energies

Article

Uncertainty Matters: Bayesian Probabilistic Forecasting for Residential Smart Meter Prediction, Segmentation, and Behavioral Measurement and Verification

Jonathan Roth [1,2], Jayashree Chadalawada [1,3], Rishee K. Jain [2] and Clayton Miller [1,*]

[1] Building and Urban Data Science (BUDS) Lab, National University of Singapore, Singapore 119007, Singapore; jonmroth.15@gmail.com (J.R.); jayashree@u.nus.edu (J.C.)
[2] Stanford Urban Informatics Lab, Civil, and Environmental Engineering, Stanford University, Stanford, CA 94305, USA; rishee.jain@stanford.edu
[3] Civil and Environmental Engineering, National University of Singapore, Singapore 119007, Singapore
[*] Correspondence: clayton@nus.edu.sg; Tel.: +65-81602452

Abstract: As new grid edge technologies emerge—such as rooftop solar panels, battery storage, and controllable water heaters—quantifying the uncertainties of building load forecasts is becoming more critical. The recent adoption of smart meter infrastructures provided new granular data streams, largely unavailable just ten years ago, that can be utilized to better forecast building-level demand. This paper uses Bayesian Structural Time Series for probabilistic load forecasting at the residential building level to capture uncertainties in forecasting. We use sub-hourly electrical submeter data from 120 residential apartments in Singapore that were part of a behavioral intervention study. The proposed model addresses several fundamental limitations through its flexibility to handle univariate and multivariate scenarios, perform feature selection, and include either static or dynamic effects, as well as its inherent applicability for measurement and verification. We highlight the benefits of this process in three main application areas: (1) Probabilistic Load Forecasting for Apartment-Level Hourly Loads; (2) Submeter Load Forecasting and Segmentation; (3) Measurement and Verification for Behavioral Demand Response. Results show the model achieves a similar performance to ARIMA, another popular time series model, when predicting individual apartment loads, and superior performance when predicting aggregate loads. Furthermore, we show that the model robustly captures uncertainties in the forecasts while providing interpretable results, indicating the importance of, for example, temperature data in its predictions. Finally, our estimates for a behavioral demand response program indicate that it achieved energy savings; however, the confidence interval provided by the probabilistic model is wide. Overall, this probabilistic forecasting model accurately measures uncertainties in forecasts and provides interpretable results that can support building managers and policymakers with the goal of reducing energy use.

Keywords: Bayesian probabilistic forecasting; measurement and verification; residential energy prediction; smart meters

Citation: Roth, J.; Chadalawada, J.; Jain, R.K.; Miller, C. Uncertainty Matters: Bayesian Probabilistic Forecasting for Residential Smart Meter Prediction, Segmentation, and Behavioral Measurement and Verification. *Energies* **2021**, *14*, 1481. https://doi.org/10.3390/en14051481

Academic Editor: Andrés Elías Feijóo Lorenzo

Received: 13 January 2021
Accepted: 1 March 2021
Published: 8 March 2021

1. Introduction

The electricity grid is becoming greener but also more unstable. The precipitous drop in renewable energy prices—coupled with decarbonization efforts and threats of climate change—is leading to a significant surge in the amount of solar and wind being installed [1]. The problem is that these renewable energy resources are intermittent and non-dispatchable. Unlike fossil-fuel-based generation, grid managers cannot control when these resources supply energy to the grid. Instead, their output is dependent on the weather, generating energy when the sun shines or the wind blows. As renewables grow and the infrastructure of poles and wires that carry electricity age, the grid manager's job to meet demand with supply will become more difficult. Compounding this problem, the demand-side of the electricity grid is also quickly changing. Building electrification, electric vehicles, and distributed energy resources (DERs)—where consumers are decentralizing power generation through, for instance, rooftop solar and behind-the-meter battery storage—are examples of changing consumer power demand profiles [2]. The electricity grid will also see increased demand from the broader electrification of the heating and transportation sector, creating even more uncertainty in meeting demand with supply. A mismatch between supply and demand in the electricity grid can have devastating and costly impacts, such as rolling blackouts [3]. Hence, predicting future load and characterizing its uncertainty is critical to maintaining this delicate balance and ensuring the reliability of the whole system [4].

1.1. Residential Smart Meter Infrastructure Deployment

Fortunately, the rapid deployment of smart meter infrastructure over the last decade has generated vast amounts of data that provide an unprecedented level of detail on global building energy demand, both temporally and spatially. The high granularity of data from smart meters enables a better understanding of when buildings—the largest consumer of electricity out of any sector—consume energy and can be used to forecast consumption patterns. By 2024, an estimated 1.2 billion smart meters will be installed worldwide, up from a modest 25 million installed in 2010 [5]. This adoption is a massive step towards digitizing the grid, eliminating the archaic method of determining customer electricity use by sending workers to read meters at every site manually. Historically, utilities were limited to measuring electricity flow at the node level and relied upon customer phone calls to know when power lines were down. Smart meters make it possible to remotely monitor the energy consumption of buildings at the grid edge. New computing techniques can be leveraged to create a smarter, more sustainable, and reliable grid.

1.2. Using Probabilistic Load Forecasting (PLF) to Capture Uncertainty

Using this large influx of data from smart meters, probabilistic load forecasting (PLF) can help energy providers better measure the uncertainty and growing volatility in the grid caused by adopting renewable energy, DERs, and electric transportation. Unlike point forecasts, which provide a single number for estimated future load, PLF provides prediction intervals—analogous to confidence intervals from inferential statistics—that capture uncertainty in the forecast and assign a probability to each forecasted outcome. In short, PLF improves on the status quo of point forecasting with smart meter data by measuring this uncertainty. Measuring this uncertainty is critical for grid operators to manage their primary responsibilities, from power system planning to unit commitment, all of which are made more difficult when uncertainty is not captured [6]. In recent decades, research efforts have primarily focused on point forecasts. Still, as electricity demand and supply become more volatile, PLF will become more critical to maintaining a reliable grid [7]. Understanding this load uncertainty is of the utmost importance because prediction errors can lead to grid instability, large monetary losses, or potentially blackouts. For example, a load forecast error of just 1% mean absolute percentage error (MAPE) can result in hundreds of thousands of dollars of losses per GW for a utility [7]. By understanding the variability in future energy consumption and not just the point estimate, energy providers can make more informed decisions, and better account for potential losses. Beyond merely using the measured uncertainty for better grid management, PLF can benefit other related practices, such as measurement and verification (M&V) and battery control systems [8,9]. Wherever understanding forecast uncertainty can lead to better energy management practices, PLF can provide value, especially when the costs associated with erroneous forecasts are asymmetrical—typical for many grid and site-level use-cases.

As probabilistic load forecasting (PLF) aims to capture and measure uncertainty, the Bayesian paradigm is aptly suited for PLF. Bayesian models are probabilistic in nature, meaning that uncertainty is automatically and exactly quantified [10]. However, these models have been largely overlooked in the probabilistic forecasting and M&V fields partly, we presume, because of the extended historical use of frequentist statistics in the power engineering and building science fields [10]. Until recently, these models have been impractical due to their computational requirements, which may have also contributed to their second-class status. Despite being less popular, Bayesian statistics are not prone to some of the same misconstructions that plague frequentist statistics, such as incoherence and the interpretation of a p-value [11,12]. Uncertainty calculations can also be much less conservative using Bayesian statistics, with one study showing a 40% reduction in uncertainty [13]. In addition, priors must be made explicit rather than concealed like frequentist methods, ensuring greater transparency and reduced likelihood of model misuse.

In this paper, we propose the Bayesian Structural Time Series (BSTS) model for probabilistic load forecasting, combining the benefits of using a time-series-based model with the Bayesian paradigm. Specifically, the benefits of this model include:

1. Flexibility—The model can work with either univariate or multivariate data, allowing the modeler to use additional time series data, such as weather or energy data from similar buildings, if available. Unlike other probabilistic forecast models, which rely on ensembles of forecasts, the BSTS model is computationally fast and can also handle non-Gaussian data;
2. Feature selection—If using more than one variable, the model can use spike-and-slab priors to reduce the number of covariates in the final model while incorporating uncertainties of the coefficient estimates. This process discovers the static effects of covariates and, by explicitly stating priors before modeling, is more transparent than comparable models;
3. Inclusion of dynamic effects—The model handles time-varying effects found in time series data, such as seasonal effects, holidays, trends, and dynamic effects from covariates;
4. Interpretable—Because the BSTS model is a state-space model, it allows for a modular design, since each state component can be added independently, thereby enhancing the user's understanding of the model. Explicit use of priors helps avoid misapplication of the model, while the output allows the user to visually inspect the underlying state components and coefficients of included covariates;
5. Use for Measurement and Verification (M&V)—Estimating energy savings from an energy conservation measure (ECM) requires predicting the load that would have occurred without the intervention, a process known as M&V. This practice also requires providing confidence bands for savings estimates; therefore, the probabilistic forecasts provided by the BSTS model naturally lends itself to M&V.

To demonstrate the proposed probabilistic load forecasting model's functionality, we apply it to three case study applications using 120 residential apartment units in Singapore with granular smart meter and submeter data. First, we highlight the BSTS model's benefits in forecasting the aggregated apartment load, simulating a large residential building. Second, we incorporate submeter data into the model to highlight its ability to handle multiple covariate data in an interpretable fashion. Third and finally, we examine its ability to be used as a measurement and verification (M&V) tool to measure the savings effect of three demand response programs administered to the same residential units. The focus on these applications is to exemplify BSTS's use on a large collection of buildings for the purposes of district or grid-scale analysis. This effort can be contrasted against the more contemporary focus of using such techniques on a single building for the purposes of simulation model calibration [14,15].

