Forecasting Building Energy Demand under Uncertainty Using Gaussian Process Regression: Feature Selection, Baseline Prediction, Parametric Analysis and a Web-based Tool

Bin Yan, Xiwang Li, Wenbo Shi, Xuan Zhang, Ali Malkawi
Harvard Center for Green Buildings and Cities, United States of America

Abstract

Gaussian process (GP) regression has drawn growing attention in the application of building energy demand prediction. This paper demonstrates the process of developing GP models for baseline prediction and parametric analysis. In particular, this study proposes the method of feature selection based on characteristic length-scale in covariance function, as well as the methods of anomaly detection and parametric analysis utilizing the predictive distribution of a GP regression. Two case studies are used to illustrate the processes in detail. The results show that successful feature selection can improve predictive accuracy and reduce computational cost. In baseline prediction, the outcome of a GP regression gives a confidence range, which provides valuable information for anomaly detection. In parametric analysis, the additional variance in the output caused by intentionally varying the range of a control variable gives an estimate of its impact on energy demand. Another contribution of this study is the development of a web-based tool, which allows users to build GP models without knowledge of the details GPs or programming skills.

Introduction

There is growing attention drawn to forecasting building energy use as it can be applied to fault detection and diagnosis (FDD), operation optimization and interactions between buildings and smart grid. Energy use forecasting has been widely used in building commissioning to determine retrofit savings (Cohen & Krarti, 1995; Kissock, Reddy, & Claridge, 1998). It is also essential in fault detection since a baseline prediction is necessary to detect excessive energy consumption (Yan & Malkawi, 2013). The ability to accurately forecast demand-side loads plays a critical role in electric power system, especially that the future power grid is expected to provide unprecedented flexibility in how energy is generated, distributed and managed (Burger & Moura, 2015; Jetcheva, Majidpour, & Chen, 2014). Grid-level forecasting is crucial to the planning and operation of utility companies, while building-level load forecasting is crucial to building owners to reduce the electricity charges by optimizing electricity purchase strategies. Forecasting both levels is important for design of microgirds and intelligent distribution systems as well as the implementation of demand response (Kwac, Flora, & Rajagopal, 2014).

The literature is rich with forecasting methods for grid-level and building-level loads. Perhaps the most attention has been given to the approaches such as Artificial Neural Networks (ANN), Support Vector Regression (SVR) and Autoregressive Integrated Moving Average (ARIMA) models (Burger & Moura, 2015). Approaches such as Multiple Linear Regression, Fuzzy Logic, Decision Trees, and k-Nearest Neighbors (k-NN) methods are also widely used (Burger & Moura, 2015; Majidpour, Qiu, Chu, Gadh, & Pota, 2015) . There has been a surge of interest in Gaussian process (GP) modeling following recent advances in the machine learning community (Neal, 1995; Rasmussen, 1996). Gaussian Processes have been successful in solving many real-word data modeling problems (MacKay, 1997). Recently Gaussian process modeling has been adopted to forecast thermal dynamics and energy consumption in buildings (Kim, Ahn, Park, & Kim, 2013; Manfren, Aste, & Moshksar, 2013; Rastogi, 2016; Yan, Kim, Ahn, & Park, 2013). Gray and Schmidt (2016) developed a Gaussian Process regression model to predict the day-ahead hourly zone temperature in a building. Their study uses synthetic data simulated by TRNSYS (A TRaNsient SYstems Simulation Program). The input variables include ambient temperature, solar gains, heating medium mass flow, number of occupants, hour of day, day of week, heating schedule and calculated zone temperature of previous days. There are 18 weeks of data for training and 10 for testing. The study compares models with different training period, 72-hour (three days), 168-hour (one week), 504-hour (three weeks) and 1008-hour (six weeks). Longer training period leads to lower day-ahead prediction error. Compared with a physics-based grey-box model, GP achieves a lower prediction error during occupied times for training period of three weeks or longer. However, in their case study, the grey-box model has a lower total prediction error, requires less training and is less sensitive to input data not present in the training dataset than GP. The sensitivity to unknown data is a limiting factor for GP prediction, especially when a reduced amount of training data is available. Heo,

Choudhary, and Augenbroe (2012) presented a Bayesian approach, which involves GP regression, to calibrate normative energy models. The mathematical model is based on the work of Kennedy and O'Hagan (2001). The Gaussian process formulation is used to compute the likelihoods of observations given model parameters, $p(y|\boldsymbol{\theta})$. In their case study, $\boldsymbol{\theta}$ includes intercept C for window opening, indoor temperature during heating, infiltration rate and discharge coefficient. The variable to be calibrated y is monthly gas consumption. Later Heo and Zavala (2012) developed a Gaussian process modeling framework to determine energy savings in measurement and verification (M&V) practices. Burkhart, Heo, and Zavala (2014) extended Heo and Zavala (2012) research by incorporating input uncertainty using a Monte Carlo expectation maximization (MCEM) framework.

There is usually discrepancy between models and actual system performance (B. Yan, Li, Malkawi, & Augenbroe, 2017). A key advantage of GP modelling is that the outcomes come in the form of probability distributions, which take uncertainties in the modelling process into account. With certain adaptions, noise in training and predictive inputs can be incorporated in the outcomes as well. To authors' best knowledge, the existing studies have not provided a comprehensive workflow of forecasting energy demand using GP regression. Technical issues such as feature selection have not been fully discussed. Moreover, the advantage of GP modelling, its capability of dealing with uncertainty, has not been fully exploited in terms of its applications in fault detection and control optimization. This study emphasizes on how to utilize the advantages of GP modelling in two applications: (1) Constructing a baseline model for anomaly detection. Anomaly detection means detecting energy consumption which does not conform to an expected pattern. For example, excessive energy consumption might indicate a fault in system operations. (2) Conducting parametric analysis in order to evaluating the impact of a control variable on energy demand. As part of the contribution of this study, a web-based forecasting tool has been developed. This tool allows users without coding background to build GP models to make predictions and evaluate the impact of certain variables.

