Layered Timeseries Analysis for Smart Grid Agents

Prashant Reddy

Machine Learning Department
Carnegie Mellon University, Pittsburgh, USA

ppr@cs.cmu.edu

Abstract

The vision of evolving the current electrical power grid towards a Smart Grid promises increased utilization of distributed renewable energy resources through more robust technical and economic control systems and active participation from customers. We study two problems within the distribution grid that employ autonomous agents to enable robust control and customer participation. First, we study the evolution of hourly electricity prices in a modern wholesale electricity market with the goal of predicting hourly forward prices so that an intermediary *broker agent* can effectively manage trading risk. We analyze real market data from the Ontario Independent Electricity System Operator (IESO) from 2002 to 2011 and corresponding real weather data from the US National Climatic Data Center (NCDC). Our analysis shows that a layered combination of a multi-class classification approach and a multiplicative seasonal ARIMA model can be used to predict the hourly forward prices with confidence. Second, we develop generative hierarchical models to simulate the demand for and supply of electricity from various types of retail customers (*e.g.,* households, universities, industrial plants). Demand and supply depend on various factors such as installed load capacity, household size, geographic locale, day of week, month of year, cloud cover and wind speed. We assume that a sample of data from real world metering or fine-grained simulation is available *a priori* for training and an additional small sample of data is made available online. We then generate more data for forward simulation from the online data, while borrowing characteristics from the training data, using a coarse-grained model based on a novel combination of ARIMA and hierarchical Bayesian methodology. We evaluate our methodology using data from a fine-grained simulation model based on real data from the MeRegio pilot project in Baden-Wurttemberg, Germany.

1 Introduction

The Smart Grid aims to apply digital communications technologies to modernize the power grid [5] [1]. The overall goals of the Smart Grid include: (i) increased production of power from renewable and distributed production facilities, (ii) shifting demand to periods of time when power is produced more cheaply, and (iii) more robust control systems to manage anomalous events. A key subgoal of Smart Grid design is to enhance the ability of distributed small-scale power producers, such as small wind farms or households with solar panels, to sell energy into the power grid. The corresponding increased complexity creates the need for new technical and economic control mechanisms to maintain the stability of the grid. Moreover, automating such control mechanisms can improve reliability and reduce the costs of operating the grid.

One approach to addressing the challenge of encouraging increased participation from distributed producers is through the introduction of *broker agents* who buy power from such producers and also sell power to consumers [10]. Broker agents interact with producers and consumers through a *tariff market*–a new market mechanism. In this mechanism, each broker agent acquires a portfolio of producers and consumers by simultaneously publishing prices to buy and sell power. The design of fees and penalties in the tariff market incentivizes broker agents to balance supply and demand within their portfolio by buying production and storage capacity from local producers instead of acquiring all supply from the national grid. This also gives them the ability compete in the market by offering prices distinct from the prices on the national grid while also helping local producers. Broker agents that are able to operate successfully in tariff markets, while earning a profit so that they continue to participate, contribute to the overall goals of the Smart Grid.

Figure 1 provides an overview of the Smart Grid tariff market domain. Tariff markets exist in the context of a national grid which carries power produced by large centralized production plants and also moves excess power across tariff markets. The *wholesale market* is an auction-based market that determines the price at which power can be bought from or sold into the national grid. *Physical Coupling* points regulate the flow of power between the national and regional grids. The tariff market, which operates over the regional grid (*a.k.a.* distribution grid), is not an operating entity like the wholesale market but instead defined by a set of participants and rules.

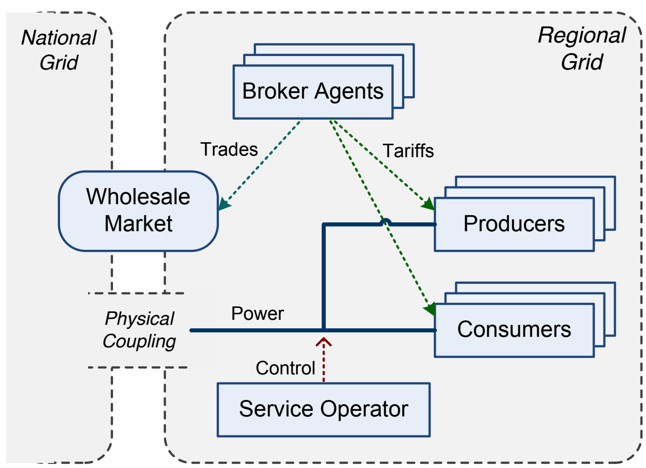

Figure 1: The Smart Grid tariff market domain which focuses on the retail distribution grid.

A tariff market consists of a set of participants $\langle \mathcal{C}, \mathcal{P}, \mathcal{B}, O \rangle$ where:

- $\mathcal{C} = \{C_n\}_{n=1}^{N}$ are the *Consumers*, $N = \mathcal{O}(10^5)$;
- $\mathcal{P} = \{P_m\}_{m=1}^{M}$ are the *Producers*, $M = \mathcal{O}(10^3)$;
- $\mathcal{B} = \{B_k\}_{k=1}^{K}$ are the *Broker Agents*; $K = \mathcal{O}(10^1)$;
- O is the *Service Operator*, a regulated regional monopoly, which manages the physical infrastructure for the grid.

Let $\mathcal{Q} = \mathcal{C} \cup \mathcal{P}$ be the combined set of *customers* from a broker agent's perspective. A *tariff* is a public contract offered by a broker agent that can be subscribed to by a subset of the customer population. At each timeslot, t, each broker agent, B_k, publishes two tariffs visible to all agents in the environment–a *producer tariff* with price $p_{t,P}^{B_k}$, and a *consumer tariff* with price $p_{t,C}^{B_k}$. Each broker agent holds a *portfolio*, $\Psi_t = \Psi_{t,C} \cup \Psi_{t,P}$, of consumers and producers who have subscribed to one of its tariffs for the current timeslot, t. Each consumer consumes some amount of power, κ, per timeslot and each producer generates $\nu\kappa$ units of power per timeslot. At each timeslot, the *profit*, $\pi_t^{B_k}$ of a broker agent is the net proceeds from consumers, $\Psi_{t,C}$, minus the net payments to producers, $\Psi_{t,P}$, and the service operator:

$$\pi_t^{B_k} = p_{t,C}^{B_k} \kappa \Psi_{t,C} - p_{t,P}^{B_k} \nu\kappa\Psi_{t,P} - \phi_t |\kappa\Psi_{t,C} - \nu\kappa\Psi_{t,P}| \tag{1}$$

The term $|\kappa\Psi_{t,C} - \nu\kappa\Psi_{t,P}|$ represents the supply-demand imbalance in the portfolio at timeslot, t. This imbalance is penalized using a substantial *balancing fee*, ϕ_t, which is specified by the service operator at each timeslot. The broker agent is therefore motivated to offset the anticipated imbalance in his portfolio by buying or selling forward contracts in the wholesale market. Conversely, if prices in the wholesale market are expected to be unfavorable, the broker agent may choose to try and alter the makeup of their portfolio by changing their published tariffs or exercising optional controls that limit supply or demand from their portfolio.

In this paper, we first analyze the evolution of hourly electricity prices in a modern wholesale electricity market with the goal of predicting hourly forward prices for at least the next 24 hours so that a broker agent can effectively manage trading risk. We base our analysis on real market data from the Ontario Independent Electricity System Operator (IESO) from 2002 to 2011 and corresponding real weather data from the US National Climatic Data Center (NCDC). We study the overall characteristics of the prices using density estimation and k-means clustering. We then restrict our study to the year 2009 and apply regression and multi-class classification methods to estimate the changes in hourly prices based on a number of market- and weather-related covariates. We find a strong correlation of prices with historical prices, so we also extend the study to a time series analysis of only the prices. Our analysis shows that a combination of the multi-class classification approach and the multiplicative seasonal ARIMA model from the time series analysis can be used to predict the hourly forward prices with confidence. We then briefly discuss the analysis methods and results.

Second, we develop generative hierarchical models to simulate the demand for and supply of electricity from various types of retail customers (*e.g.,* households, universities, industrial plants). Since Smart Grid agent technologies are difficult to test in field projects without the risk of widespread disruption [3], sophisticated agent-based simulation platforms are an essential tool for such testing. Such simulation platforms include a large variety of customer models, whose demand and supply depend on various factors such as installed load capacity, household size, geographic locale, day of week, month of year, cloud cover and wind speed. We assume that a sample of data from real world metering or fine-grained simulation is available *a priori* for training and an additional small sample of data is made available online. We then generate more data, on behalf of a coarse-grained customer model agent, that simulates forward from the online data while borrowing characteristics from the training data. We do this using a novel combination of ARIMA and hierarchical Bayesian methodology, which also allows us to generate data for heterogeneous Smart Grid customers for whom we do not have sample data by incorporating our beliefs on how those new customers are different from other customers for whom we do have sample data.

The rest of this paper is structured as follows. Section 2 presents the analysis of wholesale market prices with subsection that describe the data in more detail, present the analysis methods and results, and discuss the methods and the results. Section 3 presents the analysis of customer model capacities with subsections that correspond to Section 2. In Section 4, we summarize our analysis and briefly describe future directions.

2 Wholesale Market Prices

The primary goal of a broker agent is to maximize its cumulative profit over all timeslots, $\sum_t \pi_t^{B_k}$. In order to achieve this goal, the broker agent must minimize the imposed balancing fee, by proactively trading in the wholesale market to counteract anticipated supply-demand imbalances in its portfolio. It is therefore critical that a broker agent be able to understand the evolution of prices in the wholesale market and be able to predict them. The liberalized Ontario electricity market is a good representative model for how such wholesale markets are expected to work. Trading in the market is conducted using a periodic double auction mechanism that is cleared once every hour throughout each day. During a given hour, trading is allowed in electricity that is intended to be consumed during the next H hours (typically, H=24.) So, in fact, the wholesale market conducts H simultaneous auctions to determine the clearing price for each of the H future timeslots.

