Energy and Buildings 53 (2012) 7–18

Contents lists available at SciVerse ScienceDirect

Energy and Buildings

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

Gaussian process modeling for measurement and verification of building energy savings

Yeonsook Heo[a,*], Victor M. Zavala[b]

[a] Decision and Information Sciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA
[b] Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA

ARTICLE INFO

Article history:
Received 20 May 2012
Received in revised form 22 June 2012
Accepted 30 June 2012

Keywords:
Gaussian process modeling
Measurement and verification
Performance-based contracts
Retrofit analysis
Uncertainty

ABSTRACT

We present a Gaussian process (GP) modeling framework to determine energy savings and uncertainty levels in measurement and verification (M&V) practices. Existing M&V guidelines provide savings calculation procedures based on linear regression techniques that are limited in their predictive and uncertainty estimation capabilities. We demonstrate that, unlike linear regression, GP models can capture complex nonlinear and multivariable interactions as well as multiresolution trends of energy behavior. In addition, because GP models are developed under a Bayesian setting, they can capture different sources of uncertainty in a more systematic way. We demonstrate that these capabilities can ultimately lead to significantly less expensive M&V practices. We illustrate the developments using simulated and real data settings.

1. Introduction

Energy service companies (ESCOs) provide efficiency services to reduce energy consumption in buildings on the basis of performance-based contracts. These contracts guarantee energy or cost savings during a given contract term. If the guarantee is not realized during the term, ESCOs pay the shortfall to building owners according to their contract clauses. Moreover, their service costs are entirely or partly tied to the amount of savings achieved. Hence, ESCOs must quantify risks associated with their retrofit decisions and determine (verify) retrofit savings with high confidence. In addition, ESCOs play a significant role in achieving nationwide energy efficiency targets because they help mitigate perceived risk and speculation of investors and building owners in emerging building technologies. In the United States, the ESCO industry is expected to earn annual revenues of $7.1–7.3 billion in 2011, three quarters of which come from performance-based contracts [23,7].

Performance-based contracts follow intensive measurement and verification (M&V) procedures that seek to quantify energy savings and uncertainty levels resulting from retrofit practices. For the M&V process, ESCOs follow the International Performance Measurement and Verification Protocol (IPMVP) [13,9] and ASHRAE's Guideline 14 [2]. The IPMVP offers uncertainty estimation techniques based on measured data and linear regression models. Regression models need to be typically adjusted in ad hoc manners (e.g., domain partition) in order to capture nonlinear behavior, which arises from complex (physical) multivariable interactions between ambient conditions, occupancy levels, and building operating conditions. Regression models are also highly sensitive to data availability, so uncertainty estimates need often be adjusted accordingly [6].

We argue that the use of existing regression and uncertainty estimation techniques tends to make the entire M&V process costly because large amounts of baselining data are required in order to obtain acceptable uncertainty levels in model-predicted savings. This situation is important from a financial point of view because ESCO projects often confront limited M&V budgets for data collection, sensor deployment, and engineering analysis [13,8]. Consequently, new modeling strategies must be developed capable of reducing uncertainty levels by capturing multivariable interactions and by exploiting additional sources of information, such as hourly energy trends.

Recent work has emphasized the importance of uncertainty analysis in energy-efficiency projects and applied statistical techniques such as Monte Carlo analysis for quantitative risk management in the retrofit analysis process [17,18,16,14,24]. In addition, general Bayesian settings have recently been developed to quantify uncertainty in building energy models [11,12]. In these works, the authors demonstrate that the use of a Bayesian setting enables the quantification of three types of uncertainties (i.e., parameter uncertainty, model inadequacy, measurement error) in a systematic manner, while enhancing models to capture the actual behavior.

* Corresponding author. Tel.: +1 630 252 3568.
E-mail address: yheo@anl.gov (Y. Heo).

0378-7788/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.enbuild.2012.06.024

In this paper, we present a Gaussian process (GP) modeling framework for M&V applications. The modeling technique is entirely data-based, follows a Bayesian setting, and is a generalization of nonlinear multivariable regression. Notably, the approach does not require the modeler to specify the structural relationships between independent and dependent variables. Instead, the modeler specifies the covariance structure of the input variables in order to explain their interactions with the outputs [20,3]. As a result, GP models can capture complex behavior with fewer parameters and thus reduces the modeling effort. GP has been widely used in different application domains, including geophysics and weather forecasting, to capture spatiotemporal patterns [25,5,31,30]. We present a simulation study demonstrating that GP can enable the incorporation of multiple input variables affecting energy demand and that it can lead to important reductions of baselining data. Using a case study with real data, we demonstrate that GP leads to more rigorous uncertainty estimates compared with those of standard regression methods and can exploit the availability of high-resolution data in order to reduce uncertainty levels.

The paper is structured as follows: In Section 2 we summarize existing M&V modeling and uncertainty estimation techniques. In Section 3 we describe the general principles of GP and discuss advantages and limitations. In Section 4 we present modeling strategies to capture data with multiple time resolutions. In Section 5 we present two case studies demonstrating the capabilities of GP. The paper closes with conclusions and directions of future work.

2. Measurement and verification

In this section we present the general principles and regression techniques currently used in measurement and verification.

2.1. General principles

The IPMVP [13] and ASHRAE's Guideline 14 [2] provide three methods to determine energy savings and uncertainty levels from energy efficiency measures (EEMs). These are retrofit isolation, whole-building metering, and calibrated simulation. The retrofit isolation method evaluates the savings from an upgraded building component (e.g., chiller, boiler, lighting system) by submetering its energy demand during pre-retrofit and post-retrofit periods. The whole-building metering method compares total energy demand or cost during pre-retrofit and post-retrofit periods. These two methods can be applied when measurement data during both the pre-retrofit and the post-retrofit periods is available. When data from either period is not available, the guidelines recommend the calibrated simulation method, which develops an energy simulation model based on available metered data. In this paper, we focus on methods relying on metered data.

Fig. 1 illustrates the procedure for estimating energy savings and uncertainty levels associated with those estimates on the basis of metered data. First, one selects relevant independent (also called explanatory) variables. ASHRAE's guideline recommends including variables that are expected to change between the pre-retrofit and the post-retrofit periods and that significantly impact energy use. Common variables considered are weather and occupancy levels, and the most commonly used is outdoor dry-bulb air temperature. Second, one develops predictive baseline models based on metered pre-retrofit energy use, $\mathbf{y}_{pre}$, to predict pre-retrofit energy use, $\mathbf{y}^{P}_{pre}$. The guideline provides different modeling strategies based on linear regression techniques. Piecewise linear models are often used to capture nonlinear energy behavior in different regimes (e.g., weekday/weekend, heating/cooling seasons). In this step, one should select duration and time resolution (hourly/daily/monthly) of data for pre-retrofit and post-retrofit periods. The guideline states that measurement data should be sufficient to characterize operating periods over the full range of independent variables and to achieve prescribed uncertainty levels. Depending on the type of baseline model and resolution, large amounts of data may be necessary in order to improve accuracy. In this context, a key limitation of linear regression models is that they cannot adequately capture autocorrelations existing at higher time resolutions. Therefore, the guideline recommends that uncertainty levels be adjusted to deal with autocorrelation effects. With the baseline model developed, one calculates pre-retrofit energy use in the post-retrofit period, $\mathbf{y}^{pre,P}_{post}$, as excluding the effects of other factors such as weather conditions and changes in building usage (e.g., occupancy) patterns. Then, by subtracting energy use during the post-retrofit period from model-predicted baseline energy use during the post-retrofit period, one obtains energy-saving estimates.