2. Literature Review

As a deluge of data are becoming more readily available with the recent worldwide deployment of smart meter infrastructure, and researchers have examined ways to use this for building load forecasting. Most papers have explored various methodologies for point load forecasting, but these fail to capture forecast uncertainty. Without measuring this uncertainty, stakeholders, from facility managers to grid operators, cannot manage the risks associated with erroneous forecasts. Probabilistic load forecasting is a nascent field that aims to quantify this uncertainty to improve programs such as demand response and measurement & verification. The proposed BSTS model has additional benefits that overlap with other smart meter analytics, such as customer segmentation, which we will also review in this section.

2.1. Probabilistic Load Forecasting

Probabilistic load forecasting has many use-cases including: (i) power system planning and operations (i.e., flow analysis and reliability planning); (ii) understanding customer behavior and volatility; (iii) facilitating energy efficiency and energy management programs; (iv) revenue projection and energy trading (i.e., electricity market bidding); (v) demand response scheduling; (vi) stochastic unit commitment; (vii) probabilistic pricing; (viii) measurement and verification; (ix) battery control systems; (x) predicting of equipment failure; (xi) integrating renewable energy sources [7,16–19]. By producing a prediction interval, which is similar to, but different than, a confidence interval (though this term is often mistakenly used), PLF measures the uncertainty of the prediction by assigning a probability to each outcome [20]. Overall, PLF allows for better assessment of future uncertainty, enabling a greater ability to plan different strategies for the range of possible outcomes [21].

Load forecasting has primarily focused on the system- or bus-level loads, mainly due to the long historical need to predict demand at this level to help grid operators balance the load. However, with the more recent interest in demand-side flexibility, research on forecasting has picked up momentum at the individual building level. Forecasts at this smaller scale are critical for demand response programs and new energy management systems. Compared to system-level and aggregate loads, individual building loads, particularly residential buildings, are prone to being non-stationarity and exhibit greater volatility, making them particularly difficult to forecast [22]. Generally, the smaller the prediction level's scale, the higher the error will be—residential buildings are among the most challenging buildings to predict due to the stochastic characteristics of power demand that are primarily driven by occupant behavior, calendar effects, weather, and building efficiency.

Research on probabilistic forecasting models can generally be split into three main categories of methodologies: (i) feeding many similar, though perturbed, inputs into a deterministic model to create an ensemble of point forecasts; (ii) post-processing of point forecast residuals by employing a probability density function; (iii) developing novel probabilistic forecasting models [23]. One of the most popular PLF techniques is to use quantile regression. This technique has been leveraged in several PLF studies due to its ability to draw a conditional probability curve by modeling each quantile independently—allowing for the uncertainties in the power load to be determined—and for its general robustness to outliers [22,24–27]. Another popular model is neural networks (NN). Their growing popularity has spurred researchers to investigate their application for use in the probabilistic forecasting of energy [28,29]. Despite showing a good performance, these models are also very computationally expensive, making them unfeasible when fast predictions are necessary, like short-term load forecasting. Furthermore, NNs are often avoided in practice for regulatory purposes due to their black-box nature and lack of interpretability.

2.2. Demand Response

One primary application of harnessing smart meter data is demand response (DR), which aims to reduce and shift electricity demand by incentivizing customers to change their normal energy use patterns [30]. Already, DR has proved effective at shifting demand away from peak hours and is playing a larger role in helping to reduce network demand and volatility [31]. DR offers several key benefits, including (a) avoiding new generation capacity by reducing the number of power plants that have to be built to meet peak power demands; (b) lowering costs by avoiding purchasing power from expensive peak power plants; (c) delaying or preventing the need for costly transmission and distribution upgrades by shaving peak demand or shifting demand away from congested areas of the grid; (d) enabling more renewables on the grid by shifting demand to times with high renewable output and avoiding curtailment; (e) providing ancillary services, such as frequency regulation, voltage control, load following, and operating reserves. In the United States, load flexibility through demand response has a current capacity equivalent to 59 GW. Still, a study by the Brattle Group has identified nearly 200 GW of cost-effective DR potential by 2030—equal to 20% of US peak demand—worth more than $15 billion annually in avoided system costs [32]. DR programs will continue to grow and be bolstered by smart meter data that enable greater detail when energy is being consumed at the building (and even sub-building) scale, allowing for more targeted DR programs.

There are many demand response types, including direct load control, interruptible tariffs, demand-bidding programs, emergency programs, time-of-use pricing, critical peak pricing, and real-time pricing [33]. Some types of DR can be automated, like direct load control, and are increasing in popularity as new flexible dispatchable loads from distributed energy resources are rapidly being deployed. These resources leverage controllable internet-connected appliances such as thermostats, water heaters, and batteries to shift demand and stabilize the grid. Other types of DR rely on a behavioral response, like real-time and critical peak pricing, but quantifying user responsiveness remains difficult; inaccurate estimates can result in erroneous compensation. Obtaining accurate baseline consumption estimates—assuming DR events have not happened—is challenging, yet critical for determining the total savings obtained from the event [18]. Given that DR events typically occur on extreme weather days, when cooling or heating demand is at its peak, many baseline estimates are biased because they have been produced using limited data from similar extreme weather days. To remedy this issue, new research examines methods to select control groups of similar consumers who have not received the DR event to reduce inaccurate estimates.

2.3. Smart Meter Customer Segmentation

Historically, research on clustering methods has focused on high- and medium-voltage customers—partly due to data limitations—but the recent widespread adoption of smart meter infrastructure has enabled new research into low-voltage household consumers [34]. As the grid's demand-side continues to change rapidly with distributed energy resources, building electrification, and flexible load control, network operators are increasingly interested in understanding how residential consumers use their energy and how this affects low-voltage networks [35]. In recent years, numerous research studies have examined the application of various unsupervised machine-learning techniques on these new smart meter datasets [36–39]. The goal of these studies is, specifically, to: (i) segment consumers into different behavior groups; (ii) identify suitable candidates for demand response; (iii) detect profitable locations for energy storage; (iv) isolate network constraint violations; (v) help create more appropriate tariffs; (vi) improve forecasting methodologies; (vii) uncover potential consumers for targeted energy efficiency programs [39–42]. Despite the insights gained from many of the clustering algorithms examined in the literature, many of these models face difficulties scaling to larger and new datasets due to the highly stochastic and irregular demand from households. Current models either cluster specific features of the smart meter time series—like peak demand or volatility (e.g., entropy)—or attempt to cluster the entire time series, which is computationally expensive and leads to the curse of dimensionality [42–46]. These clustering algorithms often give varied results because they require the user to select the correct time series attributes or algorithm hyperparameters, which is difficult due to the high volatility in household demand [34]. The proposed BSTS model for PLF addresses several of the same goals as the smart meter clustering literature. Understanding the benefits and limitations of traditional customer segmentation methods can inform the modeler when using BSTS might be more appropriate.

2.4. Residential Smart Meter Measurement and Verification (M&V)

Simply put, the goal of M&V is to quantify the savings achieved from energy efficiency, demand response, and demand-side management projects (i.e., an energy conservation measure (ECM)) based on real-world measurements or energy models. However, the execution of quantifying these savings is far less straightforward and, depending on the method deployed, can cost between 1% and 5% of the project expenses [47]. It requires comparing the observed load, after the ECM intervention, to a theoretical load that would have occurred without the intervention. Creating this theoretical load for comparison is, by definition, a load we cannot observe and is inherently subject to estimation error. Ultimately, this leads to uncertainty in the amount of energy savings achieved from the ECM. As a result, impact evaluation reports provide confidence bands for their estimates, but many of these bands are based on point estimates and often misinterpreted [48–51]. Unfortunately, there is no singularly accepted methodology for calculating this theoretical load or the associated uncertainty, which leads to large differences in savings estimates. However, with the growing availability of data from smart meters, the industry is moving away from the traditional physics- and engineering-based approaches to streamline the M&V process through increased levels of automation and advanced data analytics [8,52]. These innovations are often referred to as M&V 2.0 [47].

Advanced data analytics using smart meter data are being explored to save time and money, compared to traditional M&V practices, and capture uncertainty in savings. The progress made in machine learning has spurred researchers to explore using these techniques to create baseline models for building energy load that can be used as a counterfactual to the observed load post-ECM [53]. Various papers have explored using linear regression, support vector machines (SVM), random forests, gradient boosting, artificial neural networks (ANN), kernel smoothing, and other models to forecast building energy use [54–59]. Typically, these models use lagged energy and temperature features that must be manually constructed, creating a bias in the reported results due to the variability of constructed features between models [47]. Biased models, whether caused by differences in feature construction or selected hyperparameters, lead to underestimates of savings on hot days and overestimates of savings on mild days [48]. Like SVM and ANN, many approaches are also non-interpretable and computationally expensive, which limits their applications due to regulatory purposes [7]. Most importantly, many models are not designed to provide probabilistic forecasts, and hence do not inherently provide a confidence band for energy savings when used for M&V.

3. Methodology

This section introduces the Bayesian Structural Time Series (BSTS) model for probabilistic load forecasting. Here, the mechanics of the model are explained, as well as how it can be combined with other time series data, like submeter data streams, and how it can be applied to measurement and verification (M&V) problems. Because we highlight the benefits of the BSTS model through three case-study applications, we end the section by explaining the applied dataset that we use. The following section introduces the case study applications in detail using the dataset described at the end of this section.

3.1. Bayesian Structural Time Series Models

The core component of BSTS is a state-space model that is designed to work with time series data. Unlike popular time series forecasting models, like autoregressive integrated moving average (ARIMA), the BSTS model does not rely on differencing, lags, and moving averages but instead allows the user to inspect the underlying components of the model. The BSTS model quantifies the posterior uncertainty of the individual components used to make the prediction, controls the elements' variance, and imposes priors on the model [60]. This model is particularly suited for building load time-series data due to its non-stationarity and variability, leading to high uncertainty [27].