Let $\omega_{t,t'}$ be the clearing price at a given timeslot, t, for a future trading timeslot, t', in the wholesale market. Let $q^{B_k}_{b,t,t'}$ and $q^{B_k}_{s,t,t'}$ be the quantity of electricity bought and sold respectively by broker agent B_k at timeslot t for each of the open trading hours t'. Equation 1 is then extended as:

$$\pi_t^{B_k} = p^{B_k}_{t,C}\kappa\Psi_{t,C} - p^{B_k}_{t,P}\nu\kappa\Psi_{t,P} + \sum_{t+1\leq t'\leq t+H} \omega_{t,t'}(q^{B_k}_{s,t,t'} - q^{B_k}_{b,t,t'}) - \phi_t|\kappa\Psi_{t,C} - \nu\kappa\Psi_{t,P}|$$

and the subgoal of the broker agent is to predict $\omega_{t,t'}$ under various market scenarios.

2.1 Data Description

The market data from the Independent Electricity System Operator (IESO) of Ontario, Canada from 2002 to 2011 and corresponding weather data from the US National Climatic Data Center (NCDC) contain several attributes relevant to our analysis. In addition to the actual Hourly Ontario Electricity Prices (HOEP), the attributes include: (i) hourly electricity demand, (ii) hourly wind power generation, (iii) daily operating reserve prices, (v) hourly uplift settlement charges, and (vi) hourly temperatures and wind speeds from various cities across Ontario. Samples from this dataset for the year 2009 are shown in Figure 2.

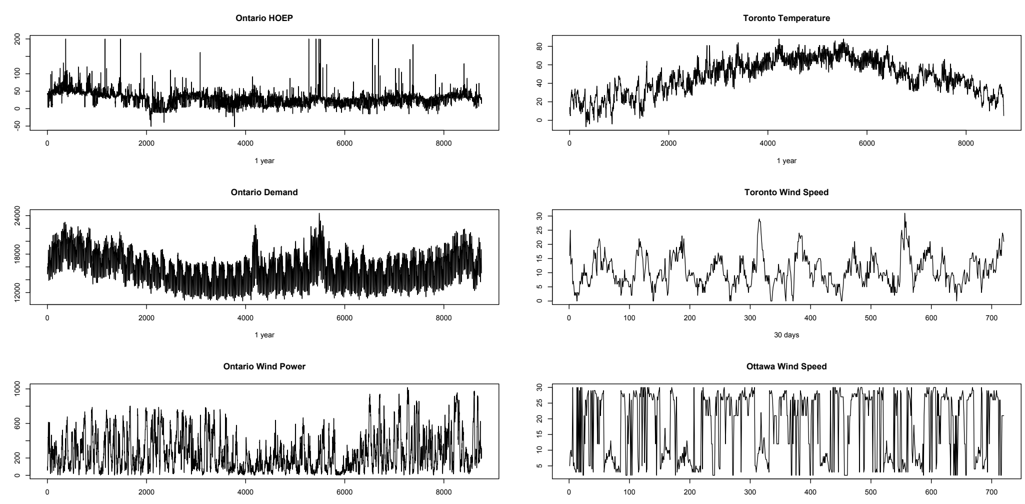

Figure 2: Samples from the cleaned data set: (i) (top-left) Ontario electricity prices (HOEP), (ii) demand, (iii) wind power generation, (iv) (top-right) Toronto temperatures, (v) Toronto wind speed, (vi) Ottawa wind speed. The first four plots are for all of 2009 while the last two are for the first 30 days of 2009.

2.2 Data Analysis

Over the past decade, the electricity markets have been going through significant restructuring. So we first attempt to characterize the prices from 2002 to 2010 to see if we can find evidence of the restructuring. We take median daily prices over this period and apply k-means clustering. Figure 3 shows the resulting clustering. We see some evidence of higher price volatility in the early years, but more notably, there is a significant downward shift in prices over the last 2-3 years.

We further characterize the range of prices using a kernel density estimate of the median prices as shown in Figure 4. We see that the prices are vastly skewed, and the skew would be even greater in the hourly, as opposed to median daily, prices. The negative prices, which we can't simply disregard as anomalous, make it difficult to apply typical transformations like taking the log of the prices to adjust for the skew.

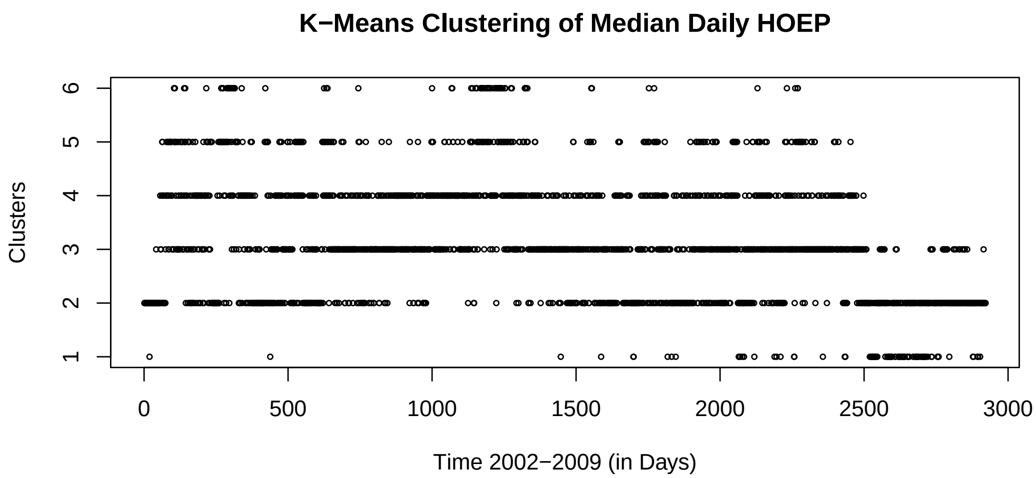

Figure 3: K-Means clustering of median daily HOEP. We see higher price volatility in the early years and lower and more stable prices in 2009.

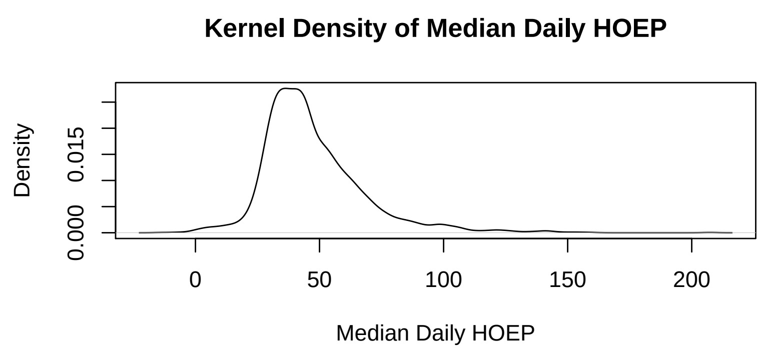

Figure 4: We see a very long right tail in the distribution, thus making it difficult to work with normality assumptions and the negative values rule out log transformation.

We restrict our subsequent analysis to the more stable price regime of hourly prices for 2009, which gives us 8760 samples. In Figure 5, we plot correlations for a subset of data features against the changes in hourly prices. The top-left plot in Figure 5 shows significant autocorrelation of the hourly price change with the hourly price change in the previous hour. We also note that Ontario demand changes and Toronto temperatures have some positive and negative correlations with historical HOEP. However, if we include all of these covariates in a multivariate regression, the effect of demand changes and Toronto temperature is subsumed by the predictive power of historical HOEP.

Since multivariate regression yielded limited success, we turn our attention to the easier problem of classification. Specifically, we aim for 3 target classes which would help inform the strategy of a broker agent in our problem domain. We label the classes as follows:

$$h(x) = \begin{cases} -1 & \text{if } x < \delta \\ +1 & \text{if } x > \delta \\ 0 & \text{otherwise} \end{cases}$$

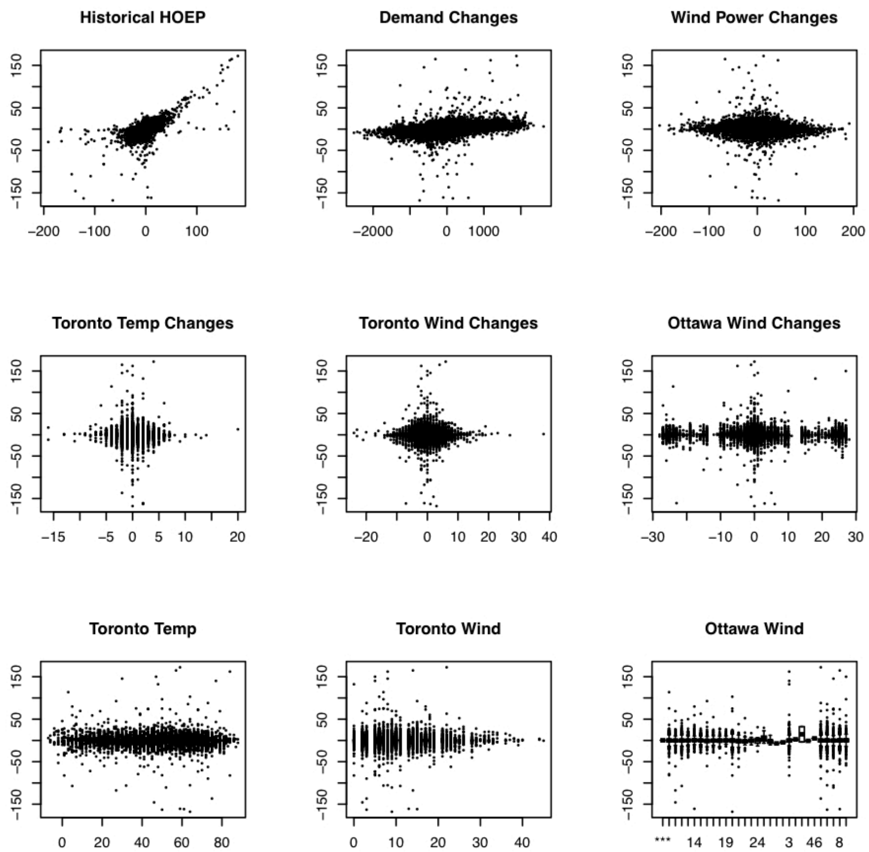

Figure 5: Correlation plots for various covariates, which include hourly changes and actual values of the raw data described in Figure 2, considered for regression.

where $x \in \mathcal{X}$ are the hourly price changes and δ is a threshold parameter that determines how much of a deviation from zero would be considered "no change" by the broker agent. This is an intuitive classification because minor price changes are common in real-time markets and are difficult to predict or explain, so they are not worth worrying about. The +1 and -1 labels focus the broker agent's attention on anomalous price predictions that likely require explicit action from the agent. For example, broker agents typically acquire some *controllable capacities* as part of their portfolio. Such capacities, *e.g.*, a household water heater, can be explicitly shut off by the service operator in response to the broker agent exercising its option to do so. When a broker agent determines that the price movement in the wholesale market is sufficiently unfavorable, it can include the exercise of such options in its overall strategy.