The guideline explicitly acknowledges the importance of uncertainty, particularly for validating analysis results. It proposes to quantify three types of uncertainty: model uncertainty, input/output measurement uncertainty, and sampling uncertainty. Model uncertainty is estimated by the statistics of the model prediction error. Input measurement uncertainty arises from independent variables (e.g., outdoor weather conditions), while output measurement uncertainty is estimated from the accuracy of an instrument used to measure energy use (e.g., chilled-water flow rate, power meter). Sampling uncertainty arises from the use of restricted sample sizes that can bias uncertainty estimates. The guideline regards measurement uncertainty and sampling uncertainty as negligible for most retrofit analysis cases and uses a simplified formula derived from a simple linear model to quantify modeling uncertainty. In this work, we focus on model uncertainty.

2.2. Technical description

ASHRAE's guideline provides equations to calculate the baseline model uncertainty as suggested by Reddy and Claridge [21]. The equations are derived on the assumption that the model is a linear parametric regression model with additive noise of the form $y = \mathbf{x}^T \cdot \mathbf{w} + \epsilon$ where $y \in \Re$ is the energy output, $\mathbf{x} \in n_x\Re$ is the input vector with n_x variables (e.g., outdoor dry-bulb temperature, humidity), $\mathbf{w} \in n_w\Re$ are the model parameters, and $\epsilon \sim \mathcal{N}(0, \sigma^2_\epsilon)$ is the random model error assumed to be normally distributed with zero mean and variance σ^2_ϵ. [1]

Consider a set of $i = 1, \ldots, n$ measured inputs $\mathbf{X}_{pre}[i,:] = \mathbf{x}_{pre}(i)$ and energy outputs $y_{pre}(i)$ during the pre-retrofit period. Similarly, consider a set of $j = 1, \ldots, m$ inputs $\mathbf{X}_{post}[j,:] = \mathbf{x}_{post}(j)$ and outputs $y_{post}(j)$ during the post-retrofit. Here, $\mathbf{X}_{pre} \in n \times n_x\Re$ and $\mathbf{X}_{post} \in m \times n_x\Re$ are the input regression matrices. On the basis of the data, one develops a linear pre-retrofit model that seeks to predict baseline energy use during the post-retrofit period. We denote this model prediction as $\mathbf{y}^{pre,P}_{post} = \mathbf{X}_{post}\mathbf{w}$. For convenience, we also define the real (unknown) baseline energy use in the post-retrofit period as $\mathbf{y}^{pre}_{post}$. The variable definitions are illustrated in Fig. 1. Using these definitions, the *predicted* and *real* energy savings aggregated over the post-retrofit period are given by

$$y^{P}_{save,m} = \sum_{j=1}^{m} y^{pre,P}_{post}(j) - \sum_{j=1}^{m} y_{post}(j) \tag{2.1a}$$

[1] Throughout the paper we will use a boldface, lowercase font to indicate column vectors $\mathbf{x}$ and a boldface, uppercase font to indicate matrices $\mathbf{X}$. We will denote the ith entry of a vector as $\mathbf{x}[i]$, the ijth entry of a matrix as $\mathbf{X}[i,j]$, and its ith row as $\mathbf{X}[i,:]$.

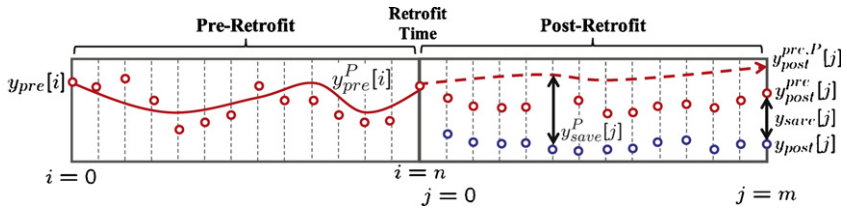

Fig. 1. Standard energy-savings calculation procedure.

$$y_{save,m} = \sum_{j=1}^{m} y^{pre}_{post}(j) - \sum_{j=1}^{m} y_{post}(j), \tag{2.1b}$$

respectively. We can find the parameters of the linear model by applying maximum likelihood estimation, which gives $\mathbf{w}^* = (\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-1}\mathbf{X}^T_{pre}\mathbf{y}_{pre}$ [6]. The variance of the parameters is

$$\mathbb{V}[\mathbf{w}^*] = (\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-1}\mathbf{X}^T_{pre}\mathbb{V}[\mathbf{y}_{pre}]\mathbf{X}_{pre}(\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-T}, \tag{2.2}$$

where $\mathbb{V}[\mathbf{x}] \in n_x \times n_x\Re$ denotes the covariance matrix of vector $\mathbf{x} \in n_x\Re$. Then,

$$\begin{aligned}\mathbb{V}[\mathbf{y}_{pre}] &= \mathbb{E}\left[(\mathbf{y}_{pre} - \mathbb{E}[\mathbf{y}_{pre}])(\mathbf{y}_{pre} - \mathbb{E}[\mathbf{y}_{pre}])^T\right] \\ &= \mathbb{V}[\boldsymbol{\epsilon}_{pre}]\end{aligned} \tag{2.3}$$

is the variance of the model error, since $\mathbb{E}[\mathbf{y}_{pre}] = \mathbf{y}^P_{pre} = \mathbf{X}_{pre}\mathbf{w}^*$. Here, $\mathbb{E}[\mathbf{x}] \in n_x\Re$ denotes the expected value of a vector $\mathbf{x} \in n_x\Re$.

For the post-retrofit period, the baseline outputs are given by $\mathbf{y}^{pre}_{post} = \mathbf{y}^{pre,P}_{post} + \boldsymbol{\epsilon}_{post}$, where ϵ_{post}, is the model error and $\mathbf{y}^{pre,P}_{post} = \mathbf{X}_{post}\mathbf{w}^*$ is the prediction. The covariance matrix of the post-retrofit baseline outputs is given by

$$\begin{aligned}\mathbb{V}[\mathbf{y}^{pre}_{post}] &= \mathbb{V}[\mathbf{y}^{pre,P}_{post}] + \mathbb{V}[\boldsymbol{\epsilon}_{post}] \\ &= \mathbf{X}_{post}(\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-1}\mathbf{X}^T_{pre}\mathbb{V}[\boldsymbol{\epsilon}_{pre}]\mathbf{X}_{pre} \\ &\quad \times (\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-T}\mathbf{X}^T_{post} + \mathbb{V}[\boldsymbol{\epsilon}_{post}].\end{aligned} \tag{2.4}$$

Typically, both $\mathbb{V}[\boldsymbol{\epsilon}_{pre}]$ and $\mathbb{V}[\boldsymbol{\epsilon}_{post}]$ matrices are assumed to be diagonal (uncorrelated model errors) with entries σ^2_ϵ. As typically done in linear regression, the variance of the model error σ^2_ϵ is approximated by

$$MSE = \frac{1}{n - n_w}\sum_{i=1}^{n}(y_{pre}(i) - y^P_{pre}(i))^2. \tag{2.5}$$

This assumption is justified by the fact that $(MSE/\sigma^2_\epsilon) \sim (\chi^2_{n-n_w}/n - n_w)$, where $\chi^2_{n-n_w}$ denotes a chi-squared distribution with $n - n_w$ degrees of freedom. This yields the prediction variance

$$\mathbb{V}[\mathbf{y}^{pre}_{post}] = MSE\left[\mathbf{X}_{post}(\mathbf{X}^T_{pre}\mathbf{X}_{pre})^{-1}\mathbf{X}^T_{post} + I\right]. \tag{2.6}$$

The total aggregated variance over the post-retrofit period is the sum of entries of the above matrix, given by

$$(\sigma^{pre}_{post,m})^2 = \mathbf{1}^T_m\mathbb{V}[\mathbf{y}^{pre}_{post}]\mathbf{1}_m = \sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{V}[\mathbf{y}^{pre}_{post}][i,j], \tag{2.7}$$

where $\mathbf{1}_m$ is a vector of ones with m entries. This variance also corresponds to the uncertainty of the post-retrofit savings $\sigma^2_{save,m} = (\sigma^{pre}_{post,m})^2$.

