1353, Jul. 2011, ISSN: 0038-092X. DOI: `10.1016/j.solener.2011.03.016`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X11000995` (visited on 09/12/2013).

[7] A. Murata, H. Yamaguchi, and K. Otani, "A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area," *Electrical Engineering in Japan*, vol. 166, no. 4, pp. 9–19, Mar. 2009, ISSN: 04247760. DOI: `10.1002/eej.20723`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/eej.20723/abstract` (visited on 02/09/2011).

[8] J. Marcos, L. Marroyo, E. Lorenzo, and M. Garca, "Smoothing of PV power fluctuations by geographical dispersion," en, *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 2, pp. 226–237, Mar. 2012, ISSN: 1099-159X. DOI: `10.1002/pip.1127`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.1127/abstract` (visited on 02/17/2016).

[9] A. E. Curtright and J. Apt, "The character of power output from utility-scale photovoltaic systems," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 3, pp. 241–247, May 2008, ISSN: 10627995. DOI: `10.1002/pip.786`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.786/abstract` (visited on 02/09/2011).

[10] G.E. Energy, "Western Wind and Solar Integration Study," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/SR-550-47434, May 2010.

[11] D. Lew, G. Brinkman, E. Ibanez, B. M. Hodge, and J. King, "Western wind and solar integration study phase 2," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.7781&rep=rep1&type=pdf` (visited on 04/22/2015).

[12] Energy and Environmental Economics (E3), "Investigating a Higher Renewables Portfolio Standard in California," Tech. Rep., Jan. 2014. [Online]. Available: `https://www.ethree.com/documents/E3_Final_RPS_Report_2014_01_06_with_appendices.pdf`.

[13] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, Special Section: Transition Pathways to a Low Carbon Economy, vol. 52, pp. 213–234, Jan. 2013, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2012.08.062`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421512007446` (visited on 01/28/2015).

[14] S. Darby, "Smart metering: What potential for householder engagement?" *Building Research & Information*, vol. 38, no. 5, pp. 442–457, 2010.

[15] The Institute for Electric Efficiency, "Utility-sacle Smart Meter Deployments," Tech. Rep., Sep. 2014.

[16] Energy Information Administration, "An Assessment of Interval Data and Their Potential Application to Residential Electricity End- Use Modeling," Tech. Rep.

[17] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat: Using occupancy sensors to save energy in homes," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, ACM, 2010, pp. 211–224. [Online]. Available: `http://dl.acm.org/citation.cfm?id=1870005` (visited on 04/23/2016).

[18] L. Lutzenhiser, H. Hus, M. Moezzi, A. Levenda, and A. Woods, "Lifestyles, Buildings and Technologies: What Matters Most?" *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA*, pp. 12–17, 2012. [Online]. Available: `http://www.aceee.org/files/proceedings/2012/data/papers/0193-000034.pdf` (visited on 03/20/2014).

[19] A. Meier, "How people actually use thermostats," 2012. [Online]. Available: `http://escholarship.org/uc/item/3vd5q0cp.pdf` (visited on 03/28/2014).

[20] Pecan Street, "Setting the Benchmark for Non-Intrusive Load Monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation," Tech. Rep., Jan. 2015.

[21] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring," en, *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844–858, Oct. 2010, ISSN: 1530-9290. DOI: `10.1111/j.1530-9290.2010.00280.x`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9290.2010.00280.x/abstract` (visited on 03/18/2016).

[22] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in *2012 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2012, pp. 190–197. DOI: `10.1109/PerCom.2012.6199866`.

[23] B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, and I. H. Rowlands, "Disaggregating categories of electrical energy end-use from whole-house hourly data," *Energy and Buildings*, vol. 50, pp. 93–102, Jul. 2012, ISSN: 0378-7788. DOI: `10.1016/j.enbuild.2012.03.025`.

[Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0378778812001739` (visited on 02/06/2015).

[24] M. Wytock and J. Z. Kolter, "Contextually Supervised Source Separation with Application to Energy Disaggregation," *ArXiv:1312.5023 [cs, math, stat]*, Dec. 2013, arXiv: 1312.5023. [Online]. Available: `http://arxiv.org/abs/1312.5023` (visited on 10/10/2014).

[25] A. Albert and R. Rajagopal, "Thermal Profiling of Residential Energy Use," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–10, 2014, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2329485`.

[26] M. E. H. Dyson, S. D. Borgeson, M. D. Tabone, and D. S. Callaway, "Using smart meter data to estimate demand response potential, with application to solar energy integration," *Energy Policy*, vol. 73, pp. 607–619, Oct. 2014, ISSN: 0301-4215. DOI: `10.1016/j.enpol.2014.05.053`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0301421514003681` (visited on 09/12/2014).

[27] M. F. Fels, "Prism: An introduction," *Energy and Buildings*, vol. 9, no. 1, pp. 5–18, 1986.

[28] K. Kissock, J. S. Haberl, and D. E. Claridge, "Development of a toolkit for calculating linear, changepoint linear and multiple-linear inverse building energy analysis models. Final Report for ASHRAE Research Project, No. 1050-RP," *Energy Systems Laboratory Report No. ESLTR-02-11-02*, 2002. [Online]. Available: `http://repository.tamu.edu/handle/1969.1/2847` (visited on 12/11/2013).

[29] M. Tabone and D. Callaway, "Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–9, 2014, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2372751`.

[30] Solar Energy Industry Association, "Solar Market Insight Report 2013 Year in Review," Solar Energy Industry Association, Tech. Rep., Mar. 2014.

[31] *The {C}alifornia {S}olar {I}nitiative*, Apr. 2014. [Online]. Available: `http://www.gosolarcalifornia.org/about/csi.php` (visited on 04/25/2014).

[32] M. Rothleder, *Track I Direct Testimony Of Mark Rothleder On Behalf Of The California Independent System Operator Corporation*, California Public Utilities Commission, 2011.

[33] M. Hummon, P. Denholm, J. Jorgenson, D. Palchak, B. Kirby, and O. Ma, "Fundamental Drivers of the Cost and Price of Operating Reserves," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep., 2013. [Online]. Available: `http://www.nrel.gov/docs/fy13osti/58491.pdf` (visited on 02/28/2014).

[34] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," *Contract*, vol. 303, pp. 275–3000, 2013. [Online]. Available: `http://www.smartgridinformation.info/pdf/5506_doc_1.pdf` (visited on 04/22/2015).

[35] M. Milligan, P. Donohoo, D. Lew, E. Ela, B. Kirby, H. Holttinen, E. Lannoye, D. Flynn, M. OMalley, and N. Miller, "Operating reserves and wind power integration: An international comparison," in *Proc. 9th International Workshop on large-scale integration of wind power into power systems*, 2010, pp. 18–29. (visited on 02/28/2014).

[36] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Gomez Lazaro, and N. Detlefsen, "Methodologies to determine operating reserves due to increased wind power," in *2013 IEEE Power and Energy Society General Meeting (PES)*, Jul. 2013, pp. 1–10. DOI: `10.1109/PESMG.2013.6673067`.

[37] M. R. Hummon, E. Ibanez, G. Brinkman, and D. Lew, "Sub-Hour Solar Data for Power System Modeling from Static Spatial Variability Analysis," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep. Pending, 2012.

[38] M. Lave and J. Kleissl, "Cloud speed impact on solar variability scaling Application to the wavelet variability model," *Solar Energy*, vol. 91, pp. 11–21, May 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2013.01.023`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X13000406` (visited on 09/12/2013).

[39] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., and T. E. Hoff, "Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance," *Solar Energy*, Progress in Solar Energy 3, vol. 86, no. 8, pp. 2170–2176, Aug. 2012, ISSN: 0038-092X. DOI: `10.1016/j.solener.2012.02.027`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X12000928` (visited on 05/14/2014).

[40] M. J. Reno and J. Stein, "Using Cloud Classification to Model Solar Variability.," Sandia National Laboratories, Tech. Rep., 2013. [Online]. Available: `http://energy.sandia.gov/wp/wp-content/gallery/uploads/SAND-2013-1692C\_ASES-CloudCategoryVariability.pdf` (visited on 08/31/2013).

[41] J. Wegener, M. Lave, J. Luoma, and J. Kleissl, "Temporal downscaling of irradiance data via Hidden Markov Models on Wavelet coefficients: Application to California Solar Initiative data," 2012. [Online]. Available: `http://solar.ucsd.edu/datasharing/doc/UCSDReport{\_1}secCSI.pdf` (visited on 12/06/2012).

[42] R. Langrock, I. L. MacDonald, and W. Zucchini, "Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models," *Journal of Empirical Finance*, vol. 19, no. 1, pp. 147–161, Jan. 2012, ISSN: 0927-5398. DOI: `10.1016/j.jempfin.2011.09.003`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0927539811000661` (visited on 04/29/2014).

[43] Y.-z. Li, L. He, and R.-Q. Nie, "Short-Term Forecast of Power Generation for Grid-Connected Photovoltaic System Based on Advanced Grey-Markov Chain," in *International Conference on Energy and Environment Technology, 2009. ICEET '09*, vol. 2, Oct. 2009, pp. 275–278. DOI: `10.1109/ICEET.2009.305`.

[44] P. Poggi, G. Notton, M. Muselli, and A. Louche, "Stochastic study of hourly total solar radiation in Corsica using a Markov model," en, *International Journal of Climatology*, vol. 20, no. 14, pp. 1843–1860, Nov. 2000, ISSN: 1097-0088. DOI: `10.1002/1097-0088(20001130)20:14<1843::AID-JOC561>3.0.CO;2-O`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/1097-0088(20001130)20:14<1843::AID-JOC561>3.0.CO;2-O/abstract` (visited on 05/08/2014).

[45] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, vol. 79, no. 5, pp. 469–482, Nov. 2005, ISSN: 0038-092X. DOI: `10.1016/j.solener.2004.12.006`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X05000204` (visited on 05/08/2014).

[46] G. M. Masters, *Renewable and Efficient Electric Power Systems*, 1st ed. Wiley-IEEE Press, Aug. 2004, ISBN: 0471280607.

[47] Y. Makarov, C. Loutan, J. Ma, and P. de Mello, “Operational Impacts of Wind Generation on California Power Systems,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039–1050, May 2009, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2009.2016364`. [Online]. Available: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4808228` (visited on 06/07/2011).

[48] E. Ibanez, G. Brinkman, M. Hummon, and D. Lew, “A Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-Hourly Variability Analysis,” in *2nd International Workshop on Integration of Solar Power in Power Systems Proceedings, Lisbon, Portugal*, 2012. [Online]. Available: `http://www.nrel.gov/docs/fy12osti/56169.pdf` (visited on 09/02/2013).

[49] C. Goebel and D. Callaway, “Using ICT-Controlled Plug-in Electric Vehicles to Supply Grid Regulation in California at Different Renewable Integration Levels,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 729–740, Jun. 2013, ISSN: 1949-3053. DOI: `10.1109/TSG.2012.2218629`.

[50] J. Wu, A. Botterud, A. Mills, Z. Zhou, B.-M. Hodge, and M. Heaney, “Integrating solar PV (photovoltaics) in utility system operations: Analytical framework and Arizona case study,” *Energy*, vol. 85, pp. 1–9, Jun. 2015, ISSN: 0360-5442. DOI: `10.1016/j.energy.2015.02.043`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S036054421500198X` (visited on 12/19/2015).

[51] D. Corbus, J. King, T. Mousseau, *et al.*, “Eastern wind integration and transmission study,” *NREL (http://www. nrel. gov/docs/fy09osti/46505. pdf), CP-550-46505*, 2010. [Online]. Available: `http://www.nyiso.com/public/webdocs/markets_operations/committees/bic_miwg/meeting_materials/2010-02-17/EWITSNYISO.pdf` (visited on 04/22/2015).

[52] X. Bai, K. Clark, G. A. Jordan, N. W. Miller, and R. J. Piwko, “Intermittency Analysis Project: Appendix B Impact of Intermittent Generation on Operation of California Power Grid,” California Energy Commission, Tech. Rep., 2007.

[53] California Public Utilities Commission, “Planning Assumptions & Scenarios For The 2016 Long Term Procurement Plan Proceeding and The CAISO 201617 Transmission Planning Process,” California Public Utilities Commission, Tech. Rep. R.13-12-010 MP6/avs, May 2016.

[54] L. M. Hinkelman, “Differences between along-wind and cross-wind solar irradiance variability on small spatial scales,” *Solar Energy*, vol. 88, pp. 192–203, Feb. 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2012.11.011`.

[Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X12004021` (visited on 05/14/2014).

[55] C. M. S. Martin, J. K. Lundquist, and M. A. Handschy, "Variability of interconnected wind plants: Correlation length and its dependence on variability time scale," en, *Environmental Research Letters*, vol. 10, no. 4, p. 044 004, Apr. 2015, ISSN: 1748-9326. DOI: `10.1088/1748-9326/10/4/044004`. [Online]. Available: `http://iopscience.iop.org/1748-9326/10/4/044004` (visited on 04/08/2015).