Modeling Methods

GP Regression

This study mainly uses the GP models described in the work of Rasmussen & Williams (2006). A Gaussian process is specified by a mean function and a covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$. The choice of mean function in this study is zero function, and the covariance function is a squared exponential kernel,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T W^{-1}(\mathbf{x}_i - \mathbf{x}_j)\right] \quad (1)$$

where

$$W = \text{diag}[w_1^2, w_2^2, \ldots, w_D^2] \quad (2)$$

with parameter w defining the characteristic length-scale in covariance function. D in Eq.2 is the number of features. The characteristic length-scales briefly define how far apart the input values $x_{i,d}$ and $x_{j,d}$ can be for the response values to become uncorrelated, where $x_{i,d}$ is the value of the dth feature of the ith data point . Inputs that are judged to be close by the covariance function are likely to have similar outputs. A prediction is made by considering the covariance between the predictive case and all the training cases $\mathbf{k}(X, \mathbf{x}^*)$ (Rasmussen, 1996). For a noise-free input $\mathbf{x}^*$, predictive posterior mean and variance are (Rasmussen & Williams, 2006)

$$\mathbb{E}[f_* \mid X, \mathbf{y}, \mathbf{x}^*] = \mathbf{k}(X, \mathbf{x}^*)^T (K + \sigma_n^2 I)^{-1} \mathbf{y} \quad (3)$$

$$\mathbb{V}[f_* \mid X, \mathbf{y}, \mathbf{x}^*] = k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}(X, \mathbf{x}^*)^T (K + \sigma_n^2 I)^{-1} \mathbf{k}(X, \mathbf{x}^*) \quad (4)$$

K is an $N \times N$ matrix of covariance functions between each pair of training inputs. σ_n^2 denotes the variance of Gaussian noise in training targets $\mathbf{y}$, I is an $N \times N$ identity matrix. σ_f, σ_n and $w_1, w_2 \ldots w_D$ are hyperparameters to be trained through gradient descent optimization algorithm.

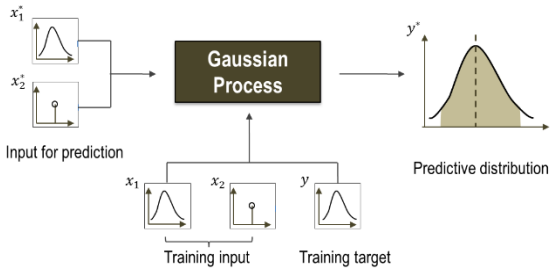

Figure 1 Diagram of Gaussian Process regression

Figure 1 describes a GP regression. A distinguished feature of Gaussian Processes is that the outcomes come in the form of probability distributions, which take uncertainty in the modeling process into account. The variance of a prediction (Eq.4) automatically includes two types of uncertainties. The first is noise in the targets. The noise might come from measurement noise. Some other sources include that the process to be modeled itself is stochastic, and thus there are random elements in $\mathbf{y}$. And/or the features in an existing model might not fully explain the variance in training targets. There might be some other important features that affect outputs. The second type of uncertainty is interpolation uncertainty. The distance between the inputs associated with a prediction and training inputs affects the magnitude of the variance. GP modeling is an interpolation method. If a new input point lies beyond the scope of the training input domain, the variance will be large in the prediction.

Predicting with Uncertain Inputs

With certain adaptations, a GP regression can also account for parametric uncertainty in training and predictive inputs as shown in Figure 1. If training inputs are noise-free while the predictive input is Gaussian $\mathbf{x}^* \sim \mathcal{N}_{\mathbf{x}^*}(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*})$, then the posterior distribution can be computed analytically as follows (Girard et al., 2003),

$$\mathbb{E}[f_*] = \mathbf{q}^T\boldsymbol{\beta} \tag{5}$$

$$\mathbb{V}[f_*] = k(\mathbf{x}^*, \mathbf{x}^*) + \mathrm{Tr}[(\boldsymbol{\beta}\boldsymbol{\beta}^T - (K + \sigma_n^2 I)^{-1}Q)] - \mathbb{E}[f_*]^2 \tag{6}$$

with

$$\boldsymbol{\beta} = (K + \sigma_n^2 I)^{-1}\mathbf{y} \tag{7}$$

$$q_i = |W^{-1}\boldsymbol{\Sigma}_{\mathbf{x}^*} + I|^{-\frac{1}{2}}\sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_{\mathbf{x}^*} - \mathbf{x}_i)^T(\boldsymbol{\Sigma}_{\mathbf{x}^*} + W)^{-1}(\boldsymbol{\mu}_{\mathbf{x}^*} - \mathbf{x}_i)\right) \tag{8}$$

$$Q_{ij} = |2W^{-1}\boldsymbol{\Sigma}_{\mathbf{x}^*} + I|^{-\frac{1}{2}}\sigma_f^2 \exp\left(-\frac{1}{2}\left(\frac{\mathbf{x}_i+\mathbf{x}_j}{2} - \boldsymbol{\mu}_{\mathbf{x}^*}\right)^T\left(\boldsymbol{\Sigma}_{\mathbf{x}^*} + \frac{1}{2}W\right)^{-1}\left(\frac{\mathbf{x}_i+\mathbf{x}_j}{2} - \boldsymbol{\mu}_{\mathbf{x}^*}\right)\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T(2W)^{-1}(\mathbf{x}_i - \mathbf{x}_j)\right) \tag{9}$$

An alternative to the analytical solution shown in Eq.5 to Eq.9 is to perform Monte Carlo analysis based on standard GP regression. When the dimension of uncertain features is low, Monte Carlo is feasible.

Feature Selection

Successful feature selection will improve prediction performance, reduce computational cost and provide a better understanding of the underlying prediction process. They are especially helpful for high-dimensional problems such as text processing of internet documents and gene expression array analysis (Guyon & Elisseeff, 2003). Although energy prediction is not a high-dimensional problem. However, if there are more than a few thousand training points, the covariance matrix gets large and computational cost could be high. For hourly prediction using one-year historical data, feature selection might be necessary in order to reduce computational cost.

Some existing feature selection methods are based on entropy (Che & Wang, 2014; Zheng & Kwoh, 2011). Following Che and Wang (2014) method, we have tested feature selection based on correlation coefficients and mutual information for GP regression. However, the method might not applicable to GP regression or the case studies in this paper. The selected features are counter-intuitive and lead to poor prediction accuracy. Here, we propose an alternative approach of feature selection specifically targeting at GP regression. The idea is to use the hyperparameters which a GP regression learns. Here are the implementation steps.