Under this classification setup, we perform several experiments using a one-vs-all 10-fold cross-validated multi-class SVM with radial basis kernels. Figure 6 shows some results. The x-axis of each panel represents *log* δ, ranging from -4 to +4, and the y-axis reports classification accuracy ranging from 30% to 100%. The SVM in the first panel only considers historical price changes as features, the second panel considers other market-based features, the third considers weather-based features and the fourth all of the features together. Each line in each panel represents the performance of the classifier on an entirely different test set. The combinations of features again confirms what we found in the correlation plots and regression–that historical prices are indeed the best indicator of future prices in the very short term future.

Note the distinctive *black swan* pattern in each panel. We find that classification accuracy dips drastically when $0 < log\ \delta < 1$. These dips show that price changes up to about 3 CAD/MWh are difficult to classify, although using just the historical prices, as in the first panel, we can achieve accuracy of about 65% even in this worst case scenario. With higher delta values, the classification accuracy climbs to full accuracy as would be expected. Therefore, if we use the classification only to identify the +1 and -1 labels for "large" price changes, we can achieve arbitrary classification accuracy by increasing the δ threshold. A suitable value for δ might be the one that gives a 95% confidence in the classification.

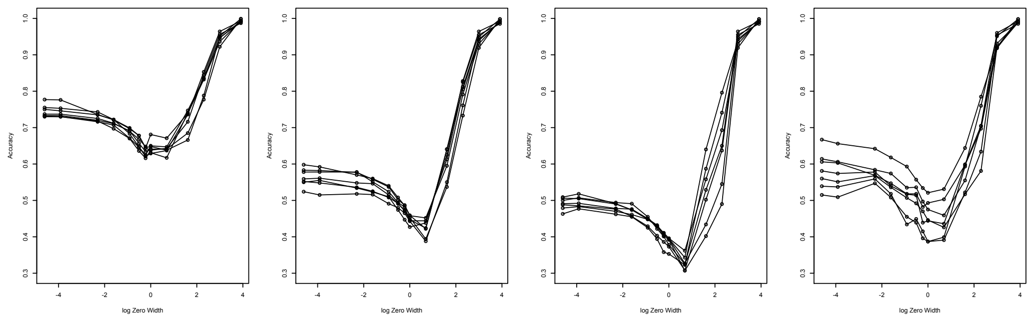

Figure 6: Classification results for three-label prediction of HOEP hourly changes vary significantly with changing δ-bands. The panels show results for different combinations of features: (i) historical HOEP changes, (ii) demand and wind power generation changes, (iii) toronto temperature and windspeed and Ottawa windspeed, (iv) all of the above.

Given that the SVM-based classifier helps identify the direction of the larger price changes, *i.e.*, the outliers, we now focus on modeling the more typical price changes. We pursue a time series analysis towards this goal. Figure 7 shows the 2009 hourly price changes and their autocorrelations.

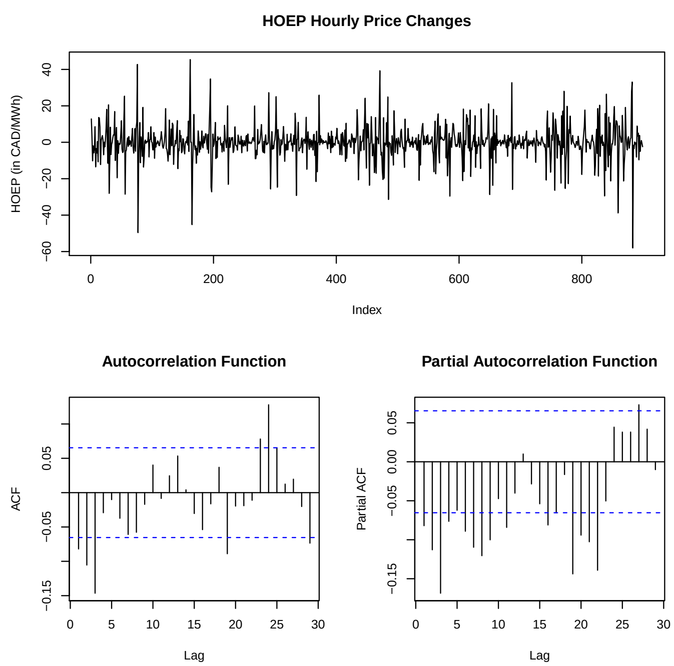

Figure 7: Time series and autocorrelations of 2009 Hourly Ontario Electricity Prices (HOEP).

We see that the series has some periods of higher variance which may warrant a GARCH model, but we will interpret this as stable variance and pursue a simpler ARIMA-only model. After some analysis we find that a ARIMA $(0,1,3) \times (0,1,1)_{24}$ multiplicative seasonal model fits fairly well. The plots for analyzing the residuals are shown in Figure 8.

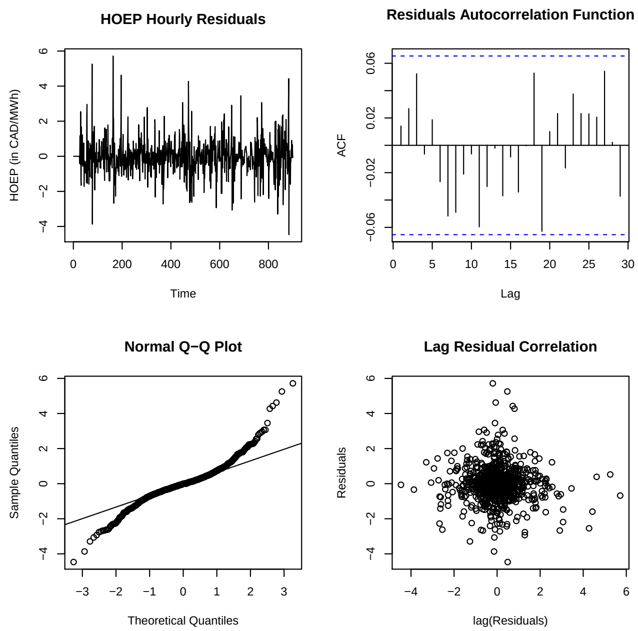

Figure 8: Residuals from Multiplicative seasonal ARIMA model fit on 2009 HOEP.

We note that the residuals at first appear to have non-constant variance but if we look at the Q-Q plot, we can see that the tails of the distribution are significantly heavier than would be expected of normally distributed errors; therefore the periods of apparent increased variance are likely just many occurrences of these *outliers*. The other two plots yield much more positive results in that we see no significant remaining autocorrelations in the residuals. We can therefore reasonably conclude that the proposed ARIMA model is a good fit when prices stay within a reasonable range, but we should not trust the model to predict the relatively frequent occurrences of larger price changes. The seasonal period of 24 hours makes intuitive sense since these are hourly prices and we expect correlations for the same hour from one day to the next. The resulting prediction model, illustrated for 72 hours in Figure 9 shows the periodic nature of the price evolution. This ARIMA prediction model can be combined with the outlier classification model to form a solution to our problem.

2.3 Discussion

We have the following ARIMA $(0,1,3) \times (0,1,1)_{24}$ model where Y_t is the hourly price at time t and e_t is the innovation or deviation at time t:

$$Y_t = Y_{t-1} + Y_{t-24} - Y_{t-25} + e_t + \varepsilon_t$$

where:

- $\varepsilon_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + \Theta_1 e_{t-24} + \theta_1 \Theta_1 e_{t-25} + \theta_2 \Theta_1 e_{t-26} + \theta_3 \Theta_1 e_{t-27}$
- $\theta_1 = -0.3034$, $\theta_2 = -0.2311$, $\theta_3 = -0.2200$
- $\Theta_1 = -0.9224$

It does a good job of explaining all the complex correlations apparent in Figure 7 and the seasonal component is logically sound because we know that we are analyzing hourly prices. However, the heavy tails in the original price series, and correspondingly in the residuals, e_t, are certainly a cause for concern. The theory of ARIMA models is based on a multivariate Gaussian assumption on the innovations, e_t, at each time step [4]. Our residuals seem to be better represented by a skewed Cauchy distribution but the theory based on Cauchy distributions is extremely difficult in

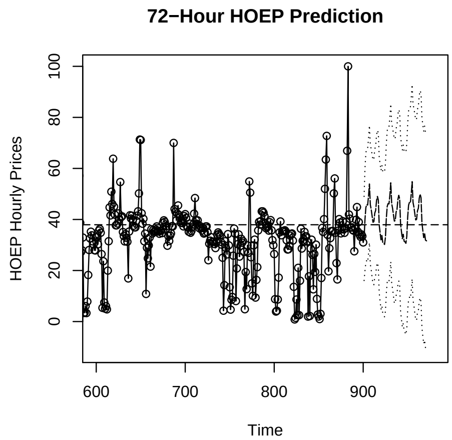

Figure 9: Typical prediction for the next 72 hours based on the ARIMA model.

such massive multivariate models. An adapted ARIMA model based on a t-distribution might also be a way to address the observed thick-tailed distribution. Conversely, we could imagine somehow thinning the tails by removing some of the outliers that we expect to be able to classify away using our SVM and then shifting and transforming the data to address the negative prices and the skewness. However, we are assuming that the spikes in the residuals are due to thick tails and not due to changes in variance, *i.e.*, heteroscedasticity. If indeed the series has non-constant variance, we would need to add a GARCH model to the innovations and look for correlations in volatility instead of simply assuming that the innovations are Gaussian white noise.