[56] M. Lave, J. Stein, J. Kleissl, A. Ellis, C. Hansen, and Y. Miyamoto, *Simulating Solar Power Plant Variability for Grid Studies: A Wavelet-based Variability Model (WVM).* Maui, HI, Oct. 2011.

[57] M. D. Tabone and D. S. Callaway, "Parameterizing Fluctuations in Solar Photovoltaic Generation Using Hidden Markov Models," in *Proceedings of the 2013 IEEE Power & Energy Society General Meeting*, Vancouver, BC, Jul. 2013.

[58] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "Power output fluctuations in large scale pv plants: One year observations with one second resolution and a derived analytic model," en, *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 2, pp. 218–227, Mar. 2011, ISSN: 1099-159X. DOI: `10.1002/pip.1016`. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1002/pip.1016/abstract` (visited on 02/17/2016).

[59] M. Tabone and D. Callaway, "Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2965–2973, Nov. 2015, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2372751`.

[60] M. Milligan, B. Kirby, J. King, and S. Beuning, "Operating reserve implication of alternative implementations of an energy imbalance service on wind integration in the Western Interconnection," in *2011 IEEE Power and Energy Society General Meeting*, Jul. 2011, pp. 1–8. DOI: `10.1109/PES.2011.6039622`.

[61] G. C. Wu, "Incorporating land-use requirements and environmental constraints in low-carbon electricity planning for California," *Environmental science & technology*, 2014. [Online]. Available: `http://pubs.acs.org/doi/abs/10.1021/es502979v` (visited on 01/29/2015).

[62] R. R. Hernandez, M. K. Hoffacker, and C. B. Field, "Land-Use Efficiency of Big Solar," *Environmental Science & Technology*, vol. 48, no. 2, pp. 1315–1323, Jan. 2014, ISSN: 0013-936X. DOI: `10.1021/es4043726`. [Online]. Available: `http://dx.doi.org/10.1021/es4043726` (visited on 10/28/2015).

[63] O. Bartholomy, T. Vargas, M. Simone, C. Hansen, S. Fitchett, and A. Pohl, "Benchmarking solar power and irradiance forecasting accuracy at Sacramento Municipal Utility District," in *Photovoltaic Specialist Conference (PVSC), 2014 IEEE 40th*, Jun. 2014, pp. 0063–0068. DOI: `10.1109/PVSC.2014.6925196`.

[64] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renn, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, vol. 84, no. 12, pp. 2161–2172, Dec. 2010, ISSN: 0038-092X. DOI: `16/j.solener.2010.08.014`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0038092X10002823` (visited on 08/29/2011).

[65] A. Olson, R. A. Jones, E. Hart, and J. Hargreaves, "Renewable Curtailment as a Power System Flexibility Resource," *The Electricity Journal*, vol. 27, no. 9, pp. 49–61, 2014. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1040619014002322` (visited on 11/23/2015).

[66] J. H. Nelson and L. M. Wisland, "Achieving 50 Percent Renewable Electricity in California," Union of Concerned Scientists, Oakland, CA, Tech. Rep., Aug. 2015. [Online]. Available: `http://www.ucsusa.org/clean-energy/california-and-western-states/achieving-50-percent-renewable-energy-in-california` (visited on 09/03/2015).

[67] L. Bird, J. Cochran, and X. Wang, "Wind and solar energy curtailment: Experience and practices in the United States," *National Renewable Energy Laboratory2014*, 2014. [Online]. Available: `http://www.nrel.gov/docs/fy14osti/60983.pdf` (visited on 11/23/2015).

[68] J. Zhang, B.-M. Hodge, A. Florita, S. Lu, H. F. Hamann, and V. Banunarayanan, "Metrics for evaluating the accuracy of solar power forecasting," in *3rd International workshop on integration of solar power into power systems, London, England*, 2013. [Online]. Available: `http://www.nrel.gov/docs/fy14osti/60142.pdf` (visited on 09/09/2015).

[69] V. P. A. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Solar Energy*, vol. 97, pp. 58–66, Nov. 2013, ISSN: 0038-092X. DOI: `10.1016/j.solener.2013.08.002`. [Online]. Available: `http:`

//www.sciencedirect.com/science/article/pii/S0038092X13003125 (visited on 05/14/2014).

[70] M. A. Cohen and D. S. Callaway, "Physical Effects of Distributed PV Generation on California's Distribution System," *ArXiv:1506.06643 [physics]*, Jun. 2015, arXiv: 1506.06643. [Online]. Available: `http://arxiv.org/abs/1506.06643` (visited on 09/17/2015).

[71] A. Todd, *Insights from Smart Meters: Identifying Specific Actions, Behaviors and Characteristics that drive savings in Behavior-Based Programs*, 2014.

[72] A. Albert, T. Gebru, J. Ku, J. Leskovec, and R. Rajagopal, "Drivers of variability in energy consumption," in *Submitted for review to European Conference on Machine Learning*, 2013. [Online]. Available: `http://www.stanford.edu/~adalbert/papers/drivers_variability.pdf` (visited on 03/20/2014).

[73] A. Todd, *Insights from Smart Meters: The Potential for Peak-Hour Savings from Behavior-Based Programs*, 2014.

[74] K. Basu, V. Debusschere, and S. Bacha, "Residential appliance identification and future usage prediction from smart meter," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov. 2013, pp. 4994–4999. DOI: `10.1109/IECON.2013.6699944`.

[75] J. St. John, *Putting Energy Disaggregation Tech to the Test : Greentech Media*, Nov. 2013. [Online]. Available: `http://www.greentechmedia.com/articles/read/putting-energy-disaggregation-tech-to-the-test` (visited on 07/22/2015).

[76] ——, *Perfecting Energy Disaggregation in the Home : Greentech Media*, Nov. 2013. [Online]. Available: `http://www.greentechmedia.com/articles/read/perfecting-energy-disaggregation-in-the-home` (visited on 07/22/2015).

[77] K. X. Perez, W. J. Cole, J. D. Rhodes, A. Ondeck, M. Webber, M. Baldea, and T. F. Edgar, "Nonintrusive disaggregation of residential air-conditioning loads from sub-hourly smart meter data," *Energy and Buildings*, vol. 81, pp. 316–325, Oct. 2014, ISSN: 0378-7788. DOI: `10.1016/j.enbuild.2014.06.031`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0378778814005131` (visited on 02/18/2016).

[78] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, “Nilmtk: An open source toolkit for non-intrusive load monitoring,” in *Proceedings of the 5th international conference on Future energy systems*, ACM, 2014, pp. 265–276. [Online]. Available: `http://dl.acm.org/citation.cfm?id=2602051` (visited on 03/08/2016).

[79] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, “Unsupervised Disaggregation of Low Frequency Power Measurements.,” in *SDM*, vol. 11, SIAM, 2011, pp. 747–758. [Online]. Available: `http://epubs.siam.org/doi/pdf/10.1137/1.9781611972818.64` (visited on 06/22/2015).

[80] R. Fazeli, M. Ruth, and B. Davidsdottir, “Temperature response functions for residential energy demand A review of models,” *Urban Climate*, vol. 15, pp. 45–59, Mar. 2016, ISSN: 2212-0955. DOI: `10.1016/j.uclim.2016.01.001`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S2212095516300013` (visited on 01/25/2016).

[81] R. Langrock, T. Kneib, A. Sohn, and S. L. DeRuiter, “Nonparametric inference in hidden Markov models using P-splines,” *Biometrics*, 2015. [Online]. Available: `http://onlinelibrary.wiley.com/doi/10.1111/biom.12282/full` (visited on 01/18/2015).

[82] M. C. Jones, J. S. Marron, and S. J. Sheather, “A brief survey of bandwidth selection for density estimation,” *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996. [Online]. Available: `http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1996.10476701` (visited on 04/29/2016).

[83] A. D. Mills and J. Seel, “Flexibility Inventory for Western Resource Planners,” Oct. 2015, <p>Lawrence Berkeley National Laboratory hosted a webinar on November 4, 2015 entitled ”Flexibility Inventory for Western Resource Planners.”</p><p>To view a recording of the webinar, click here.</p>.

[84] National Renewable Energy Laboratory, *Solar Maps*. [Online]. Available: `http://www.nrel.gov/gis/solar.html` (visited on 08/31/2013).

Appendix A

The effect of PV siting on power system flexibility needs

A.1 Locational scenarios appendix

Dispersion test scenarios

Dispersion tests are intended to more precisely measure the effects of inter-plant distance and climate on the variability and uncertainty of utility-scale PV. These scenarios consist of 100, 120 MW PV systems randomly located in areas of different size – 50 km, 100 km, 200 km, and 300 km squares. Systems in smaller areas will have shorter inter-plant distances than those in larger areas.

We also look at each of these area sizes in one climate region that is favorable for PV and one that is not. We define how favorable a climate is by three metrics: (1) frequency of sub-hourly cloud volatility, (2) frequency of hourly cloud volatility, and (3) expected annual energy production. To achieve adequate dispersion we removed the ecological constraints applied in the utility-scale scenarios. Because we relax the ecological constraints, these scenarios are meant to be illustrative but not realistic.

Figure A.1 displays the three climate metrics for all areas in California. Overlaying these maps are outlines of the eight areas used for the dispersion test scenarios. The expected annual energy production for 10km grid cells in California is identified by the National Renewable Energy Laboratory and published online [84]. We define heuristics for hourly and sub-hourly cloud variability in this paper, and use them as inputs to our model, as explained in Appendix A.4. These heuristics identify 5 classes of cloud volatility for 1km grid cells on an hourly basis for hourly volatility and on a 30-min basis for sub-hourly; where 1 is the least volatile class and 5 is the most. The percentage of time spent in classes 4 or 5 defines the frequency of hourly

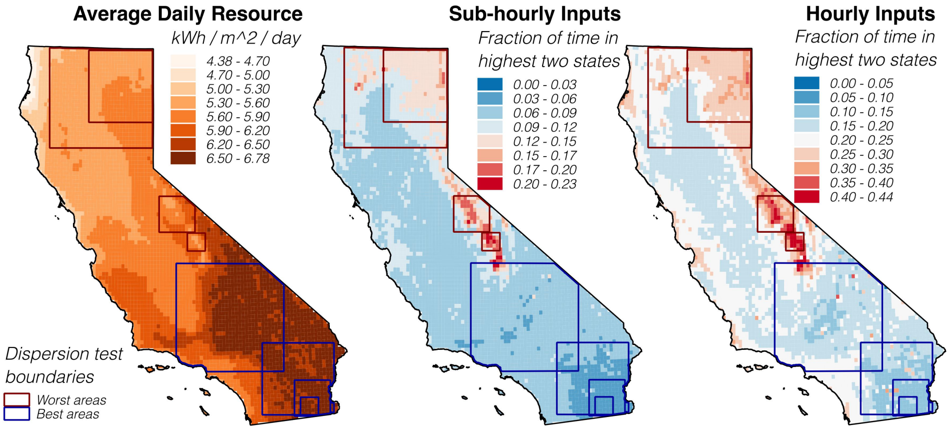

Figure A.1: Availability and variability of the solar resource in California; (left) average kWh of irradiance per day on an optimally tilted, 100% efficient PV panel [84], (middle) fraction of daylight time spent in the two most volatile (of 5) input states for the sub-hourly variability models, (left) fraction of daylight time spent in the two most volatile (of 5) input states for the hourly variability model. Red and blue boundaries outline areas for the utility-scale dispersion tests.

or sub-hourly cloud variability in a grid cell.[1]

To choose the "best" and "worst" climates we created an aggregate climate metric by normalizing each of the three metrics to the mean for all of the grid cells, and then subtracting normalized variability from normalized energy production. Eq. (A.1) displays the process of calculating the aggregate climate metric for grid cell i, C_i; where E_i, H_i, and S_i are the expected annual energy production, the frequency of hourly cloud volatility, and the frequency of sub-hourly cloud volatility for grid cell i respectively. The best areas were those for which the average climate metric is the greatest (excluding any area that isn't at least 90% covered by California land area). Table A.1 displays the average values for the climate metrics in each of the selected areas.

$$C_i = E_i \frac{N}{\sum_{j=1}^{N} E_j} - H_i \frac{N}{\sum_{j=1}^{N} H_j} - S_i \frac{N}{\sum_{j=1}^{N} S_j} \tag{A.1}$$

---

[1] Hourly and sub-hourly cloud variability heuristics are technically available for 1km grid cells, we instead define them here for 10km gird cells by using the 1km cell in the center of each 10km cell to calculate the heuristics. This is meant only as a rough estimate, and is only used to make siting decisions for the distribution test scenarios.

