Applied Energy 123 (2014) 168–178

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy

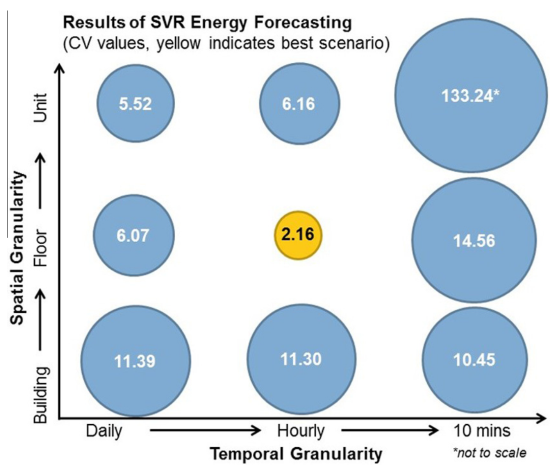
Rishee K. Jain [a,*], Kevin M. Smith [b], Patricia J. Culligan [c], John E. Taylor [d]

[a] Center for Urban Science & Progress, New York University, 1 MetroTech Center, 19th Floor, Brooklyn, NY 11201, United States
[b] Department of Earth and Environmental Engineering, Columbia University, S.W. Mudd Building, 500 West 120th Street, New York, NY 10027, United States
[c] Institute for Data Sciences & Engineering, Department of Civil Engineering and Engineering Mechanics, Columbia University, Room 626, S.W. Mudd Building, 500 West 120th Street, New York, NY 10027, United States
[d] Department of Civil and Environmental Engineering, Virginia Tech, 113B Patton Hall, Blacksburg, VA 24061, United States

HIGHLIGHTS

- We develop a building energy forecasting model using support vector regression.
- Model is applied to data from a multi-family residential building in New York City.
- We extend sensor based energy forecasting to multi-family residential buildings.
- We examine the impact temporal and spatial granularity has on model accuracy.
- Optimal granularity occurs at the *by floor* in *hourly* temporal intervals.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form 28 December 2013
Accepted 15 February 2014
Available online 15 March 2014

Keywords:
Forecasting
Machine learning
Monitoring
Prediction
Residential building
Support vector regression

ABSTRACT

Buildings are the dominant source of energy consumption and environmental emissions in urban areas. Therefore, the ability to forecast and characterize building energy consumption is vital to implementing urban energy management and efficiency initiatives required to curb emissions. Advances in smart metering technology have enabled researchers to develop "sensor based" approaches to forecast building energy consumption that necessitate less input data than traditional methods. Sensor-based forecasting utilizes machine learning techniques to infer the complex relationships between consumption and influencing variables (e.g., weather, time of day, previous consumption). While sensor-based forecasting has been studied extensively for commercial buildings, there is a paucity of research applying this data-driven approach to the multi-family residential sector. In this paper, we build a sensor-based forecasting model using Support Vector Regression (SVR), a commonly used machine learning technique, and apply it to an empirical data-set from a multi-family residential building in New York City. We expand our study to examine the impact of temporal (i.e., *daily*, *hourly*, *10 min* intervals) and spatial (i.e., *whole building*, *by floor*, *by unit*) granularity have on the predictive power of our single-step model. Results indicate

* Corresponding author. Tel.: +1 16469970545.
E-mail addresses: rishee.jain@nyu.edu (R.K. Jain), kms2227@columbia.edu (K.M. Smith), pjc2104@columbia.edu (P.J. Culligan), jet@vt.edu (J.E. Taylor).

http://dx.doi.org/10.1016/j.apenergy.2014.02.057

that sensor based forecasting models can be extended to multi-family residential buildings and that the optimal monitoring granularity occurs at the *by floor* level in *hourly* intervals. In addition to implications for the development of residential energy forecasting models, our results have practical significance for the deployment and installation of advanced smart metering devices. Ultimately, accurate and cost effective wide-scale energy prediction is a vital step towards next-generation energy efficiency initiatives, which will require not only consideration of the methods, but the scales for which data can be distilled into meaningful information.

1. Introduction

The built environment accounts for 40% of energy consumption in the United States [1] and a similar proportion in many other parts of the world. In dense urban areas like New York City, buildings account for a staggering 94% of electricity usage and 75% of greenhouse gas (GHG) emissions [2]. Therefore, characterizing, modeling and forecasting the energy consumption of buildings is crucial if urban areas are to reduce their overall energy consumption. Accurate modeling and forecasting of building energy demand enables numerous energy management and efficiency applications such as: informing early stage design decisions [3], estimating improvements to building energy performance [4], optimizing building HVAC systems [5] and urban energy infrastructure planning [6].

Traditionally, building scale demand estimates have been performed using engineering software packages (e.g., EnergyPlus) that rely on an in-depth compilation of structural, geometric, and material building properties. The difficulty in obtaining and validating such information is a hindrance to wide-scale energy forecasting. In response, there is a growing interest in alternative "sensor based" approaches that forgo the demanding input requirements of theoretical formulations in favor of a practicable set of localized historical measurements.

In a sensor based approach, data from energy smart meters, building management systems, and weather stations are fed into a machine learning algorithm in order to infer the complex relationships between energy consumption and variables of influence such as temperature, time of day and occupancy. Prior research [7,8] has established that the prediction accuracy of sensor-based approaches is comparable, and sometimes superior, to traditional engineering based energy forecasting while requiring significantly less input data from the end user. Due to the accelerated development and proliferation of inexpensive off-the-shelf options for energy metering in recent years, sensor based approaches are becoming increasingly relevant and cost effective. A comprehensive discussion regarding the advantages of sensor based energy forecasting over traditional engineering methods can be found in [7].

While sensor based forecasting has been applied extensively in the commercial building sector since the launch of the American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE) Great Energy Predictor Shootout [9], there is a dearth of research applying this method to the residential building sector. The limited body of work [7] for the residential sector has been focused on single-family homes and has largely ignored multi-family residential units. As multi-family units are the dominant residential building stock in dense urban areas, accurate forecasting of their consumption is imperative to implementing effective energy efficiency and conservation measures on an urban scale.

In this study, we extend the sensor based energy forecasting approach to a monitored multi-family residential apartment building in New York City. Additionally, we deepen our exploration of sensor based forecasting by examining the effectiveness of our prediction algorithm at making single-step forecasts on various spatial and temporal data scales. Unlike single-family buildings, energy consumption data for multi-family residential complexes are often readily available at both the unit and building aggregate level. While utilities have traditionally reported residential energy use at the monthly scale, readily available metering technology can now report energy consumption as frequently as every minute. In general, the impact of temporal scales on sensor based energy forecasting has not been explored, and the relative utility of high-resolution monitoring remains an open question. By varying the aggregation granularity of monitoring data across several temporal (i.e., *daily*, *hourly*, *every 10 min*) and spatial (i.e., *whole building*, *by floor*, *by unit*) scales we hope to gain insight into the impact that each scale has on the forecasting accuracy of our model. In the end, our goal is to determine the optimal monitoring granularity for energy consumption that optimizes forecasting accuracy with both the financial and nonfinancial costs of installing and operating monitoring equipment.

2. Related work

2.1. Sensor based energy forecasting of commercial and residential buildings

Sensor based energy forecasting has been explored by researchers for both commercial and residential buildings. However, due to a lack of data availability for residential buildings the majority of previous work has been focused on the prediction of consumption in commercial buildings. Sensor based energy forecasting gained popularity after the first Great Energy Predictor Shootout that was hosted by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) [9]. The Great Energy Predictor Shootout was a competition in which participants were asked to develop models to predict the whole building electrical (WBE) consumption of a commercial building based on historical electrical consumption and environmental data. The overall winner of the competition [10] developed a sensor based energy forecasting model and employed an intelligent machine learning algorithm that required very little explicit domain knowledge about the actual commercial building whose energy was to be predicted. The results of the competition spurred increased attention by the research community towards the use of sensor based forecasting models.