ASHRAE's guideline specializes the prediction variance (2.6) by considering the case of one input variable ($\mathbf{X}_{pre} \in n\Re$, $\mathbf{X}_{post} \in m\Re$). This yields the expression for computing point variances,

$$\sigma^2_{save}[j] = MSE\left(1 + \frac{1}{n} + \frac{(\mathbf{X}_{post}[j] - \overline{X}_{post})^2}{\sum_{i=1}^{n}(\mathbf{X}_{pre}[i] - \overline{X}_{pre})^2}\right), \tag{2.8}$$

for $j = 1, \ldots, m$. Here, $\overline{X}_{pre}, \overline{X}_{post} \in \Re$ are the averages of the entries of vectors $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$, respectively. Following the assumptions about the statistics of the model error, we can construct confidence intervals for the savings using a t-statistic, since $\mathbf{y}^{pre}_{post} \sim \mathcal{N}(\mathbf{y}^{pre,P}_{post}, \mathbb{V}[\mathbf{y}^{pre}_{post}])$ and $(MSE/\sigma^2_\epsilon) \sim (\chi^2_{n-n_w}/n - n_w)$. By expressing the two-sided t-statistic value with α confidence level and $n - n_w$ degrees of freedom ($t_{\alpha,n-n_w}$), we obtain a final expression for the point confidence intervals,

$$\pm\sigma_{save}[j] = t_{\alpha,n-n_w}\sqrt{MSE}\sqrt{1 + \frac{1}{n} + \frac{(\mathbf{X}_{post}[j] - \overline{X}_{post})^2}{\sum_{i=1}^{n}(\mathbf{X}_{pre}[i] - \overline{X}_{pre})^2}}, \tag{2.9}$$

where $j = 1, \ldots, m$. Since the errors are assumed to be uncorrelated, the point variances can be aggregated over the m post-retrofit points to obtain the total uncertainty $\sigma^2_{save,m}$ and the confidence intervals

$$\pm\sigma_{save,m} = t_{\alpha,n-n_w}\sqrt{MSE}\sqrt{m + \frac{m}{n} + \frac{\sum_{j=1}^{m}(\mathbf{X}_{post}[j] - \overline{X}_{post})^2}{\sum_{i=1}^{n}(\mathbf{X}_{pre}[i] - \overline{X}_{pre})^2}}. \tag{2.10}$$

The guideline stipulates the validity of the energy savings uncertainty levels from the above expressions: if the calculated interval is less than 50% of the predicted savings at 68% confidence level, the resulting M&V saving estimate is appropriate.

ASHRAE's guideline suggests correcting the confidence intervals in order to consider autocorrelated model errors as those arising at hourly and daily resolutions. This correction is done in a deterministic manner (i.e., it is not a natural outcome of the regression procedure) by reducing the number of independent measurements depending on the lag of autocorrelation [21].

Prevailing modeling and uncertainty estimation techniques in M&V practices have two major drawbacks. First, they need to assume relationships between independent and dependent variables and often force modelers to consider simplified model structures. Piecewise linear models are often used to capture nonlinearity and multivariable behavior arising at different regimes (heating/cooling, weekday/weekends) [2,21]. These ad hoc procedures can lead to overparameterized models with unnecessarily high uncertainty levels. Furthermore, without the ability to capture nonlinearity and variable interdependencies, regression models are of limited use in capturing dynamic behavior at higher time resolutions. Second, the uncertainty estimation techniques used are valid only in regions of close-to-linear behavior and often require large data sets. These limitations can lead to misspecifications of uncertainty levels that, in turn, can lead to erroneous estimates of the amounts of baseline data required. These drawbacks make the M&V process more expensive than necessary and misguide decision-makers.

3. Gaussian process modeling

Several data-based modeling techniques can be used for M&V, including time series [4], neural networks [10], support vector machines [27], and polynomial regression [6,3]. In this work we propose to use Gaussian process modeling [20,22,15,19], an

approach that has recently received attention in many application domains. A key advantage of this approach is that it constructs the model by specifying the structure of the covariance matrix of the explanatory variables rather than the algebraic structure of the input–output relationship itself, as is done in traditional parametric regression approaches. This feature makes the GP approach highly flexible. In addition, GP is derived from a Bayesian setting, which enables the modeler to fuse and quantify different sources of uncertainty in a natural way. In this section, we describe the computational procedures of GP, discuss advantages and limitations, and describe modeling techniques to predict multiple time resolutions.

3.1. Formulation

To illustrate the use of the GP technique, we construct a model by regressing the *explanatory* variables $\mathbf{x} \in n_x\Re$ to the *output* variable $y \in \Re$. We collect $i = 1, \ldots, n$ observations for these pairs and define the *training* matrix for the input vectors $\mathbf{X}[i, :] = \mathbf{x}(i)$ and the outputs $\mathbf{y}[i] = y(i)$. Here, matrix $\mathbf{X} \in n \times n_x\Re$ is the training input matrix, and $\mathbf{y} \in n\Re$ is the training vector.

Typically, two input vectors $\mathbf{x}(i)$, $\mathbf{x}(j)$ are assumed to be correlated through an exponential kernel function. This gives the training covariance matrix $\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta) \in n \times n\Re$, with entries

$$\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta)[i, j] = \eta_0 + \eta_1 \cdot \exp\left(-\frac{1}{\eta_2}\|\mathbf{x}(j) - \mathbf{x}(i)\|^2\right), \quad i, j = 1, \ldots, n. \tag{3.11}$$

Here, $\eta = [\eta_0, \eta_1, \eta_2]^T \in \Re$ are the covariance function hyperparameters, and $\|\cdot\|$ is the Euclidean norm. The model is trained by finding the hyperparameter values that maximize the log likelihood function

$$\log p(\mathbf{y}|\eta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}(\mathbf{y} - \mathbf{m}(\theta))^T\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta)(\mathbf{y} - \mathbf{m}(\theta)) - \frac{1}{2}\log\det(\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta)). \tag{3.12}$$

Here $\mathbf{m}(\theta)$ is the mean model vector with n_θ parameters $\theta = [\theta_1, \ldots, \theta_{n_\theta}]$. In this work, we assume that limited information about the structure of the mean model is available so we assume it to be zero.