Table A.1: Climate conditions in the dispersion test cases

Climate	Area Size	Ave. energy production (kW/m^2/day)	Freq. of sub-hourly variability	Freq. of hourly variability
Best	50km	6.6	0.046	0.128
	100km	6.6	0.049	0.138
	200km	6.5	0.057	0.156
	300km	6.4	0.071	0.188
Worst	50km	5.8	0.172	0.376
	100km	5.8	0.138	0.313
	200km	5.6	0.122	0.267
	300km	5.5	0.111	0.242

Rooftop scenario

The rooftop scenario represents a generally realistic arrangement for 12GW of rooftop PV in California, but is by no means a prediction for precisely where rooftop PV will be located. We expect that even significant changes in the concentrations and locations of PV within census tracts will have little affect on our results of the rooftop scenario.

We use data from the 2012 American Communities Survey (ACS) to estimate density of residential structures in each census tract in California. The ACS surveys samples of households throughout the US, then uses the survey data in combination with census data to estimate statistics of housing and population demographics. Data provided by the ACS are the number of housing units of each "structure type" in a census tract. Structure type is defined by the number of housing units in the structure; e.g., 1000 single family homes, 100 units in structures with 2-3 units, 200 units in structures with 4-5 units, etc. To estimate the total number of structures of each type in a tract, we divide the number of units of each type in the tract by the median number of units per structure of that type. For example, if ACS states that a tract contains 100 housing units in structures containing 6-10 units, we estimate there to be 12.5 ($= \frac{100}{8}$) structures with 6-10 units in the tract. We then divide the estimated number of structures (of all types) in each tract by the land area of the tract to find the "structure density."

Table A.2 describes how 5kW PV systems are distributed throughout California

Table A.2: Placement of PV systems in the rooftop scenario. The structure density of a census tract defines how many grid cells will contain PV at all, and how many PV systems will be placed into grid cells that do contain PV.

Struct. density (Struct./km^2)	Area covered by tracts (km^2)	Percent of housing units in tracts	Percent of area with any PV	Sys. in grid cells with PV	Total PV capacity in tracts
$(1030, 3606]$	1,000	15 %	100%	600	3.1 GW
$(602, 1030]$	4,000	33 %	75%	469	7.0 GW
$(299, 602]$	5,000	23 %	25%	200	1.3 GW
$(10, 299]$	40,000	23 %	3%	100	0.6 GW
$(0, 10]$	374,000	5 %	-	-	0 GW

in the rooftop scenario based on structure density. To understand this process, it is important to note that the geographic unit of analysis for our simulation is 1km^2 area grid cells. For each tract, we include PV in only a fraction of randomly selected grid cells—this greatly reduces the computational costs of the simulation. We also place a different number of PV systems in each grid cell that contains PV. As shown in Table A.2 both the fraction of grid cells assigned any PV, and the number of systems located in a grid cell—once assigned—are defined by the structure density of the census tract. Grid cells in high density census tracts are both more likely to contain PV, and contain more PV per unit area. We note that at no point do we assign more PV systems than the estimated number of structures in a grid cell.

Systems are homogeneously sized at 5kW of rated AC capacity. The tilt and azimuth of systems vary with the same distribution as those installed by SolarCity—SolarCity provided us with the tilts and azimuths for each rooftop mount they own and operate. Locations of systems within a cell are random.

A.2 Derivation of variability and uncertainty from reserve schedules

Prior to the adoption of the Energy Imbalance Market, the CAISO contracted generators to follow an hourly schedule, where the amount of energy contracted each hour is finalized 80 minutes in advance of an operating hour. 20 minute ramping periods at the edge of each hour allow generators time to move from one contracted amount to another. Forecasts of net-load will inevitably be incorrect. We denote $\epsilon_{i,a}$ as a vector of forecast errors from predictions of net-load averaged over an i-minute interval made a-minute prior to the center of the interval.

Eq. (A.2) mathematically decomposes the hourly schedule into the actual net-load time-series, NL, and the time series of forecast errors in this hourly schedule $\epsilon_{60,80}$.

$$HAS_t = \mathcal{R}_t^{60,20}(NL + \epsilon_{60,80}) = \mathcal{R}_t^{60,20}(NL) + \mathcal{R}_t^{60,20}(\epsilon_{60,80}), \tag{A.2}$$

where $\mathcal{R}_t^{i,j}$ is a linear function that averages a time-series over i-minutes and adds j-minute ramps between intervals (see Figure 3.3).[2]

Within an operating hour, the system is redispatched every 5 minutes to minimize operating costs; we refer to this redispatch schedule as a *real time schedule* RTS_t, depicted in Figure 3.3 (middle). We define RTS_t mathematically as follows – in a similar manner to HAS_t:

$$RTS_t = \mathcal{R}_t^{5,5}(NL + \epsilon_{5,5}) = \mathcal{R}_t^{5,5}(NL) + \mathcal{R}_t^{5,5}(\epsilon_{5,5}), \tag{A.3}$$

We now can define load following reserves in Eq. (A.4), where load following reserves provide the difference between the real time schedule and the hour ahead schedule:

$$\begin{aligned} LF_t &= RTS_t - HAS_t \\ &= \mathcal{R}_t^{5,5}(NL) + \mathcal{R}_t^{5,5}(\epsilon_{5,5}) - \mathcal{R}_t^{60,20}(NL) - \mathcal{R}_t^{60,20}(\epsilon_{60,80}) \\ &= \eta_{5,60,t} + \mathcal{R}_t^{5,5}(\epsilon_{5,5}) - \mathcal{R}_t^{60,20}(\epsilon_{60,80}) \end{aligned} \tag{A.4}$$

where we define $\eta_{5,60,t} = \mathcal{R}_t^{5,5}(NL) - \mathcal{R}_t^{60,20}(NL)$ as the 5 minute variability the system operator would face even if forecasts were perfect.

Analogously regulation reserves make up the difference between RTS_t and actual net load:

$$\begin{aligned} Reg_t &= NL_t - RTS_t \\ &= NL_t - \mathcal{R}_t^{5,5}(NL) - \mathcal{R}_t^{5,5}(\epsilon_{5,5}) \\ &= \eta^{1,5,t} - \mathcal{R}_t^{5,5}(\epsilon_{5,5}) \end{aligned} \tag{A.5}$$

---

[2] In CAISO $i = 60$ and $j = 20$ because hourly schedules are constant for the middle 40 minutes of each hour and ramps for the following 20 minutes to the next hour's forecasted net-load.

where $\eta^{1,5,t} = NL_t - \mathcal{R}_t^{5,5}(NL)$ is the sub-5 minute variability the system operator would face if the forecast were perfect. We use 1 in the first subscript to reflect the fact that we will be working with 1 minute time scale data.

Adding variability and uncertainty from load, wind, and solar.

We take it as given that the forecast errors for net load will be a composite of the errors from all forecasts at that time, shown in (A.6) and (A.7). There are many loads, wind farms, and solar plants that may have separate forecasts, and this property can be further applied by stating that the total forecast error for all solar is the sum of forecast errors from each solar farm, and so on.

$$\epsilon_{60,60,t} = \epsilon_{60,60,t}^{load} - \epsilon_{60,60,t}^{wind} - \epsilon_{60,60,t}^{solar} \tag{A.6}$$

$$\epsilon_{5,5,t} = \epsilon_{5,5,t}^{load} - \epsilon_{5,5,t}^{wind} - \epsilon_{5,5,t}^{solar} \tag{A.7}$$

Deviations are defined as the difference between schedules applied at two time scales given a perfect forecast (i.e. no forecast error.). We define 5 minute deviations for net-load in (A.8). Ramp operators (e.g., $\mathcal{R}_t^{5,5}$) are linear, and thus they can be distributed to each component of net load. We apply this property in (A.9) and (A.10), where L, S, and W represent load, wind and solar respectively – the components of net-load. Rearranging terms in (A.11) and (A.12) we show that the 5-min deviations of net load are a composition of the 5-min deviations of each component of net load. Again, there are many loads, solar farms and wind farms in California's system, and it is readily apparent that this property hold no matter how many components we use for net-load.

$$\begin{aligned}
\eta_{5,60,t} &= \mathcal{R}_t^{5,5}(NL) - \mathcal{R}_t^{60,20}(NL) && \text{(A.8)}\\
&= \mathcal{R}_t^{5,5}(L-S-W) - \mathcal{R}_t^{60,20}(L-S-W) && \text{(A.9)}\\
&= \big(\mathcal{R}_t^{5,5}(L) - \mathcal{R}_t^{5,5}(S) - \mathcal{R}_t^{5,5}(W)\big) - \big(\mathcal{R}_t^{60,20}(L) - \mathcal{R}_t^{60,20}(S) - \mathcal{R}_t^{60,20}(W)\big) && \text{(A.10)}\\
&= \big(\mathcal{R}_t^{5,5}(L) - \mathcal{R}_t^{60,20}(L)\big) - \big(\mathcal{R}_t^{5,5}(S) - \mathcal{R}_t^{60,20}(S)\big) - \big(\mathcal{R}_t^{5,5}(W) - \mathcal{R}_t^{60,20}(W)\big) && \text{(A.11)}\\
&= \eta_{5,60,t}^{load} - \eta_{5,60,t}^{solar} - \eta_{5,60,t}^{wind} && \text{(A.12)}
\end{aligned}$$

A.3 Modeling reserve requirements from PV.

We simulate regulation and load following reserve requirements using an updated version of our volatility state model, which is described, fit and validated in [59]. We then combine the variability and uncertainty from PV with those from wind and load, and use the result to calculate reserve procurement to follow net-load. Figure A.2 outlines each stage of our simulation method, along with the corresponding sections of this paper that explain the stage.

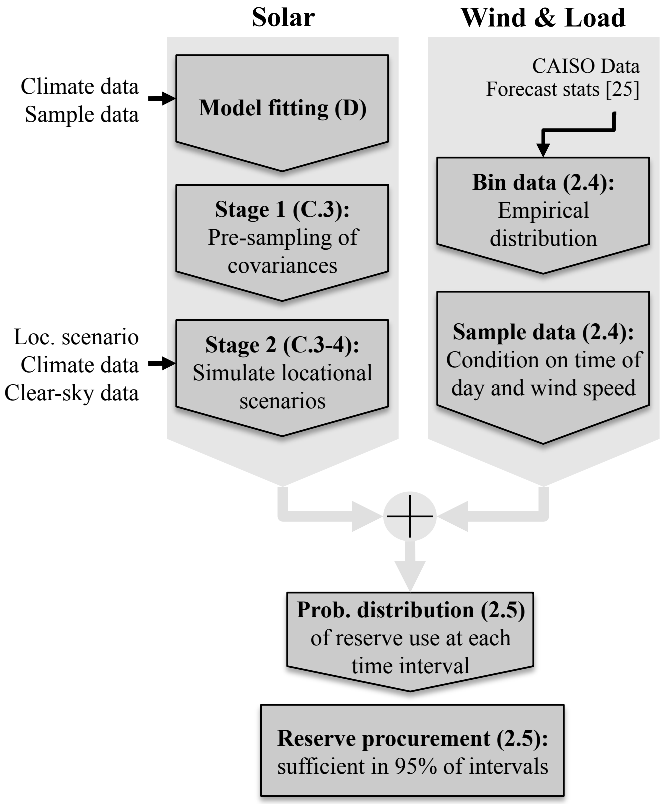

Figure A.2: Process diagram for simulation. Sections describing each stage of the simulation are included in parentheses.

Volatility state model

The volatility state model uses mixtures of spatially correlated Gaussian random variables to describe each of the four variability/uncertainty metrics from a set of N geo-located PV systems. Mixture coefficients—in this work referred to as "volatility states"—define both the variance of variability/uncertainty at a given time and the parameters of exponential spatial autocorrelation. Our earlier work [59] presents fits and validates the volatility state model, and portions of this section are reproduced from our original paper verbatim.

Figure A.3 depicts the model as a directed acyclic graph, where shaded nodes represent observed variables and unshaded nodes represent unobserved, latent variables.[3] We estimate separate model for each variability or uncertainty metric. $y(t) \in \mathbb{R}^{N_s}$, represents normalized variability or uncertainty of N_s systems, i.e. either $\eta_{1,5}$, $\epsilon_{5,5}$, $\eta_{5,60}$, or $\epsilon_{60,60}$. $x(t) \in \{1, 2, ...5\}^{N_g}$ represents the inputs to the model, which in our case are volatility heuristics for each of N_g grid cells encompassing the PV systems. $v(t) \in \{1, 2, ...M\}^{N_s}$ is a vector of unobserved volatility states, i.e. each system at each time is in one of M volatility states.