Subsequently, several studies further explored the use of sensor based energy forecasting in commercial buildings [11–16] and to a lesser extent residential buildings [7,17]. The limited work examining residential building energy forecasting has been narrowly focused on single-family homes [7]. While energy forecasting of single-family homes is certainly valuable, in the context of dense urban areas multi-family residential units constitute a significant portion of the residential building stock and, thus, energy consumption associated with this sector. Therefore, the first contribution of this study is to investigate whether previously established sensor based forecasting models will yield acceptable levels of accuracy when applied to multi-family residential buildings. We aim to demonstrate that sensor based forecasting methods can be extended to multi-family residential buildings. Multi-family residential buildings by definition are comprised of numerous independent apartment units. Each independent unit has their own characteristics, occupancy levels and, most importantly, energy consumption patterns. For this reason, forecasting multi-family residential buildings represents a unique problem in the sensor based forecasting literature. Sensor based forecasting relies on machine learning algorithms to infer the relationship between such consumption patterns and the specific characteristics and occupancy levels to construct the most powerful predictive model [7]. In theory, a sensor based forecasting model built with spatially granular data would predict energy consumption better than a comprehensive building model since it would be able to account for each individual unit's characteristics, occupancy levels and consumption patterns. The unique composition of a multi-family residential building raises the question: *does spatial granularity impact the predictive power of sensor based energy forecasting models?*

Among the body of work regarding sensor based energy forecasting, studies have been conducted at varying temporal scales. Energy consumption modeling and forecasting was conducted in [12,17] using monthly data and in [13,16] using daily data. Energy consumption data has been increasingly more accessible to researchers and as a result studies have been conducted on an hourly scale [14,15,18] and even in intervals of 15 min [7]. While the trend has been towards utilizing more temporally granular data sets in sensor based forecasting applications, the impact of temporal granularity has yet to be analyzed using a single comprehensive and high resolution energy consumption data set. For this reason, we ask the question: *does temporal granularity impact the predictive power of sensor based energy forecasting models?*

2.2. Machine learning algorithms in sensor based energy forecasting

Machine learning algorithms form the basis for sensor based energy forecasting models. The algorithms train on a historical data set to infer the complex relationship between current energy consumption and influencing variables (e.g., previous consumption, temperature, day of the week). In the literature, the two most common machine learning algorithms utilized for sensor based energy forecasting are:

- Artificial Neural Networks (ANN) utilized in [11,13,16,18].
- Support Vector Machines (SVM) utilized in [12,19–22].

A version of SVM for regression estimation is known as Support Vector Regression (SVR) and was introduced in [23]. Prior work [14,15] has indicated SVR outperforms ANN methods when aiming to predict building energy loads associated with cooling. Moreover, Kavaklioglu [19] concluded that electricity consumption can be modeled and future consumption predicted using SVR. A comparative analysis by Ruas et al. [24] found SVR more accurately predicts electrical energy demand than ANN. ANN based methods have been shown to have significantly slower running speeds than SVR methods on large high resolution sets [8]. Based on these previous findings, we chose to build this initial application of sensor based forecasting for multi-family residential buildings using Support Vector Regression (SVR). The sparsity of the support vectors in SVR allow for the model to be scalable to our large high resolution empirical dataset. A more detailed explanation on the mechanics of SVR is provided in Section 3.2.1.

3. Methodology

Our primary objectives are to assess whether sensor based single-step forecasting of energy consumption can be successfully extended to the multi-family residential sector and to ascertain the impact various spatial and temporal scales have on the predictive power of our single-step forecasting model. We chose to develop a model based on Support Vector Regression (SVR) using the Scikit-learn module [25], which provides a Python front-end to LIBSVM, a widely cited Support Vector Machine library [26]. We then validated the model using data from the Great Energy Predictor Shootout and tested various temporal and spatial scenarios using our empirical data set, which was collected from a multi-family residential building in New York City. This section is organized as follows: Section 3.1 provides a definition of the model's inputs and the performance metrics utilized to assess our algorithm's predictive power, Section 3.2 presents the mechanics of SVR and the optimization techniques for the SVR parameters employed, Section 3.3 provides an overview of the algorithm structure, Section 3.4 describes how we validated our algorithm using the Great Energy Predictor Shootout data set and Section 3.5 introduces our empirical data set, the spatial and temporal scenarios that we tested and the method used to compare scenarios.

3.1. Model definition and performance metrics

We define two sets of model inputs: $M1$ for validation on the Great Energy Predictor Shootout data and $M2$ for testing on our own empirical data set. $M1$ is consistent with inputs presented in [7] and is defined as the following:

$$M1 : \overrightarrow{x(t)} = [y(t-1),\ y(t-2),\ T(t),\ s,\ sh,\ ch] \tag{1}$$

where $y(t-1)$ and $y(t-2)$ represent known electrical consumption values for the previous two time steps, $T(t)$ is the current temperature, $S(t)$ is the current solar flux, s is an indicator variable that denotes weekend/holiday or weekday, sh is the sine of the current hour and ch is the cosine of the current hour.

Based on previous work [13,18] and a lack of availability of solar flux data, we refine the model inputs as $M2$ for testing on the Watt Hall data set as:

$$M2 : \overrightarrow{x(t)} = [y(t-1),\ y(t-2),\ T(t),\ s,\ sh,\ ch] \tag{2}$$

We utilize the coefficient of variation (CV) as the primary performance metric in our analysis. The CV metric was chosen to be consistent with the metric used in the ASHRAE Great Energy Predictor Shootout [9] and several other studies [7,13,18] to measure the performance of sensor based forecasting models. Such consistency will facilitate comparison of our results to previous forecasting work both in the residential [7] and commercial building [13,18] sectors. The CV metric is defined by:

$$\mathrm{CV} = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\bar{y}} * 100 \tag{3}$$

where $\hat{y}_i$ is the predicted value, y_i is the observed value, $\bar{y}$ is the mean of the observed values and N is the total number of observations.

3.2. Support vector regression and model optimization

3.2.1. Mechanics of support vector regression

Support Vector Regression (SVR) distinguishes itself from other methods of predicting continuous variables by exhibiting a high degree of generalization when introduced to previously unseen data. SVR is also able to achieve high degrees of consistency by relying upon only a subset of the training observations known as the support vectors. In SVR, the support vectors are distinguished from the remainder of the training observations by a discriminating loss function that does not penalize residuals less than a tolerance ε. As a consequence, for a given hypothesis and ε, the observations constrained to the "ε-tube" bounding the hypothesis have no influence on the predictions. The "ε-tube" is illustrated in Fig. 1. In accordance with the principle of Structural Risk Minimization, for a given ε, SVR attempts to find a hypothesis with not only a small structural risk, but also a reduced complexity. The primal objective function is formally given as:

$$\textbf{Minimize}: \frac{1}{2}||\omega||^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{4}$$

where ω is the vector of feature weights. A smaller Euclidean norm of ω is suggestive of a "flatter" hypothesis with reduced complexity. ξ_i and ξ_i^* capture the magnitude of residuals beyond the prescribed tolerance ε and serve to guarantee a solution for all ε (illustrated in Fig. 1). C is a regularization term that determines the degree of the linear penalty applied to the residual excess $\xi_i^{(*)}$. In order to facilitate a computationally efficient non-linear case, SVR often employs the "kernel trick" that implicitly maps the input space into a higher dimensional feature space using a kernel function ϕ.