(1) Normalize all features and the target to [-1, 1].

(2) Divide the dataset into training, validation and test sets. Use training and validation sets for feature selection. If the dataset is large and computational cost is a concern, select smaller training and validation sets for the purpose of feature selection.

(3) Include all the features to train a GP regression and derive the characteristic length-scale *w* that corresponds to each feature from the training set.

(4) Calculate the indicator of each feature as follows.

$$g(\mathbf{x}_i) = \frac{w_i}{\mathrm{std}(\mathbf{x}_i)} \tag{10}$$

where $\mathbf{x}_i$ is the input vector of a certain feature, $i = 1,2,\dots d$ and d is the number of features. The indicator $g(\mathbf{x}_i)$ is defined as the characteristic length-scale of a feature divided by the standard deviation of all inputs on that feature dimension. As mentioned in the Modeling Methods section, the characteristic length-scale determines how close two points have to be to influence each other. As it is normalized by the variation of the training inputs on that feature dimension, it indicates how significant that feature is. If $g(\mathbf{x}_i)$ is small, it means small changes in input value of a certain feature will have a significant impact on output value. Therefore, a relatively small $g(\mathbf{x}_i)$ indicates that the corresponding feature is important in GP regression.

(5) Select $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots \mathbf{x}_{(d^*)}$ as follows.

$$\begin{aligned} \mathbf{x}_{(1)} &= \arg\min_{X}[g(\mathbf{x}_i)] \\ \mathbf{x}_{(2)} &= \arg\min_{X-\{\mathbf{x}_{(1)}\}}[g(\mathbf{x}_i)] \\ &\vdots \\ \mathbf{x}_{(d^*)} &= \arg\min_{X-\{\mathbf{x}_{(1)},\mathbf{x}_{(2)},\dots\mathbf{x}_{(d^*-1)},\}}[g(\mathbf{x}_i)] \end{aligned} \tag{11}$$

The feature with the smallest $g(\mathbf{x}_i)$ is the first feature to select. The number of the features d^* depends on training and validation accuracies, as well as computational cost. In general, if the training and/or validation accuracy is much lower than that when using all features, include more features in the model. Otherwise, test with fewer features to further reduce computational time.

Implementation

GP regression is implemented through GP Regression and Classification Toolbox version 3.6 (Rasmussen & Nickisch, 2015) with certain modification to incorporate noisy inputs in Eq.5 to Eq.9. Data wrangling and preliminary analysis is implemented through Pandas libray (McKinney, 2011).

Case Studies

This paper presents two case studies. The first case study illustrates the process of developing a GP model for baseline prediction, which is useful in fault detection and building commissioning. The second case study demonstrates how to use a GP regression to evaluate the impact of certain variables on energy demand.

Baseline Prediction and Anomaly Detection

Baseline predication is crucial in fault detection and diagnostics. Most existing fault detection tools on the market provide this basic function. The energy consumption is compared with a baseline in order to evaluate whether there is unexpected energy use pattern, which might indicate a fault in operations. Figure 9 shows the screenshots of three existing commercialized fault detection tools. These tools normalize energy consumption by weather and compare monthly energy consumption of this year with that of last year, or daily or

hourly energy consumption of this week with last week. Their comparisons give an estimate of the change in energy consumption, as well as throw out a warning when energy consumption is noticeably higher than the baseline. This function can be improved in two perspectives. First, the existing tools only normalize the energy consumption by weather. It is necessary to consider other factors such as occupancy. For example, when comparing the energy consumption of Christmas week with the previous week, merely normalizing by weather is not sufficient. Second, the difference between the baseline and the measured consumption can be caused by modelling uncertainty. It would be useful if the baseline prediction gives a confidence range to take modeling uncertainty into account. As mentioned in the Modeling Methods section, GP regression is able to construct a baseline with uncertainty range. It has the flexibility to include other factors in the model as well as to take parametric uncertainty into account. The first case study demonstrates the baseline prediction and anomaly detection.

This case study uses energy data of a campus building located in Cambridge, Massachusetts, United Sates. The total area of this building is 15067m^2. It provides studio and office areas to approximately 500 students and more than 100 faculty and staff. There are also lecture and seminar rooms, a cafeteria, an auditorium and a library in the building. Its HVAC (heating, ventilation and air-conditioning) system consists of 9 air handling units (AHU), variable air volume (VAV) boxes and fan coil units (FCU) as terminal units. Three types of energy use are metered, electric energy, chilled water and steam. Electric energy use consists of lighting, plug load, fans and pumps of the HVAC system. Chilled water is supplied from the campus chiller plant and it is used for cooling. Steam is used for heating and domestic hot water. The data resolution is 5-minute and aggregated to daily energy use.

Weather data is collected from a weather station installed on the roof. Dry-bulb air temperature t (°C), relative humidity ϕ (%), atmospheric pressure p (mbar), wind direction w_{dir}, wind speed w_s (m/s), gust speed g_s (m/s) and solar radiation q (W/m^2) are sampled at 5-minute interval. Apart from these directly measured variables, we calculate dew-point temperature t_d (°C) and specific humidity h (kg/kg) to include in the potential feature set. We also derive the following variables for cooling and heating forecasting.

Dehumidification h_d (kg/kg):

$$h_d = \begin{cases} h - 0.0087, & h > 0.0087 \\ 0, & h \le 0.0087 \end{cases} \quad (12)$$

Cooling degrees t_c (°C):

$$t_c = \begin{cases} t - 12, & t > 12 \\ 0, & t \le 12 \end{cases} \quad (13)$$

Heating degrees t_h (°C):

$$t_h = \begin{cases} 0, & t \ge 15 \\ 15 - t, & t < 15 \end{cases} \quad (14)$$

If the time interval is daily, then

$$h_d = \sum_{i=0}^{23} h_{d,i} \quad (15)$$

where $h_{d,i}$ is hourly sample. Daily t_c and t_h should be calculated in the same way. Other daily weather variables are simply the average of hourly values.