After the optimal hyperparameters η^* are computed, we can determine mean predictions $\boldsymbol{\mu}_{test} \in m\Re$ and associated covariance matrix $\mathbf{V}_{test} \in m \times m\Re$ for a set of m input test points contained in the test input matrix $\mathbf{X}_{test} \in m \times n_x\Re$ with entries $\mathbf{X}_{test}[j, :] = x_{test}(j)$ and where $x_{test}(j)$, $j = 1, \ldots, m$ are the testing points. The GP predictions seek to predict the *true* observed outputs, which we denote as $y_{test}(j)$, $i = 1, \ldots, m$.

The mean and covariance of the predictive *posterior* distribution $p(\mathbf{y}_{test}|\mathbf{y}, \mathbf{X}, \mathbf{X}_{test}, \eta^*)$ are given by

$$\mu_{test} = \mathbb{E}[\mathbf{y}_{test}] = \mathbf{V}(\mathbf{X}_{test}, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta^*)\mathbf{y} \tag{3.13a}$$

$$\mathbf{V}_{test} = \mathbf{V}(\mathbf{X}_{test}, \mathbf{X}_{test}, \eta^*) - \mathbf{V}(\mathbf{X}_{test}, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta^*)\mathbf{V}(\mathbf{X}, \mathbf{X}_{test}, \eta^*). \tag{3.13b}$$

Here $\mathbf{V}(\mathbf{X}_{test}, \mathbf{X}, \eta^*) \in m \times n\Re$ is the cross covariance. With this, we can define an abstract GP model of the form

$$(\mu_{test}, \mathbf{V}_{test}) = \mathcal{M}(\mathbf{X}_{test}, \eta^*). \tag{3.14}$$

3.2. Strengths and limitations of GP

GP can capture strong nonlinearities and multivariable interactions in a systematic way. To explain the reason, we consider the nonlinear regression model

$$y = \sum_{j=1}^{n_w} w[j]\phi_j(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}), \tag{3.15}$$

where $\mathbf{x} \in n_x\Re$ are input vectors, $\phi(\cdot) \in n_w\Re$ are nonlinear basis functions, and $\mathbf{w} \in n_w\Re$ are the model weight parameters. Assume for now noise-free outputs $\mathbf{y} \in n\Re$ so that $\mathbf{y} = \Phi\mathbf{w}$, where $\Phi := \Phi(\mathbf{x}) \in n \times n_w\Re$. If we assume a normal prior distribution for the weights $\mathbf{w} \sim \mathcal{N}(0, \mathbb{V}[\mathbf{w}])$ with diagonal covariance $\mathbb{V}[\mathbf{w}] \in n_w \times n_w\Re$ with entries σ_w^2, the mean and covariance of the probability distribution of $\mathbf{y}$ are given by

$$\mathbb{E}[\mathbf{y}] = \Phi\mathbb{E}[\mathbf{w}] = 0 \tag{3.16a}$$

$$\mathbb{V}[\mathbf{y}] = \Phi^T\mathbb{E}[\mathbf{y}\mathbf{y}^T]\Phi^T = \Phi\mathbb{V}[\mathbf{w}]\Phi^T = \mathbf{K}, \tag{3.16b}$$

where $\mathbf{K} \in n \times n\Re$ is a matrix with elements $\mathbf{K}[i, j] = k(\mathbf{x}(i), \mathbf{x}(j)) = \sigma_w^2\phi(\mathbf{x}(i))^T\phi(\mathbf{x}(j))$ and where $k(\cdot, \cdot)$ is the kernel function. From these relations, one can see that *it is possible to obtain the distribution of* $\mathbf{y}$ *without explicit knowledge of the weight values* $\mathbf{w}$. All that is needed is the structure kernel function. In Eq. (3.11), for instance, the kernel function is the Gaussian kernel $k(\mathbf{x}(i), \mathbf{x}(j)) = \eta_1\exp(-\frac{1}{\eta_2}\|\mathbf{x}(i) - \mathbf{x}(j)\|) + \eta_0$ with $\eta_1 \leftarrow \sigma_w^2$, and η_0 is the measurement noise for the outputs $\mathbf{y}$. Consequently, we can think of the GP model as a nonlinear regression model with Gaussian basis functions. Note that since the weight parameters are not needed, GP models typically require significantly fewer parameters. A wide range of kernel functions and combinations exist to capture different forms of nonlinearity and correlation behavior. For more details, we refer the reader to [20,3].

Another critical feature of GP is that it provides direct information on the posterior distribution for the predictions. As in a Bayesian setting, one needs only to compute the conditional distribution $p(\mathbf{y}_{test}|\mathbf{y}, \mathbf{X}, \mathbf{X}_{test}, \eta^*)$. Since the distribution is assumed to be Gaussian, this computation can be done by using a Schur complement approach, leading to (3.13). In other words, the framework implicitly estimates the model error structure (e.g., additive, multiplicative) as part of the maximum-likelihood procedure. This makes the approach more systematic and adaptive to data.

The predictive qualities of the GP model are strongly influenced by the range covered by the training and testing set. As suggested by the structure of the kernel function (3.11), as the distance between training and testing sets increases, so does their correlation; and the GP model is required to extrapolate outside the training domain.

Poor input variable scaling can also lead to biasing preference to certain variables. To ameliorate this effect, we normalize all input and output variables using upper and lower bounds as

$$\mathbf{x} \leftarrow \frac{\mathbf{x} - \mathbf{x}_L}{\mathbf{x}_U - \mathbf{x}L}, \quad y \leftarrow \frac{y - y_L}{y_U - y_L}. \tag{3.17}$$

Since the number of hyperparameters used in GP is typically small, the maximization procedure can be carried out by using gradient-based or pattern-search algorithms. A key challenge presented by the log likelihood function is that it is strongly nonconvex and exhibits multiple local minima. Consequently, if gradient-based approaches are used, multiple starting points are often needed to eliminate spurious local minima.

We also note that each evaluation of the log likelihood function is expensive, since it requires the computation of a log-determinant term and of the inverse of the covariance matrix $\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta) \in n \times n\Re$. The inverse is typically computed indirectly by Cholesky factorization, which scales cubically with the number of training

points n [20]. Recent approaches try to overcome this complexity by sparsifying the covariance matrix [20] or by applying iterative linear algebra techniques [1,26].

4. Modeling strategies for M&V

In the M&V context, the explanatory variables are forcings of the building system such as ambient conditions (temperature, humidity, enthalpy, wind speed, radiation), occupancy level, and set-points of the control system that can be captured at different time resolutions, depending on the M&V objective at hand and the available data. The output variables are state variables such as energy demand of a given device (air-handling unit (AHU), chiller, steam) or the total building energy demand, but they can also be some other state variables that implicitly represent energy consumption (e.g., fan flow rate). In this section, we describe how to use GP modeling to capture trends with different time resolutions, and we discuss the need for incorporating more explanatory variables.

4.1. Time resolution

Selecting appropriate data time resolutions is important because autocorrelations arise as time scales become shorter. These can affect the predictions if not considered in the model construction. Resolution also has implications for the complexity of the models because the number of explanatory variables needs to be increased to capture autocorrelations. In addition, autocorrelations tend to increase the range of extrapolation of the model, since the model becomes state-dependent. On the other hand, as time scales become shorter, the availability of data increases. For instance, daily ambient and energy variations can provide useful information for M&V. We emphasize that, since GP enables the model to capture larger number of explanatory variables, it can be used to exploit higher data resolutions.