Accordingly, Eq. (4) is subject to the following constraints:

$$y_i - \omega \cdot \phi(x_i) - b \leqslant \varepsilon + \xi_i \tag{5}$$

$$\omega \cdot \phi(x_i) + b - y_i \leqslant \varepsilon + \xi_i^*, i = 1, \ldots, l \tag{6}$$

$$\xi_i^{(*)} \geqslant 0 \tag{7}$$

It should be noted that we chose to utilize an epsilon-SVR model in our analysis. We chose an epsilon-SVR model over a nu-SVR or Least Squares SVR model as to not restrict the number of support vectors in the model, to maintain sparsity and to be consistent with previous literature [12,14,15,24] so that comparisons could be directly made between our results and previous work. The Gaussian radial basis function (RBF) is one of the most widely used kernel functions because of its ability to generalize non-linear functions and efficiently manage large data-sets [20]. The RBF kernel is formulated as follows:

$$\varphi(x, x') = \exp(-\gamma|x - x'|^2), \quad \gamma > 0 \tag{8}$$

where γ is the kernel parameter. Intuitively, γ defines the radius of influence for each data point. It is important to note that with the RBF implementation C, ε and γ are user-defined variables that each have a significant influence on the SVR outcome.

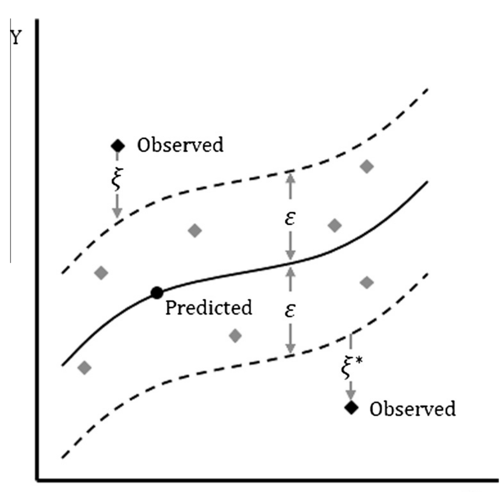


Fig. 1. Parameters of SVR, adapted from [12,15,27].

3.2.2. Model optimization

As stated earlier, the objective function of SVR utilizes two user-defined parameters: the training tolerance ε, and the regularization term C. Additionally, utilization of the Gaussian radial basis function (RBF) requires a value for γ, which determines the radius of influence of each support vector. We found that for the Great Energy Predictor Shootout and our empirical datasets, our model was relatively insensitive to values of ε in the range ($10^{-5} \leqslant \varepsilon \leqslant 10^{-1}$) whereas both γ and C exhibited sensitivity and necessitated independent fitting. While some heuristic approaches exist, small-dimension parameter selection is typically performed by brute force methods that exhaustively search a grid of plausible parameter combinations scoring each case based on model performance. For each combination of C and γ, k-folds cross validation was performed to limit overfitting. The k-folds method of cross validation [28] splits the training set into k sequential subsets of near-equal size, holding out each in turn as a validation set, while using the other k-1 subsets to train. The determinate nature of the k-folds approach, as compared to randomized cross validation strategies, means that the results of each combination of C and γ are directly comparable. As such, the mean value of the CV metric across each of the folds (with $k = 5$) was used to establish relative performance. We implemented a recursive routine to iterate through increasingly finer grids so long as $|\Delta CV_{min}|$ between iterations was greater than a defined threshold (i.e., $|\Delta CV_{min}| > \varphi$ where $\varphi = 0.05$).

3.3. Algorithm overview

In this section, we provide a complete overview of the general algorithm used for model selection and evaluation, as shown in Fig. 2. The algorithm is general in the sense that it is invariant to the temporal and spatial granularity of the input dataset, D, so as to preclude errors made through inconsistent procedures across scales. Practically we approached this by pre-formatting the observations in D to the appropriate timescale (i.e., *daily*, *hourly*, *every 10 min*), and by defining the set of families within the building $f \in B$ that will be used to predict the whole building electricity usage (WBE). At the *by unit* scale, each unit is modeled as an independent family and at the *by floor* scale, each floor is modeled as an independent family, whereas at the *whole building* scale the behavior is modeled as a single entity $B = \{f_1\}$. The model is fully defined by a set of predictor variables X, as the other parameters (C, γ) are set by a grid-search procedure using k-folds cross validation, as shown in Fig. 3. The initial pre-processing step of the algorithm creates time lagged observations $y(t-1)$ and $y(t-2)$ as specified in Eq. (2), and applies relevant scaling or normalizing transformations to the data.

The next step is parameter selection for D and γ and is performed independently for each modeled family f The algorithm for the parameter selection sub-process is provided in Fig. 3. As there is no *a priori* basis for choosing optimum parameter values, models are fit and scored for a large grid G of plausible combinations of C and γ. To prevent overfitting during this process, k-folds cross validation is employed. As noted above, the k-folds cross validation technique splits the input data into k consecutive sections $S_{1\rightarrow k}$ of near-equal size. For each combination j of these folds, the model is trained on all folds $S_{1\rightarrow k}$ except S_j which is held out for validation. The score recorded for each combination of parameters C and γ is the mean value of the CV function across the folds. For the grid G, the best performing combination of parameters (C', γ') is chosen by having the lowest score. This process is repeated with a smaller but finer grid G' centered at (C', γ') for as long as the marginal change in $|\Delta CV_{min}|$ between iterations is greater than the defined threshold φ.

After parameters are set for a family f, models are continuously trained and evaluated through a bootstrapping process that gives

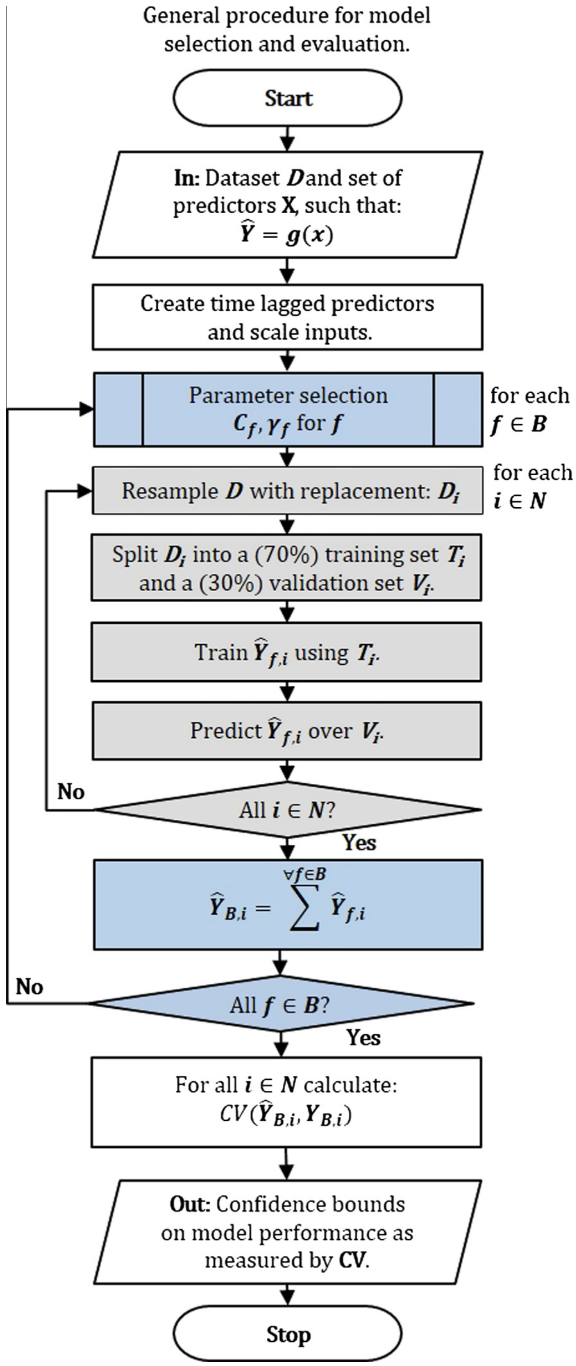


Fig. 2. General algorithm used for model selection and validation.