The reasoning of transforming t to t_c and h to h_d is related to the HVAC system type and its controls. If the supply air temperature set-point is fixed in an air handling unit, ASHRAE (2009) recommends the set-point to be 12.8°C (55°F), which would satisfy the humidity control in the meantime. Taking the temperature rise due to fan heat into consideration, if outdoor temperature is below 12°C, in most cases no chilled water consumption is needed. Free cooling can be utilized even when there is cooling demand for interior zones. When $t = 12.8$ °C and $\phi = 95\%$, the corresponding specific humidity $h = 0.0087$ kg/kg. Therefore, when $h \le 0.0087$ kg/kg, there is no dehumidification demand. If using t_c in GP regression, all the data points are "similar" in temperatures when $t < 12$ °C and "similar" in humidity if $h \le 0.0087$ kg/kg. Although the supply air temperature set-points of most of the AHUs in this building are not exactly 12°C, h_d and t_c might still serve as better features for cooling demand.

Time related features, day of week, day of year and week of year are included in the feature set. A factor between 0 and 1 that indicates occupancy is estimated from academic calendar. For instance, a normal weekday during the semester is assigned as 1, university holidays for staff but not for students are assigned as 0.5 and Christmas is assigned as 0 as the University is completely closed.

In this case study, metered energy consumptions from January 2012 to June 2016 are available. Data from January 2012 to September 2013 is used as training set, and data from October 2013 to December 2014 is used as validation set for feature selection using the proposed indicator according to Eq.10. Feature selection results are shown in Table 1.

Table 2 compares training accuracy, validation accuracy and computational cost between using all features and using selected features. The number of iterations used to search for optimal hyperparameters is considered as a measure of computational cost. By selecting five features out of fifteen, training accuracy is slightly lower while validation accuracy is slightly higher. The computational cost has been reduced to less than one-tenth of that when using all features. Figure 2 illustrates detailed feature selection process of daily electric energy consumption prediction according to Eq.10 and Eq.11. The model performs poorly on both training data and validation data when there are fewer than three features. One or two features are not expressive enough to capture the relationship between inputs and target. The model is

underfitting the training data. Training accuracy keeps increasing with more features, although very marginally after four features have been selected. Validation accuracy reaches highest when there are four features and decreases slightly afterwards. After ten features are included in the model, validation accuracy almost keeps the same. When irrelevant or redundant features are included in the model, their corresponding characteristic length-scale w will be large so that they have minimal impact on the model. In this way, GP regression is robust to prevent overfitting. However, prediction accuracy might still be optimized by selecting the right features. More important, removing irrelevant or redundant features can significantly speed up the process.

Table 1 Features selected for daily prediction

	Features selected
Electric energy	Day of year, week of year, weekday, occupancy, heating degrees
Chilled water	Cooling degrees, solar radiation, weekday, dehumidification, occupancy
Steam	Day of year, week of year, heating degrees, weekday, occupancy

Table 2 Comparison of training accuracy, test accuracy and computational cost before and after feature selection

	Features included	Training accuracy	Validation accuracy	Iterations
Electric	All	0.9676	0.8554	2965
	Selected	0.9087	0.8614	150
Chilled water	All	0.9393	0.8230	2281
	Selected	0.9284	0.9000	224
Steam	All	0.9839	0.8928	1390
	Selected	0.9666	0.9117	138

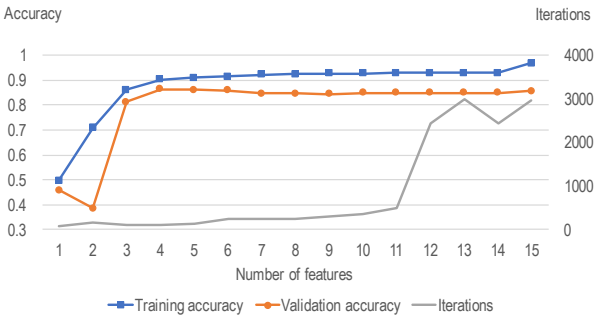

Figure 2 Model accuracy and computational cost versus number of features selected for daily electric energy use prediction

Time features play an important role in the prediction of electric energy consumption. Figure 3 shows clear difference of electric energy use during day and night, weekdays and weekends. The accumulated weekly consumption (Figure 10) reveals interesting study patterns of the students in this building. It looks like during each semester, electric energy use ramps up toward a peak at finals, perhaps because the students are working day and night in that week. Additionally, the students are working harder and harder toward finals. Then there is a dip after semesters end, including Christmas vacation. The electric energy consumption is relatively low during January, summer terms, and spring break, when campus can be relatively empty. This explains why the four most significant features are all time features. Including these four features alone leads to highest validation accuracy. The fifth feature $\mathbf{x}_{(5)}$ is a weather related variable, heating degrees. The electric energy consumption of the investigated building includes consumption of HVAC fans and pumps, which is affected by weather. Although the validation accuracy of adding the fifth feature is slightly lower, it might be beneficial to include a weather feature for unseen data. The decision to select this weather feature is based on domain knowledge.

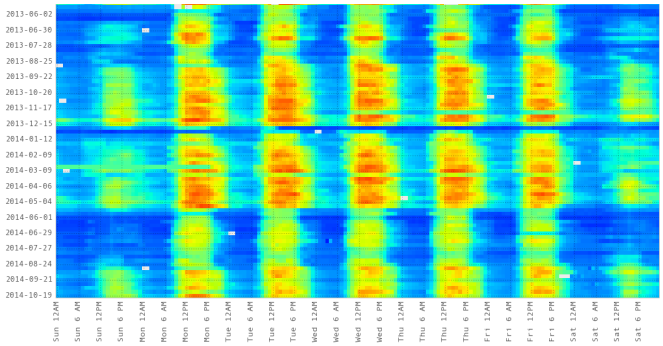

Figure 3 Heat map of hourly electric energy consumption

The features for chilled water prediction consist of three weather variables and two time variables, and weather features play a more important role. Temperature, solar radiation and humidity all affect chilled water consumption prediction, which is consistent to domain knowledge. The features for steam prediction are the same as those for electric energy. The difference is that heating degrees ranks higher in steam prediction.