State-space models are defined by two equations: the observation and state equations is shown in Equations (1) and (2), respectively.

$$y_t = Z_t^T \alpha_t + \epsilon_t \tag{1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \tag{2}$$

Here, the error terms $\epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$ and $\eta_t \sim \mathrm{N}(0, Q_t)$ are independent of all other unknowns, while $Z_t \in \mathbb{R}^d$ is the output vector, $T_t \in \mathbb{R}^{d,d}$ is the transition matrix, $R_t \in \mathbb{R}^{d,q}$ is the control matrix, and $Q_t \in \mathbb{R}^{q,q}$ is the state-diffusion matrix with $q \leq d$. The observation equation (Equation 1) links the observed data y_t to a latent d-dimensional state vector α_t, which is governed by the state equation (Equation (2)) [61]. The Z_t, T_t, R_t, and Q_t model matrices can be assembled from a library of sub-models, enabling a modular BSTS model construction that can capture other important features and trends in the data, such as seasonality, effects of holidays, and other similar building loads.

One of the state-components is the *local linear trend* and it is defined by Equations (3) and (4), where $\eta_{\mu,t} \sim \mathrm{N}\left(0, \sigma_\mu^2\right)$ and $\eta_{\delta,t} \sim \mathrm{N}(0, \sigma_\delta^2)$.

$$\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t} \tag{3}$$

$$\delta_{t+1} = \delta_t + \eta_{\delta,t} \tag{4}$$

The μ_t component is the trend of t, where δ_t is the expected increase in the next time step. This component adapts quickly to local variation and is, therefore, highly useful for predicting hourly building energy demand. However, this component should not be used for predicting longer-term energy demand, and should be substituted out by a long-term trend, because such predictions would produce large prediction intervals.

Another state-component captures seasonality and is shown in Equation (5), where S represents the number of seasons, $\eta_{\gamma,t}$ is a scalar, and γ_t denotes the joint contribution to y_t.

$$\gamma_{t+1} = -\sum_{s=0}^{S-2} \gamma_{t-s} + \eta_{\gamma,t} \tag{5}$$

This component includes the $S-1$ most recent seasonal effects, and the mean of γ_{t+1} gives a total seasonal effect of zero when summed over all S seasons. This component can be generalized to incorporate multiple seasonal components with different periods. For example, we can include two seasonal components when modeling hourly data, one with $S = 24$ for hour-of-day effect and another with $S = 168$ for hour-of-week effect. The seasonal state-component is included in the transition matrix T_t and is a $S-1$ by $S-1$ matrix, where the top row is filled with, $-1's$ along the subdiagonal, and $0'$ s everywhere else.

Contemporaneous covariates can also be captured in another state-component, with static coefficients, as shown in Equation (6), where $\alpha_t = 1$

$$Z_t = \beta^T x_t \tag{6}$$

or this state-component can be expressed with dynamic coefficients, as shown in Equations (7) and (8), where $\eta_{\beta,j,t} \sim \mathrm{N}\left(0, \sigma_{\beta_j}^2\right)$, $\beta_{j,t+1}$ is the coefficient of the j^{th} control series, and σ_{β_j} is the standard deviation of the associated random walk.

$$x_t^T \beta_t = \sum_{j=1}^{J} x_{j,t} \beta_{j,t} \tag{7}$$

$$\beta_{j,t+1} = \beta_{j,t} + \eta_{\beta,j,t} \tag{8}$$

For this dynamic state, $Z_t = x_t$ and $\alpha_t = \beta_t$ where the associated part of the transition matrix is set to $T_t = I_{j \times j}$ with $Q_t = \text{diag}\left(\sigma^2_{\beta_j}\right)$. This state component is useful for capturing, for example, hourly weather effects on building energy demand and other buildings with similar load profiles, which will be explored in more depth in Section 3.3. Because only one of these contemporaneous state components can be used, the static coefficients should be used when the relationship between the treated unit and regressors is stable. In contrast, the dynamic coefficients should be used when their linear relationship changes over time.

Once the state-components are independently assembled, a Bayesian approach is used to estimate the model parameters θ, where $\alpha = (\alpha_1, \ldots, \alpha_m)$ represents the full state sequence. The prior distribution for the model parameters $p(\theta)$ and initial state values $p(\alpha_0 \mid \theta)$ are specified, and $p(\alpha, \theta \mid y)$ is then sampled using a Markov chain Monte Carlo (MCMC) [62]. We use the Gamma distribution as the prior where the sample variance is scaled so that we can model the data in its original scale.

3.2. Feature Selection Using Spike-and-Slab Method

A spike-and-slab prior—a Bayesian feature selection technique—is placed over the coefficients to reduce the number of contemporaneous covariates in our state-component. The "spike" is the probability that a coefficient in the model is zero (i.e., excluded), and its prior distribution is the product of independent Bernoulli distributions, as shown in Equation (9), where $\varrho = \left(\varrho_1, \ldots, \varrho_J\right)$, and $\varrho_j = 1$ if $\beta \neq 0$ and $\varrho_j = 0$ otherwise [63].

$$p(\varrho) = \prod_{j=1}^{J} \rho_j^{\rho_j} \left(1 - \pi_j\right)^{1-\varrho_j} \tag{9}$$

The prior probability of regressor j being included in the model is represented as $\pi_j = M/J$, where M is the user set expected model size.

The "slab" is the prior distribution of the regression coefficient values and is modeled as the conjugate normal-inverse Gamma distribution, as shown in Equations (10) and (11), where $\boldsymbol{b}$ is a vector of the prior expectation of the coefficient value for each β, and is typically set to zero.

$$\beta_\varrho \mid \sigma^2_\varepsilon \sim \text{N}\left(\boldsymbol{b}_\varrho, \sigma^2_\varepsilon \left(\Sigma^{-1}_\varepsilon\right)^{-1}\right) \tag{10}$$

$$\frac{1}{\sigma^2_\varepsilon} \sim G\left(\frac{v_\varepsilon}{2}, \frac{s_\varepsilon}{2}\right) \tag{11}$$

β_{rho} is the nonzero elements of vector β, where Σ^{-1}_ϱ is the rows and columns of Σ^{-1}, associated with the nonzero entries of ϱ; the Σ^{-1} is the prior precision over β in the full model (i.e., when all variables are included). The number of observations to weight the prior is denoted by v_ε where $s_\varepsilon = v_\varepsilon(1 - R^2)s^2_y$ and is set by a user-defined $R^2 \in [0, 1]$ value.

Finally, the spike and slab priors are combined and factorized, as shown in Equation (12), where the "spike" selects the included covariates and the "slab" tunes the complementary set of nonzero coefficients.

$$p\left(\varrho, \beta, 1/\sigma^2_\varepsilon\right) = p(\varrho)p\left(\sigma^2_\varepsilon \mid \varrho\right)p\left(\beta_\rho \mid \varrho, \sigma^2_\varepsilon\right) \tag{12}$$

Although the spike-and-slab helps us determine the most essential features, we use Bayesian model averaging to compute the probabilistic forecasts, which prevents overfitting by not committing to the point estimates of the coefficients.

3.3. Application for Measurement and Verification

Much like current M&V practice, this model can measure energy savings by using a control set of customers similar to the one that received the intervention (i.e., treatment). This practice eliminates bias introduced by manually selecting these similar customers, as the model can choose them automatically, as described in Section 3.2. To add these potential similar customer candidates, a regression component (Equation (6)) is used to produce counterfactual predictions by constructing a synthetic control based on a selected combination of customers that were untreated. This process allows the model to explain variance components in the treated market that are uncaptured when solely using seasonal sub-models. Since we are interested in the posterior probabilistic forecasts, the posterior predictive density is defined in Equation (13) as the coherent joint distribution over all counterfactual datapoints rather than as a collection of pointwise univariate distributions. This process allows us to provide summary statistics, such as the cumulative effect of the treatment intervention.

$$p\left(\widehat{\boldsymbol{y}}_{n+1:m} \mid \boldsymbol{y}_{1:n}, \boldsymbol{x}_{1:m}\right) \tag{13}$$

To evaluate the treatment's impact, samples from the posterior predictive distribution are drawn and compared to the counterfactual activity, as shown in Equation (14), where τ is the draw from the distribution and $t = n + 1, \ldots, m$.

$$\phi_t^{(\tau)} := y_t - \hat{y}_t^{(\tau)} \tag{14}$$

We can also understand the cumulative impact of an event by summing over t for each draw of τ.

3.4. Data

To showcase the capabilities of BSTS for residential probabilistic forecasting, we used energy data collected from smart meters and plug load sensors for 120 apartment units in Singapore. Specifically, granular energy consumption data at hourly time steps were collected between March 2018 and August 2019 (1.5 years) using plug load sensors installed at each of the 120 apartment units. The granular consumption data measured by the plug load sensors are classified into seven categories: main (i.e., whole-unit), air-conditioners, water heaters, fans & lights, washing machine & drier, other lights, miscellaneous loads (kitchen appliances, TV, iron, dehumidifiers, etc.). The plug load meter system uploaded the collected data in near-real-time using 3G data connectivity to a virtual cloud platform that retrieves the data using an API access key every week.

The overall study had the main purpose of understanding how behavioral interventions shape apartment dwellers' energy and water consumption behaviors. The experimental results of the larger study and detailed information about the experimental setup can be found in a complementary publication [64]. The participants in the study were divided into four treatment groups (control, T2, T3, T4), with approximately 30 units in each, based on their average energy and water from the previous year (2017). In this paper, we use a subset of this collected data to exemplify the BSTS model's benefits. The design of the two behavioral experiments based on individual goal setting consisted of the following phases:

- A baseline period measures every apartment unit's energy use, preceding any intervention, from February 2019 to April 2019 (2 months);
- For the treatment groups, individual goal-setting, both with and without incentives, was initiated for April 2019 to July 2019 (3 months). The three treatment groups received emails at the beginning of the month explaining their goal (and incentive, if applicable). The control group also received a message simply indicating that their energy consumption would continue to be monitored. At the end of each month, all four groups received their utility bill along with an update indicating whether or not they were currently meeting their goal;
- After the goal-setting treatments, post-treatment effects were measured for July 2019 to September 2019 (2 months).