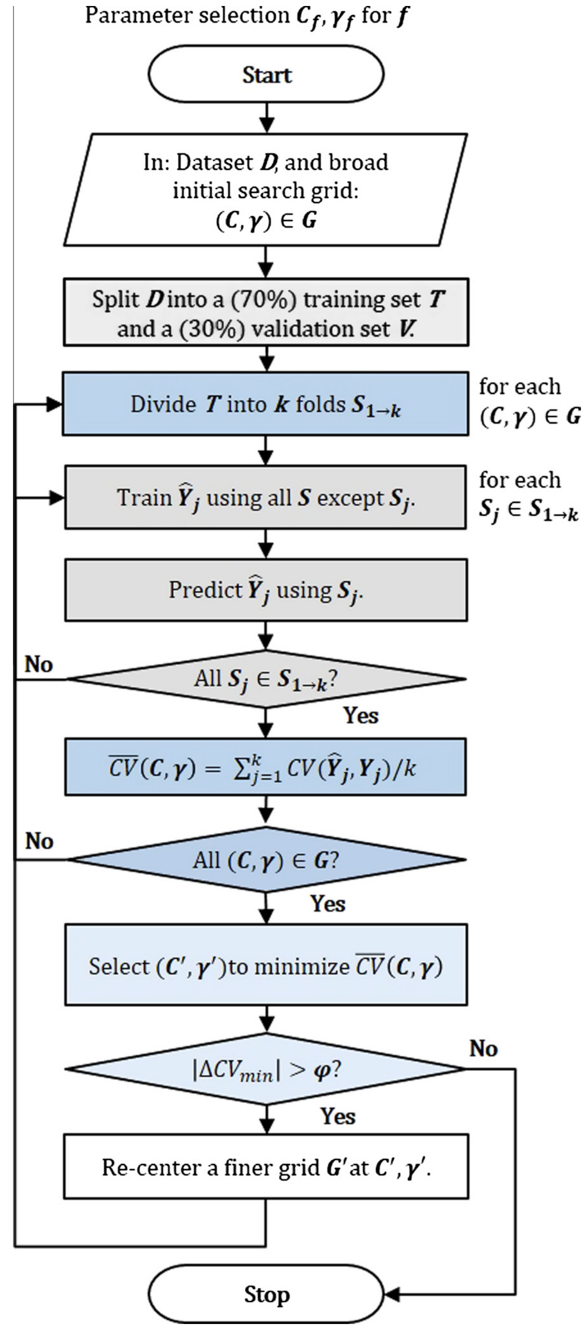


Fig. 3. Algorithm to describe parameter selection sub-process.

bounds on our confidence in model performance as measured by the CV value. For each family, the dataset is resampled N times. Each bootstrap resampling i of D is split into a unique training set T_i and a validation set V_i that are 70% and 30% of D_i respectively. The bootstrapping process for each family is initiated with a common random seed so that the predictions $\widehat{Y}_{f,i}$ can be aggregated into a WBE prediction $\widehat{Y}_{B,i}$. This aggregated prediction is then evaluated against the observed WBE values $\widehat{Y}_{B,i}$ in the CV function for each $i \in N$.

3.4. Algorithm validation

In order to validate our algorithm, we executed it using the Great Energy Predictor Shootout (GEPS) data set from [9] and compared the results to [7]. The GEPS data set consists of hourly data on the whole building electricity usage (WBE), temperature, insolation, humidity and wind speed from September 1, 1989 to February 23, 1990. Following the approach taken by the authors in [7] (as described in Eq. (1)), humidity and wind speed were ignored as inputs.

Consistent with previous studies we utilized 75% of the GEPS data set to train our model and 25% to assess the results. Our grid-search optimization techniques, as described in the previous section, returned values of 10,000 and 1E-5 for C and γ, respectively. The results are illustrated in Fig. 4.

The results of the predictions are shown in Figs. 5 and 6 with the original data shown in black and the predicted values in red (dashed line). The resulting CV value for our model was 3.30%, which is lower than the reported results for the GEPS data set in [7], thus validating our approach.

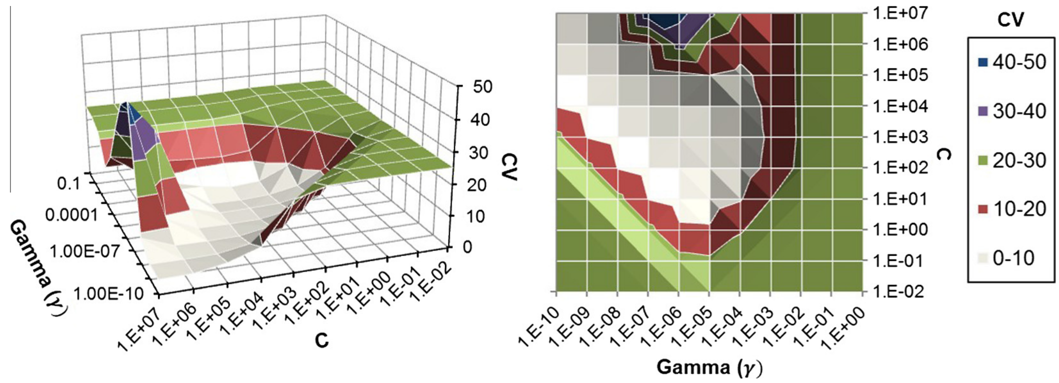


Fig. 4. Grid-search optimization for GEPS data.

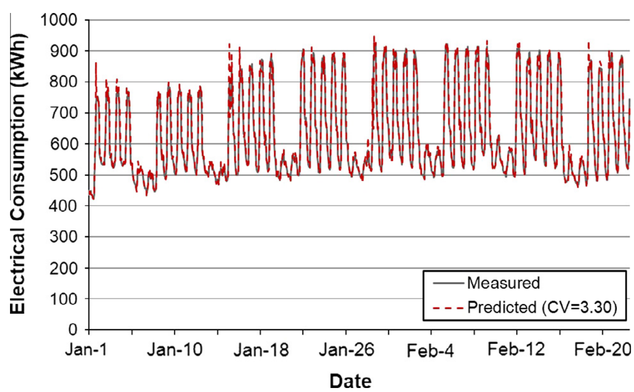


Fig. 5. Results of our model on the GEPS data set (C = 10,000; γ = 1e-5).

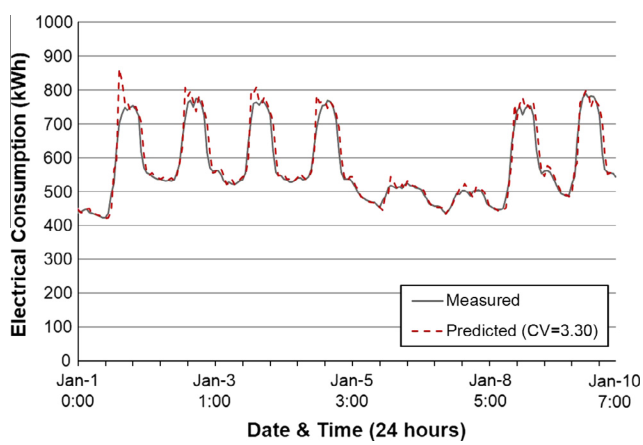


Fig. 6. Detailed results of our model on the GEPS data set on a small time interval (C = 10,000; γ = 1e-5).

3.5. Scenario testing on empirical data

3.5.1. Empirical data set

We collected empirical data from a building named Watt Hall, which is a six story multi-family residential building located on the Columbia University campus in New York City. The building is a pre-war construction and has high ceilings and thick plaster walls. All units have access to natural light either via a courtyard or the street. Each unit contains a kitchen, bathroom, living and bedroom areas.

The Watt Hall data set consists of electrical consumption data in 10 min intervals for 21 units in the building that were not subjected to any energy efficiency interventions. The data set spans from August 27, 2012 to December 19, 2012. A detailed description of the energy consumption monitoring and data acquisition system can be found in [28]. The electrical consumption data were supplemented with hourly outdoor temperature data downloaded from the National Oceanic and Atmospheric Association (NOAA) Central Park Weather Station. Additionally, university-wide holidays were encoded into the model's *s* indicator variable.