After completing feature selection, we use the entire set of data from 2012 to 2014 to construct a baseline for year 2015 and 2016. Figure 4 shows the baseline prediction of steam consumption for two months. GP regression outputs a baseline with a mean value and a standard deviation. Denote the predictive mean $\mathbb{E}[f_* \mid X, \mathbf{y}, \mathbf{x}^*]$ as μ and standard deviation $\sqrt{\mathbb{V}[f_* \mid X, \mathbf{y}, \mathbf{x}^*]}$ as σ. In Figure 4, the red dots are measure values, the blue line with dots shows the predictive mean of the baseline, and the blue area is 95% confidence region. Here we can interpret it as if a building performs in the same way as it does during the time period of the baseline, the measured energy consumption will fall within $\mu \pm 1.96\sigma$ with 95% chance. The baseline prediction includes modelling uncertainty. If the measure energy consumption falls out of the 95% confidence region, it is likely that the building is performing differently. The occupancy level is different, e.g. an event takes place. It could also indicate a fault in the operations.

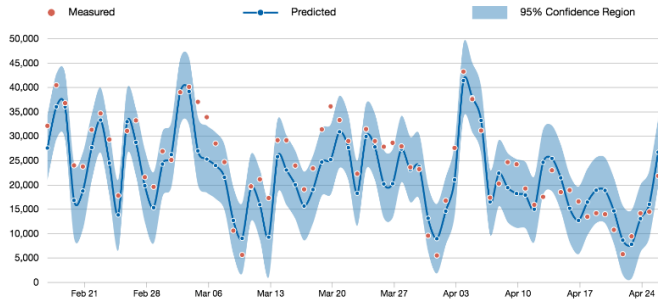

Figure 4 Measured steam consumption (lbs.) versus predicted baseline with uncertainty range from February 17th to April 25th, 2016

Define deviation $h(y^*, \mu, \sigma)$ at the point $(\mathbf{x}^*, y^*)$ as

$$h = \frac{y^* - \mu}{\sigma}$$

Figure 5 and Table 3 show the distribution of deviation from baseline prediction of three types of energy consumption from February 2015 to June 2016. 12.7% of measured chilled water consumption and 9.5% of measured steam consumption deviates away from predicted mean more than $1.96\,\sigma$. It requires further investigation regarding whether there are actual anomalies or it is a false alarm. One possibility of the deviation from the prediction is that occupancy estimation is uncertain. Occupancy is a feature selected for all three types of energy consumption, which is a factor between 0 and 1 estimated based on academic calendar. It is possible that, during certain holidays, more students than estimated still study in the building, or more people are in the buildings during certain days due to events. Assume there is uncertainty in occupancy estimation and its standard deviation $\sigma_{\text{occ}} = 0.1$. Then use Eq.5 to Eq.9 to calculate the predictive distribution with noisy inputs. Anomalies (defined as measured energy consumption falls out of the confidence range) occur 12% for chilled water consumption and 2% for steam consumption. If assume $\sigma_{\text{occ}} = 0.2$, then 9% for chilled water and 0% for steam.

Table 3 Percentage of measured energy consumption that falls out of confidence region

	$h > 1.96$			$h < -1.96$		
Occupancy uncertainty	$\sigma_{\text{occ}}=0$	$\sigma_{\text{occ}}=0.1$	$\sigma_{\text{occ}}=0.2$	$\sigma_{\text{occ}}=0$	$\sigma_{\text{occ}}=0.1$	$\sigma_{\text{occ}}=0.2$
Electric	5%	1%	1%	37%	18%	18%
Chilled water	4%	4%	3%	9%	8%	6%
Steam	7%	2%	0%	2%	0%	0%

There is 37% of measure electric energy consumption falls below $\mu - 1.96\sigma$ assuming noise-free occupancy. The percentage is 18% assuming $\sigma_{\text{occ}} = 0.1$ and $\sigma_{\text{occ}} = 0.2$. Since there is no significant change in occupancy level and equipment use, most likely the anomaly indicates a fault. After investigation, the electric meter was malfunctioning. The electric energy consumption of the investigated building is the sum of readings from two meters. One meter stopped functioning for a few days in October and November 2015. It completely stopped since then. As a result, the metered value is lower than its actual consumption. Figure 6 shows that the measured electric consumption falls below the confidence region most of the time in 2016.

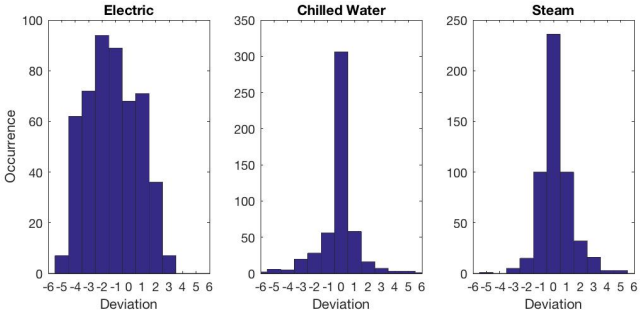

Figure 5 Distribution of deviation from baseline prediction

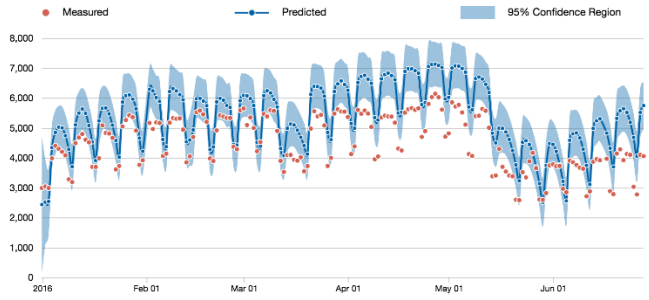

Figure 6 Comparison of baseline prediction and measured electric energy consumption

Parametric Analysis

In the first case study, we have modeled parametric uncertainty since there is embedded uncertainty in occupancy estimation. In some cases, it is desired to investigate the impact of inputs on outputs by intentionally allowing inputs to vary in their domains. In this case study, we examine the impact of an HVAC control variable on energy consumption using GP regression. The investigated building is an office building located in Philadelphia, PA, USA. It has three air-cooled condensing units which supply refrigerant to three direct expansion AHUs. The terminal units are VAV boxes with reheat. As the HVAC system of this building is also used for research purpose, its energy consumption is sub-metered. The electric energy consumption of three condensing units and the supply fans in three AHUs are metered individually at 15-min interval. AHU sensor readings of supply air temperature and return air temperature are available from June to August 2012 at hourly interval. Total building electric energy consumption consists of both HVAC, lighting and plug load consumption. The gas consumption for heating and domestic hot water is also metered but not considered in this case study, since we only look into summer months when AHU sensor readings are available.