Lee, Lin and Wahba [11] present an interesting approach to single machine multi-category classification based on the asymptotic convergence of the binary SVM classifier to the Bayes-optimal solution; i.e., the minimizer of the loss functional, $E[(1 - yf(x))_+)]$ tends to (*sign(p(x)* - $1/2$) as the number of data points goes to infinity. They claim that this single machine approach results in fewer support vectors in the derived margin classifier. On the other hand, Rifkin and Klautau [14] present a vigorous defense of one-vs-all classification. They point out that the claim of convergence to the Bayes-optimal solution of Lee, *et al*, ignores the regularization term, $\lambda\|f\|_K^2$, and the requirement that the functional f must be contained in the RKHS $\mathcal{H}_K$. They also present empirical results showing that one-vs-all classification with well-tuned binary classifiers performs as well or better than single machine multi-category classifiers and also other approaches such as those based on error-correcting codes.

3 Customer Model Simulation

A key challenge in implementing new technologies for the Smart Grid is the difficulty of testing such technologies on the real grid. Therefore, it is necessary to implement sophisticated agent-based simulation environments that can model various aspects of the Smart Grid truthfully and efficiently [9]. The scale of Smart Grid simulations is often a road block for efficient simulation. Typically, such simulations need to represent tens or hundreds of thousands of customers consuming or producing power under various tariffs offered by competitive broker agents in the market. Customers can vary along the following dimensions:

- ModelType $\in$ {Individual, Population}
- EntityType $\in$ {Residential, Commercial, Industrial}
- CapacityType $\in$ {Consumption, Production, Storage}

The need for truthful simulation mandates the use of fine-grained agent representations that model individual persons and/or appliances explicitly. However, such fine granularity means that large-scale simulations cannot be run without very large amounts of computing hardware and sophisticated programming. It would be highly beneficial to replicate the aggregate behavior of large groups of customers using higher level statistical models while maintaining similarity to the emergent behavior of corresponding fine-grained customer models.

Moreover, it would be beneficial to develop a reusable statistical framework that can be instantiated differently and fitted with different parameters to approximate the behavior of many heterogeneous customers without having to program each model from scratch. For example, we would like to model (i) an office building that only consumes power, (ii) a university campus that consumes power and also has some storage capacity, and (iii) a chemical plant that consumes power and also has some power production capability, using essentially the same model or code configured with different parameters.

We develop such a framework using a novel combination of ARIMA and hierarchical Bayesian [15] [2] methodology. We apply the framework to a specific sample data set from a fine-grained household customer model and study the process of fitting a model to that data. We also analyze the data generated by the resulting coarse-grained model to compare it with the sample data using several divergence metrics. Finally, we demonstrate some examples of how the coarse-grained model can then be reconfigured with subjective beliefs to generate alternate customer models.

3.1 Data Description

Figure 10 presents the available data; it consists of two time series, each representing the consumption or demand capacity of two villages. The two series are generated using a fine-grained household customer model (HCM) which is based on real survey and metering data from the MeRegio project, a limited-scale early Smart Grid deployment in Baden-Wurttemberg, Germany [8][7]. The HCM represents individual persons and appliances of each household as discrete stochastic nodes and generates the aggregate series of Figure 10 by drawing from each of those nodes independently.

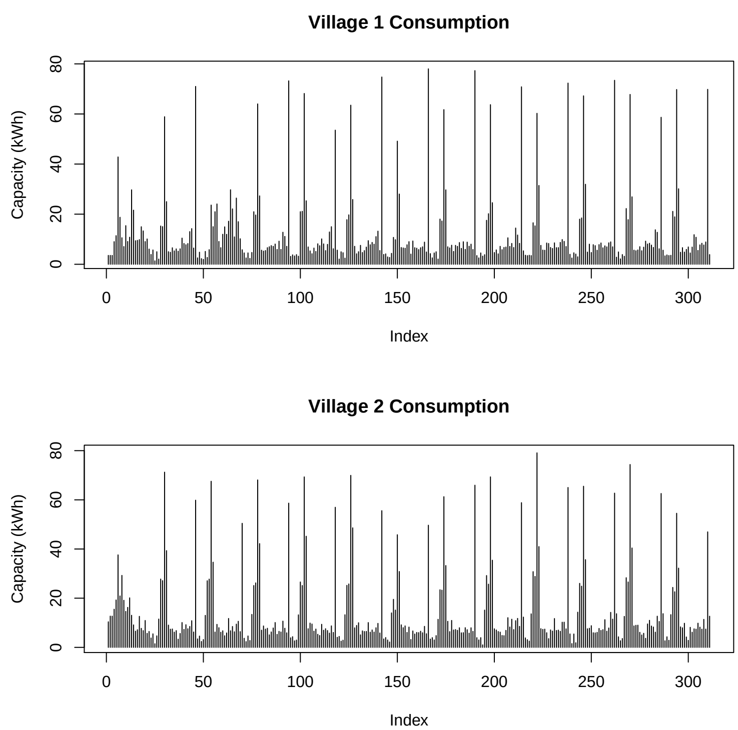

Figure 10: Demand from a fine-grained household customer model simulating two small villages.

The time series for each village includes discrete consumption capacity measurements over 312 hours, *i.e.*, 13 days. Note the presence of two daily spikes in the morning and evening hours of each day in both series. Each discrete timeslot is also labeled with appropriate hour-of-day and day-of-week labels. These labels form two natural groupings: (i) hours labeled 1 to 24 with 13 samples per label, and (ii) days labeled 1 to 7 with only 1 or 2 samples per label. We explore these groupings further for random effects in our data analysis.

A Box-Cox power transformation analysis suggests that the data be transformed using *log()* or *sqrt()* to achieve normality. We choose the *log()* transformation based on the Q-Q plots for each of the two transformations. Figure 11 summarizes the log-transformed data using kernel densities and boxplots.

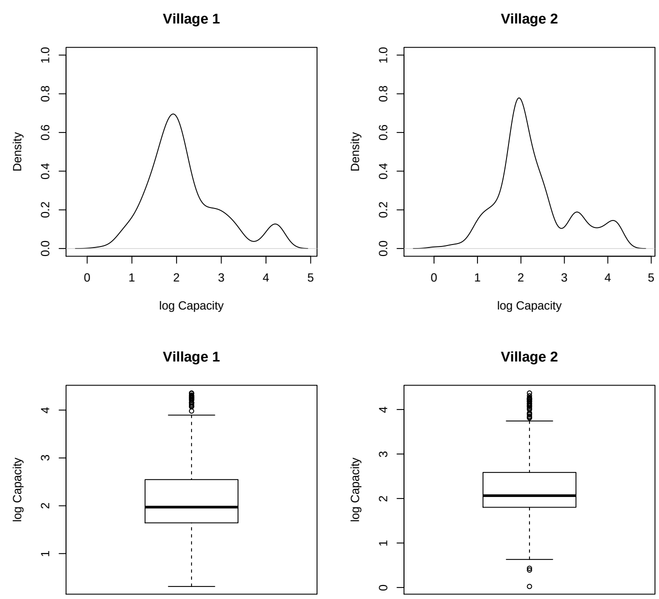

Figure 11: Kernel density plots and boxplots comparing the demand capacity from the two villages.

Furthermore, we evaluate the time series characteristics of the data in Figure 12. We see that the log-transformed series appears stationary with some seasonality. The ACF and PACF highlight significant coefficients for moving average (MA) components at lags of 1 and 24 and similarly the PACF highlights significant autoregression (AR) components also at lags 1 and 24.

We find that a multiplicative seasonal ARIMA model of the form $(1, 0, 1) \times (1, 0, 1)_{24}$ best fits the log-transformed data under the ARIMA methodology. The resulting diagnostic plots in Figure 13 show reasonably normal residuals with slightly higher variance in the first half of the series. The resulting ACF on the residuals shows no significant autocorrelation and the Ljung-Box statistic rejects the hypothesis that the lags are correlated up to lag 21 but not for further lags.

3.2 Data Analysis

We assume that we are to generate data for a simulation based on the full time series for village 1 and a small initial fragment of the time series for village 2. We will then test the generated data against the remainder of the village 2 time series for prediction accuracy. This is analogous to a situation where the village 1 time series is used to train our model *offline* and the initial subset of the village 2 time series is used to bootstrap the online prediction part of the model. Our goal here is to provide long range forecasting for simulation; therefore there is no ground truth available online to periodically recalibrate the forecast.

Concretely, let Y be a training time series representing the log-transformed production or consumption capacity of a Smart Grid customer model over N timeslots. Let $Z = Z_t$, where $t = 1..M$ and $M \ll N$, be a partial time series available to bootstrap our prediction model. Let Z_t^* for $t > M$

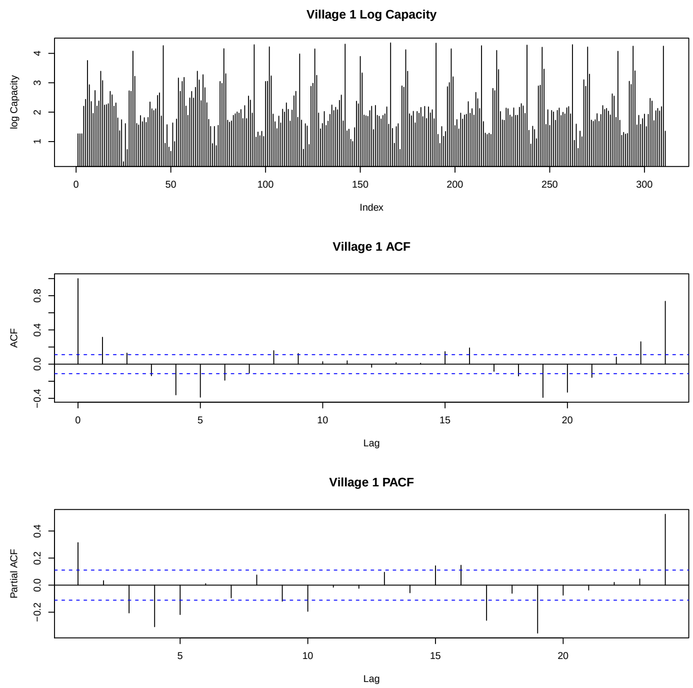

Figure 12: Time series characteristics of the demand capacity of village 1.