Table 1
Spatial and temporal scenarios tested.

Temporal granularity	Spatial granularity		
	Whole building	By floor	By unit
Daily	*1A*	*1B*	*1C*
Hourly	*2A*	*2B*	*2C*
10 min	*3A*	*3B*	*3C*

3.5.2. Scenarios tested

In order to examine the impact spatial and temporal granularity has on sensor based forecasting models, we tested the following scenarios described in Table 1.

For whole building scenarios *1A*, *2A* and *3A*, the SVR's user defined input parameters (C, γ) are optimized for the aggregated energy consumption patterns of the individual units (i.e., a summation of the energy consumption of all 21 units included in this study). For scenarios *1B*, *2B* and *3B* the input parameters (C, γ) are optimized for the aggregated energy consumption patterns of each floor (i.e., a summation of the energy consumption of units on each of the six floors with floor 1 consisting of 4 units, floor 2 consisting of 3 units, floors 3–5 consisting of 4 units and floor 6 consisting of 2 units). For scenarios *1C*, *2C* and *3C* each individual unit's energy consumption patterns as described in Section 3.3. It should be noted that the time of day variables (*sh*, *ch*) are not utilized in the *daily* scenarios (*1A, 1B, 1C*) and hourly temperature readings are utilized in the *10 min* scenarios (*3A, 3B, 3C*) due to lack of availability of publically accessible NOAA temperature data at time scales under one day. Additionally, the coefficient of variation (CV) metric for the *by floor* (*1B, 2B, 3B*) and *by unit* (*1C, 2C, 3C*) cases was calculated by summing the predicted consumption values for all units at each time step and comparing it to the summed real consumption of all floors or units. This allows for accurate representation of how the predictive model performed in terms of the total building consumption and enables a direct comparison of all the scenarios. Accuracy bounds for each of the spatial and temporal scenarios were obtained using the bootstrapping resampling method. The mean and standard error[1] was computed for each scenario from 100 resamples.

[1] The standard error is defined as the standard deviation of the results obtained using the bootstrapping resampling method.

4. Results and discussion

4.1. Overall results

The mean and standard error of the CV is provided in Table 2 for all scenarios and a baseline persistence model. A baseline persistence model simply takes the previous time step's consumption as the predicted value and has been utilized in previous work [31] to contextualize model performance. Plots of the whole building, by floor, by unit predictive results are provided in Figs. 7–9. All scenarios, except for scenario *3C* (*by unit, 10 min*), returned CV values that were at least 50% lower than CV values reported in previous work that applied sensor based forecasting to residential buildings [7]. In the commercial building sector where forecasting is notably easier due to less variability in occupant behavior, a sensor based energy forecasting model is considered acceptable if the model returns CV values fall between 2% and 13% [7,9]. For all scenarios except *3C*, the results of our single-step predictive model were within this acceptable accuracy range. Thus, we are able to demonstrate that sensor based forecasting techniques can be applied to multi-family residential buildings successfully and extend prior residential energy forecasting research [7,17]. A comparison of the results indicated that scenario *2B* (*by floor, hourly*) yielded the lowest CV values and was statistically distinct (i.e., at least one standard error apart) than all other scenarios. Overall, the *by floor* at *hourly* temporal interval (scenario 2B) performed the best with a CV value of 2.16. The SVR model performed significantly better than the persistence model for *hourly* and *daily* temporal intervals. However, for *10 min* temporal intervals the SVR model performed only marginally better using *whole building* granular data. The *by floor* and *by unit* models (*B*/*C* scenarios) performed better than the *whole building* models (*A* scenarios) with the exception of the most granular *10 min* scenarios (*3A*, *3B, 3C*). For this temporal interval, the *by unit* model (*3C*) performed an order of magnitude worse than the *whole building* model (*3A*). A detailed analysis and discussion of the impact spatial and temporal granularity has on the results is provided in Sections 4.2 and 4.3.

4.2. Spatial granularity analysis

In order to ascertain the impact spatial granularity has on the predictive power of our sensor based forecasting model, we conducted a comparative spatial analysis for each temporal interval The *by floor* scenarios (*B*) perform statistically better than the *whole building* and *by unit* cenarios (*A, C*) for *hourly* predictions. However, for temporally granular *10 min* intervals the *whole building* scenario performs the best. Previous research [30,32] has concluded that a significant portion of the variability associated with residential energy consumption can be traced to occupant behavior and therefore we postulate that this improvement in CV values is the result of the model's ability to more successfully characterize the erratic consumption behavior at *daily* and *hourly* intervals of each unit's residential occupants. The *by unit* (*C*) scenario models accomplish this by constructing, optimizing and training an SVR to each unit's unique energy consumption signature. The subtleties of these dynamics are lost in models that rely on past observations

Fig. 7. Daily forecasting results for whole building, by floor, by unit and baseline persistence model.

of WBE data, and thus they tend to provide a less accurate WBE prediction than the aggregated predictions of the independently modeled units. Interestingly, at the *hourly* temporal interval the *by floor* spatial aggregation allows for the model to accurately capture changes in occupant consumption behavior without overfitting to spikes in consumption. Our results suggest that the optimal clustering of units for single-step *hourly* energy forecasting occurs somewhere between one unit and the whole building. Thus, there may be a considerable advantage to monitoring consumption and constructing a sensor based forecasting model at granularities smaller than the whole building level.

Comparative consumption plots for the *daily*, *hourly* and *10 min* intervals are provided in Figs. 10–12. A sample floor (2nd floor) is given for the by floor scenarios (*1B, 2B, 3B*) and sample unit (3I) is given for the *by unit* scenarios (*1C, 2C, 3C*). It can be seen that in the *by floor* and *by unit* scenarios the individual models are able to account for the larger variability that occurs and perform better than the single *whole building* model. Interestingly, both the *by floor* and *by unit* models have higher CV values than the *whole building* model when assessed on their ability to predict unit level consumption. However, when all units are aggregated they are statistically better at predicting the entire building's consumption than the *whole building* model as noted in Table 2. We postulate that this is due to the fact that the aggregation process significantly dampens the effect of random prediction errors at the *by floor* and *by unit* level through cancellation.

Results of the *10 min* temporal interval are strikingly different than the two smaller temporal intervals (i.e., *hourly*, *daily*). The model performs an order of magnitude better in the *whole building* and by *floor* scenarios (*3A, 3B*) than in the *by unit* scenario (*3C*) and even more indicative of the model's poor performance is the large standard error observed through the bootstrapping resampling method (10.10%). A more detailed examination of the plots in Fig. 13 reveal that the poor performance of the model on the *by unit* level stems from the model trying to capture the spikes in energy consumption that occur for short periods of time in a specific unit. These spikes in consumption can be attributed to the on/off cycling of large appliances such as a refrigerator or other short-term

Table 2
Summary of results (CV values and standard error in%) for scenarios. Bold indicates the best overall result.

Temporal granularity	Spatial granularity			Baseline model
	Whole building	By floor	By unit	Persistence
Daily	*11.39 ± 2.73*	*6.07 ± 1.79*	*5.52 ± 1.94*	*11.23*
Hourly	*11.30 ± 0.65*	***2.16 ± 0.28***	*6.16 ± 1.06*	*12.40*
10 min	*10.47 ± 0.29*	*14.56 ± 2.36*	*133.24 ± 10.10*	*11.17*

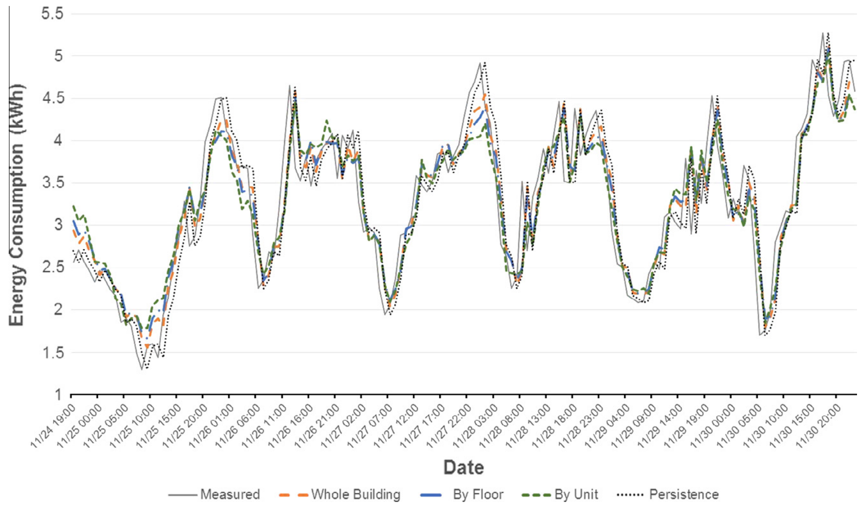


Fig. 8. Hourly forecasting results for whole building, by floor, by unit and baseline persistence model.