The supply air temperature (SAT) set-point of an AHU is fixed. Different fixed set-point values are used for different seasons. As AHU SAT does not vary according to outdoor air temperature or zone thermal load, it is very

likely to be suboptimal. As shown in Figure 7, AHU return air temperatures (RAT), which are close to zone average temperatures, vary according to outdoor air temperature (OAT). When OAT is low, zone air temperatures are far below the upper bound of indoor comfort level (ASHRAE, 2004; Cao et al., 2012), which indicates excessive cooling, in particular for AHU1 and AHU2. Obviously, AHU SAT can be further optimized. In building commissioning/retrofitting projects, it is worth knowing the cost-effectiveness beforehand. An estimate the impact of optimizing AHU SAT can assist decision-making. Physics-based simulation, e.g. EnergyPlus (DOE, 2005), is usually used to evaluate the importance of control variables and estimate energy savings potential of optimization. The problem is that physics-based simulation requires detailed information of a building and it systems. Consequently, it is labor-intensive and time-consuming. Our question is without knowing the details of a building and its system, is there a fast way to estimate the impact of a control variable on energy consumption? If historical data of system operations is available, it is possible to develop a data-driven model to replace the physics-based simulation and to give an estimate of impact.

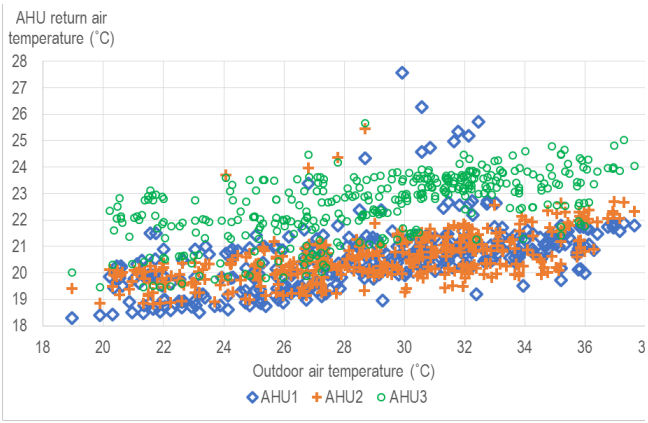

Figure 7 AHU return air temperatures versus outdoor air temperature

Table 4 Distribution of AHU SAT and impact factor

	Mean value (°C/°F)	Standard deviation (°C/°F)	Impact factor
AHU1	14.6 / 58.4	0.52 / 0.93	3.0%
AHU2	12.6 / 54.6	0.95 / 1.71	4.4%
AHU3	15.0 / 59.0	0.32 / 0.58	4.7%

Table 4 shows the distribution of measured AHU SATs. Although the set-points are fixed, there is some variance in actual SAT control. Poor PID control, or insufficient or excessive cooling supply might account for the variance. The variation in actual SATs allows training a surrogate model based on historical data using GP regression. First use the model described in Eq.1 to Eq.4 to get the distribution of energy consumption based on measured SATs. Then plug a variance in SAT(s) in Eq.5 to Eq.9 to get new distributions. The additional variance in energy consumption caused by SAT variation provides an estimate of the impact of SAT on energy consumption.

Hourly data of outdoor air temperature, humidity, AHU SATs, RATs, sub-metered and total building electric energy consumption is available from June 12th to August 15th 2012. After removing unoccupied hours, there are 370 data points left. The first step is to use measured outdoor air temperature, humidity and AHU SATs as the features, and total building electric energy consumption as the target to train a GP regression. In this study, the training R^2 is 0.96. Then use the same set of data to calculate the predictive distribution of each training target and get the corresponding variance $\mathbb{V}_{y_i}$ of each point according to Eq. 4. The second step is to add a noise of $\sigma_{\mathrm{SAT}} = 0.56$ °C (1°F) in the SAT of an AHU using Eq.6 and get the $\mathbb{V}_{y_{\mathrm{noisy},i}}$ that accounts for noisy AHU SAT. Lastly, define the impact factor as the sum of additional output variation caused by the input noise divided by the sum of predictive mean.

$$\text{Impact factor} = \frac{\sum_{i=1}^{n}\left(\sqrt{\mathbb{V}_{y_{\mathrm{noisy},i}} - \mathbb{V}_{y_i}}\right)}{\sum_{i=1}^{n} y_i}$$

The impact factor gives an estimate of the impact of the input variance on the output. A large value indicates larger impact of the input. In this case study, the impact factors of SATs of three AHUs are listed in Table 4. These factors give an estimate of the impact of a small change in each AHU SAT on total building electric energy consumption. SATs of AHU2 and AHU3 have larger impact than AHU1 SAT. This is consistent to sub-metered electric energy consumption of the AHUs and condensing units. As shown in Figure 8, the electric energy consumption of AHU2 and AHU3 during the investigated time period is significantly larger than that of AHU1. Intuitively, the magnitude of impact factor of SAT is related to the energy consumption of that AHU, although not exactly proportional. Further validation requires detailed physics-based simulation or on-site experiments. The proposed method using GP regression intends to provide a rapid estimate in order to assist early-decision making without the necessity of acquiring detailed information of a building and its system. The accuracy of the estimate depends on data quality, which directly affects the accuracy of surrogate model using GP regression.