Modeling at low time resolutions (days, weeks, and months) can be performed by aggregating (or averaging) $t=0, \ldots, T$ high-resolution (hours, minutes) signals over low-resolution time periods of length N. If we have a set of input measurements $\mathbf{x}(t) \in n_x\Re$ at a set of times $t=0, \ldots, T$, we partition the data in $k=0, \ldots, N_{ag}$ blocks, each with N entries, and where $n=(T-1)/N$ is the number of aggregated baselining points. We create the average $\mathbf{x}(k) \leftarrow \mathbb{E}[\mathbf{x}(k), \mathbf{x}(k+1)\ldots\mathbf{x}(k+N)]$, $k=1, \ldots, n$. The same approach is used for the output variables $y(t)$, $t=0, \ldots, T$. The GP model is then created by defining the input matrix $\mathbf{X}[k, :]=\mathbf{x}(k)$ and the output vector as $\mathbf{y}[k]=y(k)$, $k=1, \ldots, n$. To test the model, we use a set of $\tau=0, \ldots, T_\tau$ testing points, aggregating by periods of length N to generate $m=(T_\tau-1)/N$ points. This gives the input matrix and output vector $\mathbf{X}_{test}[k, :]=\mathbf{x}(k)$, $\mathbf{y}_{test}[k]=y(k)$, $k=1, \ldots, m$.

For M&V purposes, it is of interest to compute the total aggregated energy savings and uncertainty over the testing (post-retrofit) period. Since $\mathbf{y}_{test} \sim \mathcal{N}(\mu_{test}, \mathbf{V}_{test})$, the total energy predicted in the testing period is normally distributed with mean and variance:

$$\mu_{test,m} = \mathbf{1}_m^T \mu_{test} = \sum_{j=1}^{m} \boldsymbol{\mu}_{test}[j] \tag{4.18a}$$

$$\sigma^2_{test,m} = \mathbf{1}_m^T \mathbf{V}_{test} \mathbf{1}_m = \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbf{V}_{test}[i,j]. \tag{4.18b}$$

Here following the notation of Section 2, we have $\mu_{test} = \mathbf{y}^{pred,P}_{post}$ and $\mathbf{y}_{test} = \mathbf{y}^{pred}_{post}$. The mean and variance of the total energy savings are then

$$\begin{aligned} y^{sav,P}_{test,m} &= \mu_{test,m} - \mathbf{1}_m^T \mathbf{y}_{post} \\ &= \sum_{j=1}^{m} (\mu_{test}(j) - y_{post}(j)). \end{aligned} \tag{4.19}$$

We also have $(\sigma^{sav}_{test,m})^2 = \sigma^2_{test,m}$. To benchmark the accuracy of the GP model, we also define the *true* (unknown) energy savings,

$$y^{sav}_{test,m} = \sum_{j=1}^{m} (y_{test}(j) - y_{post}(j)). \tag{4.20}$$

With this definition, we can define the total energy savings prediction error,

$$\begin{aligned} \epsilon_{test,m} &= \mu^{sav}_{test,m} - y^{sav}_{test,m} \\ &= \mathbf{1}_m^T(\boldsymbol{\mu}_{test} - \mathbf{y}_{post}) - \mathbf{1}_m^T(\mathbf{y}_{test} - \mathbf{y}_{post}) \\ &= \mathbf{1}_m^T(\mu_{test} - \mathbf{y}_{test}), \end{aligned} \tag{4.21}$$

and the total sum-of-squares error $\epsilon^{SSE}_{test,m} = (\mu_{test} - \mathbf{y}_{test})^T(\mu_{test} - \mathbf{y}_{test})$. We note that aggregation can lead to wrong perceptions of the quality of the model. For instance, the individual errors in (4.21) can cancel out and give a small $\epsilon_{test,m}$, giving the erroneous perception that a bad model is *globally* accurate. We know that $\epsilon^{SSE}_{test,m} \geq \epsilon^2_{test,m}$, so a model with lower $\epsilon^{SSE}_{test,m}$ than another one does not necessarily yield a lower $\epsilon_{test,m}$. Because of these possible misconceptions, it is important to monitor the entire set of point errors $\epsilon_{test}(j), j=1,..,m$, in order to determine points at which the model is more accurate and certain.

4.1.1. Autocorrelation

When autocorrelations need to be included in the GP model, we can consider two approaches to generate predictive models. We assume we have data for explanatory $\mathbf{x}(t) \in n_x\Re$ and output $y(t) \in \Re$ variables spread over the horizon $t=0, \ldots, T$. We assume a period of autocorrelation (lag) of $N<T$ steps, and we define the input variable vector at time t as $\mathbf{x}^T(t) \leftarrow [\mathbf{x}^T(t), y(t), \mathbf{x}^T(t-1), y(t-1), \ldots, \mathbf{x}^T(t-N), y(t-N)]$ for $t=N, \ldots, T$ so that $\mathbf{x}(t) \in (n_x+n_y)\cdot N\Re$. We seek to create a dynamic model of the form $y(t+1) = \mathcal{M}(\mathbf{x}(t))$, $t=N, \ldots, T-1$. We define the entries of the training input matrix as $\mathbf{X}[t, :]=\mathbf{x}(t)$ and the output vector as $\mathbf{y}[t]=y(t+1)$ with $t=N, \ldots, T-1$, and we note that $n=T-N-1$ is the number of training points.

For the testing procedure, we differentiate two settings: one-step-ahead (off-line) and multi-step-ahead (on-line). In the off-line approach, we construct the test matrix entries $\mathbf{X}_{test}[\tau, :]=\mathbf{x}_{test}(\tau)$, $\tau=N, \ldots, T_\tau-1$, where T_τ is the number of testing points and the input vector is constructed by using the testing inputs $\mathbf{x}^T_{test}(\tau) \leftarrow [\mathbf{x}^T(\tau), y[\tau], \mathbf{x}^T(\tau-1), y(\tau-1)\ldots., \mathbf{x}^T(\tau-N), y(\tau-N)]$. We note that this approach relies on the existence of the testing observations and thus is useful only in off-line M&V procedures and/or on-line approaches that require only one-step-ahead predictions.

For on-line approaches requiring multi-step-ahead predictions, it is necessary to generate sophisticated recursions by using a mixture of observations of GP model predictions [31]. Note that at current time t, we have measurements to compute only the single-step prediction $y_{test}(t+1)$. To obtain multi-step predictions $y_{test}(t+j)$, $j=1, \ldots, N_F$, where N_F is the number of forecast points, we must propagate the GP predictions recursively. We use the following algorithm.

Assume we have trained hyperparameters η^* and input covariance $\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta^*)$ with data collected up to time t. This creates the model $\boldsymbol{\mu}_{test} = \mathcal{M}(\mathbf{x}_{test}, \eta^*)$.

1. **Prediction mean computation:** For $j = 1, \ldots, N_F$,
   (a) Set $\mathbf{X}_{test}[j, :] \leftarrow [\mathbf{x}^T(t), y(t), \mathbf{x}^T(t-1), y(t-1) \ldots, \mathbf{x}^T(t-N), y(t-N)]$.
   (b) Compute $\boldsymbol{\mu}_{test}[j] = \mathcal{M}(\mathbf{X}_{test}[j, :], \eta^*)$.
   (c) Set $y(t+1) \leftarrow \boldsymbol{\mu}_{test}[j]$ and obtain forcing $\mathbf{x}(t+1)$.
   (d) Drop oldest observation $\mathbf{x}(t-N)$, $y(t-N)$ and set $t \leftarrow t+1$.
2. **Prediction covariance computation:** Compute self-covariance $\mathbf{V}(\mathbf{X}_{test}, \mathbf{X}_{test}, \eta^*)$ and cross-covariance $\mathbf{V}(\mathbf{X}_{test}, \mathbf{X}, \eta^*)$. Compute forecast covariance $\mathbf{V}_{test}$ from (3.13).