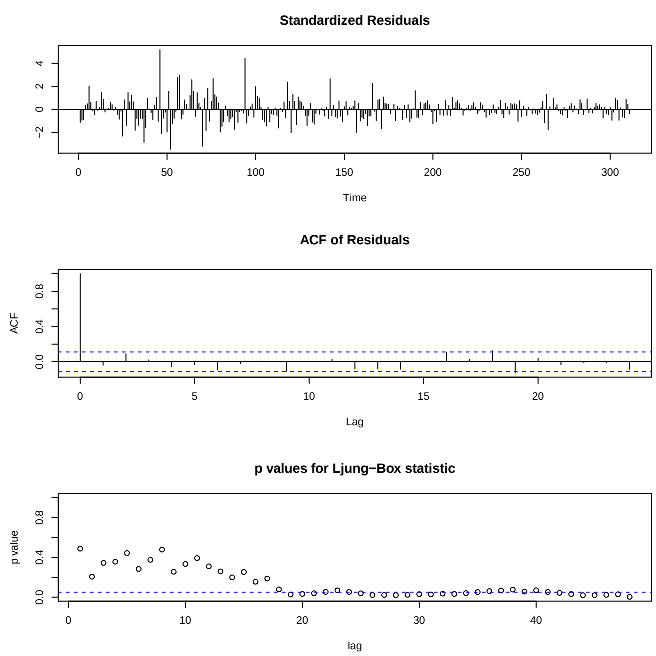

Figure 13: Residual diagnostics for a $(1, 0, 1) \times (1, 0, 1)_{24}$ fit of the demand capacity of village 1.

be the remainder of the Z series to be used for evaluating prediction accuracy. Let D be the vector identifying the day of the week for each timeslot and let H be the vector identifying the hour of the day for each timeslot.

3.2.1 ARIMA Analysis

The $(1, 0, 1) \times (1, 0, 1)_{24}$ seasonal ARIMA model is of the form:

$$Y_t = Y_0 + \phi_1 Y_{t-1} + \Phi_1 Y_{t-24} + e_t + \theta_1 e_{t-1} + \Theta_1 e_{t-24} + \theta_1 \Theta_1 e_{t-25}$$

The fitted ARIMA coefficients are as follows:

	Y_0	ϕ_1	Φ_1	θ_1	Θ_1	$\theta_1\Theta_1$
SARIMA	2.1661	0.6162	-0.4473	0.9888	-0.7343	0.3285

When we use the fitted ARIMA model for prediction, we encounter a characteristic problem with ARIMA forecasting whereby the autoregressive components of the fitted model deteriorate over time leading to forecasts that revert to the mean eventually. This behavior is apparent in the top panel of Figure 14, which forecasts based on the full Y_t series, and more emphatic in the bottom panel, where the available time series, Z_t, is short and also doesn't capture the two-spiked daily cycle. We would like to use daily spike information from Y_t as a *prior* in predicting from Z_t but the mechanism for doing so is not obvious in ARIMA methodology. We could concatenate the two series but that can lead to inconsistencies at the concatenation point, so we do not pursue that approach.

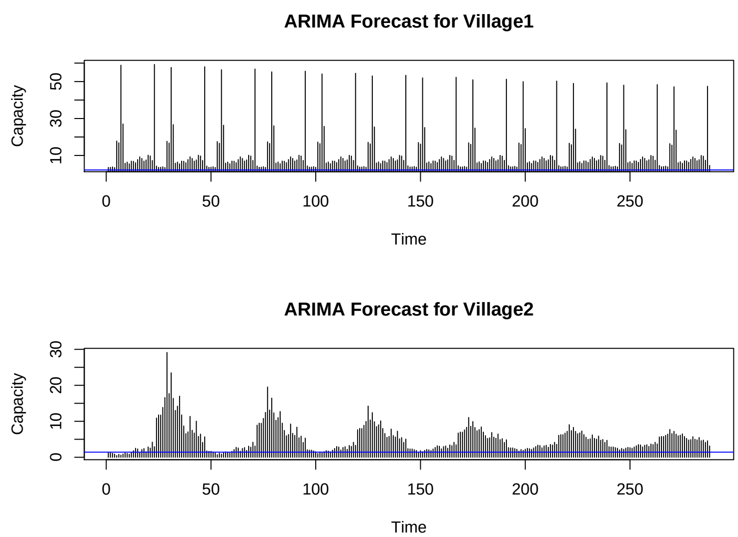

Figure 14: Forecasts from ARIMA models based on the full Village 1 time series and the training subset of the Village 2 time series respectively.

3.2.2 LM/LMER Analysis

For comparison, we fit coefficients to terms corresponding to those identified in the seasonal ARIMA model using plain linear regression. Furthermore, we also define a multilevel model, using LMER [12], with random effects for the day of week and hour of day.

$$Y_t \sim 1 + Y_{t-1} + Y_{t-24} + e_{t-1} + e_{t-24} + e_{t-25} + (1|D) + (1|H)$$

We find that the fixed effects coefficients, shown in the table below, vary significantly amongst the three models. Forecasts using the coefficients from LM and LMER worsen the problem of reversal to the mean that we encountered with the seasonal ARIMA model.

	1	Y_{t-1}	Y_{t-24}	e_{t-1}	e_{t-24}	e_{t-25}
SARIMA	2.1661	0.6162	-0.4473	0.9888	-0.7343	0.3285
LM	0.3061	0.3102	0.5489	-0.1775	0.2485	0.1991
LMER	1.4635	0.3390	-0.0180	-0.1303	-0.0405	0.0836

We further find that only a few of the hour of day random effects are statistically significant and the day of week random effects are forced to zero. As an alternative approach to estimating the random effects, we pursue a model based on hierarchical Bayesian methodology.

3.2.3 Hierarchical Bayesian Model

Given that we want to use information from the Y series to help predict from the Z series, a Bayesian approach seems natural. The hierarchical Bayesian model [6] we fit using Gibbs sampling is described below in Equations 2-17, but we note that this forms only a subset of the overall model we propose to address the goals set out in Section 3.

$$Y_t \sim N(\widehat{Y1}_t, \sigma^2);\ t = 27..N \tag{2}$$
$$Y2_t \sim N(\widehat{Y2}_t, \sigma^2);\ t = 27..N \tag{3}$$
$$\widehat{Y1}_t \leftarrow Y_0 + Y_d[D[t]] + AR_t + MA_t \tag{4}$$
$$\widehat{Y2}_t \leftarrow Y_0 + Y_h[H[t]] + AR_t + MA_t \tag{5}$$
$$AR_t \leftarrow \phi_1 Y_{t-1} + \Phi_1 Y_{t-1} \tag{6}$$
$$MA_t \leftarrow \theta_1(Y_{t-1} - Y_{t-2}) + \Theta_1(Y_{t-24} - Y_{t-25}) + \theta_1\Theta_1(Y_{t-25} - Y_{t-26}) \tag{7}$$

$$Y_0 \sim N(\widetilde{Y_0}, \sigma^2_{Y_0}) \tag{8}$$
$$\phi_1 \sim N(\widetilde{\phi_1}, \sigma^2_{\phi_1}) \tag{9}$$
$$\Phi_1 \sim N(\widetilde{\Phi_1}, \sigma^2_{\Phi_1}) \tag{10}$$
$$\theta_1 \sim N(\widetilde{\theta_1}, \sigma^2_{\theta_1}) \tag{11}$$
$$\Theta_1 \sim N(\widetilde{\Theta_1}, \sigma^2_{\Theta_1}) \tag{12}$$
$$\sigma \sim \mathit{Unif}(0, 100) \tag{13}$$

$$Y_d[d] \sim N(0, \tau^2);\ d = 1..7 \tag{14}$$
$$\tau \sim \mathit{Unif}(0, 100) \tag{15}$$

$$Y_h[h] \sim N(0, \eta^2);\ h = 1..24 \tag{16}$$
$$\eta \sim \mathit{Unif}(0, 100) \tag{17}$$

$$Y^*_t \sim N(\widehat{Y1}_t, \sigma^2);\ t = 27..N \tag{18}$$

Before presenting the results of using this model, we briefly describe the steps taken to arrive at this model. Our initial step was to fit coefficients, using Gibbs sampling, for the ARIMA features, *i.e.*, $Y_{t-1}, Y_{t-24}, e_{t-1}, e_{t-24}, e_{t-25}$ with a single intercept. We note that the fitted coefficients match those fitted by `LM` if the e_{t-25} term is omitted in the `LM` model.

We then introduced the Y_d and Y_h terms using the following in place of Equations 4 and 5:

$$\widehat{Y1}_t \leftarrow Y_0 + Y_d[D[t]] + Y_h[H[t]] + AR_t + MA_t$$

However, this approach does not work well because of the labeling ambiguity between the Y_d and Y_h terms, thus leading to very high standard deviations for the coefficients of both of those terms. So, instead we employ the mechanism of Equations 4 and 5, where we duplicate the original time time series Y into a replica $Y2$. We then estimate coefficients for Y_d and Y_h separately using each of the two replicas. We acknowledge that we may be overestimating the coefficients for Y_d and Y_h by fitting them separately, but we address that concern further below when we describe how we utilize those coefficients.

Y, $Y2$, D, H and N are provided as input to the model. Most of the remaining symbols in Equations 2-17 are self-describing based on earlier discussion, except the prior means and variances on the saved variables $Y0$, ϕ_1, Φ_1, θ_1 and Θ_1. The prior means are set equal to the coefficients estimated by the seasonal ARIMA model described in Section 3.2.1. For the corresponding variances, we try two models: (1) the variances are also set equal to those estimated by the seasonal ARIMA Model, and (2) the variances are set be very high thus making the priors quite vague. We compare the results of these two alternatives further below.

As a posterior predictive check for our fitted hierarchical Bayesian model of Equations 2-17, we use the Y_t^* simulated time series generated in Equation 18. We use two statistics, kernel density estimates and IQR/boxplots, to see if the original series, Y_t, could have been generated by this model. The results are shown in Figure 15. We can plainly see that the kernel density for Y is significantly higher in the 5-15 log Capacity (in kWh) range than for Y^*. We also see in the boxplot in the bottom panel that Y has lower variance and smaller range than Y^*. So, we conclude that the model of Equations 2-17 is not sufficient to meet our goals.