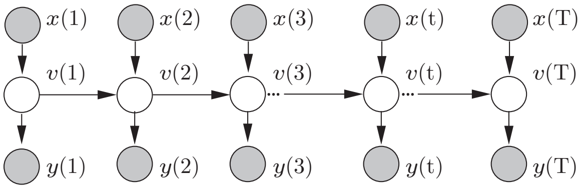

Figure A.3: Hidden Markov model for stochastic volatility represented as a directed acyclic graph

Eq. (A.13) defines $y(t)$ as the normalization of a variability or uncertainty metric for the modeled set of PV systems, shown for 1 minute deviations but equivalent for any of the four metrics. $\mathbb{S}$ is a matrix containing the generation from each observed PV system at each observed time, and $\eta_{1,5,t}(\cdot)$ is an operator returning an $N \times 1$ vector containing the 1 minute deviation at time t from each of the N systems contained in $\mathbb{S}$. $\eta_{1,5,t}(\mathbb{CL})$ is the 1 minute deviation given clear sky signals, $\mathbb{CL}$, instead of actual generation signals. Eq. (A.14) defines M_t as a diagonal normalization matrix

---

[3] Figure A.3 differs from the directed acyclic graph presented in Figure 3.4. In Figure A.3 volatility states are a Markov process where states at time t are dependent on the states in time $t-1$. For fitting the model, it is important that we include this dependence when estimating the latent states. This time dependence is not important for simulation, for which the model predicts times that are far into the future and the chain is well mixed. As a result we can use the stationary distribution the above Markov process, for which there is no dependence on prior states.

where the i^{th} element is the maximum of the clear sky signal in the hour containing t. Variability resulting from the clear sky signal is subtracted during normalization, and the M matrix normalizes variability or uncertainty so that it is expected to be the same magnitude throughout the day.

$$y(t) = M_t^{-1}(\eta_{1,5,t}(\mathbb{S}) - \eta_{1,5,t}(\mathbb{CL})) \tag{A.13}$$

$$M_{t,ii} = \max_{j \in hout(t)} \mathbb{CL}_{i,j} \tag{A.14}$$

Eq. (A.15) defines the distribution of $y(t)$ as a multivariate Gaussian. Eq. (A.16) defines the covariance matrix of $y(t)$ to be dependent on the volatility state of each system, $v_i(t)$. Volatility states define the variance of the systems as well as the parameters of the spatial correlation function. Eq. (A.17) describes this spatial correlation function; where i and j denote the volatility states of the two systems. $a_{i,j}$ defines the correlation between systems separated by a distance of 0; $a_{i,j}$ can be less than one in order to account for the fact the systems reported as having the same location are actually a small but unknown distance apart. $\tau_{i,j}$ is an exponential decay rate, in units of kilometers.

$$y(t) \sim MVG\left(0, \mathbf{\Sigma}(v(t), d; \phi)\right) \tag{A.15}$$

$$\mathbf{\Sigma}_{ij}(v(t); \sigma^2, \phi) = \begin{cases} \sigma^2_{v_{i(t)}} & i = j \\ \sigma_{v_{i(t)}}\sigma_{v_{j(t)}}\rho(v_i(t), v_j(t); d_{i,j}, \phi) & i \neq j \end{cases} \tag{A.16}$$

$$\rho(m, n; d_{i,j}, \phi) = a_{m,n} \cdot \exp\left\{-d_{i,j}/\tau_{m,n}\right\} \tag{A.17}$$

The probability of being in a volatility state is conditionally dependent on the volatility state at the previous time step and on the input heuristic from the grid cell containing the system, $x_g(t)$ where g indexes the grid cell. Eq. (A.18) shows a set of Markov chain transition matrices that govern the progression of the volatility state for each system; $\mathbf{A}^{(k)} \in \mathbb{R}^{M \times M}$, k indexes the input heuristic. Eq. (A.19) describes each matrix element. For simulation purposes we sample volatility state from the stationary probabilities of these transition matrices given the input state.

$$\mathbb{A} = \{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots \mathbf{A}^{(K)}\} \tag{A.18}$$

$$\mathbf{A}^{(k)}_{m,n} = P(v_i(t) = n | v_i(t-1) = m, x_g(t) = k) \tag{A.19}$$

Direct simulation is computationally infeasible

Simulating the sum of forecast errors or deviations from all PV in a balancing area is completed by summing a de-normalized y vector, which contains one element for each of the N systems in the area. Eqs. in (A.20) describe the distribution for the total 1-minute deviations of PV in a balancing area however could be equivalently written for any of the metrics. Eq. A.20a defines the variable S to be the sum of generation from all systems contained in $\mathbb{S}$ at each time. Eq. A.20b shows the de-normalization and summing of the vector y to obtain 1-minute deviations at time t for the sum of generation from all systems. Eq. A.20c shows expected value of this sum, which is dependent only on the clear sky signal because $E[y] = 0$. Eq. A.20d shows variance of this sum, which is dependent only on covariance matrix of y because the clear sky signal is non-random.

$$S_t = \sum_{i=1}^{N} \mathbb{S}_{i,t} \tag{A.20a}$$

$$\eta_{1,5,t}(S) = \sum_{i=1}^{N} \eta_{1,5,t}(\mathbb{S}_i) = u'(M_t' y_i + \eta_{1,5,t}(\mathbb{CL})) \tag{A.20b}$$

$$\mathbb{E}\left[\eta_{1,5,t}(S)\right] = u'\eta_{1,5,t}(\mathbb{CL}) \tag{A.20c}$$

$$Var\left(\eta_{1,5,t}(S)\right) = u' M_t' \Sigma(v(t), d; \phi) M_t u \tag{A.20d}$$

The variance of $\eta_{1,5,t}(S)$ is a random variable, resulting from its dependence on the volatility states of each of the N systems at time t, $v(t)$. Sampling from the distribution of $\eta_{1,5,t}(S)$ can be completed by the following process

- Sample a set of volatility states for each system in the balancing area given the climate data $x(t)$.
- Use (A.20d) to calculate the variance given these volatility states.
- Use (A.20c) to calculate the mean.
- Draw a sample from a normal distribution with the calculated mean and variance.

While it is theoretically possible to sample in this manner, it is computationally infeasible. The covariance matrix contains N^2 elements which need to be calculated and summed for each sampled value, and multiple samples are desired per interval. For a study period of one year, intervals of 30 minutes, 2.4 million systems, 200 samples per interval this leads to sampling and summing on the order of 10^{19} covariance matrix elements!

Two Stage Simulation

To avoid having to sample the entire $N \times N$ covariance matrix at each time interval, we create a simulation method that runs in two stages.

The premise of the two stage method is that we can reduce the $N \times N$ covariance matrix into a smaller number of identically distributed intermediate components that have a finite (and tractable) number of possible input conditions. We simulate a conditional distribution for these components in Stage 1, and sample from this distribution in Stage 2. In our case the intermediate component is an element of the covariance matrix aggregated to the resolution of 1 km^2 grid cells—as opposed to the final result, which is aggregated for the entire balancing area. Stage 1 samples 200 values of the aggregate covariance between pairs of cells under every possible set of input conditions. The result is a distribution of equally likely covariances conditional on system arrangements and input data. Stage 2 then uses the results from Stage 1 to sample a more tractable, $G \times G$, covariance matrix using spatial and climate data specific to each locational scenario; where G is the number of grid cells contained PV the locational scenario.

Table A.3 defines five possible arrangements of systems within a 1 km^2 grid cell. The locational scenarios presented in Section 3.2 define the vector a where a_m is the arrangement of systems in grid cell m.

Eqs. (A.21a) - (A.21d) show this two stage decomposition mathematically. Eq. (A.21a) shows the variance of $\eta_{1,5,t}(S)$ calculated with an intermediate resolution of 1 km^2 grid cells; where H is an $N \times G$ assignment matrix for which $H_{n,g}$ equals 1 if system n is in grid cell g and 0 otherwise, and u is now a $G \times 1$ vector of ones. Thus $H'M'\Sigma MH$ is the lower resolution, $G \times G$, covariance matrix.

Eq. (A.21b) applies *Approximation 1*—explained below—which allows the normalization constants, M_t, to be moved outside the $G \times G$ matrix. Eq (A.21c) applies *Approximation 2*—also explained below—which describes the distance matrix in terms the arrangement of systems in each grid cell, a, and binned distances between grid cells, d^G. Finally, Eq. (A.21d) replaces the $G \times G$ covariance matrix with the intermediate random variable R. Stage 1 simulates a conditional distribution of each element of R conditional on $x(t)$, a, and d^G. Stage 2 samples 10 realizations of R for each time and locational scenario.

$$\begin{aligned}
var(\eta_{1,5,t}(S)) &= u'H'M_t'\Sigma(v(t)|x(t),d;\phi)M_tHu && \text{(A.21a)}\\
&\approx u'm_t'K'\Sigma(v(t)|x(t),d;\phi)Km_tu && \text{(A.21b)}\\
&\approx u'm_t'K'\Sigma(v(t)|x(t),a,d^G;\phi)Km_tu && \text{(A.21c)}\\
&\approx u'm_t'R(x(t),a,d^G)m_tu && \text{(A.21d)}
\end{aligned}$$

(Approxomation 1)

To the reduce number of conditions that need to be sampled in Stage 1, it is helpful to approximate the normalization values, M_t, as something that can be moved outside of the $G \times G$, covariance matrix. M_t contains the hourly maxima of the clear sky signals for each system at time t, it depends on location, system geometry, and time.

Eq. (A.22) defines *Approximation 1* in which we approximate each system's clear sky generation to be a fixed proportion of the aggregate clear sky signal from all system is a grid cell. In Eq. (A.22), K is an $N \times G$ column stochastic matrix for which $K_{n,g}$ defines the fraction of capacity that system n contributes to total capacity contained in cell g. m_t is a $G \times G$ diagonal matrix where $m_{g,g}$ is the hourly maximum of the clear sky signal from all PV systems in grid cell g.

This approximation is identical to the truth only when all PV systems in a grid cell have the same location and geometry; however we do not expect the relaxation to greatly affect our results. Locations of systems cannot vary significantly within a grid cell, and geometries of systems are identical in the utility-scale scenarios. In the rooftop scenarios, geometries vary, but not much.

$$M_tH \approx Km_t \tag{A.22}$$

(Approximation 2)

Eq. (A.23) defines *Approximation 2* in which the distance matrix between cells m and n is dependent only on the arrangement of systems, $\{a_m, a_n\}$, within each cell and the distance between cells, $d^G_{m,n}$. The distance between cells is only applied in the horizontal direction when in reality it can be in any direction. For the following derivation, Let $p_x^{(i)}$, and $p_y^{(i)}$] be horizontal and vertical positions of each system in arrangement i relative to the center of the cell, let $d^G_{m,n}$ be the distance between the centers of cells m and n, and let $\hat{d}^{(m,n)}$ be a matrix containing the actual distances between systems in cell m from systems in cell n.

$$\hat{d}^{(m,n)}_{i,j} \approx \sqrt{\left(p^{(a_m)}_{x,i} - p^{(a_n)}_{x,j} + d^G_{m,n}\right)^2 + \left(p^{(a_m)}_{y,i} - p^{(a_n)}_{y,j}\right)^2} \tag{A.23}$$

Table A.3: Pre-defined arrangements of systems within 1 km grid cells

	Name	# of Sys.	Sys. Cap	Placement	Total Cap
1	Utility-scale	6000	5 KW	Grid	30 MW
2	Rooftop 1	600	5 KW	Random	3 MW
3	Rooftop 2	469	5 KW	Random	2.3 MW
4	Rooftop 3	200	5 KW	Random	1 MW
5	Rooftop 4	100	5 KW	Random	500 kW

The purpose of accounting for distance while ignoring direction is to reduce the possible number of conditions under which we must simulate R. We do not expect this relaxation to significantly affect our results because systems are placed either symmetrically or randomly within each grid cell, thus there is no reason to believe that vertical displacement will differ much from horizontal. Diagonal displacement will be systematically different from horizontal, however this effect is reduced at longer distances. As shown in our results, aggregate variability is driven by hourly forecast errors which has long decay rate relative to the size of a cell, thus at short distances, errors in the assumed distance matrix will only have a small effect on the results.

$d^G_{m,n}$ has an infinite number of possible values, and we desire a finite and tractable number for Stage 1. We further this approximation by binning elements of d^G into 50 discrete values, spaced exponentially such that the differences in covariance between adjacent bins are equal[4]. The closest bin center is always 100 meters, the furthest is always where the covariance has decayed to 0.5% of the covariance at 100 m[5]. We place bin edges at the midpoint of each interval, and the final bin edge is placed such that the final value is in the center of the final bin. We assume the covariance at distances further than the furthest edge to be 0.

(Stage 1: Sampling covariances between grid cells)
(A.24a) and (A.24b) define R as the intermediate $G \times G$ normalized covariance matrix; where k_n is the n^{th} column of the matrix k. In Stage 1, we construct distributions of elements for this this matrix, $R_{m,n}$, conditioned on input data $x(t)$, and locational information a, d^G.