This procedure generates the mean $\boldsymbol{\mu}_{test} = [y(t+1), \ldots, y(t+F)]$ and covariance $\mathbf{V}_{test}$. If necessary, the model can also be retrained as on-line information arrives.

4.1.2. Data differencing

The autocorrelation approach creates a time-series type of model. An alternative approach is to calculate a discrete-time approximation of the dynamic model:

$$\frac{dy}{dt}(t) = f(y(t), \mathbf{x}(t)), \tag{4.22}$$

where $y(t) \in \Re$ is the desired state and $\mathbf{x}(t) \in n_x\Re$ are the forcings. An explicit Euler discretization of the dynamic model results in

$$y(t + \Delta) = y(t) + \Delta \cdot f(y(t), \mathbf{x}(t)), \tag{4.23}$$

where Δ is a fixed discretization time. In this approach, we seek to directly approximate the mapping $f(\cdot, \cdot)$ by regressing the output point $\Delta y(t) \leftarrow y(t+\Delta) - y(t)$ directly with $y(t)$ and $\mathbf{x}(t)$. The motivation behind this approach is that if the mapping $f(\cdot, \cdot)$ is known, then the dynamic evolution of the system can be predicted. In addition, the approach can be used to predict the steady-state conditions of the system.

We can see that the data differencing approach is equivalent to the autocorrelation approach with $N = 1$ if we set $\mathcal{M}(y(t), \mathbf{x}(t)) \leftarrow y(t) + \Delta \cdot f(y(t), \mathbf{x}(t))$. The key difference, however, is in the interpretation of the model. For instance, $\mathcal{M}(y(t), \mathbf{x}(t))$ is not necessarily zero when $f(x(t), \mathbf{y}(t))$ is zero.

4.2. Spatial resolution

M&V can be performed at a coarse *spatial* resolution by regressing total building energy demand to some exogenous explanatory variables such as ambient conditions. Sometimes, however, it might be necessary to perform M&V at a finer spatial granularity, for instance, by disaggregating air-handling energy from fan energy consumption or electricity from natural gas consumption. Performing M&V at a finer spatial resolution might require regressing output variables to exogenous variables as well as to endogenous explanatory variables such as set-points and control variables of the particular spatial domain of interest. Similar to the time resolution case, predicting energy demands at higher spatial resolutions increases the number of explanatory variables and nonlinearity. GP can be used to capture these higher resolutions.

5. Case studies

We demonstrate the M&V capabilities of GP using two case studies. The first study is a simulation setting where we perform a regression error analysis of a control system using different explanatory variables and numbers of baselining points. In the second study, we use real data to demonstrate that GP can capture nonlinear trends more consistently compared with standard regression methods. In addition, we demonstrate that GP models can be used to exploit data at higher time resolutions.

5.1. Case study I

We consider a simulated scenario in which an advanced model predictive control (MPC) system is deployed at a typical office building and for which we seek to baseline its energy performance. The MPC system uses all the available degrees of freedom of the building to compute optimal operating policies in order to minimize energy while satisfying comfort and operational constraints. The abstract controller formulation has the form

$$\min_{\mathbf{u}(\cdot)} \quad \int_t^{t+T} y(\tau) d\tau \tag{5.24a}$$

$$\text{s.t.} \quad \frac{d\mathbf{x}}{dt}(\tau) = f_x(\mathbf{x}(\tau), \mathbf{z}(\tau), \mathbf{u}(\tau), \boldsymbol{\omega}(\tau)) \tag{5.24b}$$

$$0 = f_z(\mathbf{x}(\tau), \mathbf{z}(\tau), \mathbf{u}(\tau), \boldsymbol{\omega}(\tau)) \tag{5.24c}$$

$$0 \le g(\mathbf{x}(\tau), \mathbf{z}(\tau), \mathbf{u}(\tau), \boldsymbol{\omega}(\tau)) \tag{5.24d}$$

$$y(\tau) = \Xi(\mathbf{x}(\tau), \mathbf{z}(\tau), \mathbf{u}(\tau)) \tag{5.24e}$$

$$\mathbf{x}(t) = \text{given}, \quad \tau \in [t, t+T]. \tag{5.24f}$$

In this formulation, t is the current time, T is the prediction horizon, $\mathbf{x}(\cdot)$ are the dynamic states of the building (zone temperature, CO_2 and H_2O concentrations, building air mass), $\mathbf{z}(\cdot)$ are the algebraic states (relative humidity, supply air temperature), $\mathbf{u}(\cdot)$ are the degrees of freedom or controls (AHU electrical energy, ambient air flow, recirculation flow), and $\boldsymbol{\omega}(\tau)$ are disturbances such as occupancy heat and CO_2/H_2O loads and ambient conditions (CO_2, temperature, and relative humidity). Variable $y(\tau)$ represents the total AHU electrical energy in the cooling/heating system. The controller runs over a receding horizon of 24 h with time steps of 1 h. We used real weather data in the Chicago area to perform year-long closed-loop simulations (see Fig. 2). A detailed formulation of the controller is presented in [29,28].

Traditional control systems track fixed set-points during occupied times and set-back set-points during unoccupied times. The advantages of the energy-oriented MPC controller, Eq. (5.24), are that it adapts set-points according to prevailing internal and external ambient conditions, it can exploit the flexibility of the thermal and air quality comfort zone, and it can exploit physical multivariable interactions to enable the indirect control of output variables with nontraditional control variables. This approach has been shown to dramatically reduce energy consumption, reaching savings of over 30% [28].

We first demonstrate that GP modeling can be used to accurately predict the energy outputs of the MPC controller. This is of relevance because the MPC controller introduces a high degree of multivariable interaction. We evaluate the incorporation of different sets of explanatory variables: ambient temperature, relative humidity, and occupancy count. In addition, we incorporate the supply air temperature, which is an important variable (endogenous) optimized by the MPC controller to save energy. The controller is run over the entire year with a prediction horizon of 24 hours. We collected energy $y(t)$, ambient temperature $T_{amb}(t)$, occupancy $O(t)$, relative humidity $RH(t)$, and supply temperature $T_{sup}(t)$ data with time increments of one hour. A total of $T = 8736$ points was available, and we aggregated daily to generate $N = 364$ data points. In this first case study, we used $n = 122$ samples to train the GP model and used the rest for testing $m = 242$. This gives a baselining data ratio of $N/n = 3$. The baseline points are chosen at every N/n points along the domain $t = 0, \ldots, N$.