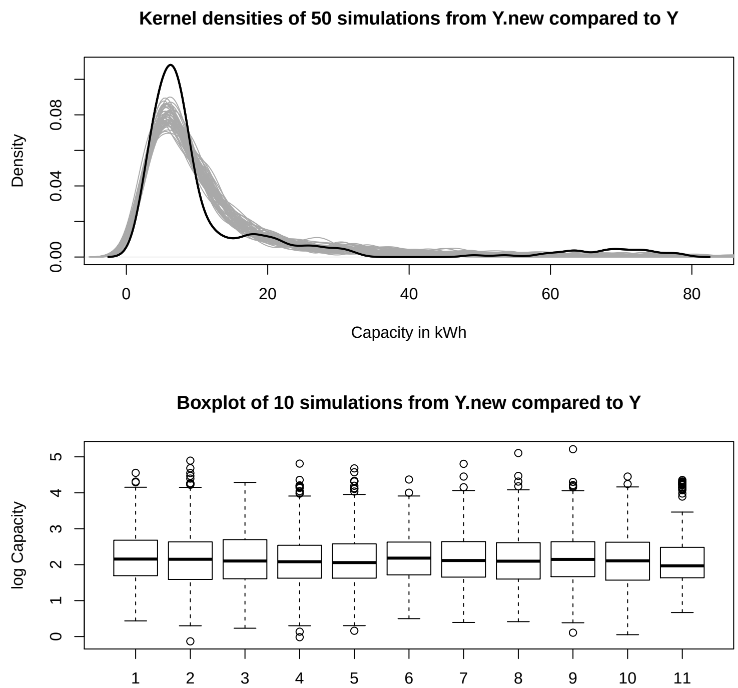

Figure 15: Kernel density plots and boxplots comparing the original time series Y to simulated samples Y^* show that the model of Equations 2-17 is a poor fit.

We can also conclude that the model is poor or insufficient given the plots shown in Figure 16. The top panel shows the original time series Y. The middle panel shows the predicted values for $\hat{Y}$ given the true historical values for Y. The bottom panel shows the forecast $Y_f = \hat{Y}$ derived by using the forecast Y_f values as historical values for $t > 26$ instead of the true Y values in Equations 6 and 7:

$$AR_t \leftarrow \phi_1 \widehat{Y}_{t-1} + \Phi_1 \widehat{Y}_{t-1} \tag{19}$$

$$MA_t \leftarrow \theta_1(\widehat{Y}_{t-1} - \widehat{Y}_{t-2}) + \Theta_1(\widehat{Y}_{t-24} - \widehat{Y}_{t-25}) + \theta_1\Theta_1(\widehat{Y}_{t-25} - \widehat{Y}_{t-26}) \tag{20}$$

3.2.4 Augmented Bayes Forecast

Comparing the forecasts from the ARIMA methodology in Figure 14 and from the hierarchical Bayesian model in the bottom panel of Figure 16, it might seem like we have not gained much, and

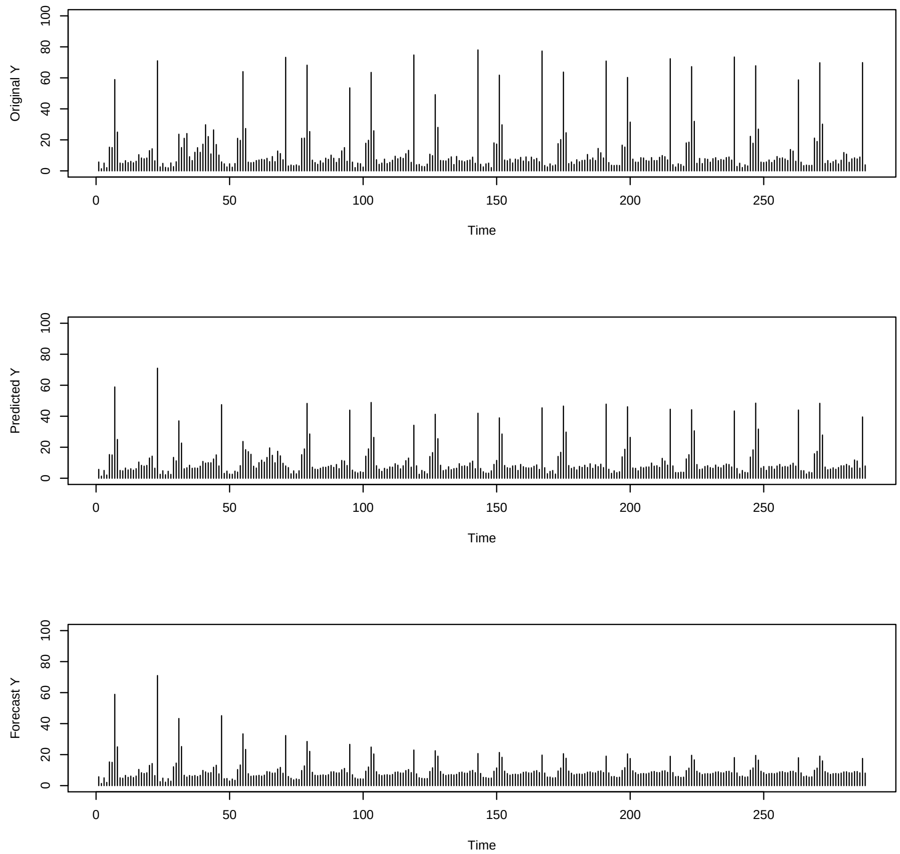

Figure 16: The original time series Y is shown in the top panel. The middle panel shows predictions from the hierarchical model using true historical values of Y. The bottom panel shows the forecast from Y using the forecast values themselves as historical values.

lost significantly in computational complexity, by using the latter methodology. However, applying the hierarchical Bayesian has given us some tools that we can use to further enhance the model.

Specifically, we now have the fitted coefficients for Y_d and Y_h which capture random effects or group intercepts for daily and hourly variations in the data. From visually examining Figure 16 and realizing that the fitted coefficients for the AR components of the model are lower than those in the ARIMA model, we hypothesize that *augmenting* the daily and hourly variations in the hierarchical Bayesian model will improve our forecast.

Furthermore, we note that the forecast deteriorates exponentially with time. So, we propose *augmenting* Y_f, our forecast from the hierarchical Bayesian model, by adding a log term that includes a weighted combination of the daily and hourly intercepts, Y_d and Y_h. The rationale for using a weighted combination takes into account our previously noted observation that Y_d and Y_h are both likely to be over-estimates given that we estimated them separately, thus treating the intercepts from the other grouping as equal to zero. Equation 21 summarizes the model described in detail in Equations 2-17. Equations 22-25 then describe how to extend the model to include the log term which utilizes Y_d, Y_h and timeslot t.

$$Y_t^f \leftarrow Y_0 + Y_d[D[t]] + Y_h[H[t]] + AR_t + MA_t \tag{21}$$

$$Y_t^{bf} \leftarrow Y_t^f + \lambda \frac{log(t-26)}{log(N1-26)}((1-\gamma)Y_d[D[t]] + \gamma Y_h[H[t]]) \tag{22}$$

$$\lambda^*, \gamma^* \leftarrow \underset{\lambda,\gamma}{\text{argmin}}\ KL_D(f_K(Y), f_K(Y^{bf})) \tag{23}$$

$$\lambda^*, \gamma^* \leftarrow \underset{\lambda,\gamma}{\text{argmin}}\ \sum_t (Y_t^{bf} - Y_t)^2 \tag{24}$$

$$Z_t^{bf} \leftarrow Z_t^f + \lambda^* \frac{log(t-26)}{log(N2-26)}((1-\gamma^*)Y_d[D[t]] + \gamma^* Y_h[H[t]]) + N(0,\sigma^2) \tag{25}$$

The weighted combination of Y_d and Y_h is determined by a parameter γ. The combination is then multiplied with a normalized timeslot factor, $log(t-26)/log(N1-26)$, where N1 is the length of Y, further multiplied by a scaling factor λ.

We then define choosing the right values for λ and γ as an optimization problem that minimizes the "distance" between Y and the augmented model forecast Y^{bf}. The distance need not be symmetric, so we can use an asymmetric measure such as KL-divergence or a symmetric distance such as a point-wise sum of squares over the timeslots in the forecast. In order to be able to use KL-divergence, we need probabilities to compare, so we obtain kernel density estimations, f_k, for Y and Y^{bf} and use those in the comparison. Equations 23 and 24 represent these two options.

In our experiments, we use a brute-force search over combinations of λ and γ:

$$\lambda \leftarrow i = 1...100;\ i \in \mathcal{I} \tag{26}$$

$$\gamma \leftarrow 0.01j;\ j = 1...100;\ j \in \mathcal{I} \tag{27}$$

Figure 17 shows the two surfaces representing the divergence metric computed by the two methods. We observe that the surface in the right panel corresponding to the sum of least squares method is smoother and therefore a candidate for a more efficient optimization method such as gradient descent or the Gauss-Newton method. In our experiments, both methods take approximately the same amount of computational time and yield nearly the same values for the optimal parameters, λ^* and γ^*. However, we anticipate that other training samples for Y may yield results with greater difference and therefore present both options here. It is also possible to combine the estimates from both options to compute the values for λ^* and γ^*.

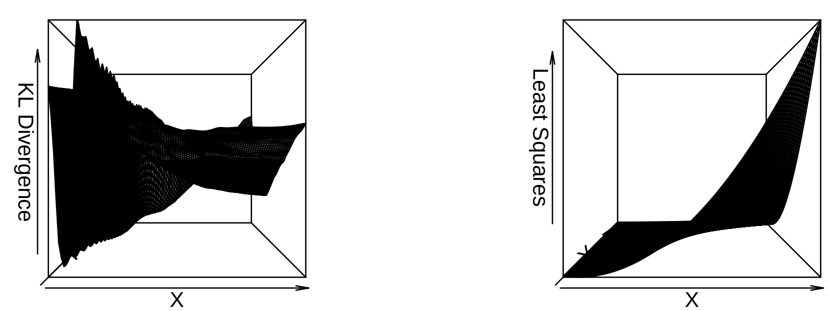

Figure 17: Surfaces representing the "distance" between the original series and the forecast using (i) KL-divergence (left panel), and (ii) sum of least squares over forecast timeslots (right panel).