---

[4] The differences are defined by the average of all estimated decay rates.

[5] Computed using an average decay rate from all exponential fits.

$$R = K'\Sigma(v(t)|x(t), a, d^G; \phi)K \tag{A.24a}$$

$$R_{m,n} = k_m'\Sigma(v(t)|x(t), a, d^G; \phi)k_n \tag{A.24b}$$

$$\tag{A.24c}$$

Recall that $K_{i,n}$ defines the amount of total PV capacity in grid cell i contained in system n, thus only elements of k_n pertaining to systems in cell n will be non-zero. As a result, we can define the distribution of $R_{m,n}$ in terms of only information relating to both cells m and n, of which there are 5:

- $\{x_m(t), x_n(t)\}$, The pair of input states for each cell, of which there are 25 possible pairs of input states.
- $\{a_m, a_n\}$, The arrangements of PV systems contained within each cell, we use 5 possible arrangements of PV systems within a grid cell, defined in Table A.3. There are 15 unique pairs of arrangements.
- $d^G_{m,n}$, The distance separating the pair of cells, which we bin into one of 50 discrete values.

Thus there are $15 \cdot 25 \cdot 50 = 18,750$ possible conditions.

For each set of conditions we simulate 200 sets of volatility states of systems contained in the pair of grid cells. These states are denoted $v^{(i)}|\{x1, x2\}$, where the superscript (i) indicates the sample number between 1 and 200, and $\{x1, x2\}$ indicate the input states for cell 1 and cell 2. We sample volatility states from the stationary distributions of the transition matrices, which are dependent on input state. We then apply the conservative assumption defined in our original paper [59] for which states are sorted such that sample number 1 contains all of the most volatile states and sample number 200 contains all of the least volatile, thus assuming complete correlation of volatility states.

Eq. A.25 defines a sampled value for elements of R conditioned on all available information; where $r^{(i)}|\{x1, x2, a1, a2, d\}$ is the $i^{(th)}$ sampled value of $R_{m,n}(t)$ given that $x_m(t) = x1$, $x_n(t) = x2$, $a_n = a1$, $a_m = a2$, and $d^G_{m,n} = d$. Because the volatility states are sorted, so are the values of r such that the 1^{st} sample results from the most volatile states and the 200^{th} results from the least.

$$r^{(i)}|\{x1, x2, a1, a2, d\} = k_{a1}'\Sigma(v^{(i)}|\{x1, x2\}, [a1, a2], d; \phi)k_{a2} \tag{A.25}$$

(Stage 2: Sampling covariances between grid cells) Scenarios, outlined in Section 3.2, define the vector of arrangements, a, and the distance matrix, d^G, by placing predefined arrangements of systems in specific grid cells in California. The utility-scale scenarios use only the grid arrangement and the rooftop scenario uses only the random arrangements.

Input data for each grid cell at each time, $x(t)$, are derived as defined in Section A.4. The time resolution of input data is 30 minutes for the sub-hourly metrics and 60 minutes for the hourly forecast errors.

For each interval of the input data, we sample 10 values of R by drawing 10 random integers between 1 and 200, $s_{t,i}$. Eq. (A.26) defines $R^{(i)}(t)$ as the i^{th} sample of $R(t)$. The same random sample number is used for each element, $s_{t,i}$, thus maintaining complete correlation of volatility states among all systems in the BA. Eq. (A.27) then parallels Eq. (A.21d) to define the i^{th} sampled variance for all PV in the balancing area at time t.

$$R^{(i)}_{m,n}(t) = r^{(s_{t,i})}|\{x_m(t), x_n(t), a_m, a_n, d^G_{m,n}\} \tag{A.26}$$

$$\sigma^2_{\eta_{1,5,t}(S)}(t,i) = u' m_t' R^{(i)}(t) m_t u \tag{A.27}$$

PV's contribution to reserve use.

From the output of Stage 2, we obtain climatically conditioned distributions for each variability and uncertainty metric of PV for each time in each locational scenario. These distributions are Gaussian mixtures, with one mean and ten equally likely variances at each time.

As an example, Eq (A.28) describes the distribution of hourly forecast errors of PV at time t as a Gaussian mixture with 10 states. This distribution could be equivalently written for each of the metrics. $\mu_{\epsilon_{60,60}}(t)$ is the mean hourly forecast error at time t. This mean is the same for all mixture components.[6] $\sigma^2_{\epsilon_{60,60}}(t,n)$ is the variance of the n^{th} mixture component at time t. Components are ordered such that the 1^{st} component results from the most volatile simulated hidden states and the 10^{th} component results from the least volatile. u is a vector of ones such that $u/10$ defines each mixture state to be equally likely.

$$\epsilon_{60,60}(t) \sim GMM\left(\mu_{\epsilon_{60,60}}(t), \left[\sigma^2_{\epsilon_{60,60}}(t,1), \sigma^2_{\epsilon_{60,60}}(t,2), \ldots \sigma^2_{\epsilon_{60,60}}(t,10)\right], u/10\right) \tag{A.28}$$

Equations (A.29a-c) and (A.30a-c) show the probability distributions for PVs contribution to regulation and load following reserves respectively. The variance

---

[6] For forecast errors these means are always 0 – i.e. forecasts are unbiased – for deviations they are the deviation expected at time t given a clear sky.

calculation assumes no correlation between forecast errors and deviations given the mixture states, i.e. there are no covariance terms between uncertainty and variability. The method *does* assume correlation between mixture states: the variance of the n^{th} component results from the sum of the variances from the n^{th} components' of the forecast errors and deviations. Thus forecast errors and deviations are more likely to be volatile at the same time than if they were independent, but are not more likely to fluctuate in the same direction.

$$Reg^{PV}(t) \sim GMM\left(\mu_{Reg^{PV}}(t), \left[\sigma^2_{Reg^{PV}}(t,1), \sigma^2_{Reg^{PV}}(t,2), \dots \sigma^2_{Reg^{PV}}(t,10)\right], u/10\right) \tag{A.29a}$$

$$\mu_{Reg^{PV}}(t) = \mu_{\epsilon_{5,5}}(t) + \mu_{\eta_{1,5}}(t) \tag{A.29b}$$

$$\sigma^2_{Reg^{PV}}(t,n) = \sigma^2_{\epsilon_{5,5}}(t,n) + \sigma^2_{\eta_{1,5}}(t,n) \tag{A.29c}$$

$$LF^{PV}(t) \sim GMM\left(\mu_{LF^{PV}}(t), \left[\sigma^2_{LF^{PV}}(t,1), \sigma^2_{LF^{PV}}(t,2), \dots \sigma^2_{LF^{PV}}(t,10)\right], u/10\right) \tag{A.30a}$$

$$\mu_{LF^{PV}}(t) = \mu_{\epsilon_{60,60}}(t) + \mu_{\eta_{5,60}}(t) + \mu_{\epsilon_{5,5}}(t) \tag{A.30b}$$

$$\sigma^2_{LF^{PV}}(t,n) = \sigma^2_{\epsilon_{60,60}}(t,n) + \sigma^2_{\eta_{5,60}}(t,n) + \sigma^2_{\epsilon_{5,5}}(t,n) \tag{A.30c}$$

Validation

We validate this method by using the short time-scale data from model fitting. We use our model to simulate probability distributions for load following and regulation reserves that are observed in our short time-scale data, but withheld from the model fitting process – referred to as the test data. We then standardize the test data using the simulated distributions. As a result, if the simulated distributions are correct, the standardized data will exhibit a standard normal distribution (mean = 0, standard deviation = 1).

Figure A.4 shows quantile-quantile plots of the standardized test data against a standard normal distribution. Should the simulated distributions match the empirically observed distribution of the data, points will lie along the $y = x$ line. As shown, our model predicts the distribution of PV's contribution to load following reserves well. However our model predicts heavy tails for the distribution of regulation reserves, i.e. our model will over-predict extreme events. In our previous paper (and in Appendix A.4) we find that our model generally over-predicts 5 minute forecast errors, likely causing this result for regulation reserves. Load following reserves also contain 5 minute forecast errors, but their distribution is dominated by hourly forecast errors and thus is less affected. As a result, our model serves as an effective predictor of load following need, and a conservative predictor of regulation need.

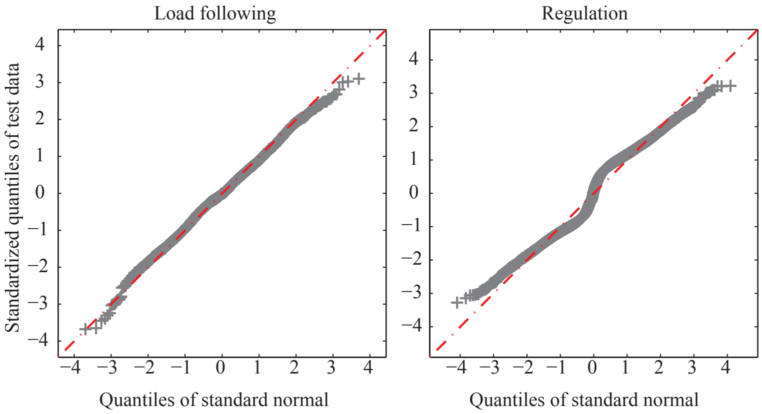

Figure A.4: QQ plots of standardized observations load following and regulation reserve use versus a standard normal.

A.4 Updated Parameter Fitting and Validation

The volatility state model, described in [59], simulates each of the four variability/uncertainty metrics as vectors of spatially correlated random variables: Hour-ahead, hourly forecast errors, $\epsilon_{60,60}(t)$; 5 minute deviations, $\eta_{5,60}(t)$; 5 minute forecast errors, $\epsilon_{5,5}(t)$; 1 minute deviations, $\eta_{1,5}(t)$. In this appendix, we update the results from our original paper to include two changes: conditioning on satellite-derived irradiance data, and including hourly forecast errors.

In [59], we condition the volatility state model on 15 min, 2 km resolution generation from distributed PV systems. However these data were only available for a limited number of locations and times. In this work, we condition on 30 min, 1 km resolution satellite derived irradiance data that are available for any location California and for any year after 1998 from Clean Power Research.

Also in [59], we fit models only for sub-hourly metrics of variability and uncertainty, however hourly forecast errors also exhibit stochastic volatility and spatial correlation and we include them in this analysis. This appendix describes our method for forecasting PV generation—from which we collect hour-ahead forecast errors— and the creation of hourly input data. It also includes results from parameter fitting and validation of the volatility state model fit to hourly forecast errors.

Data

The PV integrator SolarCity provide instantaneous voltage and current measurements at 15 minute resolution for over 6000 PV systems in California, we fit models for hourly forecast errors to one year of data from a subset of 1000 PV systems—250 from each of 4 regions within a 40 km radii of Los Angeles, San Francisco, Sacramento, and Fresno California. These regions represent a variety of coastal and inland climates in California as well as contain high densities of PV systems. All systems provide data for more than 80% of daytime hours throughout the year.

To study variability at faster timescales, SolarCity increased the sampling rate to once per minute at a small subset of systems. Figure A.5 shows locations for sources of 1-minute resolution generation data in the final dataset. We chose these systems to be in one of two 256 km^2 areas, each representative of different types of weather in California: the central valley (CV), and the Los Angeles coast (LA). We sampled 100 systems in each area from about 500 available using an algorithm that combined (1) quota sampling for distances between pairs of locations and (2) geographically random sampling of site pairs. Systems were monitored from mid-June to the end of August 2012. We used only systems for which there we no gaps in data over a period of 30 days, leaving us with data from 39 system inverters in LA and 55 in CV.

The data in our sample cover a number of extreme events; half of the systems experience a 1 minute duration ramp rate of more than 58% of their nameplate capacity, and half of the systems experience a 5-min average ramp of more than 34% of their nameplate capacity.

Input climate data are estimates of horizontal solar irradiance, derived from satellite images at resolutions of 30 minutes and 0.01 decimal degrees—roughly 1 km^2 Data can be downloaded for free from Clean Power Research (CPR) https://www.solaranywhere.com, CPR provided us with access to an API which allowed us to automated downloading.

An empirically adjusted a solar-earth geometry model gives clear sky signals, $CL_i(t)$, both for generation from PV systems and for the satellite derived irradiance. The clear sky signal is the generation or irradiance that would have occurred in absence of clouds and is assumed to be predictable for any PV system. An explanation of this method is contained in Section II-A of [59].

Forecasting hourly average generation

Eq. (A.31) describes a linear model that predicts hourly average generation from a PV system or hourly average irradiance in a grid cell, $G_i(t)$, given a clear sky signal, $CL_i(t)$, and measurements of PV generation from preceding hours; where $\gamma_{HOD,MOY}(t)$ is a vector of 144 indicator variables identifying a combination of hours of the day and 2-month intervals (Jan-Feb, Mar-Apr, etc.). $\hat{G}(t, t-h)$ is a prediction of generation in hour t assuming a persistence of the clear sky index from time $t-h$, defined in Eq. (A.32).