It is important to note that for this specific scenario, we determined that having a short baseline period of just several months was sound because Singapore is located near the equator, and thus has similar weather all year round. If this experiment were to be conducted at another location, we would recommend creating a more extended baseline period that better accounts for weather differences between months.

4. Case Study Applications

To highlight the benefits of the BSTS model for probabilistic load forecasting, we explore three major applications of its use through a series of case studies using the data outlined above. The variety of the case studies—and how the data are processed in each application—highlight several contexts where the BSTS model can be used to better understand electric loads in buildings, capture uncertainty in forecasts, and lead to improved demand charge management programs: (1) *Probabilistic Load Forecasting for Apartment-level Hourly Loads*. This application was specifically chosen due to the importance of forecasting residential loads for operating various distributed energy resources (DERs). For example, using rooftop solar panels and batteries requires a forecasted load of the building's energy demand. The system can determine when to charge the battery, pull electricity from the grid, discharge the battery, and give electricity back to the grid. More specifically, understanding the uncertainty in this forecasted load can help the system operate more efficiently by reducing times when the system is sub-optimally in one of these states when it should be in another. (2) *Using Submeter Data for Load Forecasting*. In most typical applications, utilities do not have submeter data, but only smart meter data that measure whole-unit energy use and temperature data. However, understanding how residential customers use their energy—like using major appliances at different times of day—can help utilities better design demand response programs that lead to higher energy savings through improved customer segmentation and targeting. (3) *Measurement & Verification for Behavioral Demand Response*. Utilities running demand response programs need to measure how effective they are in adequately compensating those who have participated and effectively plan for future events. However, measuring how much energy was potentially saved during an event is non-trivial and requires building a counterfactual of what would have happened if this did not occur. Since this counterfactual is unobservable but rather measured, inherent uncertainty should be captured to provide more robust estimates of energy savings and, therefore, improved compensation and planning processes.

In the following three subsections, we explore using BSTS for residential load forecasting to understand how this model measures uncertainty, captures trends in submeter loads and applies to measurement & verification. Each subsection focuses on one application, as outlined above, to exemplify the benefits of the BSTS model for three major applications. The diversity in case study applications serves to highlight the flexibility and numerous benefits of the model. Utilities can better assess risk for demand response programs, target customers based upon expected submeter loads, and identify savings from their programs. Facility managers can benefit from these same features for demand charge management and energy efficiency initiatives at the building level. Throughout this section, we use a static regression component, instead of a dynamic element, with a spike-and-slab prior with model size M = 3 (when applicable for multivariate cases), an expected explained variance of $R^2 = 0.80$ and 50 prior df. Because the covariates are other households from a randomized experiment, we expect them to account for any local linear trends and seasonal variation in the response variable.

4.1. Application 1: Probabilistic Load Forecasting for Apartment-Level Hourly Load

In this section, we examine how the BSTS model captures forecasting uncertainty for residential loads. At the building level, measuring forecasting uncertainty is particularly important for integrating battery storage, rooftop solar, and electric vehicles into the building's operations. For example, batteries are often installed in buildings to shift demand to times of day when electricity prices are cheaper and to reduce demand charges on utility bills that charge customers for their peak demand over the course of a billing cycle (typically about 30 days). These demand charge costs can approach 50% of the total utility bill for large commercial and industrial buildings and act as a mechanism for utilities to transfer costly improvements to their infrastructure—from upgrading power lines and transformers to handle higher loads—onto customers. When operating these behind-the-meter batteries (i.e., installed in buildings rather than as large stand-alone systems connected to transmission lines), battery systems must forecast future building load to identify when to charge and discharge to maximally reduce demand charges. When forecasts are incorrect, battery systems may discharge too early, leaving them unprepared for future hours of higher than expected demand. Producing 24–72 h forecasts with captured uncertainty allows for better self-scheduling of battery control systems (or EV charging with rooftop solar), as their performance depends mainly on forecasting uncertainty. Directly incorporating uncertainty into battery control systems—potentially through stochastic programming methods—can lead to better performance by adjusting the charging and discharging times of batteries to reduce customer bills better.

Provided the potential benefits of PLF, in this application, we examine how the BSTS model can capture forecasting uncertainty for residential building loads. From our original dataset, as described in Section 3, we examine the efficacy of BSTS to capture forecasting uncertainty for 49 apartment units. Because only larger buildings pay demand charge costs, we aggregate the load for all 49 apartment units into one time series to simulate a larger apartment building. We build the BSTS model on this aggregated time series and compare it to the popular time series forecasting model, the autoregressive integrated moving average (ARIMA) model. Using other modeling alternatives for benchmarking, such as gradient boosting, support vector regressions, or other techniques, falls outside the scope of this publication, as we seek to show a simplified comparison. An overview of the other benchmarking modeling options is available in the literature [65]. We used the auto.arima model from the forecast package in R to help us obtain the optimal hyperparameters for the number of time lags p, the degree of differencing d, and the order of the moving-average q. This model uses a stepwise Hyndman–Khandakar algorithm to select the hyperparameters based on the AIC, AICc, and BIC values [66]. We were using the auto.arima function to ensure that we obtained the best ARIMA model possible through a standardized and objective process. We also compare these results to forecasts obtained when modeling just one apartment unit to show how aggregating apartment units results in lower forecasting error and uncertainty. Using the hourly data, we train our models using two weeks of data—from March 17 to March 31—and aim to predict the next 24 h of load, therefore simulating applications relevant to a battery and solar control systems. We choose a small training dataset, with only 49 aggregated apartment units, to demonstrate the model's efficacy on highly variable residential load data and to mimic scenarios where limited data are available, such as shortly after a smart meter installation.

After building the BSTS and ARIMA model on the 49 aggregated apartment units, we found a mean absolute percentage error (MAPE) of 0.127 for the BSTS model compared to a 0.206 error for the ARIMA model, meaning the BSTS model has better performance. Figure 1 shows the forecasted load for both models, with the BSTS forecasting results on top (in blue) and the ARIMA results on the bottom (in red). The entire 336 h of training data are not shown, just the last 86 h, to provide greater clarity and better highlight the probabilistic forecast, which predicts the subsequent 24-hour load. This figure visually shows how the BSTS model better captures the trend in the aggregate load and provides prediction intervals. The three blue bands included in the figure represent the 60th, 80th, and 90th percentile prediction intervals, represented by the dark, medium, and light blue colors, respectively. By producing prediction intervals in addition to final forecasts, the BSTS model can highlight at which hours its forecasts are more uncertain and provide a distribution of what the uncertainty looks like.

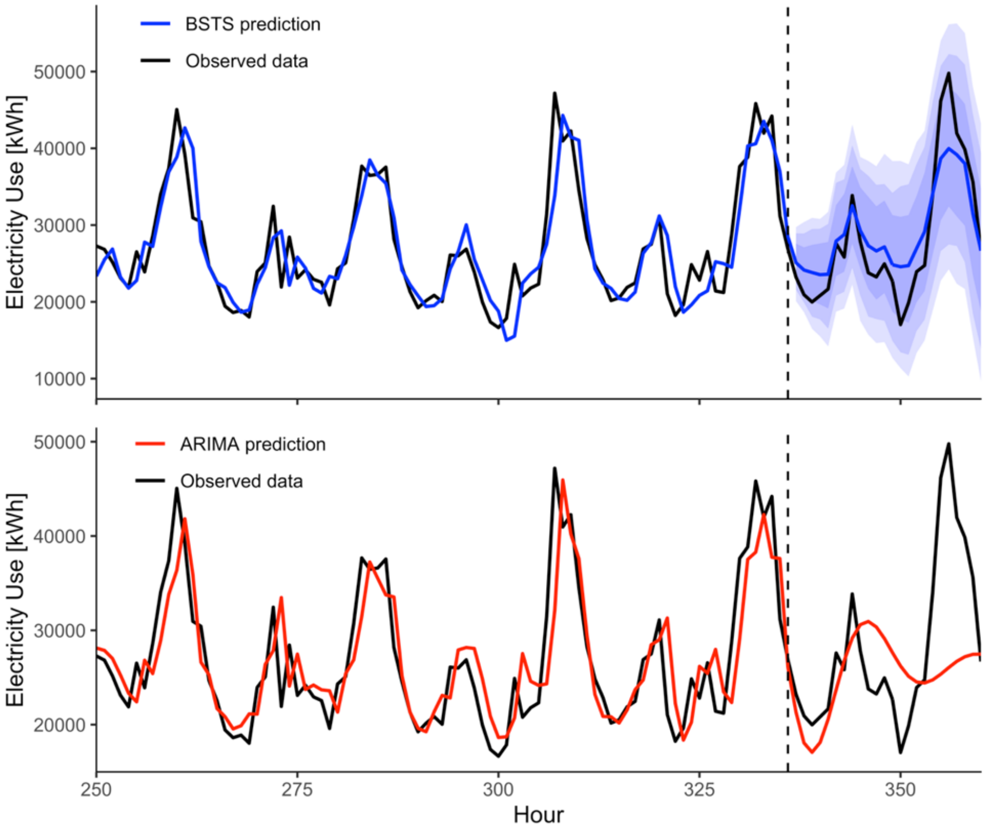

Figure 1. Forecasting results for the BSTS model (on top in blue) and the ARIMA model (at the bottom in red). In the top graph, the BSTS model produces 60th, 80th, and 90th percentile prediction intervals as represented by the dark, medium, and light blue bands, respectively.

Figure 2 illustrates the density plots of the resulting forecasts for the first 6 h to better demonstrate the prediction intervals produced from the model. The vertical red lines are the observed demand values, while the dotted black lines are the 5th and 95th prediction percentiles, where the difference between the two gives the 90th percentile prediction interval (PI Range). Examining the distributions, they are similar to Gaussian distributions; however, they also show some variation in shape. Furthermore, the distributions show how future loads are more difficult to predict since the PI range increases from 9617 to 18,126 between a 1 and 6-h forecast. More importantly, the observed demand values all fall within the PI range and are in the middle of the probability density curves, meaning that similar values appear in the forecasting distributions with high probability. Figures 1 and 2 show how the BSTS model provides a complete probability distribution for each hour of predicted load, highlighting the advantages of the BSTS model in quantifying uncertainty and improving prediction accuracy over the current popular ARIMA time-series model.