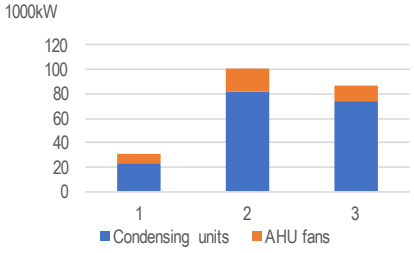

Figure 8 Sub-metered energy consumption of condensing units and AHUs

Web-based GP Forecasting Tool

A web-based forecasting tool (http://harvardcgbc.org/cgbc-launches-online-gaussian-processes-forecasting-tool-to-analyze-building-energy-consumption/) that allows users without coding background to train a GP model has been developed and currently hosted on AWS (Amazon Web Services) Cloud. This tool is designed for but not limited to building energy demand forecasting. The backend of this web application consists of two parts, algorithm and database. The GP regression algorithm is implemented through Gaussian Process Regression and Classification Toolbox (Rasmussen & Nickisch, 2015) in Matlab code and then converted to Java using MATLAB Complier SDK. MongoDB (MongoDB, 2016) is used to store data, models and prediction results. The architecture of this web application adopts modern REST web service technology to connect the frontend with the backend. The frontend interface is written in HTML and Javascript. The visualization is implemented using D3.js (Bostock, 2015).

The web application consists of the following pages, upload, configure, train, predict, explore and download. Users can upload their data in csv format with the feature and target names as header. In model configuration page as shown in Figure 11, they can view different variables, select a target, choose the features to be included in the model and select a training time period. The remaining time period will be used for testing purpose. After users have configured and submitted the model, they can view training accuracy in the train page. Ten-fold cross-validation will be performed, if selected, to further validate the model. If the training accuracy is low, users can go back to the configure page, try different feature sets and/or different training time period. When users get a more reliable model, they can check the predicting accuracy in the predict page. Additionally, they have the option to further explore the model. The Explore page allows users to evaluate the impact of an individual feature or a combination of features on the output target. Users can select the variables they want to investigate, and then input a relative or absolute uncertainty range. The output is an impact factor as described in the second case study. A variable with a high impact factor indicates that it is a crucial feature. All the prediction results, such as predictive mean and variance of each point, training, validation and test accuracy, as well as trained hyperparameters can be downloaded in csv format. This web-based tool can perform modeling for the first case study. Figure 4 and Figure 6 are generated by this tool.

Discussion and Conclusion

This paper has demonstrated building energy demand forecasting under uncertainty using GP regression. In particular, this study has proposed methods for feature selection, baseline prediction along with anomaly detection, and parametric analysis. Two case studies are used to illustrate the applications.

The first case study demonstrates the procedures of developing a baseline GP model for anomaly detection. It discusses feature selection for GP regression in detail. GP regression is robust in terms of handling irrelevant and redundant features. However, removing useless features will not only improve predictive accuracy but also significantly reduce computational cost for large datasets. The results also show that simple transformation of weather features might be useful. In the case study, temperature is transformed to cooling and heating degrees and humidity is transformed to dehumidification. According to the feature selection results, transformed features are better than the original ones. Feature transformation requires domain knowledge. Different HVAC system types may need different forms of transformation. The first case study also shows the advantage of using GP regression for baseline prediction. The output is a distribution instead of point estimation. It takes modelling uncertainty into consideration and provides confidence range, which is crucial for anomaly detection.

The second case study explores the process of estimating the impact of a control variable using a GP regression as surrogate model. The proposed impact factor gives an estimate of how significant the impact of the investigated parameter is on total energy consumption. This rapid estimate could be useful in early decision-making in building commissioning/retrofitting projects. For example, engineers can use this information to select control variables that have large impact to optimize. The proposed GP modeling process is a direct and rapid modeling method based on actual data. It avoids input configuration and calibration.

As part of the contribution of this study, a web-based forecasting tool has been developed. This tool allows more users to develop their GP models without the necessity for programming.

Acknowledgement

The authors would like to express our gratitude to Kevin Cahill, the director of facilities management of Harvard Graduate School of Design, who has been supporting our research and teaching since 2013. The authors would also like to thank Dr. Peter Henstock for the discussion and comments.

Reference

ASHRAE. (2004). Standard 55-2004—Thermal Environmental Conditions for Human Occupancy. Atlanta, GA: ASHRAE Inc.

ASHRAE. (2009). *ASHRAE Handbook Fundamentals* Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers Inc.

Bostock, M. (2015). D3, Data-Driven Documents. Retrieved from https://d3js.org/

Burger, E. M., & Moura, S. J. (2015). Gated ensemble learning method for demand-side electricity load forecasting. *Energy and Buildings, 109*, 23-34.

Burkhart, M. C., Heo, Y., & Zavala, V. M. (2014). Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. *Energy and Buildings, 75*, 189-198.

Cao, B., Ouyang, Q., Zhu, Y., Huang, L., Hu, H., & Deng, G. (2012). Development of a multivariate regression model for overall satisfaction in public buildings based on field studies in Beijing and Shanghai. *Building and environment, 47*, 394-399.

Che, J., & Wang, J. (2014). Short-term load forecasting using a kernel-based support vector regression combination model. *Applied energy, 132*, 602-609.

Cohen, D., & Krarti, M. (1995). *A neural network modeling approach applied to energy conservation retrofits.* Paper presented at the Building Simulation Fourth International Conference.

DOE. (2005). EnergyPlus Engineering reference: DOE.

EnerNOC. (2016). Retrieved from https://www.enernoc.com/products

Gray, F. M., & Schmidt, M. (2016). Thermal building modelling using Gaussian processes. *Energy and Buildings, 119*, 119-128.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.

Heo, Y., Choudhary, R., & Augenbroe, G. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings, 47*, 550-560.

Heo, Y., & Zavala, V. M. (2012). Gaussian process modeling for measurement and verification of building energy savings. *Energy and Buildings, 53*, 7-18.

Jetcheva, J. G., Majidpour, M., & Chen, W.-P. (2014). Neural network model ensembles for building-level electricity load forecasts. *Energy and Buildings, 84*, 214-223.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(3), 425-464.

KGS. (2016). Retrieved from http://www.kgsbuildings.com/clockworks

Kim, Y.-J., Ahn, K.-U., Park, C., & Kim, I.-H. (2013). *Gaussian emulator for stochastic optimal design of a double glazing system.* Paper presented at the Proceedings of the 13th IBPSA Conference, August.

Kissock, J. K., Reddy, T. A., & Claridge, D. E. (1998). Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. *Journal of Solar Energy Engineering, 120*(3), 168-176.

Kwac, J., Flora, J., & Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid, 5*(1), 420-430.

MacKay, D. J. (1997). Gaussian processes-a replacement for supervised neural networks? *Tutorial lecture notes for NIPS 1997*.