We fit a model to each of the thousand systems using a linear regression that minimizes square errors. The hourly forecast errors for each system, $\epsilon_{60,60,i}(t)$, are the resulting residuals from each model.

$$G_i(t) = \gamma_{HOD,MOY}(t)\beta_{HOD} + \sum_{i=1}^{3} \hat{G}_i(t, t-i)\beta_i + \hat{G}_i(t, t-24)\beta_{24} + \epsilon_{60,60,i}(t) \quad \text{(A.31)}$$

$$\hat{G}_i(t, t-h) = CL_i(t)\frac{G_i(t-h)}{CL_i(t-h)} \quad \text{(A.32)}$$

Climate data: Sub-hourly cloud volatility

As we explain in Section A.3, we condition the model's volatility states on discrete climate data that are specific to the modeled times and locations, $x(t)$. For sub-hourly metrics, we use a heuristic of cloud volatility from the satellite-derived irradiance data, which is observed at resolutions of 1 km^2 and 30 minutes. This section describes how we compute the heuristic.

First, we calculate a moving standard deviation as follows: where $S_i(t)$ is the irradiance derived for grid cell i at time t, $CL_i(t)$ is the clear-sky irradiance for cell i at time t, $\mu_i(t)$ is a 2-hour centered moving average of irradiance from cell i at time t, and m is the number of intervals for the moving window—$m = 4$ resulting in a 2-hour window.

$$\sigma_i(t) = \frac{1}{m+1}\left[\sum_{j=(t-\frac{m}{2})}^{t+\frac{m}{2}}\left(\frac{S_i(j)}{CL_i(j)} - \mu_i(j)\right)^2\right]^{\frac{1}{2}} \tag{A.33}$$

We place each standard deviation reading into one of 5 bins, resulting in the vector of data inputs for the model:

$$x(t) \in \{1, 2, \ldots 5\}^G, \tag{A.34}$$

Binning the data is necessary because the model is conditioned on discrete (not continuous) inputs, as explained in Section A.3 of the supporting information. We defined the bin edges using equally spaced exponential intervals: 0, $e^{-3.5}$, $e^{-2.83}$, $e^{-2.16}$, and $e^{-1.5}$.

Panel C of Figure A.5 shows one day of generation at 15-min resolution along with the moving standard deviation of this signal; panel B shows and the resulting volatility heuristics (conditioning inputs) for this day.

Climate data: hourly cloud volatility

We desire a metric of cloud volatility such that it relates to hourly forecast errors. To this end we base our hourly cloud volatility heuristics on errors in an hourly irradiance forecast. These heuristics are similar in form to those for sub-hourly cloud volatility: there is one value for each of T, 60 minute intervals and each of G 1 km^2 grid cells, and inputs take on 5 discrete states, $x(t) \in \{1, 2, \ldots 5\}^{T,G}$.

To compute the heuristic, we first average the irradiance data from a 30 minute resolution to 60 minute. Second, we fit a separate forecast model to predict hourly irradiance for each grid cell, and collect the residuals from these fits as forecast errors; we use the same forecast model form described in Section A.4. Third, we find a moving standard deviation of the forecast errors using a 5 hour moving window. Finally, we bin these moving standard deviations into 5 bins. Bin edges are defined such that the least volatile bin contains 40% of all readings and each subsequent bin contains 15% of all readings.

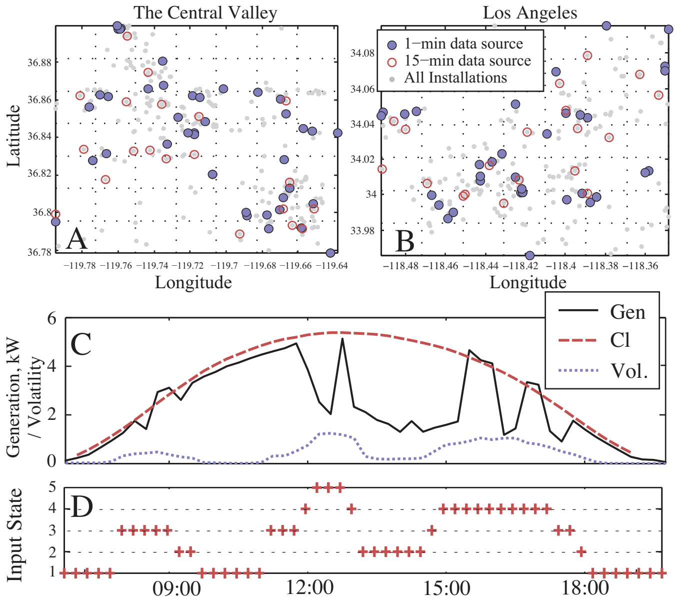

Figure A.5: A: 15-min generation, clear sky signal, and moving standard deviation for one system. B: Volatility heuristic based on the moving standard deviation.

Parameter estimation

Tables A.4-A.7 display the log-likelihood of the model data (data used to fit the model) and the test data (reserved data for testing) fit to each metric, along with the variance of each state estimates and the frequency with which that state is encountered. For each metric, there is a number of states above which the log-likelihood decreases or increases negligibly, and/or additional states are encountered very infrequently ($< 1\%$). We use this number of states for subsequent validation: $M = 3$ for hourly forecast errors, $M = 7$ for 5-minute persistence forecast errors, and $M = 5$ for both 1 and 5 minute deviations.

Tables A.11 - A.8 show estimates of the correlation parameters, $a_{m,n}$ and $\tau_{m,n}$, for the sub-hourly metrics, where a higher volatility state index corresponds to higher standard deviation. For load following variability, a (correlation at a distance of 0 m) generally increases with volatility state. Trends in τ suggest a non-monotonic relationship with volatility state, where the decay range is short for high and low variance states, and long for moderate variance.

Table A.4: Summary of model fitting to hourly forecast errors, $\epsilon_{60,60}(t)$

	M = 2		**M = 3**		**M = 4**		**M = 5**	
ELL Model	9.396e+05		9.457e+05		9.325e+05		9.9e+05	
ELL Test	1.329e+06		1.338e+06		1.317e+06		1.405e+06	
State #	σ^2	f	σ^2	f	σ^2	f	σ^2	f
1	0.0045	0.65	0.004	0.61	0.0054	0.66	0.00091	0.42
2	0.047	0.35	0.025	0.09	0.036	0.25	0.018	0.37
3	-	-	0.048	0.3	0.051	0.068	0.037	0.13
4	-	-	-	-	0.14	0.021	0.08	0.067
5	-	-	-	-	-	-	15	9.1e-05

Table A.5: Summary of model fitting to 5-minute forecast errors, $\epsilon_{5,5}(t)$

	M = 3		**M = 4**		**M = 5**		**M = 6**		**M = 7**		**M = 8**	
ELL Model	9.805e+05		9.936e+05		1.013e+06		1.017e+06		1.02e+06		1.02e+06	
ELL Test	3.074e+05		3.108e+05		3.167e+05		3.181e+05		3.189e+05		3.189e+05	
State #	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f
1	3.1e-06	0.55	3e-06	0.53	1.2e-06	0.33	7.3e-07	0.23	5.7e-07	0.18	5.6e-07	0.18
2	0.00031	0.33	0.00016	0.27	1e-05	0.28	4.9e-06	0.3	3.1e-06	0.28	3.1e-06	0.27
3	0.0099	0.11	0.0017	0.15	0.00022	0.22	5.6e-05	0.19	2.1e-05	0.18	2e-05	0.18
4	-	-	0.022	0.043	0.0023	0.14	0.00054	0.16	0.00019	0.17	0.00018	0.17
5	-	-	-	-	0.027	0.032	0.0042	0.095	0.0013	0.12	0.0012	0.12
6	-	-	-	-	-	-	0.038	0.019	0.007	0.059	0.0055	0.0014
7	-	-	-	-	-	-	-	-	0.049	0.013	0.0067	0.061
8	-	-	-	-	-	-	-	-	-	-	0.048	0.013

Table A.6: Summary of model fitting to 5-minute deviations, $\eta_{5,60}(t)$

	M = 3		**M = 4**		**M = 5**		**M = 6**		**M = 7**		**M = 8**	
ELL Model	7.389e+05		7.485e+05		7.528e+05		7.533e+05		7.547e+05		7.549e+05	
ELL Test	2.109e+05		2.137e+05		2.151e+05		2.152e+05		2.156e+05		2.156e+05	
State #	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f
1	2.2e-05	0.52	1.7e-05	0.46	1.4e-05	0.38	1.4e-05	0.37	5.8e-06	0.1	5.9e-06	0.11
2	0.00052	0.26	0.00017	0.22	7.5e-05	0.22	7.3e-05	0.22	2e-05	0.32	2e-05	0.32
3	0.013	0.22	0.0024	0.2	0.00067	0.16	0.00064	0.16	9.1e-05	0.18	9.2e-05	0.18
4	-	-	0.021	0.13	0.005	0.16	0.0046	0.15	0.00069	0.15	0.00066	0.15
5	-	-	-	-	0.027	0.08	0.022	0.087	0.0046	0.15	0.0029	0.0017
6	-	-	-	-	-	-	0.083	0.0044	0.021	0.088	0.0045	0.15
7	-	-	-	-	-	-	-	-	0.078	0.0052	0.021	0.089
8	-	-	-	-	-	-	-	-	-	-	0.075	0.0058

Table A.7: Summary of model fitting to 1-minute deviations, $\eta_{1,5}(t)$

	M = 3		**M = 4**		**M = 5**		**M = 6**		**M = 7**		**M = 8**	
ELL Model	4.597e+06		4.722e+06		4.751e+06		4.754e+06		4.761e+06		4.761e+06	
ELL Test	1.471e+06		1.51e+06		1.518e+06		1.519e+06		1.521e+06		1.521e+06	
State #	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f	σ^2	f
1	1e-05	0.64	3.4e-06	0.33	3.3e-06	0.31	3.2e-06	0.31	3.1e-06	0.3	3.1e-06	0.3
2	0.00027	0.23	2.5e-05	0.38	2.1e-05	0.36	2e-05	0.36	1.7e-05	0.33	1.7e-05	0.33
3	0.015	0.14	0.00055	0.18	0.0002	0.15	0.00018	0.14	8.1e-05	0.13	8.1e-05	0.13
4	-	-	0.018	0.11	0.0022	0.11	0.0018	0.12	0.00064	0.12	0.00064	0.12
5	-	-	-	-	0.026	0.068	0.018	0.01	0.0048	0.073	0.0048	0.073
6	-	-	-	-	-	-	0.026	0.063	0.023	0.0041	0.022	0.00083
7	-	-	-	-	-	-	-	-	0.032	0.047	0.023	0.0034
8	-	-	-	-	-	-	-	-	-	-	0.032	0.047

Table A.8: Parameters for spatial correlation functions fit to $\eta_{60,60}(t)$.

M = 3

	a			τ		
S#	1	2	3	1	2	3
1	0.44	0.29	0.50	55.46	59.04	52.18
2	-	0.30	0.28	-	62.50	48.52
3	-	-	0.62	-	-	46.21

Table A.9: Parameters for spatial correlation functions fit to $\epsilon_{5,5}(t)$

M = 7

	a							τ						
S#	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	0.53	0.48	0.30	0.02	0.09	0.03	0.00	0.30	0.47	0.59	60.68	9.72	31.31	0.10
2	-	0.44	0.13	0.06	0.12	0.08	0.11	-	0.52	3.72	7.42	5.21	5.91	0.11
3	-	-	0.37	0.31	0.22	0.19	0.19	-	-	4.40	6.70	7.18	3.85	1.83
4	-	-	-	0.54	0.53	0.42	0.15	-	-	-	5.40	4.41	3.84	10.95
5	-	-	-	-	0.71	0.67	0.56	-	-	-	-	3.33	3.08	3.09
6	-	-	-	-	-	0.75	0.68	-	-	-	-	-	2.39	2.32
7	-	-	-	-	-	-	0.86	-	-	-	-	-	-	2.86

Table A.10: Parameters for spatial correlation functions fit to $\eta_{5,60}(t)$

M = 5

	a					τ				
S#	1	2	3	4	5	1	2	3	4	5
1	0.18	0.18	0.07	0.08	0.00	13.26	15.15	7.48	0.53	0.10
2	-	0.27	0.19	0.01	0.00	-	9.42	5.51	0.87	0.10
3	-	-	0.52	0.37	0.08	-	-	5.46	7.99	20.19
4	-	-	-	0.66	0.60	-	-	-	7.80	4.45
5	-	-	-	-	0.92	-	-	-	-	4.85

Table A.11: Parameters for spatial correlation functions fit to $\eta_{1,5}(t)$.