Finally, comparing the above results to those when modeling just one apartment unit at a time, Figure 3 shows the MAPE when forecasting all 49 apartment units independently for both the BSTS and ARIMA models, thus creating a histogram of 49 MAPE for each model. Figure 3 shows the high variability of MAPE for both models when predicting a residential building load that is hard to capture in a forecasting model. Errors for both models are relatively high, with nearly the same MAPE for all the units, where the BSTS model has a marginally better MAPE of 0.886 compared to ARIMA's MAPE of 0.888. This situation confirms the difficulty in predicting residential energy use for individual units and how aggregating loads from apartment units can lead to better forecasting accuracy, as expected based upon previous research [35,56]. By disaggregating the load into individual apartment-unit loads, the increased irregularity and stochasticity of the time series data makes them more unpredictable, leading to worse forecasting errors. In other words, the aggregated apartment load has more structure to it, which leads to better forecasts. The reduced forecasting accuracy for predicting individual apartment-unit loads supports the notion that larger buildings are more likely to benefit from installing behind-the-meter batteries. The resulting utility bill demand-charge savings will be more significant when peak loads are higher, and forecasting accuracy is improved.

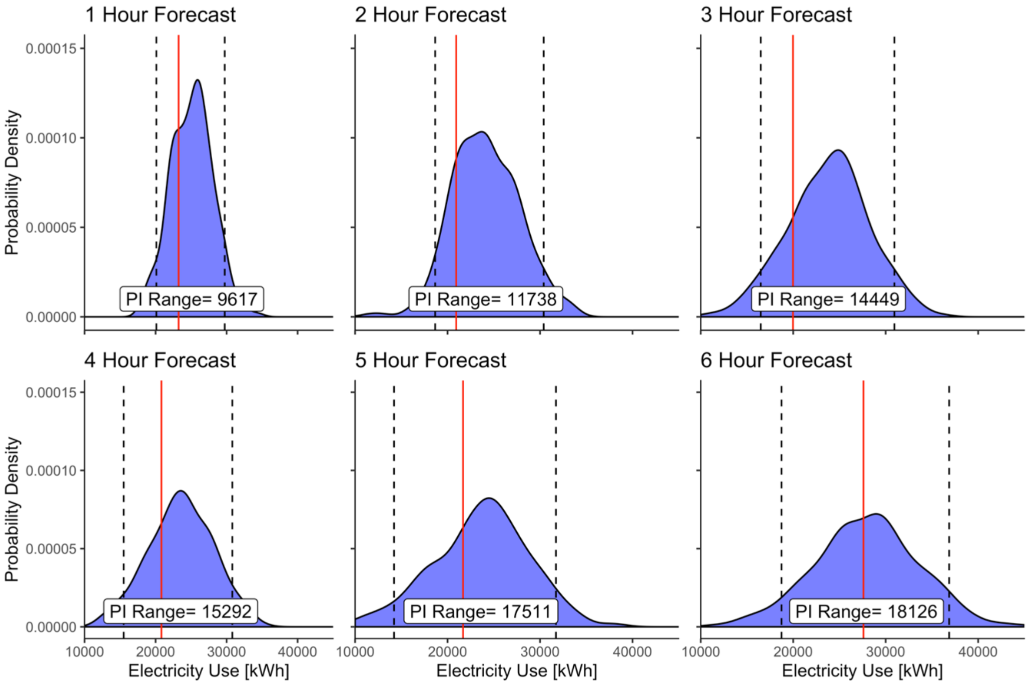

Figure 2. Probability density plots created by the BSTS model for the first 6 h. The red lines are the observed load, while the black dotted lines are the 5th and 95th prediction percentiles, where the difference between the two gives the 90th percentile prediction interval (PI Range).

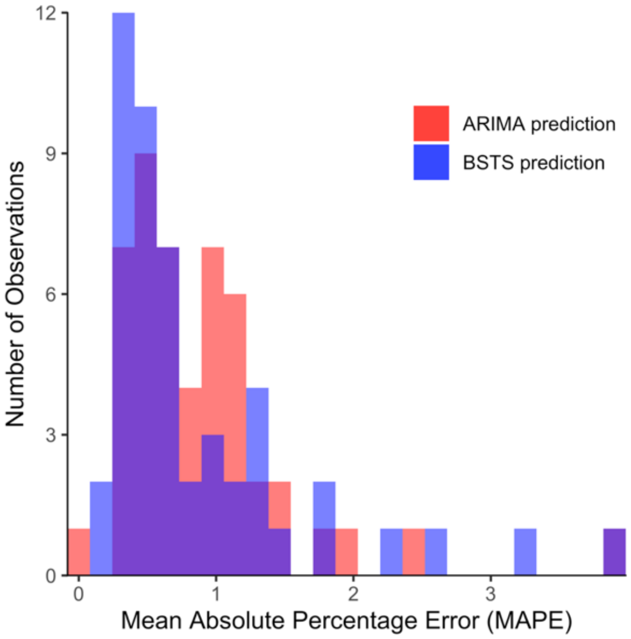

Figure 3. Histogram of the mean absolute percentage error (MAPE) for the BSTS and ARIMA models for each of the 49 apartment units when modeled independently. The ARIMA (red) and BSTS (blue) and overlapping (purple) indicate the difference between the MAPE for the two techniques.

4.2. Application 2: Using Submeter Data for Load Forecasting

In this application, we use submeter data produced from the meters to improve the forecasting accuracy of the BSTS model. Specifically, we examine how different combinations of submeter loads can be used to improve accuracy and how these submeter data can be forecasted. As more smart devices become available for the home—like smart thermostats, water heaters, and electric vehicle charges—we need a better understanding of how their energy-related data can be used to improve the accuracy of these forecasting models. As discussed in the previous section, with improved forecasting accuracy comes better-distributed energy resource (DER) operations and control. For example, improved forecasting accuracy stemming from these data's inclusion can lead to better behind-the-meter battery operations and, therefore, higher utility bill savings. Furthermore, as more smart devices reach consumers' homes, DER aggregators will combine their capabilities and provide utilities with bulk power by integrating many smaller loads. Such services can help utilities defer or replace costly investments to traditional grid infrastructure while providing customers with cheaper electricity prices in exchange for the limited flexibility of one or several home devices. Properly aggregating many smaller loads to offer these services in an economical, reliable, and unobtrusive fashion requires a better understanding of how these devices use energy and improved quantification of forecasting uncertainty.

In this subsection, we first forecast apartment-level electric loads using different submeter time series data produced from the meters to show how predictive power is altered and which devices (i.e., submeters) result in the most significant reduction in forecasting error. We only examine individual apartments and not the aggregated load because each apartment has different types of submeter loads. After this analysis, we examine how a DER aggregator might combine many smaller loads from one type of device—in this case, air-conditioning units—to produce probabilistic forecasts that they can use to offer utilities bulk power at a price. Similar to application 1 in Section 4.1, we train our models using two weeks of data—from 17 March to 31 March—and aim to predict the next 24 h of load.

First, we predicted the apartment-level load for one unit with eight different BSTS models, where each model included one more submeter load than the last, starting with a univariate case and ending with all submeter loads in the last model. In other words, submeters (i.e., covariates) were added sequentially, starting with a univariate model 1, until all seven were included in the final model 8. See Table 1 for a description of which submeters were included in each of these models. Figure 4 shows the resulting cumulative absolute error when training each of these eight models, showing that including the water heater and dryer and washer result in the most considerable reductions in error. This plot indicates specifically where each model encountered trouble, rather than just providing a single number describing the model's accuracy. For example, the inclusion of the water heater variable in model 2 did little to decrease the error until about hour 70, when the errors begin to diverge. Furthermore, including the AC unit submeter did little to decrease model error when only included alone, despite being a high-energy-use device. Given the hot climate in Singapore, this resident may have kept the AC unit on while not home, thereby making the submeter a poor predictor for the whole apartment load, since the other submeters more closely relate to being home. Each subsequent model used another submeter from the apartment, resulting in the cumulative absolute error decreasing until there is near zero error in model 8 once all meters were used. The inclusion of each additional submeter results in a larger fraction of the total energy load being accounted for, resulting in a lower error and a decrease in prediction uncertainty. In most typical applications, we would not have submeter data but only smart meter and temperature data. Understanding how residential customers use their energy—like using major appliances at different times of the day—can help utilities better design demand response programs that lead to higher energy savings through customer segmentation and targeting.

Table 1. The models created and their included features (i.e., submeters) for application 2 that correspond to Figures 4 and 5.

Model #	Feature Names	# of Features
1	None (univariate)	0
2	AC (air conditioning)	1
3	AC (air conditioning), WH.2 (water heater)	2
4	AC (air conditioning), WH.2 (water heater), Drier.wash (drier & washer)	3
5	AC (air conditioning), WH.2 (water heater), Drier.wash (drier & washer), wash.DW (dishwasher)	4
6	AC (air conditioning), WH.2 (water heater), Drier.wash (drier & washer), wash.DW (dishwasher), fans_lights (fans & lights)	5
7	AC (air conditioning), WH.2 (water heater), Drier.wash (drier & washer), wash.DW (dishwasher), fans_lights (fans & lights), P.H.K. (other plugs)	6
8	AC (air conditioning), WH.2 (water heater), Drier.wash (drier & washer), wash.DW (dishwasher), fans_lights (fans & lights), P.H.K. (other plugs), wetbulb temperature	7

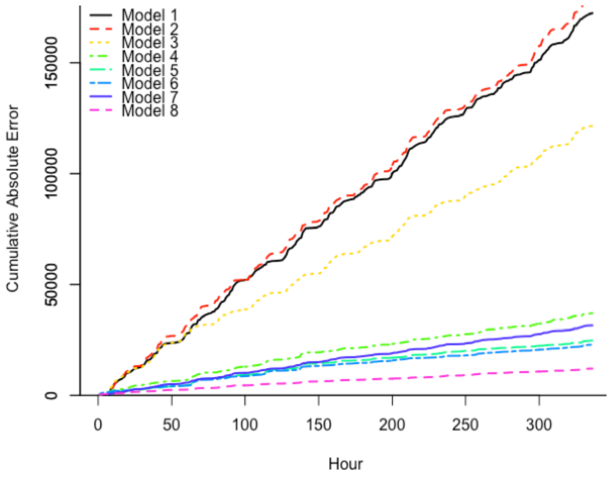

Figure 4. The cumulative absolute error over the training period for each of the eight models constructed.