The GP model predictions *at the testing points* using T_{amb} as the only explanatory variable and the corresponding 95% confidence

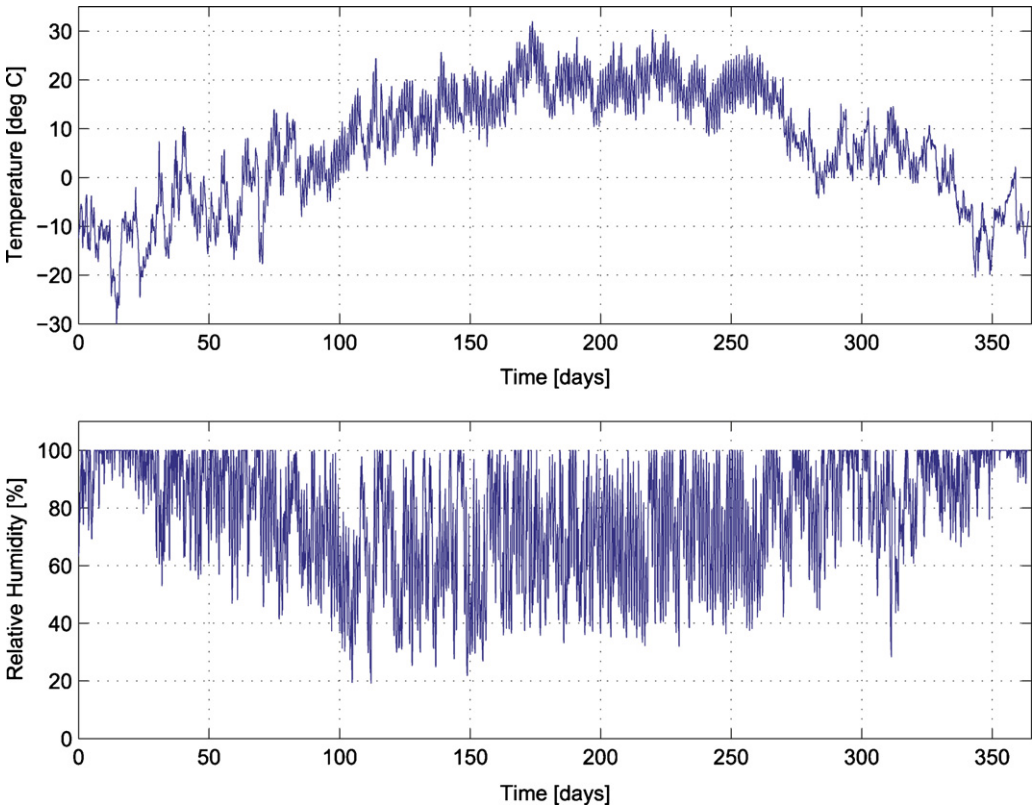

Fig. 2. Hourly ambient temperature and relative humidity profiles for Chicago area, 2006.

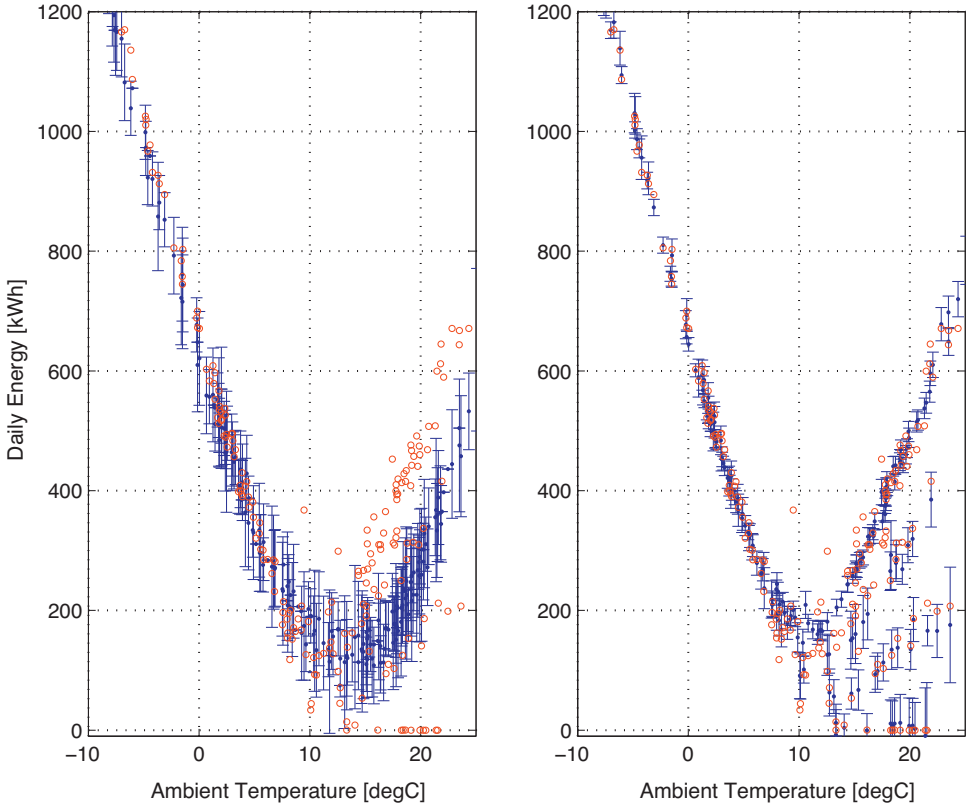

Fig. 3. Effect of explanatory variables on GP model predictions at testing points. Ambient temperature and occupancy are explanatory variables (left). Ambient temperature, occupancy, and supply temperature are explanatory variables (right).

Table 1
Summary of test predictions for different sets of explanatory variables.

Explanatory variables	$\epsilon^{SSE}_{test,m}$	$\lvert\epsilon_{test,m}\rvert$	$\pm 2\cdot\sigma_{test,m}$
T_{amb}	3.6×10^6	497	±4560
$T_{amb}+OC$	3.4×10^5	345	±1530
$T_{amb}+T_{sup}$	2.2×10^6	2123	±3980
$T_{amb}+OC+T_{sup}$	2.7×10^5	139	±1810
$T_{amb}+OC+RH_{amb}$	5.1×10^5	31	±2290
$T_{amb}+OC+T_{sup}+RH_{amb}$	3.6×10^5	633	±1980

levels $\pm 2\sigma_{test}$ are presented in the left graph of Fig. 3. We note that the energy profile follows a quadratic trend in the ambient temperature. However, a significant number of points exist outside the main trend that T_{amb} alone cannot explain. We also see that the uncertainty levels are inconsistent and seem to be underestimated. This is because the error is the result of a structural model inconsistency rather than noise (which is not added in this study). This observation implies that an additional explanatory variable is needed in the model. From Table 1, we see that the total SSE at the test points ($\epsilon^{SSE}_{test,m}$) for the basic GP model is 3.6×10^6. After the occupancy load is introduced as an additional variable, the SSE is reduced by an order of magnitude to a value of 3.4×10^5. If instead the supply temperature is added as an explanatory variable, then the SSE reduces only to 2.2×10^6. These results indicate that occupancy is the second most important explanatory variable and can explain the points outside the main trend. When the three variables are combined, the SSE is further reduced to 2.7×10^5. We also considered a case using ambient relative humidity instead of supply temperature. As can be seen from Table 1, both the model accuracy and the variance are better for the supply temperature model, thus indicating the importance of adding endogenous variables when operating conditions of the building are adapted dynamically as in MPC.

We present the predictions and confidence levels for the GP model using T_{amb}, OC, T_{sup} as explanatory variables in the right graph of Fig. 3. We can see that the fit of the variables outside the main curve is significantly better and that the uncertainty levels are consistent. We highlight that the GP model can factor occupancy directly into the model without having to split the model into a weekday and a weekend/holiday model, as is the standard practice. In other words, a single GP model captures all the multivariable interactions. In addition, we do not need to the specify structural relationships between variables. Thus, the modeling effort and the number of parameters needed are significantly reduced.