λ^* and γ^* are then used in Equation 25 to generate an augmented forecast for the online time series Z. Equation 25 replicates the structure of Equation 21 with $N2$ representing the forecast horizon or length of Z^*. It also adds Gaussian noise with variance σ^2 where σ is a parameter fitted by the hierarchical Bayesian model in Equations 2 and 3.

In Figure 18, we analyze the fit of the forecast Y^{bf} using kernel density plots and boxplots as we did in Figure 15. We now see that the original series Y appears to fit well with the simulations for the forecast model for Y^{bf}. Furthermore, Figure 19 shows how the augmented Bayes method we propose improves upon the forecast from the plain hierarchical Bayes method. The top panel shows the true series, Z, the middle panel shows the forecast, Z^f, from the unaugmented hierarchical Bayes method and the bottom panel, which shows the augmented forecast Z^{bf}, illustrates the absence of the reversal to the mean pattern that we see with Z^f.

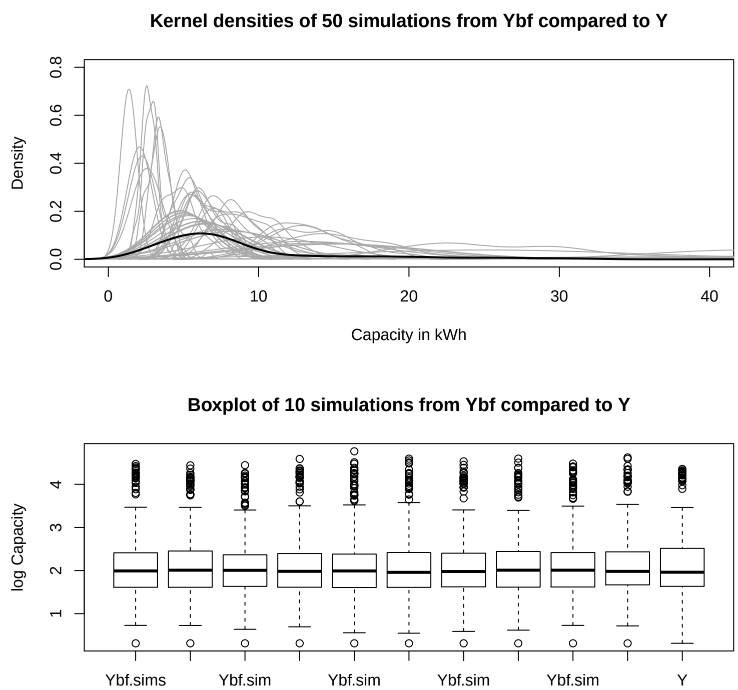

Figure 18: Kernel density plots and boxplots comparing the original time series Y to simulated samples from our forecast model, Y^{bf}, show that the extended model of Equations 1-23 fits well.

Figure 20 shows the difference in the forecast from the training series Y versus the forecast from the test series Z. We see that the forecasts for the two series disagree more in the earlier timeslots where Z exerts greater influence while the difference disappears over time as Y exerts more influence later in the forecast. This demonstrates that we meet the goals set out in Section 3 whereby we wanted to use information from the training series Y as a prior but also take into account the available portion of the online test series Z in making a forecast for Z.

Figure 21 shows the performance of the various forecasting methods discussed thus far using a different metric. The x-axis plots *tolerance*, a threshold for the percentage error beyond which the forecast consumption capacity for a future timeslot is considered wrong. For example, at the 20% tolerance level, given that the true capacity at some timeslot t' is $Z^*_{t'}$, then a forecast that equals $1.2Z^*_{t'}$ would be classified as 1 whereas $1.21Z^*_{t'}$ would be classified as 0. The y-axis measures *accuracy* which is the percentage of timeslots over the forecast horizon which are classified as 1.

The solid line shows the forecast from the online test series Z using conventional ARIMA methodology as presented in Section 3.2.1 and serves as a reference benchmark. The four remaining lines

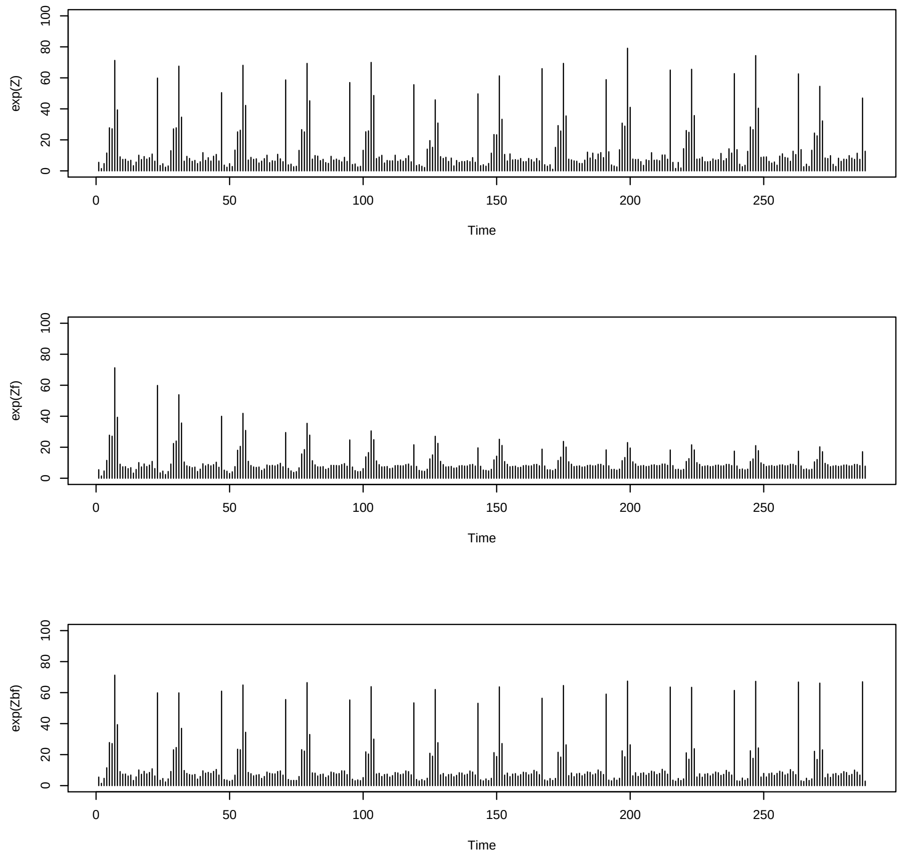

Figure 19: The top panel shows the test series, Z. The middle panel demonstrates the characteristic reversal to the mean of the ARIMA model that also afflicts the plain hierarchical Bayesian model. The bottom panel shows the elimination of that effect using the augmentation method of Equations 21-24.

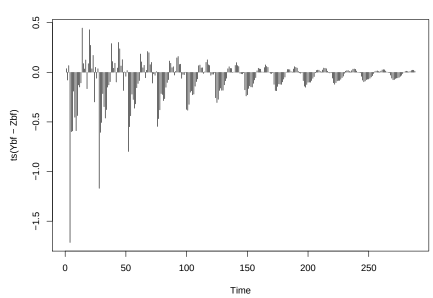

Figure 20: Difference in the forecast from the training series Y versus that from the test series Z.

are all based on the hierarchical Bayesian model of Equations 2-17. The red dashed line presents the accuracy plot for the training forecast, Y^f, using the variances from the seasonal ARIMA model fit as the variances for the priors in the hierarchical Bayesian model in Equations 8-12. The blue dotted line presents the equivalent plot for Y^f with those variances set at 10^4 thus providing very little information in the priors. We can see that the ARIMA priors perform slightly better at lower

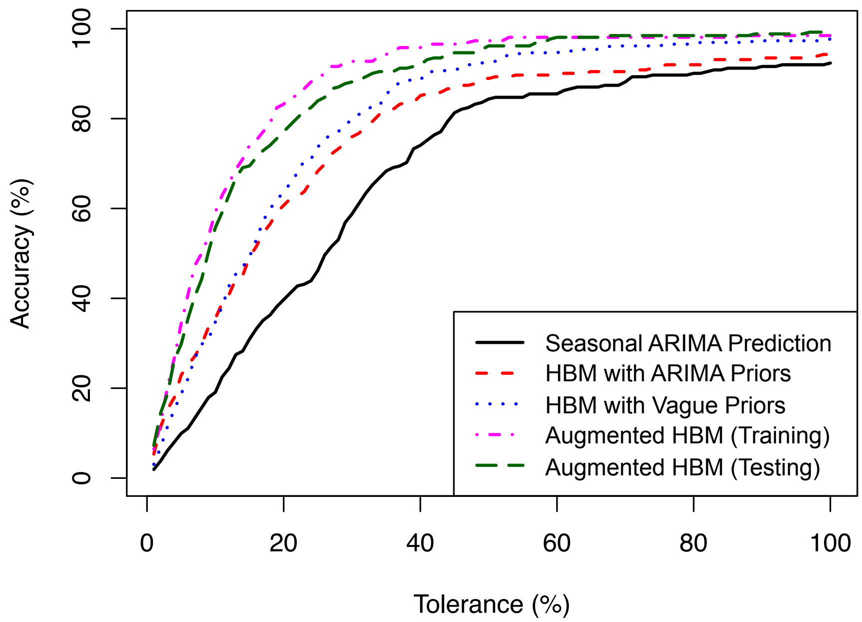

Figure 21: Plots showing accuracy of the various forecsasting methods at a range of error-tolerance levels; we see that the augmented Bayes method of Section 3.2.4 outperforms the other methods.

tolerance levels and the vague priors perform better at higher tolerance levels but the difference is marginal. The magenta dash-dotted line presents the accuracy plot for Y^{bf} using the augmentation methodology of Equations 22-25; it universally performs better than the three previous methods. Finally, the green dashed line presents the accuracy plot for the testing forecast Z^{bf} using the augmentation methodology; this performs slightly worse than the corresponding training forecast Y^{bf} as is to be expected, but performs universally better than the other three methods with the largest differences in the critical 10-40% tolerance range.