However, when only several submeters are included, the included variables' coefficients can provide us with some additional insights. Figure 5 shows the coefficient values for the included submeters (i.e., covariates) for model 6 and model 7, both when the submeters time series are standardized (although there are several types of data normalization, we use the most conventional practice of standardization by transforming each feature to have mean zero with a standard deviation of one) (top row) and non-standardized (bottom row) when used as inputs into the models. Unlike some other regression models—where using non-standardized input data can cause convergence issues, numerical instability, or slower training times—the BSTS model is agnostic to feature scale. When using the non-standardized data, model 6 assigns the fans/lights an average coefficient value of 1.58 when the other four submeters have a value near 1. In model 7, once the submeter that accounts for plug-loads was added, the fans/lights' coefficient is reduced to 1.23 while the coefficient for the plug-loads received a value near 1, like the other submeters. This evidence shows that using the unstandardized submeter data provides us with coefficients that point towards submeter loads that better correlate with the unaccounted-for load.

However, unstandardized submeter data do not provide a sense of the relative scale of each submeter load, which the standardized submeter data better show, as seen in the bottom row of Figure 5. The drier/washer, air-conditioner, and water-heater use far more energy than the fanslights or plug-load, as reflected in the coefficient in these two plots, which exhibit a far less dramatic change in variable coefficients between model 6 and 7. Despite achieving the same prediction accuracy without standardizing the input data, this exercise highlights that the coefficients with standardization provide different insights—like a sense of scale between the covariates—than when the covariates are non-standardized. This situation is advantageous when doing M&V, as only a few submeter loads could be prioritized and measured to increase performance.

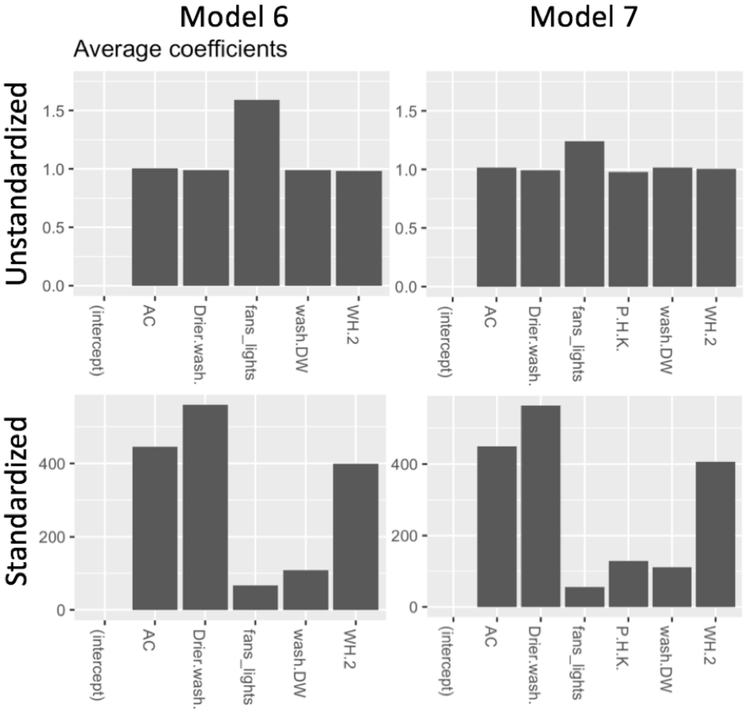

Figure 5. The coefficients for model 6 and model 7 both standardized and unstandardized.

4.3. Application 3: Measurement & Verification for Behavioral Demand Response

In this section, we examine how the BSTS model can be used for measurement and verification (M&V) to estimate energy savings from an energy conservation measure (ECM). Specifically, we study the impact of a behavioral demand response program that attempted to influence participants in apartments to save energy over two months. As described in detail in Section 3.4, 120 participants participated in an experiment, where about 90 participants received a goal to reduce energy use by 10% compared to their baseline levels, of which about 60 also received a monetary incentive. About 30 participants were kept as the control group and received neither goal-setting reduction targets nor incentives. Please see Section 3.4 and the complimentary publication about the behavior intervention results for more information about this data-collection process and the messaging the participants received throughout the experiment [64]. In this section, we look at the impact of this experiment on each treatment group. Here, we end up solely using treatment group 3 (T3), which contains about 30 apartments.

To measure the demand response program's energy savings, we must construct a counterfactual of what we expect would have happened if the event did not occur. Using this counterfactual, we can then compute the energy savings by comparing this to the observed load. However, constructing this counterfactual is difficult because we have to forecast energy use affected by many factors, including seasons, schedules, and more. Because all those apartment units in T3 received the same messaging, incentives, and goals simultaneously, we aggregate all the loads in the group to create one time series containing energy-use data before and after the program started. As shown in Section 4.1, predicting energy use for one apartment is much more complicated than aggregating many units together; therefore, aggregating the treatment group should give us better estimates of energy savings. We also aggregated the entire control group into one time series. Finally, because we are interested in measuring the behavioral changes from the combined goal setting and incentives over two months, we also aggregate the smart meter data to the daily level; hourly data are too noisy for the application of measuring savings over this timeframe. Therefore, our constructed model predicts everyday energy use for the first two months of the intervention period—acting as the counterfactual required to adequately assess energy savings—and compare it to the observed load, which was influenced by the interventions.

Figure 6 shows the probabilistic forecasts, pointwise differences between observed and counterfactual forecast, and cumulative impact for treatment group 3 (T3). The figure does not offer a distinctive effect of the treatment on saving energy use in this group when compared to the control group, which was used to help build the model. The model estimated a relative effect of a reduction in energy use of 7.3%; however, this estimate was not statistically significant, with a 95% confidence interval from [−51%, 41%] and a posterior tail-area probability p of 0.348 (higher than the commonly accepted 0.05 needed to reject the null hypothesis and conclude a statistically significant effect is present).

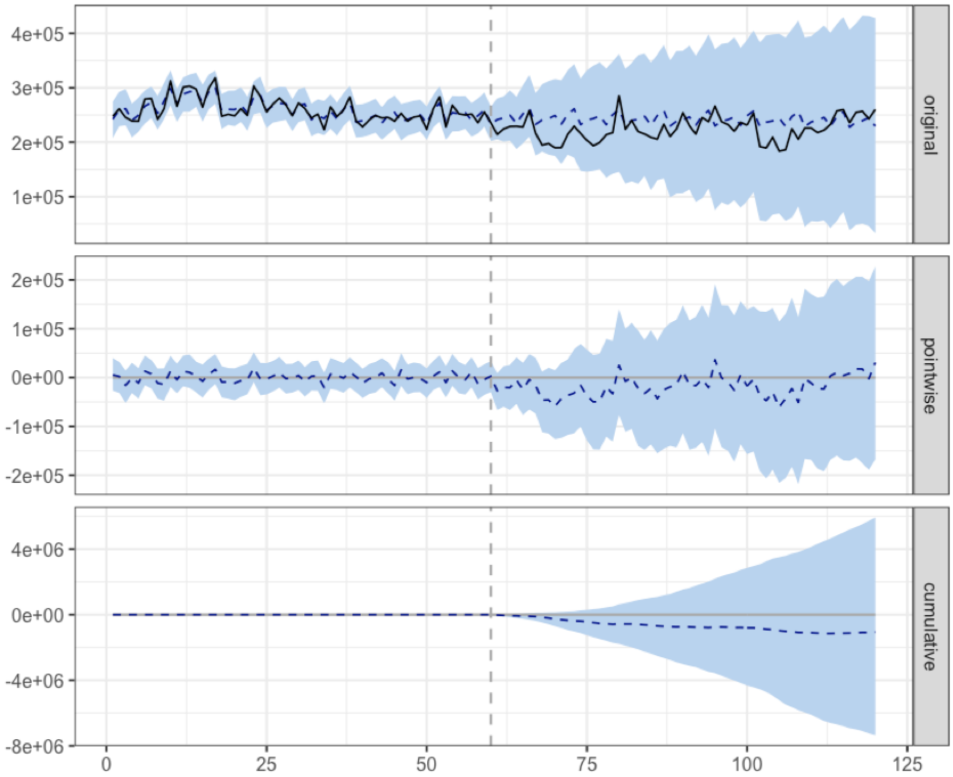

Figure 6. The probabilistic forecasts, the pointwise difference between observed and counterfactual (forecast), and the cumulative impact for treatment group 3.

When constructing the BSTS model, the two covariates we used—control group and mean temperature—were both included and produced a similar distribution of fitted coefficient values, as shown in Figure 7. In applications where more relevant covariates are present, the BSTS model might choose to eliminate them from the model. However, in this case, both covariates were included and with similar coefficient values when the data were standardized, indicating that both variables had about equal weight in being used to predict future load. With a control group present experiencing the same weather as the treatment group, it is surprising that the control group's coefficient value is roughly the same as the mean temperature variable. We believe that given the number of participants in the study and the long duration time, it is challenging to build a robust counterfactual, regardless of which model is used.