M = 5

	a					τ				
S#	1	2	3	4	5	1	2	3	4	5
1	0.41	0.17	0.05	0.01	0.00	0.21	0.40	0.56	0.23	0.10
2	-	0.22	0.14	0.03	0.01	-	1.39	1.98	2.21	0.10
3	-	-	0.54	0.38	0.03	-	-	1.23	1.08	0.15
4	-	-	-	0.69	0.49	-	-	-	0.92	0.49
5	-	-	-	-	0.62	-	-	-	-	0.41

Validation by Simulation

To validate the model we compare observations to simulated distributions, which require estimates of the volatility states. We simulate volatility states with the following methods:

$v(t)$ *Simulation Method 1*: Set $v_i(t)$ equal to the most likely state as estimated during model fitting. This requires that we restrict simulations to times and locations for which we have one minute data. Because we ultimately want to use the model to estimate reserves in the absence of one minute data, this method is only a baseline.

$v(t)$ *Simulation Method 2*: For each system at each time, simulate N_p samples of volatility states using the stationary probabilities of the transition matrices in $\mathbb{A}$. In this paper we use $N_p = 40$ samples. We model the distribution for each system at each time as a Gaussian mixture with N_p equally likely components, one for each sampled covariance matrix. This method neglects correlation between volatility states.

$v(t)$ *Simulation Method 3*: Simulate N_p volatility states per site as in method 2, then independently sort the N_p volatility states for each system at each time from highest to lowest variance, such that $v_i^{(1)}(t)$ and $v_i^{(N_p)}(t)$ contain the highest and lowest standard deviations, respectively; this maximizes correlation of volatility states.

Figure A.6 shows quantile-quantile (QQ) plots that compare standardized quantiles of the observed data to quantiles of a standard normal. We standardize the quantiles by first computing the position of the observed data within the simulated model's cumulative density function for all observations in the study period, and then taking the standard normal inverse CDF of the result. If the model and its parameterization predict the empirical distribution, the standardized quantiles should be normally distributed, and points in the QQ plot will lie along the $y = x$ line.

Column 1 of Figure A.6 shows results from the baseline scenario where volatility states are known (method 1); simulated distribution tails are slightly lighter than the data for deviations, slightly heavier than the data for hourly and 5-minute forecast errors. This result indicates that if the volatility state distribution across sites is well characterized, the model will work well in times and periods for which one minute data are unavailable.

Column 2 of Figure A.6 shows that method 2 results are worse than method 1. Because method 2 does not model volatility state correlation across systems, the

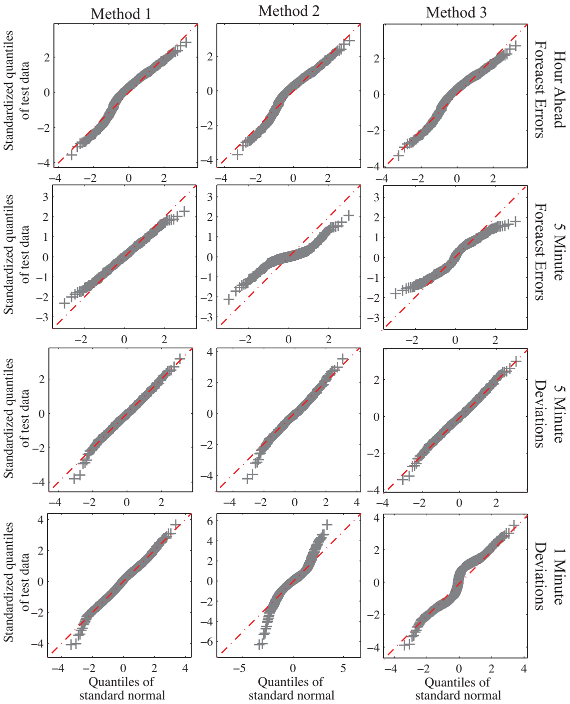

Figure A.6: QQ plots of test data pseudo-residuals using each method of volatility state simulation.

probability that multiple systems will be in a high volatility state simultaneously is relatively low, and we expect this simulations to under-predict extreme events – i.e. the observed data tails will be heavy. For 1- and 5-min deviations the observed tails are heavy, as expected. For 5-min persistence forecast errors the tails are light, though trend toward crossing the $y = x$ line back to heavy. For Hourly forecast Method 2 performs similarly to Method 1.

Column 3 of Figure A.6 shows results from method 3, which maximizes volatility state correlation across systems; the effect is evident if one compares the tails of the QQ plots between methods 2 and 3. Method 3 generates simulated tails that are heavier than Methods 2 and 1, and also the observed data; making it an ideal conservative assumption.

Predicting maximum events

Eq. (A.35) shows a method for finding the probability that all observations within a time period fall below some threshold, x, assuming independent observations; where

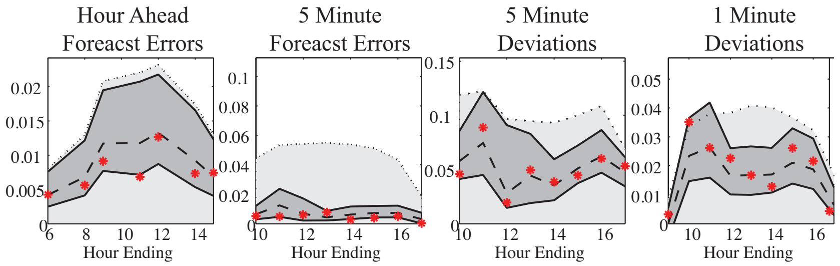

Figure A.7: Predicted distributions of maxima observed during each hour ending in the test data. Rows represent each metric ($\epsilon_{60,60}(t)$, $\epsilon_{5,5}(t)$, $\eta_{5,60}(t)$, $\eta_{1,5}(t)$). The dark gray boundaries signify a 95% confidence interval found using the baseline scenario where the most likely volatility states predicted by EM is used, the light gray boundary signifies the 97.5^{th} percentile of the predicted distribution of maxima using the sorting based worst case assumption. Red stars signify observed maxima.

T_{HOD} is a specified hour of day. The 95% confidence interval for the maximum requirement corresponds to probabilities of 0.975, and 0.025.

$$p\{\eta_{1,5}(t) \leq x : t \in T_{HOD}\} = \prod_{t \in T_{HOD}} p\left\{\eta_{1,5}(t) \leq x\right\} \tag{A.35}$$

Figure A.7 shows 95% confidence intervals for the maximum reserve requirement estimated for the test data, stratified by each hour of day. The dark gray boundaries are those calculated with known volatility states. The light gray (dotted line) boundary is 97.5% bound (i.e. the upper bound of a 95% confidence interval) of the predicted distribution using method 3, the worst-case assumption for correlation of volatility states. The light gray boundaries are those that are predicted by our simulation model.

In expectation, observed maxima should fall above the 97.5% confidence bound between 0 and 1 times for 32 observations. For Method 3 there is 1 observation above the bound, for Method 1 there are no observations above the bound. For 1 minute deviations, the Method 3 upper bound dips below the upper bound for Method 1 for a few hours; this results from small differences between stationary hidden state probabilities from method 3 versus the most likely volatility states predicted in method 1.

A.5 References

[59] M. Tabone and D. Callaway, "Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2965–2973, Nov. 2015, ISSN: 0885-8950. DOI: `10.1109/TPWRS.2014.2372751`.

[84] National Renewable Energy Laboratory, *Solar Maps*. [Online]. Available: `http://www.nrel.gov/gis/solar.html` (visited on 08/31/2013).

Appendix B

Learning HVAC Energy Use from Smart Meters

B.1 Parameter fitting

For model fitting purposes, we enforce a rule that one and only one latent state must be active at any given time. I.e., $\sum_m C_{m,t} = 1, \forall t$. We account for times when there is no heating or cooling by adding a state $c_{off,t}$ for which $X_{off,t}$ is empty and $\epsilon_{off,t} = 0$. We also necessarily assume that heating and cooling are never simultaneously active. We could correct for this by creating an additional state for when they are on simultaneously. This is common practice for factorial hidden Markov models which may include many types of devices [79]. For residential buildings, we believe that it safe to assume that heating and cooling are mutually exclusive.

Eq. (B.1) defines model residuals for predictions of whole home energy consumption given the active latent state, $\tilde{e}_{m,t}$. The tilde distinguishes this residual from the residual of the a model that predicts only cooling or heating energy, defined in Eq. (B.2). Because we don't observe heating or cooling energy in AMI data, this later residual is not observed. However, for the off state it can only contain non-HVAC equipment, and thus equals $\tilde{e}_{off,t} = e_{o,t}$. We use the distribution of these residuals as an estiamte for the distribution of disturbances, where $\tilde{f}_m(e) = p(\tilde{\epsilon}_m = e)$, and $f_m(e) = p(\epsilon_m = e)$. In the the following section, B.2, we focus on recovering estimates for e_o and e_m from observations of $\tilde{e}_m$.

$$\tilde{e}_{m,t} = P_t - Z\Theta - X_{m,t}\beta_m = e_{o,t} + e_{m,t} \tag{B.1}$$

$$e_{m,t} = P_{m,t} - X_{m,t}\beta_m \tag{B.2}$$

Expectation-Maximization

We fit the parameters to each model using the Expectation-Maximization (EM) algorithm. EM is gradient ascent method that maximizes the expected logged likelihood of observations given model parameters, and as a result also maximizes the actual model likelihood. Equations (B.3a)-(B.3c) derive the expected log likelihood. (B.3a) defines the complete log likelihood of data given the model parameters and the values of latent states. (B.3b) takes the log of this likelihood function, and (B.3c) takes the expected value of the result with respect to the latent states.

$$L_{com} = \prod_{t=1}^{T} \prod_{m=1}^{M} \left(p\{Y_t = Z_t\Theta + X_{m,t}\beta_m\} p\{C_{m,t} = 1\} \right)^{C_{m,t}} \tag{B.3a}$$

$$\log L_{com} = \sum_{t=1}^{T} \sum_{m=1}^{M} C_{m,t} \left(\log p\{Y_t = Z_t\Theta + X_{m,t}\beta_m\} + \log p\{C_{m,t} = 1\} \right) \tag{B.3b}$$

$$\mathbb{E}\left[\log L_{com}\right] = \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{m,t} \log p\{Y_t = Z_t\Theta + X_{m,t}\beta_m\} + \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{m,t} \log p\{C_{m,t} = 1\} \tag{B.3c}$$

EM is an iterative method, alternating between updating the expected values of hidden states given parameters (E-Step), and updating parameters by maximizing the ELL (M-Step). In the following section we explain each update.

A note on estimating the linear model

A proper implementation of EM would maximize the first term of ELL—which contains both variances, $\tilde{\sigma}_m$, and coefficients, β and Θ—in each M-step. We approximate this maximization by updating the variances and coefficients separately, conditioning on the most recent value. This iterative process is common for fitting linear models with heteroskedastic disturbances: (1) fit coefficients given homoskedastic disturbances; (2) estimate the variance of residuals by regressing squared residuals onto some exogenous variable, in this case the latent state; (3) fit coefficients again, weighting observations by the inverse of their estimated variance. This is actually an implementation of EM in itself, though most empiricists stop after only one iteration. This this maximization is a separate implementation of EM within our larger implementation of EM. To ensure that both EM optimizations converge, we run the iteration loop four times after convergence of the latent states.

E-M Parameter Updates

Update β, Θ
Eq. B.4 updates β and θ given all other parameters using weighted ordinary least squares. It is trivial to show that this amounts to maximizing the ELL over β and θ given normally distributed disturbances with known variances $\tilde{\sigma}_m$.

We use the same weighted OLS fitting models with a KDE for the distribution of disturbances. Though disturbances are not normally distributed in this case, it has been shown that OLS is a consistent estimator for linear models with non-normal disturbances given enough uncorrelated observations. Such proofs rely on the the central limit theorem (CLT).