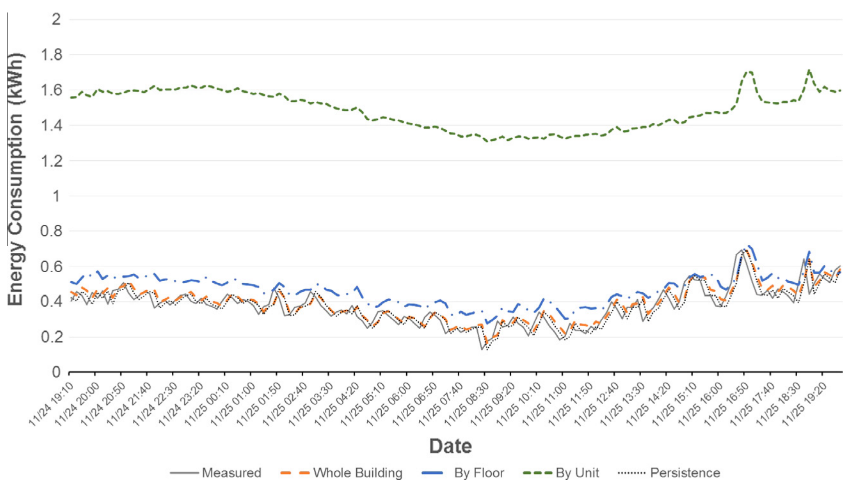


Fig. 9. 10 min forecasting results for whole building, by floor, by unit and baseline persistence model.

energy intensive activities. The result is an SVR model that consistently predicts consumption higher than the actual measured value. In order to ameliorate this issue, previous work [17,32] has employed the use of data transformations that can even out the spikes in an energy consumption dataset and prevent the model from overfitting to a small number of extreme points. We implemented a simple log transformation on our dataset to understand if predictions can be improved for the highly granular *by unit, 10 min* scenario (*3C*). The log transformation is as follows:

$$y_t = \log(y + 1) \tag{9}$$

where y_t is the transformed consumption value that maps to ($y = 0$, y_t=0) and y is the original consumption value. The transformation marginally improved the predictive power of the model and reduced the CV value from 133.24% to 119.54% and reduced the standard error from 10.09% to 9.27%. The model's ability to better account for spikes in consumption on a *10 min* temporal interval in each unit is visible in the comparison of the two plots in Fig. 11. The predicted values have shifted downward and are more in sync with the majority of the consumption profile. While future research is necessary to explore data transformations in greater detail, this example points to the possible potential of data transformation to improve the predictive power of highly spatial and temporally granular sensor based energy forecasting models.

4.3. Temporal granularity analysis

Overall, our results indicate that temporal granularity has less of an impact on the predictive power of our sensor based forecasting model than spatial granularity. However, this type of direct comparison of CV values across temporal scenarios does not account for the additional difficulty required to predict consumption behavior in the next 10 min, as compared to predicting gross behavior for the next day. In other words, it is inherently harder to predict at smaller time intervals due to the stochastic nature of energy consumption but also more valuable to do so for applications in HVAC optimization [5] and demand response [33]. We propose a modified CV performance metric that evaluates the results of the model based on the prediction accuracy over the largest time interval (in our case one day). This modified CV metric assess how well a model predicts the total building energy consumption over a 24-hour window and is given in the following equation:

$$\mathrm{CV}_{mod} = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \sum_{t=1}^{d}\widehat{y_{it}}\right)^2}}{\bar{y}} * 100$$

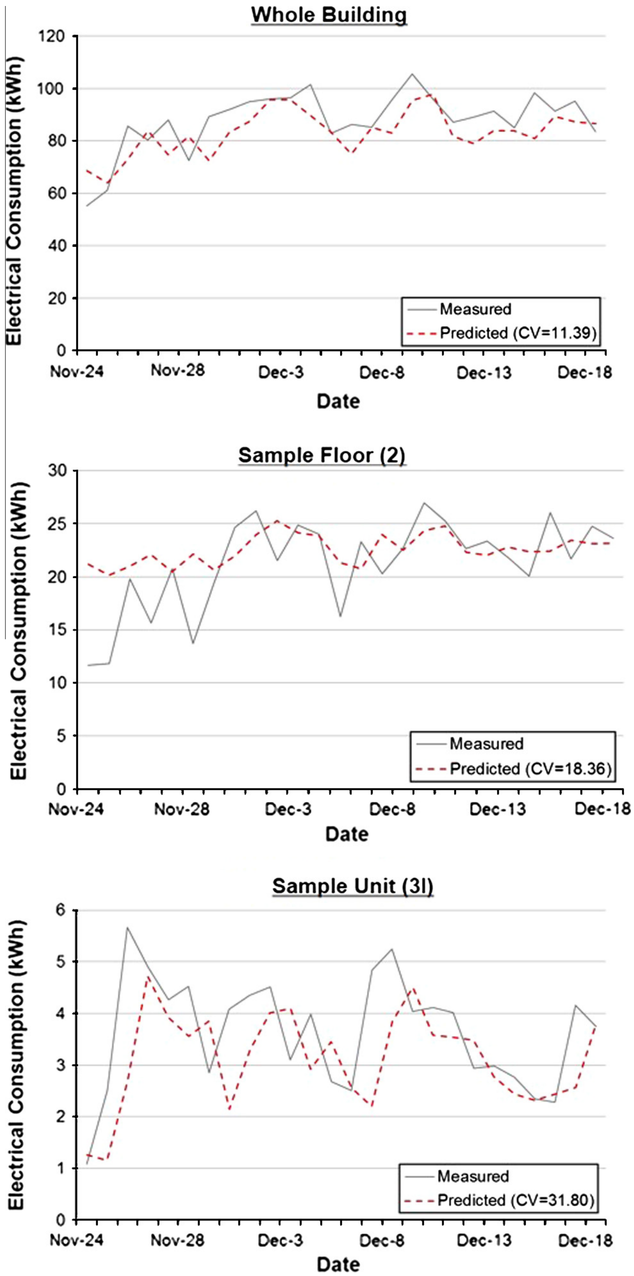


Fig. 10. Daily forecasting results for whole building, sample floor and a sample unit.