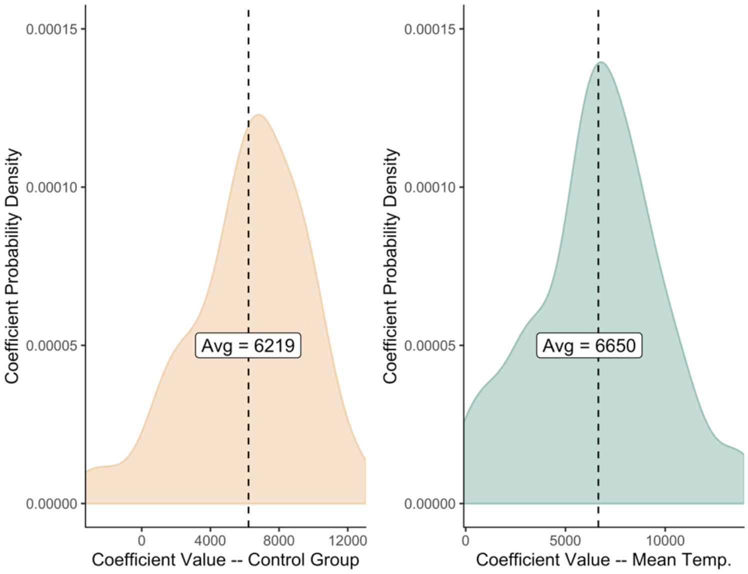

Figure 7. The coefficient values for the two included covariates in the BSTS model built to produce the counterfactual used for M&V of the demand response program.

5. Discussion

Evaluating the performance of energy forecasts is essential when assessing which models to use in practice. For point forecasts, this assessment process is straightforward, with many widely accepted error metrics that measure the discrepancy between predicted and measured values. However, evaluating probabilistic forecasts is more difficult due to the captured information on the uncertainty, that cannot be directly compared with measured values [67]. In this paper, we compared the forecasting differences of the proposed BSTS model with ARIMA models in three example case-study applications. Given the novelty of probabilistic forecasting in the building energy field and the lack of widely accepted assessment criteria for this type of forecasting, we did not directly evaluate the measures of uncertainty provided by the BSTS model. However, we did attempt to give a sense of the model's ability to capture uncertainty through the provided figures, showing that the model can overcome deficiencies of pointwise models that fail to quantify forecast uncertainty robustly. Provided these considerations, our future work will further evaluate the efficacy of the captured uncertainty in the BSTS model with other probabilistic models used in the field. Furthermore, we would like to examine how to measure the forecasting accuracy of the probabilistic load forecasting through tests, such as the Diebold–Mariano and the Giacomini–White test, to provide greater context into how well the uncertainty is captured in the BSTS model.

Beyond evaluating better probabilistic forecast metrics, we aim to more directly use the uncertainty measurements from the BSTS model for decision-making in the building energy domain. With the increased penetration of battery storage in residential and commercial buildings, probabilistic forecasting will become more critical for these batteries' operations. Bloomberg New Energy Finance projects worldwide energy storage to increase from 17 GWh in 2018 to 2850 GWh by 2040, catalyzing new research in battery control systems that will become more salient as utilities change how they charge customers [68]. Time-of-use (TOU) pricing is beginning to be implemented across rate tariffs, and demand charges on utility bills for commercial and industrial (C&I) customers already often account for nearly 50% of the total bill. This time-dependent cost of energy is what gives behind-the-meter batteries their value. By shifting demand to low-cost hours and shaving peak demand consumption, batteries' intelligent control can reduce the monthly utility costs for buildings. This control system is based on two parts: forecasting and optimization. Typically, point forecasts estimate the building's consumption—between 24 and 96 h into the future—and are used as inputs into a deterministic optimization algorithm that will create the dispatch schedule for the battery. Forecast uncertainty is therefore uncaptured, leading to suboptimal battery schedules and uncaptured savings. In contrast, probabilistic forecasts capture this uncertainty and can be used as inputs into stochastic optimization algorithms, leading to more significant savings than traditional point forecasts and deterministic optimization methodologies.

6. Conclusions

As the adoption of renewables causes electricity generation to become more variable and the demand side of the grid changes rapidly—with the adoption of electric vehicles, building electrification, and grid edge technologies—understanding uncertainties in electricity forecasting is becoming more valuable. This paper proposes the Bayesian Structural Time Series model for probabilistic load forecasting (PLF) at the building level to capture these uncertainties. The proposed model addresses limitations of other PLF models by being flexible to univariate or multivariate data, handling feature selection, utilizing either static or dynamic effects, and providing interpretable results. Many other PLF techniques rely on creating an ensemble of point forecasts and estimating a probabilistic forecast from this. Still, these techniques are often computationally intensive, and therefore only work in specific applications. The BSTS model is computationally light, running on the order of seconds. Our results show similar performance compared to standard ARIMA models, but they are more transparent—by not relying on differencing, lags, and moving averages—and can elegantly provide uncertainty. Furthermore, we show the model's ability to be used for measurement and verification applications and how measurements of savings naturally arise when using a probabilistic model. The model results' interpretability can help building managers and policymakers glean relevant insights into relationships between customers, temperature, and submeter loads. The flexibility and abundance of information from the model allow these stakeholders to make more informed decisions.

Because erroneous forecasts have enormous financial costs for utilities, and as a result, these entities are attempting to measure forecast uncertainty better. Better PLFs allow grid operators to reduce their dependence on costly and polluting standby plants by proactively managing the grid's changing supply side. Furthermore, on the demand side, PLFs will also become more valuable as grid edge technologies proliferate, such as battery storage, which relies on forecasts to optimally control charging and discharging. Overall, through many different energy industry applications, probabilistic forecasting can help various decision-makers better capture the uncertainties inherent in forecasting, leading to improved risk-management and increased energy savings.

Author Contributions: Conceptualization, J.R. and C.M.; Data curation, J.R. and J.C.; Formal analysis, J.R.; Funding acquisition, C.M.; Investigation, J.R. and J.C.; Methodology, J.R.; Project administration, C.M.; Resources, C.M.; Supervision, R.K.J. and C.M.; Writing—original draft, J.R.; Writing—review and editing, J.R., J.C., R.K.J. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Singapore National Research Foundation (NRF) BSEWWT (Behavioural Studies in the Energy, Water, Waste and Transport) Grant through the Future Resilient Systems (FRS) Laboratory at the Singapore ETH Centre (SEC). This material is also based upon the work supported in part by the Precourt Institute for Energy and the United States National Science Foundation under Grant No. 1941695. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Institutional Review Board Statement: This publication is a secondary analysis of a study that is covered by IRB-2017-09-034 dated 27 September 2017 from the Nanyang Technological University (NTU) under the project titled: Using social comparison to promote energy and water efficiency in Singapore 2017–2019.

Informed Consent Statement: As this is secondary analysis of deidentified and aggregated data (Project Title: Using social comparison to promote energy and water efficiency in Singapore 2017–2019), informed consent is not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the economics research team at the Future Resilient Systems (FRS) Laboratory for the design and deployment of the field behavioral experiment model. The authors would also like to acknowledge the cooperation offered by the Office of Housing and Auxiliary Services (OHAS) and the Energy Research Institute at the Nanyang Technological University (NTU). This research contributes to the body of work for the International Energy Agency (IEA) Energy in Building and Communities (EBC) Annex 79-Occupant-Centric Building Design and Operation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McKinsey. *How Climate Change is Challenging the Power Industry*; McKinsey: Summit, NJ, USA, 2019.
2. Jain, R.K.; Qin, J.; Rajagopal, R. Data-driven planning of distributed energy resources amidst socio-technical complexities. *Nat. Energy* **2017**, *2*, 17112. [CrossRef]
3. Roth, J.; Martin, A.; Miller, C.; Jain, R. SynCity: Using open data to create a synthetic city of hourly building energy estimates by integrating data-driven and physics-based methods. *Appl. Energy* **2020**, *280*, 115981. [CrossRef]
4. Wang, Y.; Gan, D.; Zhang, N.; Xie, L.; Kang, C. Feature selection for probabilistic load forecasting via sparse penalized quantile regression. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 1200–1209. [CrossRef]
5. Greentech Media. *WoodMac: Smart Meter Installations to Surge Globally Over Next 5 Years*; Greentech Media: Boston, MA, USA, 2019.
6. Yang, Y.; Hong, W.; Li, S. Deep ensemble learning based probabilistic load forecasting in smart grids. *Energy* **2019**, *189*, 116324. [CrossRef]
7. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
8. FEMP. *M&V Guidelines: Measurement and Verification for Performance-Based Contracts—Version 4.0*; Technical Report November; Federal Energy Management Program (FEMP): Washington, DC, USA, 2015.
9. Kumar, R.; Wenzel, M.J.; Ellis, M.J.; Elbsat, M.N.; Drees, K.H.; Zavala, V.M. A Stochastic Model Predictive Control Framework for Stationary Battery Systems. *IEEE Trans. Power Syst.* **2018**, *33*, 4397–4406. [CrossRef]
10. Carstens, H.; Xia, X.; Yadavalli, S. Bayesian Energy Measurement and Verification Analysis. *Energies* **2018**, *11*, 380. [CrossRef]
11. Robert, C.P. *The Bayesian Choice From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2017 .
12. Wasserstein, R.L.; Lazar, N.A. The ASA's statement on p-values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129–133. [CrossRef]
13. Shonder, J.A.; Im P. Bayesian Analysis of Savings from Retrofit Projects (SA-12-003). In Proceedings of the 2012 ASHRAE Annual Conference, San Antonio, TX, USA, 23–27 June 2012.
14. Chong, A.; Lam, K.P.; Pozzi, M.; Yang, J. Bayesian calibration of building energy models with large datasets. *Energy Build.* **2017**, *154*, 343–355. [CrossRef]
15. Chong, A.; Menberg, K. Guidelines for the Bayesian calibration of building energy models. *Energy Build.* **2018**, *174*, 527–547. [CrossRef]
16. Hong, T.; Pinson, P.; Fan, S. Global energy forecasting competition 2012. *Int. J. Forecast.* **2014**, *30*, 257–363. [CrossRef]
17. Sun, M.; Wang, Y.; Strbac, G.; Kang, C. Probabilistic Peak Load Estimation in Smart Cities Using Smart Meter Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 1608–1618. [CrossRef]
18. Sun, M.; Wang, Y.; Teng, F.; Ye, Y.; Strbac, G.; Kang, C. Clustering-Based Residential Baseline Estimation: A Probabilistic Perspective. *IEEE Trans. Smart Grid* **2019**, 10, 6014–6028. [CrossRef]
19. Sun, M.; Zhang, T.; Wang, Y.; Strbac, G.; Kang, C. Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting. *IEEE Trans. Power Syst.* **2019**, *35*, 188–201. [CrossRef]
20. Hong, T. *Load Forecasting Case Study*; Technical report; University of North Carolina: Chapel Hill, NC, USA, 2015.
21. Nowotarski, J.; Weron, R. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Comput. Stat.* **2015**, *30*, 791–803. [CrossRef]
22. Wang, Y.; Gan, D.; Sun, M.; Zhang, N.; Lu, Z.; Kang, C. Probabilistic individual load forecasting using pinball loss guided LSTM. *Appl. Energy* **2019**, *235*, 10–20. [CrossRef]
23. Xie, J.; Hong, T.; Laing, T.; Kang, C. On Normality Assumption in Residual Simulation for Probabilistic Load Forecasting. *IEEE Trans. Smart Grid* **2017**, *8*, 1046–1053. [CrossRef]
24. Roth, J.; Rajagopal, R. Benchmarking building energy efficiency using quantile regression. *Energy* **2018**, *152*, 866–876. [CrossRef]