We expect that the added value of using a KDE for disturbances will result when classifying observations—which we do below—while a normal distribution is effective for estimating coefficients. The coefficients of linear models average effects among thousands of observations in each dataset; for which CLT is likely to hold. Classifying one observation relies on the error distribution with no averaging, thus CLT will not apply.

$$\Theta^*, \beta^* = \arg\min_{\Theta,\beta} \sum_{t=1}^{T} \sum_{m=1}^{M} \frac{\gamma_{m,t}}{\tilde{\sigma}_m^2} ||\tilde{e}_{m,t}||_2^2 \tag{B.4}$$

Update $f_m(e)$
In the normally distributed case, $\tilde{f}_m(e) \sim \mathcal{N}(0, \tilde{\sigma}_m^2)$. Eq. (B.5) presents values of $\tilde{\sigma}_m^2$ that maximize ELL; they are found by an average of square residuals from the model given a cooling state, weighted by the expected value that each observation is in that cooling state.

$$\tilde{\sigma}_m^{2*} = \frac{1}{\sum_{t=1}^{T} \gamma_{m,t}} \sum_{t=1}^{T} \tilde{e}_{m,t}^2 \gamma_{m,t} \tag{B.5}$$

Eq. (B.6) updates the kernel density estimate for disturbances using residuals and expected values of each latent state, and normally distributed kernels. The only parameter in these equations is $\sigma_{bw,m}^2$, which is the bandwidth variance of the kernel. We use Scott's rule-of-thumb for selecting kernel bandwidths [82]. Scott's rule is computationally efficient and is considered an upper bound for smoothing in kernel densities. This allows our kernel density to empirically widen the tails of the distribution with respect to the normal case—which is our desired effect—without potentially over-fitting the distribution to a few observations which would be the result of choosing too small a bandwidth. As a result, this method may miss some nuance in the center of the distribution that could be found with less conservative bandwidths. Because the residuals to which the KDE will change as the coefficients change, we believe that this additional smoothing is desirable.

$$\tilde{f}_m(e) = \frac{1}{\sum_{t=1}^{T} \gamma_{m,t}} \sum_{t=1}^{T} \gamma_{m,t} \phi(e, \tilde{e}_{m,t}, \sigma^2_{bw,m}) \tag{B.6}$$

Update w_c

For latent state models, Eq. (B.7) updates w_c: the probability of a cooling state given only the observed values in X and Z (i.e., not P_t). We update w_c by finding the total expected number of hours in a cooling state, divided by the total possible number of hours that state. We note here that our model could be adapted to enforce an expectation that cooling is more likely to be on or off for a given day by identifying a different $w_{c,d}$ depending on the day.

In Eq. (B.8) we define an extra variable for simplicity later derivations: $w_{c,t}$ is the probability of a cooling state at each time, given only outdoor temperature. $w_{c,t}$ equals w_c when outdoor temperature is above the cooling change point and 0 otherwise. It would work analogously for heating states.

The probability that a the system is in the off state is defined simply as one minus the probability that cooling is on, $w_{off,t} = 1 - w_{c,t}$. This can be simply adapted to include multiple HVAC states: $w_{off,t} = 1 - \sum_{k \neq off} w_{k,t}$

$$w_c = \frac{\sum_{t=1}^{T} \gamma_{c,t}}{\sum_{t=1}^{T} \mathbb{I}\{T_t > T_{cp,c}\}} \tag{B.7}$$

$$w_{c,t} = \begin{cases} w_c & T_{a,t} > T_{cp,c} \\ 0 & T_{a,t} \leq T_{cp,c} \end{cases} \tag{B.8}$$

For the piecewise linear models, we restrict w_m to be 1.

Update γ

Eq. (B.9) updates the expected values for the presence of heating and cooling at each reading. The preceeding equations derive this expected value using Bayes rule, where we use the term *Rest* to refer to all other parameters in the model.

$$\begin{aligned} p\{C_{c,t}|Y_t, Rest\} &= \frac{p\{C_{c,t}, Y_t|Rest\}}{p\{Y_t|Rest\}} \\ &= \frac{p\{Y_t|Rest, C_{m,t}\}p\{C_{m,t}|Rest\}}{\sum_{k=1}^{M} p\{Y_t|Rest, C_{k,t}\}p\{C_{k,t}|Rest\}} \\ &= \frac{f(\tilde{e}_{t,m})w_{m,t}}{\sum_{k=1}^{M} f(\tilde{e}_{t,k})w_{k,t}} \end{aligned} \tag{B.9}$$

Selection of Change Point

We use a grid search to select heating and cooling change points. Separate models are fit to cooling change points within a range of 50F to 90F by increments of 5F. We then select the model with the greatest likelihood of observations given parameters.

Initialization and Stopping Criteria

We initialize each model with a coarse estimate of $\gamma_{c,t}$: it equals 1 if $T_{a,t}$ is greater than the change point temperature and P_t is is above the 50^{th} percentile of readings, and 0 otherwise.

We use changes in all $\gamma_{m,t}$ as a stopping metric; where we stop updates when the average change in each value of $\gamma_{m,t}$ is less than 0.05%. Because our model also requires convergence of σ, we require at least 4 iterations.

B.2 Disaggregation

In this section we identify a maximum likelihood method for disaggregating the residuals of each component model, which are unobserved, from the residual of the total model, which is observed. We derive our result for any number of component models; where $e \in \mathbb{R}^N$ is a vector of residuals for each of N components. The sum of residuals from components must equal the residual from the total model: $\tilde{e} = u'e$; where u is a vector of ones.

We define disaggregation for the special case where e is a multivariate normal random variable with mean 0 and covariance matrix Σ. Σ may contain covariance terms between residuals, represented by off-diagonal elements. In our main analysis, we assume each component's residual to be independent, thus all off-diagonal terms are 0. Deriving disaggregation for a more general covariance matrix allows us to identify how correlations between model residuals may affect our results.

Equations (B.10a)-(B.10c) each show the maximum likelihood disaggregation problem. (B.10a) chooses values of e that maximize the probability density function of the multivariate normal subject to the linear constraint on the sum of residuals. (B.10b) rewrites the objective using the negative logged likelihood. (B.10c) brings the constraint into the objective function using a Lagrange multiplier, λ.

$$e^* = \arg\max_e (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left\{\frac{-e'\Sigma^{-1}e}{2}\right\} \qquad s.t.: u'e = \tilde{e} \tag{B.10a}$$

$$e^* = \arg\min_e e'\Sigma^{-1}e \qquad s.t.: u'e = \tilde{e} \tag{B.10b}$$

$$e^*, \lambda^* = \arg\min_{e,\lambda} e'\Sigma^{-1}e + \lambda(u'e - \tilde{e}) \tag{B.10c}$$

Equation (B.11a) shows the first order condition of optimality that results from differentiating the objective function with respect to e. From this we obtain a solution to the optimization that is dependent on the optimal Lagrange multiplier, λ^*. Equation (B.11b) uses this result in conjunction with the other first order condition—differentiating with respect to λ—to define a closed form solution for λ^*.

$$2\Sigma^{-1}e^* + \lambda^* u = 0 \implies \qquad e^* = \frac{-\lambda^*}{2}\Sigma u \tag{B.11a}$$

$$u'e^* = \tilde{e} = \frac{-\lambda^*}{2}u'\Sigma u \implies \qquad \lambda^* = \frac{-2\tilde{e}}{u'\Sigma u} \tag{B.11b}$$

Equation (B.12a) shows the resulting solution to the disaggregation problem; where each component's residual is the total residual multiplied by a ratio defined by the covariance matrix, Σ: the sum of one row of Σ divided by the sum of all elements of Σ. Diagonal elements of Σ are the variance of each residual, and thus must be positive. However off diagonal elements can be negative if two residuals are anti-correlated; in this case the proportion could be negative, or greater than one.

$$e^* = \tilde{e}\frac{\Sigma u}{u'\Sigma u} \tag{B.12a}$$

In our analysis, we assume that all disturbances are independent. Under this condition, the covariances matrix becomes the diagonal matrix shown in Eq. B.13.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ & & \ddots & \\ 0 & 0 & 0 & \sigma_N^2 \end{bmatrix} \tag{B.13}$$

Eq. (B.14) shows the solution under the assumption of independent residuals, in which the total residual is allocated by the proportion of each component's variance to the total variance.

$$e_i^* = \tilde{e}\frac{\sigma_i^2}{\sum_{k=1}^{N}\sigma_k^2} \tag{B.14}$$

Relating disaggregation of residuals to CSSS

In Eq. (B.10a), we define the disaggregation problem as the allocation of a lumped residual to the residuals for each component model. In Wytock and Kolter's CSSS model, they define the same disaggregation problem differently, by applying a constraint that HVAC consumption and other consumption add to the total consumption [24]. In a separate paper, submitted to *Sustainable Energy, Grids and Networks*, my co-author's and I show that the two framings are equivalent under a certain set of conditions. In this section, I briefly summarize the results of this contribution.

The general formulation for CSSS is to recover a set of signals, Y_i when only the sum of all signals is observed $\bar{Y} = \sum_i Y_i$, presuming that each each signal is linearly related to a separate set of regressors, X_i. Eq. (B.15) presents a special case of CSSS; where the objective function is a minimization of square residuals and there is no regularization.

$$\begin{aligned} \underset{Y_i,\Theta_i}{\text{minimize}} \quad & \alpha_{\mathbf{i}}(Y_i - X_i\Theta_i)^2 \\ \text{subject to} \quad & \bar{Y} = \sum_{i=0}^{L} Y_i \end{aligned} \tag{B.15}$$

In this case we find that the optimal values of Θ are independent of the coefficients, α, and can be found by ordinary least squares. Once Θ is known, the source separation problem can be solved analytically, shown in Eq. (B.16).

$$\underbrace{Y_i^* - X_i\Theta_i}_{e_i} = \frac{\alpha_i^{-1}}{\sum_k \alpha_i^{-1}} \underbrace{\left(\bar{Y} - \sum_k X_i\Theta_k\right)}_{\tilde{e}} \tag{B.16}$$

By comparing Eq. (B.16) to Eq. (B.14) we see that the two are identical when $\alpha_i^{-1} = \sigma_i^2$. We also could have drawn this conclusion by comparing the maximum likelihood formulation in (B.10b) to the CSSS formulation in (B.15). By interpreting the weights as inverse variances, the maximum likelihood approach provides an insight in how to best tune a CSSS problem.

Effect of correlation between component residuals

We use the $N = 2$ case to study how correlation between residuals affects disaggregation: where $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. Table B.1 shows the disaggregation solution under three bounding cases: uncorrelated, perfectly correlated, and perfectly anti-correlated.

Table B.1: Three bounding cases for disaggregation of correlated residuals

ρ	$e_1/\tilde{e}$
0	$\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$
1	$\frac{\sigma_1}{\sigma_1+\sigma_2}$
-1	$\frac{\sigma_1}{\sigma_1-\sigma_2}$

In the uncorrelated case, residuals are disaggregated using ratio of their variance to the total variance. In the cases of perfect correlation, they are disaggregated using standard deviations. The anti-correlated case is particularly insidious, in which the

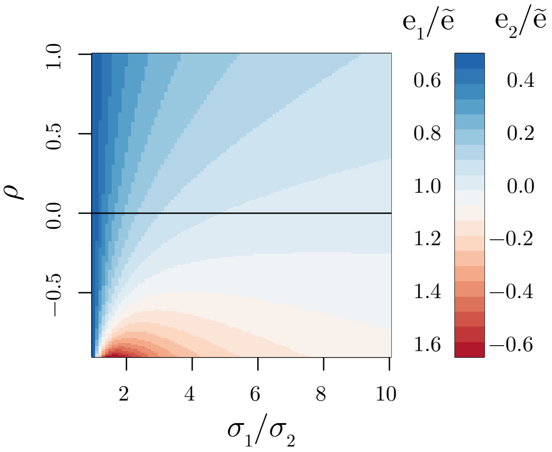

Figure B.1: Effect of correlation between residuals of component models on disaggregation.

denominator is the difference of the two standard deviations and will be negative or zero for at least one component. A negative denominator represents diverging residuals: we expect that errors in the total model to be the result of larger errors in each component that cancel each other, thus one residual will be in the opposite direction of the total residual. A zero denominator results when both residuals have the same variance and are perfectly anti-correlated thus they will perfectly cancel each other at all times. If this were true, the total residual would always be zero and disaggregation would be impossible.

Figure B.1 plots the solution to the disaggregation problem at a range of correlations and variances. Due to the divide-by-zero effect of perfectly anti-correlated residuals, we limit the analysis to a minimum value of $\rho = -0.9$. Note that when both residuals have the same variance $\tilde{e}$ is split evenly among the components regardless of correlation. The effect of correlation on the disaggregation problem is most pronounced when variances are very different, or when correlation is below -0.5. We expect correlation between non-HVAC energy use and AC energy use to be slightly positive, and we know that their variances are similar (below a factor of 2). In this region, correlation has a small effect on disaggregation and we do not expect it to greatly affect our estimates.

B.3 References

[24] M. Wytock and J. Z. Kolter, “Contextually Supervised Source Separation with Application to Energy Disaggregation,” *ArXiv:1312.5023 [cs, math, stat]*, Dec. 2013, arXiv: 1312.5023. [Online]. Available: `http://arxiv.org/abs/1312.5023` (visited on 10/10/2014).

[79] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, “Unsupervised Disaggregation of Low Frequency Power Measurements.,” in *SDM*, vol. 11, SIAM, 2011, pp. 747–758. [Online]. Available: `http://epubs.siam.org/doi/pdf/10.1137/1.9781611972818.64` (visited on 06/22/2015).

[82] M. C. Jones, J. S. Marron, and S. J. Sheather, “A brief survey of bandwidth selection for density estimation,” *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996. [Online]. Available: `http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1996.10476701` (visited on 04/29/2016).