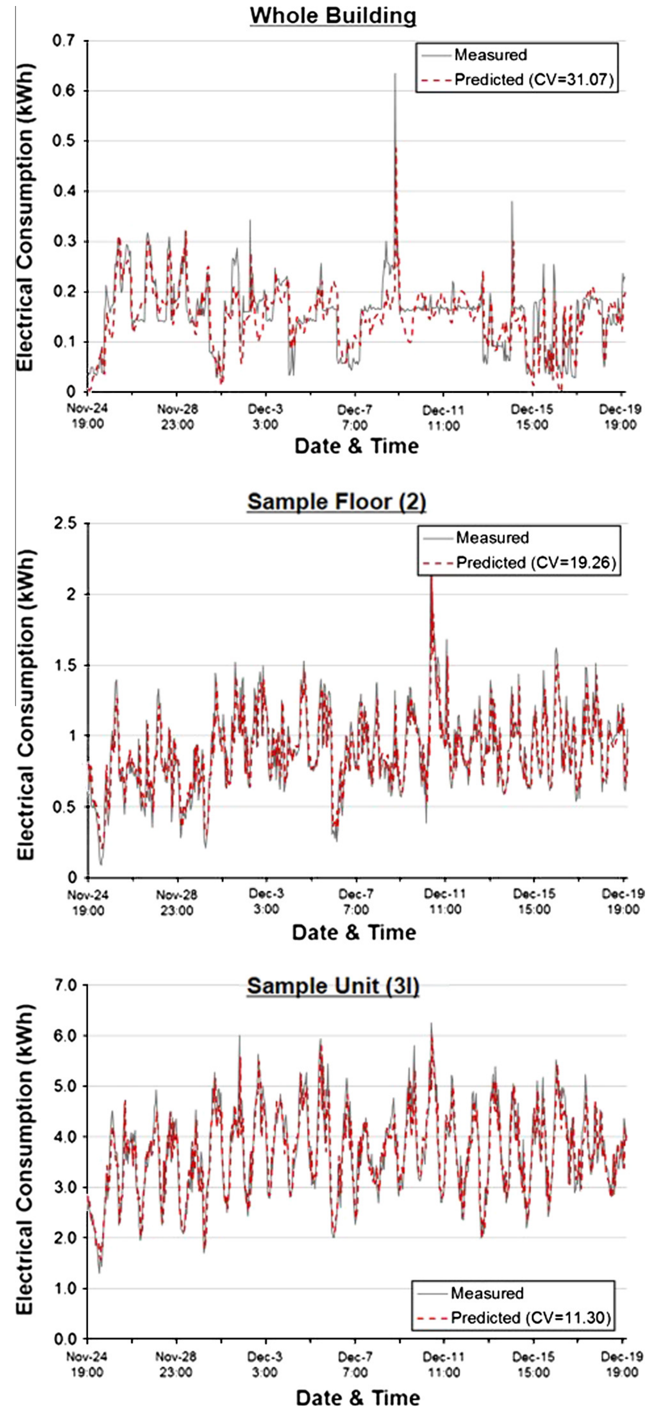


Fig. 11. Hourly forecasting results for whole building, sample floor and a sample unit.

where d is set for each time interval ($d = 1$ for daily, $d = 24$ for hourly, $d = 144$ for every 10 min), $\hat{y}_i$ is the predicted value, y_i are the observed daily consumption values, $\bar{y}$ is the mean of all daily consumption values in the dataset and N is the total number of observations. The goal of this modified CV performance metric is to allow for a more analogous comparison across the temporal scenarios.

Results of the modified CV calculation are provided in Table 3 and are significantly smaller than the original CV values. However, it is important to note that these the modified values should be compared relative to other scenarios with the same spatial granularity (i.e., *1A* vs. *2A* vs. *3A*; *1B* vs. *2B* vs. *3B*, *1C* vs. *2C* vs. *3C*) and not with the values obtained from the conventional CV calculation. The results indicate that the best performing temporal interval is *10 min* for the *whole building* model and hourly for the more spatially granular *by floor* and *by unit* scenarios. These results aligned with the findings from the conventional CV calculation but are now statistically distinct with the proposed modified CV calculation.

5. Limitations

We acknowledge that conducting our analysis with a data sample from a single multi-family residential building is a limitation of this study. However, this data sample was adequate to achieve our primary objective of demonstrating the applicability of sensor based energy forecasting to multi-family residential buildings

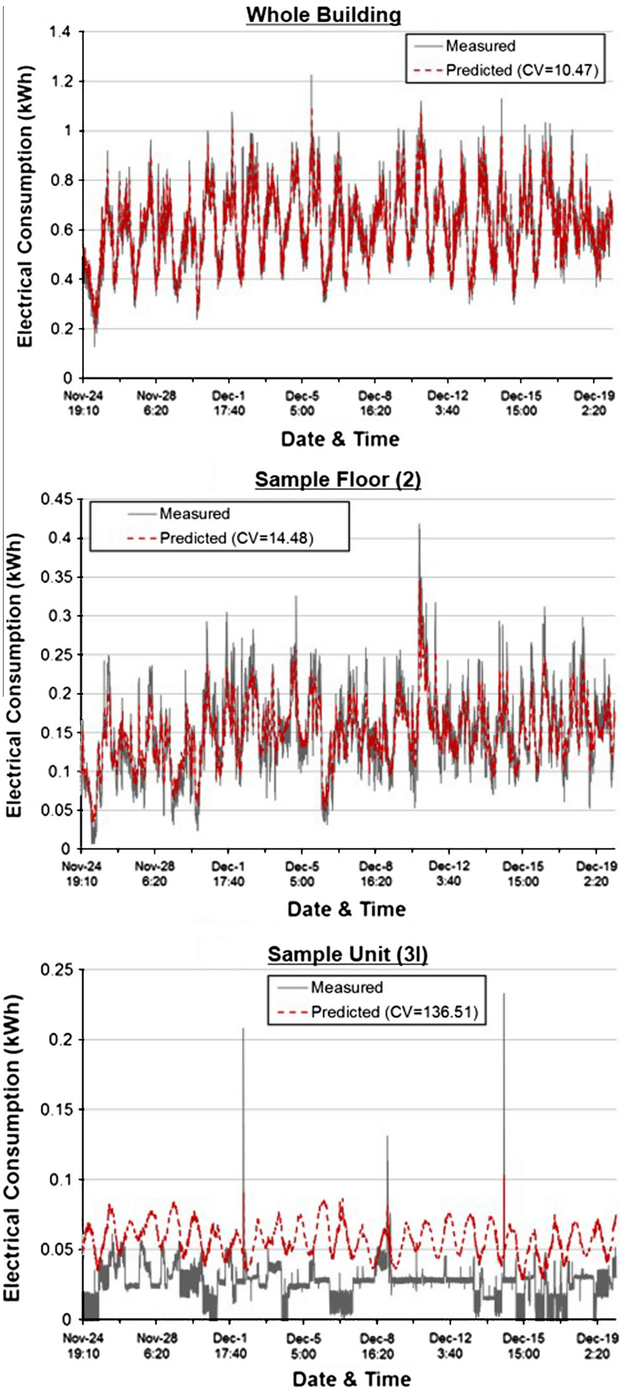


Fig. 12. 10 min forecasting results for whole building, sample floor and a sample unit.

and to examine the impact the spatial and temporal granularity can have on such forecasting models. Additionally, we acknowledge that further analysis could have been conducted on additional temporal and spatial granularities. We chose the temporal and spatial granularities in accordance with previous work and results were adequate to provide insight into the impact spatial and temporal data granularity can have on sensor based forecasting models. In the end, we hope this study will spur future work and discussion in the research community regarding sensor based forecasting for multi-family residential buildings and encourage similar work on datasets from other multi-family residential buildings.