Majidpour, M., Qiu, C., Chu, P., Gadh, R., & Pota, H. R. (2015). Fast prediction for sparse time series: Demand forecast of EV charging stations for cell phone applications. *IEEE Transactions on Industrial Informatics, 11*(1), 242-250.

Manfren, M., Aste, N., & Moshksar, R. (2013). Calibration and uncertainty analysis for computer models–a meta-model based approach for integrated building energy simulation. *Applied energy, 103*, 627-641.

McKinney, W. (2011). *Pandas: a foundational Python library for data analysis and statistics*. Paper presented at the Python for High Performance and Scientific Computing.

MongoDB, I. (2016). Retrieved from https://www.mongodb.com/

Neal, R. M. (1995). *Bayesian learning for neural networks.* (PhD PhD thesis), University of Toronto.

Rasmussen, C. E. (1996). *Evaluation of Gaussian processes and other methods for non-linear regression.* (PhD PhD thesis), University of Toronto.

Rasmussen, C. E., & Nickisch, H. (2015). Gaussian Process Regression and Classification Toolbox version 3.6. Retrieved from http://www.gaussianprocess.org/gpml/code/matlab/doc/

Rastogi, P. (2016). *On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption.* (PhD Thesis), ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE.

Yan, B., Li, X., Malkawi, A. M., & Augenbroe, G. (2017). Quantifying uncertainty in outdoor air flow control and its impacts on building performance simulation and fault detection. *Energy and Buildings, 134*, 115-128.

Yan, B., & Malkawi, A. M. (2013). *A Bayesian approach for predicting building cooling and heating consumption.* Paper presented at the 13th International Building Performance Simulation Association Conference.

Yan, J., Kim, Y.-J., Ahn, K.-U., & Park, C.-S. (2013). *Gaussian process emulator for optimal operation of a high rise office building.* Paper presented at the Proceedings of 13th International Building Performance Simulation Association Conference.

Zheng, Y., & Kwoh, C. K. (2011). A feature subset selection method based on high-dimensional mutual information. *Entropy, 13*(4), 860-901.

Figure 9 Screenshots of existing fault detection tools (EnerNOC, 2016; KGS, 2016)

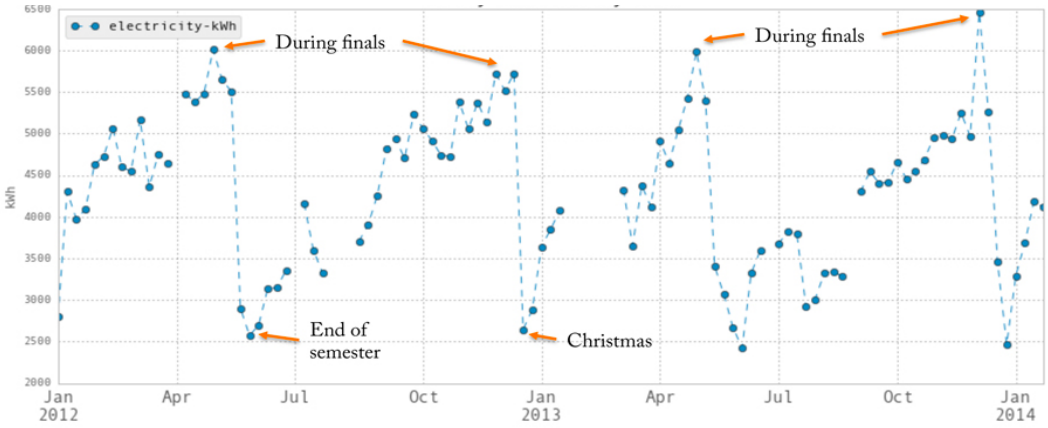

Figure 10 Weekly electric energy consumption

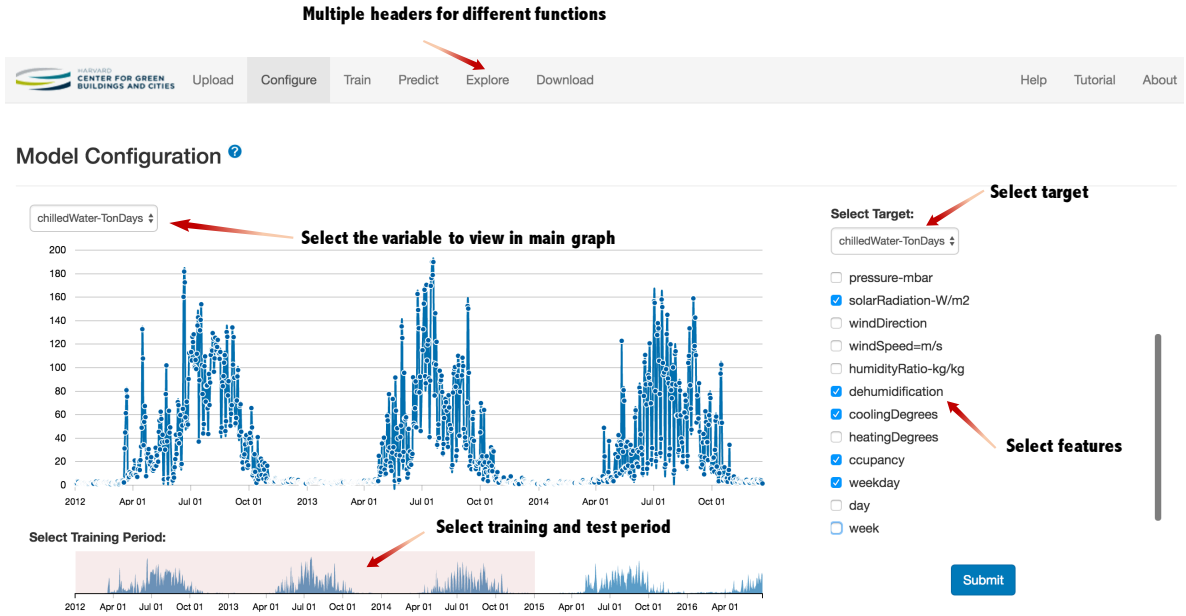

Figure 11 A screen shot of the web-based forecasting tool (Configure page)