3.3 Discussion

A significant benefit of using the hierarchical Bayesian methodology for our framework is that we can provide additional subjective beliefs on the expected forecast for a given time series. We can provide these bias factors as priors on the daily and hourly weights, Y_d and Y_h, for example. Alternately, we can impose them *a posteriori* as shown in Equation 28 where the subjective forecast, Z^{sf}, skews the augmented forecast, Z^{bf}, using hour-, day- and month-specific weights relative to the capacity at the start, t_0, of the time series.

$$Z_t^{sf} \leftarrow Z_t^{bf} \; \omega_H(t) \; \omega_D(t) \; \omega_M(t) \tag{28}$$

These weights may be computed taking into account additional factors not described in our model such as the typical occupancy profile of households in a region or daily temperature variations. Figure 22 illustrates a concrete example where the top panel shows a static set of *prior* hourly weights based primarily on the occupancy profiles of households in the region represented by the model. The bottom panel shows a modified set of weights that are based on the prior weights of the top panel and also the daily variations of temperature in the region. The hourly temperature values can be expected values for the time of the year or dynamic values predicted for the next hour by a real-time weather service. Similarly, we can provide daily weights that predict higher than average capacity on Saturdays but lower than average on Sundays, for example. Figure 23 shows the output of such a skewed forecast in the middle panel; the top panel shows the unskewed model forecast.

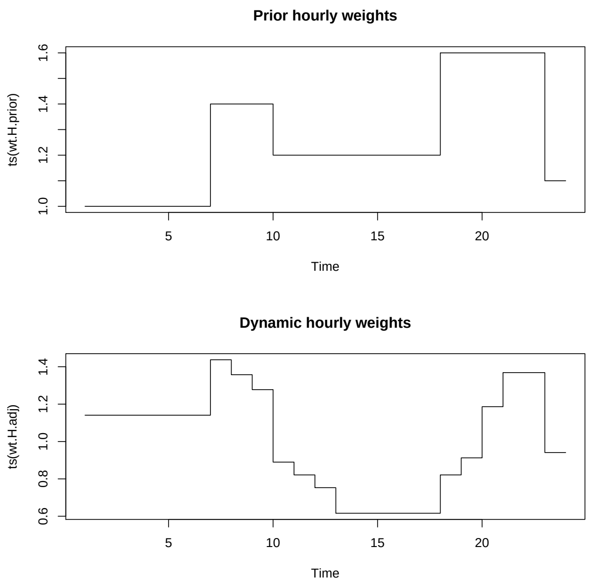

Figure 22: Subjective hourly weights that can be used as priors or imposed *a posteriori* on the forecast, Z^{bf}, to further skew the forecast based on exogenous factors. The top panel shows a static prior profile of weights which can be modified dynamically using real-time information.

Another example concerns exogenous factors that we do not expect to be captured by the training series, Y. In our scenario, the consumption capacity of a set of households is likely dependent on the *tariff rates* or electricity prices that they are subject to. In other words, we may expect that if tariff rates are lower, the households will consume more electricity on average. We can reflect such an assumption by modeling the *elasticity* [13] of consumption capacity to various electricity rates, V, relative to a benchmark rate, V^*, as shown in Equation 29.

$$Z_t^{ef} \leftarrow Z_t^{sf} + elasticity(t,\, V_t,\, V^*) \tag{29}$$

Note that the elasticity factor may be negative, for example if electricity rates become less favorable. The bottom panel of Figure 23 shows the layered model of Equation 29 where consumption is lower in the second half of the series because of higher tariff rates. Over time, we would then endogenize these additional factors into the Bayesian model and apply the associated biases *a priori* so that they can be modulated by the data.

4 Conclusion

We have analyzed data related to two interesting problems within the Smart Grid domain using timeseries methods layered with additional machine learning techniques. We first analyzed real data from a representative wholesale electricity market along with corresponding weather data to build a prediction model for hourly price changes for over 24 hours into the future. This prediction model uses a multiplicative seasonal ARIMA model for usual price patterns and a 3-label SVM-based classification model to predict the likelihood of larger price changes in the positive or negative direction. While the magnitude of these larger changes cannot be predicted by the SVM, we claim that such predictions are difficult to model without more extensive knowledge of exogenous non-repeating factors that may be causing those spikes. We then analyzed simulated data from a fine-grained household consumption model to learn how a coarse-grained model can approximately simulate a timeseries that replicates essential characteristics of the given data. We emphasize that our focus in this work is not to recover the true parameters that were used to generate the fine-grained data and

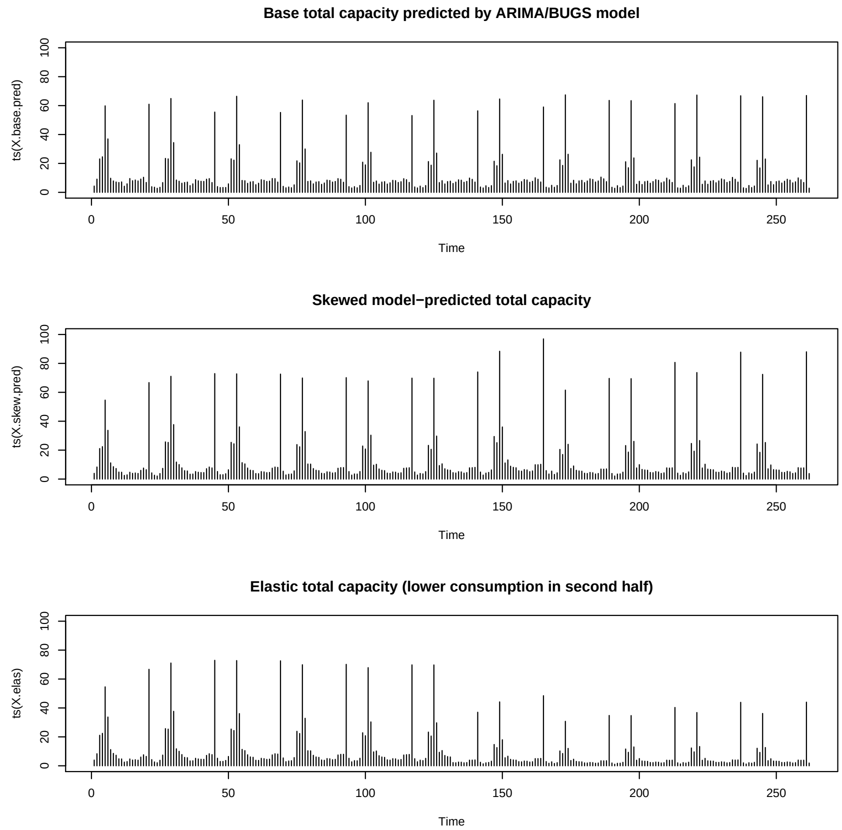

Figure 23: Factors not considered in the augmented model can be used to inform the priors of the hierarchical Bayesian model, as with the hourly weights and daily weights demonstrated in the middle panel, or layered on top as with the elasticity to tariff rates as shown in the bottom panel.

neither is our focus on trying to model parameters representing specific individual person or appliance behaviors in the population. We instead focus on (i) modeling generic higher-level features that can represent a broad set of customer models, (ii) determining appropriate hierarchical models based on those features, and (iii) fitting the coefficients or distribution assumptions for those features based on specific data that we are trying to replicate. We propose that our Bayesian timeseries simulation methodology achieves these goals.

We note that we seem to have encountered two fundamental model mismatches in these analyses. In the first analysis, we assume Gaussian innovations whereas the data seems to indicate thicker tails that may be better modeled using t-distributions or Cauchy distributions. In the second analysis, while it is reasonable to expect the predicted mean to decay with increasing uncertainty, we also see that the maximum capacities in the predictive distribution also decay; this may be alleviated if we used longer time series for training or we could consider spectral time-series learning algorithms.

In future work, we intend to incorporate the wholesale market price-prediction model developed here into an autonomous broker agent that can successfully participate in a Smart Grid tariff market by hedging its hourly portfolio balancing risk using trades in the wholesale market based on the prediction model. We also intend to enrich the Bayesian timeseries simulation model presented here using additional factors that affect the consumption or production capacity of various types of Smart Grid customers. We have focused here on a single residential customer model; we intend to test our model further on additional data sets from this model, additional residential customer models, and commercial and industrial customer models.

References

[1] M. Amin and B. Wollenberg. Toward a smart grid: Power delivery for the 21st century. *IEEE Power and Energy Magazine, 3(5):3441*, 2005.

[2] L. M. Berliner. Hierarchical Bayesian Time Series Models. In *XV^th Workshop on Maximum Entropy and Bayesian Methods*, 1995.

[3] S. Borenstein. The Trouble With Electricity Markets: Understanding California's Restructuring Disaster. *Journal of Economic Perspectives*, 2002.

[4] J. Cryer and K. Chan. *Time Series Analysis with Applications in R.* Springer, 2008.

[5] C. Gellings, M. Samotyj, and B. Howe. The futures power delivery system. *IEEE Power Energy Magazine, 2(5):4048*, 2004.

[6] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, 2007.

[7] S. Gottwalt, W. Ketter, C. Block, J. Collins, and C. Weinhardt. Demand side management - A simulation of household behavior under variable prices. *Energy Policy 39, 8163-8174*, 2011.

[8] C. Hirsch, L. Hillemacher, C. Block, and A. Schuller. Simulations in the smart grid field study MeRegio. *Information Technology 52 (2), 100106*, 2010.

[9] W. Ketter, C. Collins, P. Reddy, and C. Flath. The Power Trading Agent Competition. Technical Report ERS-2011-011-LIS, Erasmus Research Institute of Management, 2011.

[10] W. Ketter, J. Collins, and C. Block. Smart Grid Economics: Policy Guidance through Competitive Simulation. Technical report, ERS-2010-043-LIS, RSM Erasmus University, 2010.

[11] Y. Lee, Y. Lin, and G. Wahba. Multicategory support vector machines. *Journal of the American Statistical Association, 99(465): 67-81*, 2004.

[12] lme4 package for R. http://cran.r-project.org/web/packages/lme4/, 2011.

[13] R. Pindyck and D. Rubinfeld. *Microeconomics, 6th. Edition*. Pearson Prentice Hall, 2004.

[14] R. Rifkin and A. Klautau. In Defense of One-Vs-All Classification. *The Journal of Machine Learning Research, Vol. 5*, 2004.

[15] M. West and P.J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 2nd ed., 1997.