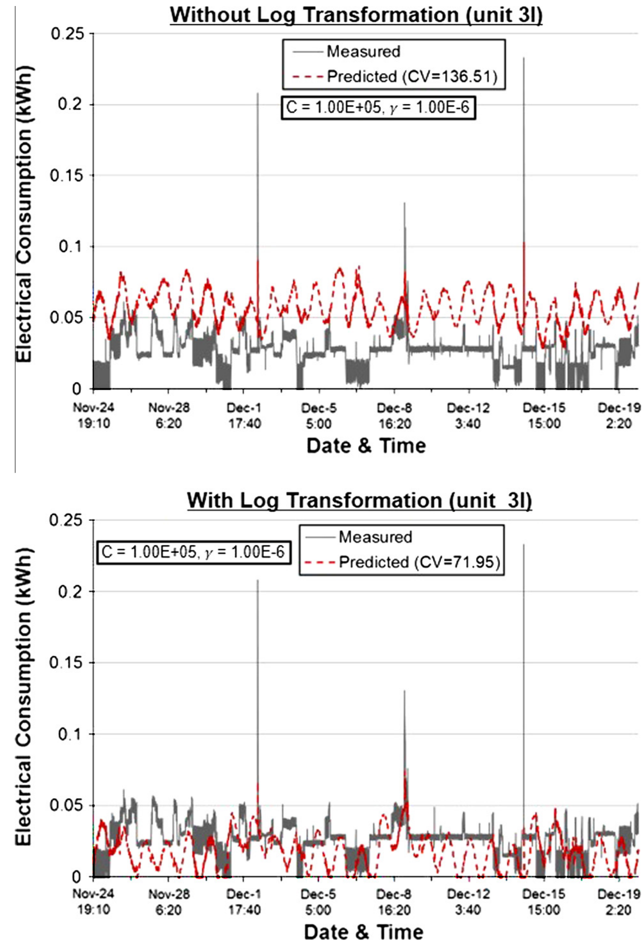


Fig. 13. Comparison of results with and without log data transformation (scenario 3C).

Table 3
Modified CV values and standard error (in%) for scenarios. Bold indicates the best results for each spatial granularity.

Temporal granularity	Spatial granularity		
	Whole building	By floor	By unit
Daily	*11.39 ± 2.73*	*6.07 ± 1.79*	*5.52 ± 1.94*
Hourly	*0.54 ± 0.13*	***0.13 ± 0.04***	***0.44 ± 0.22***
10 min	***0.08 ± 0.01***	*0.68 ± 0.29*	*9.64 ± 0.86*[a]

[a] Log transformed.

6. Conclusions, implications and future work

The results of this study demonstrated the applicability of sensor based energy forecasting models to multi-family residential buildings and extend prior research [7,17] in the area of residential energy forecasting beyond single-family buildings. Additionally, we broadened our exploration to examine the impact spatial and temporal granularity has on sensor based forecasting models. Results indicate that the most effective models are built with *hourly* consumption at the *floor* level. Specifically, spatial granularity was shown to have a substantial impact on the predictive power of sensor based forecasting models with more granular data at the floor and individual unit levels producing better predictions. Comparative analysis of the modified CV values revealed that temporal granularity can also have an impact on a model's predictive power. This work extends prior energy forecasting research [7,11–18] by determining the optimal monitoring granularity required to maximize the predictive power of an energy forecasting model. Results also provide practical implications for the deployment and installation of advanced smart metering devices that are necessary to acquire high resolution and granular energy consumption data used in the best performing forecasting models.

This study represents an important first step in the area of sensor based energy forecasting for multi-family residential buildings. More assiduous study is required to further understand and improve forecasting methods for multi-family residential buildings. Future research could build on this work by: applying additional machine learning techniques (e.g., Artificial Neural Networks) to the area of multi-family residential energy forecasting, extending this analysis to multi-step forecasts, implementing various feature selection techniques (e.g., The Lasso) to determine the best subset of exogenous features, exploring the use of data transformation to improve the predictive power of models trained on spatial and temporal granular data streams, developing and proposing alternative metrics that more accurately reflect model performance, broadening the investigation of sensor based forecasting models to other multi-family residential buildings and geographic regions and analyzing how providing occupants with eco-feedback (i.e., feedback on their energy usage) impacts forecasting models.

As the world's population continues to urbanize, multi-family residential housing and the subsequent consumption associated with such buildings will continue to grow and account for a significant portion of our environmental emissions. Accurately forecasting the energy consumption of such buildings is critical to the success of emerging energy efficiency technologies and to reaping a wide array of environmental benefits, including reduced emissions. Meeting this challenge will require research of both the methods available for prediction and scales at which data collection can be cost effectively employed.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant Nos. 1142379 and 0903597. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Rimas Gulbinas and Amy Tang for their support in the development and launch of this study.

References

[1] U.S. Department of Energy. Buildings energy data Book; 2013. Retrieved from: http://buildingsdatabook.eren.doe.gov/.
[2] PlaNYC – Green buildings & energy efficiency. The City of New York; 2013.
[3] Petersen S, Svendsen S. Method and simulation program informed decisions in the early stages of building design. Energy Build 2010;42:1113–9.
[4] Virote J, Neves-silva R. Stochastic models for building energy prediction based on occupant behavior assessment. Energy Build 2012;53:183–93.
[5] Kusiak A, Xu G. Modeling and optimization of HVAC systems using a dynamic neural network. Energy 2012;42:241–50.
[6] Howard B, Parshall L, Thompson J, Hammer S, Dickinson J, Modi V. Spatial distribution of urban building energy consumption by end use. Energy Build 2012;45:141–51.
[7] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: a machine learning case study. Energy Build 2012;49:591–603.
[8] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92.
[9] Kreider JF, Haberl JS. Predicting hourly building energy usage. United States: ASHRAE Journal (American Society of Heating, Refrigerating and Air-Conditioning Engineers); 1994. p. 36.
[10] Mackay DJC. Bayesian nonlinear modeling for the prediction competition. ASHRAE Trans 1994;100:1053–62.
[11] Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. Energy Build 2005;37:1250–9.
[12] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37:545–53.
[13] Karatasou S, Santamouris M, Geros V. Modeling and predicting building's energy use with artificial neural networks: methods and results. Energy Build 2006;38:949–58.
[14] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Convers Manage 2009;50:90–6.
[15] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86:2249–56.
[16] Wong SL, Wan KKW, Lam TNT. Artificial neural networks for energy analysis of office buildings with daylighting. Appl Energy 2010;87:551–7.
[17] Kolter JZ, Ferreira J. A large-scale study on predicting and contextualizing building energy usage. In: Proceedings of the twenty-fifth AAAI conference on artificial intelligence: AAAI. p. 1349–56.
[18] Li K, Su H, Chu J. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study. Energy Build 2011;43:2893–9.
[19] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using support vector regression. Appl Energy 2011;88:368–75.
[20] Zhao H-x, Magoulès F. Feature selection for predicting building energy consumption based on statistical learning method. J Algor Comput Technol 2012;6:59–78.
[21] Zhang WY, Hong W-C, Dong Y, Tsai G, Sung J-T, Fan G-f. Application of SVR with chaotic GASA algorithm in cyclic electric load forecasting. Energy 2012;45:850–8.
[22] Espinoza M, Espinoza M, Suykens JAK, Moor BD. Fixed-size least squares support vector machines: a large scale application in electrical load forecasting. CMS 2006;3:113–29.
[23] Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Adv Neural Inform Process Syst 1997:155–61.
[24] Ruas GIS, Bragatto TAC, Lamar MV, Aoki AR, de Rocco SM. Electrical energy demand prediction using artificial neural networks and support vector regression. In: 3rd International Symposium on Communications, Control and Signal Processing, ISCCSP 2008, IEEE; 2008.
[25] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.
[26] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2011;2:27.
[27] Vapnik V. The nature of statistical learning theory. Springer; 2000.
[28] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI 1995. p. 1137–45.
[29] Gulbinas R, Jain R, Taylor J, Peschiera G, Golparvar-Fard M. Network eco-informatics: development of a social eco-feedback system to drive energy efficiency in residential buildings. ASCE J Comput Civil Eng 2014;28(1):89–98.
[30] Emery AF, Kippenhan CJ. A long term study of residential home heating consumption and the effect of occupant behavior on homes in the Pacific Northwest constructed according to improved thermal standards. Energy 2006;31:677–93.
[31] Lü X, Lu T, Kibert CJ, Viljanen M. A novel dynamic modeling approach for predicting building energy performance. Appl Energy 2014;114:91–103.
[32] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy Build 2011;43:1409–17.
[33] Venkatesan N, Solanki J, Solanki SK. Residential demand response model and impact on voltage profile and losses of an electric distribution network. Appl Energy 2012;96:84–91.