25. Wang, Y.; Zhang, N.; Tan, Y.; Hong, T.; Kirschen, D.S.; Kang, C. Combining Probabilistic Load Forecasts. *IEEE Trans. Smart Grid* **2019**. [CrossRef]
26. Ben Taieb, S.; Huser, R.; Hyndman, R.J.; Genton, M.G. Forecasting Uncertainty in Electricity Smart Meter Data by Boosting Additive Quantile Regression. *IEEE Trans. Smart Grid* **2016**, *7*, 2448–2455. [CrossRef]
27. Yang, Y.; Li, S.; Li, W.; Qu, M. Power load probability density forecasting using Gaussian process quantile regression. *Appl. Energy* **2018**, *213*, 499–509. [CrossRef]
28. Quan, H.; Srinivasan, D.; Khosravi, A. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Trans. Neural Networks Learn. Syst.* **2014**, *25*, 303–315. [CrossRef] [PubMed]
29. Khosravi, A.; Nahavandi, S.; Creighton, D. Construction of optimal prediction intervals for load forecasting problems. *IEEE Trans. Power Syst.* **2010**, *25*, 1496–1503. [CrossRef]
30. Hussain, M.; Gao, Y. A review of demand response in an efficient smart grid environment. *Electr. J.* **2018**, *31*, 55–63. [CrossRef]
31. Muratori, M.; Rizzoni, G. Residential Demand Response: Dynamic Energy Management and Time-Varying Electricity Pricing. *IEEE Trans. Power Syst.* **2016**, *31*, 1108–1117. [CrossRef]
32. Hledik, R.; Faruqui, A.; Lee, T.; Higham, J. The National Potential for Load Flexibility Value and Market Potential Through 2030 Prepared By The Brattle Group. In Proceedings of the 2020 ASHRAE Virtual Conference, 29 June–2 July 2020. Available online: https://www.ashrae.org/conferences/2020-virtual-annual-conference (accessed on 1 March 2021).
33. Nilsson, A.; Lazarevic, D.; Brandt, N.; Kordas, O. Household responsiveness to residential demand response strategies: Results and policy implications from a Swedish field study. *Energy Policy* **2018**, *122*, 273–286. [CrossRef]
34. Haben, S.; Singleton, C.; Grindrod, P. Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data. *IEEE Trans. Smart Grid* **2016**, *7*, 136–144. [CrossRef]
35. Sevlian, R.; Rajagopal, R. A scaling law for short term load forecasting on varying levels of aggregation. *Int. J. Electr. Power Energy Syst.* **2018**, *98*, 350–361. [CrossRef]
36. Miller, C.; Nagy, Z.; Schlueter, A. A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1365–1377. [CrossRef]
37. Wang, Y.; Chen, Q.; Hong, T.; Kang, C. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Trans. Smart Grid* **2019**, *10*, 3125–3148. [CrossRef]
38. Kwac, J.; Flora, J.; Rajagopal, R. Lifestyle Segmentation Based on Energy Consumption Data. *IEEE Trans. Smart Grid* **2018**, *9*, 2409–2418. [CrossRef]
39. Miller, C.; Nagy, Z.; Schlueter, A. Automated daily pattern filtering of measured building performance data. *Autom. Constr.* **2015**, *49*, 1–17. [CrossRef]
40. Quilumba, F.L.; Lee, W.J.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. *IEEE Trans. Smart Grid* **2015**, *6*, 911–918. [CrossRef]
41. Flath, C.; Nicolay, D.; Conte, T.; Van Dinther, C.; Filipova-Neumann, L. Cluster analysis of smart metering data: An implementation in practice. *Bus. Inf. Syst. Eng.* **2012**, *4*, 31–39. [CrossRef]
42. Roth, J.; Jain, R.K. Data-Driven, Multi-metric, and Time-Varying (DMT) Building Energy Benchmarking Using Smart Meter Data. In Proceedings of the Workshop of the European Group for Intelligent Computing in Engineering, Lausanne, Switzerland, 10–13 June 2018; pp. 568–593.
43. Räsänen, T.; Kolehmainen, M. Feature-based clustering for electricity use time series data. In Proceedings of the International Conference on Adaptive and Natural Computing Algorithms, Kuopio, Finland, 23–25 April 2009; Volume 5495 LNCS, pp. 401–412. [CrossRef]
44. Dent, I.; Aickelin, U.; Rodden, T.; Craig, T. Finding the Creatures of Habit; Clustering Households Based on Their Flexibility in Using Electricity. *SSRN Electron. J.* **2012**. [CrossRef]
45. Wang, E.; Alp, N.; Shi, J.; Wang, C.; Zhang, X.; Chen, H. Multi-criteria building energy performance benchmarking through variable clustering based compromise TOPSIS with objective entropy weighting. *Energy* **2017**, *125*, 197–210. [CrossRef]
46. Albert, A.; Rajagopal, R. Finding the right consumers for thermal demand-response: An experimental evaluation. *IEEE Trans. Smart Grid* **2016**, *9*, 564–572. [CrossRef]
47. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543. [CrossRef]
48. Miriam, L.G.; Agnew, G.K. *Measurement and Verification for Demand Response*; Technical report; US Department of Energy: Washington, DC, USA, 2013.
49. ISO. *ISO/IEC Guide 98-3:2008—Uncertainty of Measurement—Part 3: Guide to the Expression of Uncertainty in Measurement (GUM:1995)*; ISO: London UK, 2008.
50. *What is ASHRAE Guideline 14 and How Does It Affect Your M&V?* EnergyWatch: New York, NY, USA, 2014.
51. *Uniform Methods Project for Determining Energy Efficiency Program Savings*; Department of Energy: Washington, DC, USA, 2017.
52. Granderson, J.; Touzani, S.; Fernandes, S.; Taylor, C. Application of automated measurement and verification to utility energy efficiency program data. *Energy Build.* **2017**, *142*, 191–199. [CrossRef]
53. Ke, M.T.; Yeh, C.H.; Jian, J.T. Analysis of building energy consumption parameters and energy savings measurement and verification by applying eQUEST software. *Energy Build.* **2013**, *61*, 100–107. [CrossRef]

54. Ertugrul, Ö.F. Forecasting electricity load by a novel recurrent extreme learning machines approach. *Int. J. Electr. Power Energy Syst.* **2016**, *78*, 429–435. [CrossRef]
55. Zhao, H.x.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. [CrossRef]
56. Jain, R.K.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl. Energy* **2014**, *123*, 168–178. [CrossRef]
57. Heo, Y.; Zavala, V.M. Gaussian process modeling for measurement and verification of building energy savings. *Energy Build.* **2012**, *53*, 7–18. [CrossRef]
58. Arora, S.; Taylor, J.W. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega* **2016**, *59*, 47–59. [CrossRef]
59. Miller, C.; Meggers, F. Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. *Energy Build.* **2017**, *156*, 360–373. [CrossRef]
60. Brodersen, K.H.; Gallusser, F.; Koehler, J.; Remy, N.; Scott, S.L. Inferring causal impact using bayesian structural time-series models. *Ann. Appl. Stat.* **2015**, *9*, 247–274. [CrossRef]
61. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications: With R Examples*; Springer: Berlin/Heidelberg, Germany, 2006.
62. Kim, C.J.; Nelson, C.R. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*; MIT Press: Cambridge, MA, USA, 1999; Volume 1.
63. Ishwaran, H.; Rao, J.S. Spike and slab variable selection: Frequentist and bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [CrossRef]
64. Schubert, R.; Schmitz, J.; Tiefenbeck, V.; Borzino, N. Energy Conservation—Assessing the Impact of Goals and Financial Incentives to Foster Resilience. In Proceedings of the World Congress on Resilience, Reliability and Asset Management (WCRRAM 2019), Singapore, 28–31 July 2019; pp. 1–4.
65. Hong, T.; Pinson, P.; Wang, Y.; Weron, R.; Yang, D.; Zareipour, H. Energy Forecasting: A Review and Outlook. *IEEE Open Access J. Power Energy* **2020**, *7*, 376–388. [CrossRef]
66. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [CrossRef]
67. Zhang, Y.; Wang, J.; Wang, X. Review on probabilistic forecasting of wind power generation. *Renew. Sustain. Energy Rev.* **2014**, *32*, 255–70. [CrossRef]
68. BloombergNEF. 31 July 2019. Available online: https://about.bnef.com/blog/energy-storage-investments-boom-battery-costs-halve-next-decade (accessed on 1 February 2021).