From the confidence values presented in Table 1, we see that incorporating occupancy in the model reduced the confidence interval significantly, from ±4560 to ±1530, but incorporating ambient relative humidity increased it to ±2290. This effect suggests that there is not enough variability in relative humidity in the training data to capture the complete range of interactions between the explanatory variables. To confirm this hypothesis, we increased the number of training points to $n=182$. The confidence level is ±1011 for the model using T_{amb}, OC and ±1162 for the model using T_{amb}, OC, RH_{amb}. As the operation regime is fully explored, the effect is much less pronounced. These results also illustrate the typical trade-off arising in modeling exercises, where adding model sophistication typically comes at the expense of additional data requirements. Moreover, the results highlight the fact that, in nonlinear models, accuracy and uncertainty levels are not necessary related, as suggested by deterministic uncertainty estimation techniques used in standard M&V practices. In the case presented here, in particular, the modeler needs to decide whether it is preferred to have a model with higher accuracy and reduced uncertainty levels at the expense of additional data.

In Table 1 we also present the absolute value of the total error in the testing points $\lvert\epsilon_{test,m}\rvert$ and the corresponding confidence intervals. As shown in Eq. (4.21), this value can be computed by aggregating the errors and covariances of all the points. The total energy consumed in the testing points is 1.4×10^5 kWh, and the energy savings for a typical *upgraded* MPC system are 10–30%, or $O(10^4)$ kWh. Consequently, the error and confidence levels are at least one order of magnitude smaller than the total savings, indicating that the model is accurate. As observed in Section 4, however, the error is small for all models even if the point predictions are off (reflected by a large value of $\epsilon^{SSE}_{test,m}$), as in the case of the model using only ambient temperature. This observation illustrates the danger of using the aggregated error as an indicator of model quality. A similar inconsistency is reflected for the $T_{amb}+OC+RH_{amb}$ model, which has a high value of $\epsilon^{SSE}_{test,m}$ but a very small value of $\epsilon_{test,m}$.

In Table 2 we illustrate the effect of local minima on the performance of the GP model. As can be seen, two entirely different models result from the different minima of the log likelihood function. The suboptimal model has more than twice the SSE of the optimal one, and the variance is underestimated. Thus, the total error lies outside the 95% confidence intervals (i.e., the variance is underestimated). In our study, we used a gradient-based algorithm to perform the search and initialized from different starting points to avoid this issue.

As our final study, we analyzed the effect of the number of training (baselining) points on the performance of the GP model. To do so, we consider the baselining ratio $N/n=1, 2, 3, 4, 5$. Because these ratios indicate that every N/nth point is used as a baselining point, the number of baselining points increases with decreasing N/n. The case with $N/n=1$ represents the maximum possible performance, where all points are used for baselining. The effect on the test SSE $\epsilon^{SSE}_{test,m}$ and variance $\sigma^2_{test,m}$ is plotted in Fig. 4 for the $T_{amb}+OC+RH_{amb}$ model. The results show that the variance is much more sensitive than is SSE to the number of baselining points. The variance changes, however, are always within the same order of magnitude in the range of 364–122 baselining points. If the number of baselining points is reduced below this threshold, the variance and SSE increase dramatically. This result is of critical importance from the point of view of M&V cost because it clearly shows that *accurate baselining models can be obtained with GP with significant reductions in the amounts of baselining data*. In our case study, for instance, this indicates that if 122 baselining points are used, an updated MPC control system can potentially be run in the remaining $N-n=m=242$ days. If the updated MPC control system saves an average fraction of α_{MPC} per day, then the total fraction of savings is $\alpha=(m/N)\alpha_{MPC}$. In our study, the total baseline energy demand is 1000 MWh, and the average savings per day for the upgraded MPC system is $\alpha_{MPC}=0.3$. If 364 training points are used, then $\alpha=0$, equivalent to 0 MWh of savings. If 182 points are used for baselining, then the savings are $\alpha=0.15$ or 150 MWh. If the number of baselining points is further reduced to 122, then $\alpha=0.2$ or 200 MWh, while the prediction accuracy and variance remain at similar levels. If we further decrease the number of training points, then the savings can go as high as 300 MWh, but the prediction uncertainty explodes. These results clearly illustrate the importance of having a better modeling capability that can enable the reduction of baselining points.

5.2. Case study II

We apply GP to verify energy savings obtained from a supervisory control system implemented at the Advanced Photon Source (APS) office building at Argonne National Laboratory. The building was opened in 1996, covers a total floor space of 190,000 sq. ft., and has a maximum occupancy of 430 occupants. The building is conditioned by two central AHUs and zone reheat systems (Fig. 5). The cooling coils in the two AHUs operate to meet cooling demands using chiller water from a central chiller plant, and reheat coils

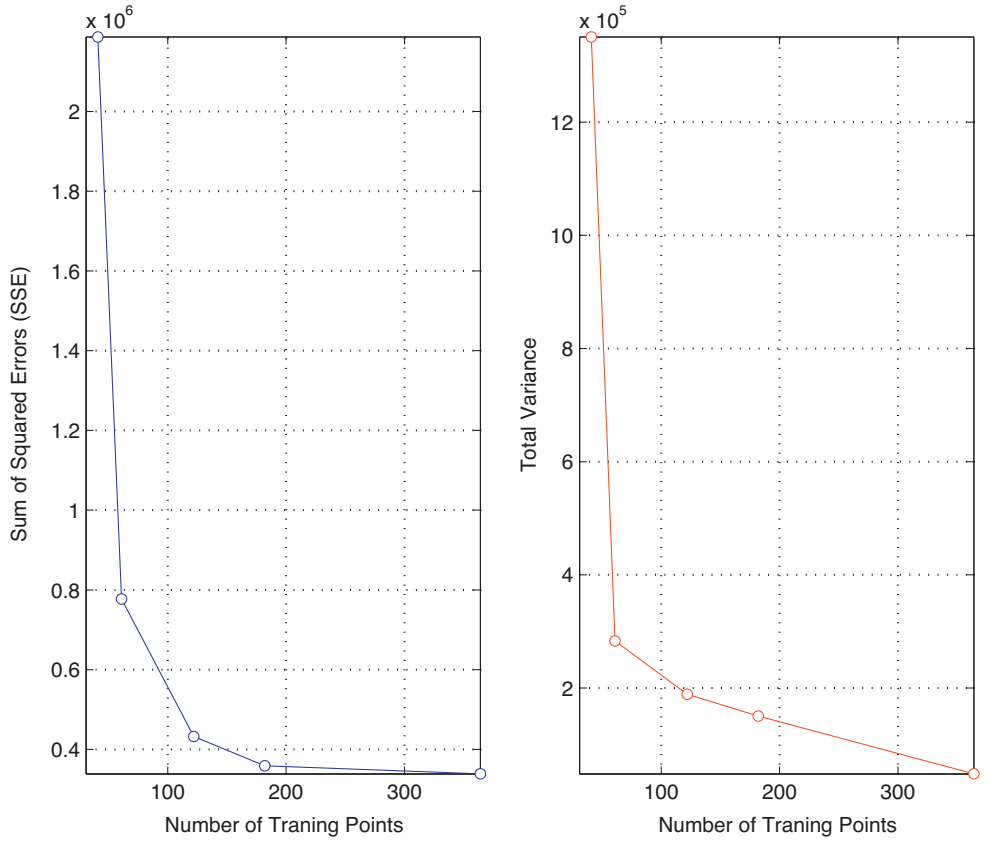

Fig. 4. Effect of number of baselining points on GP model prediction accuracy.

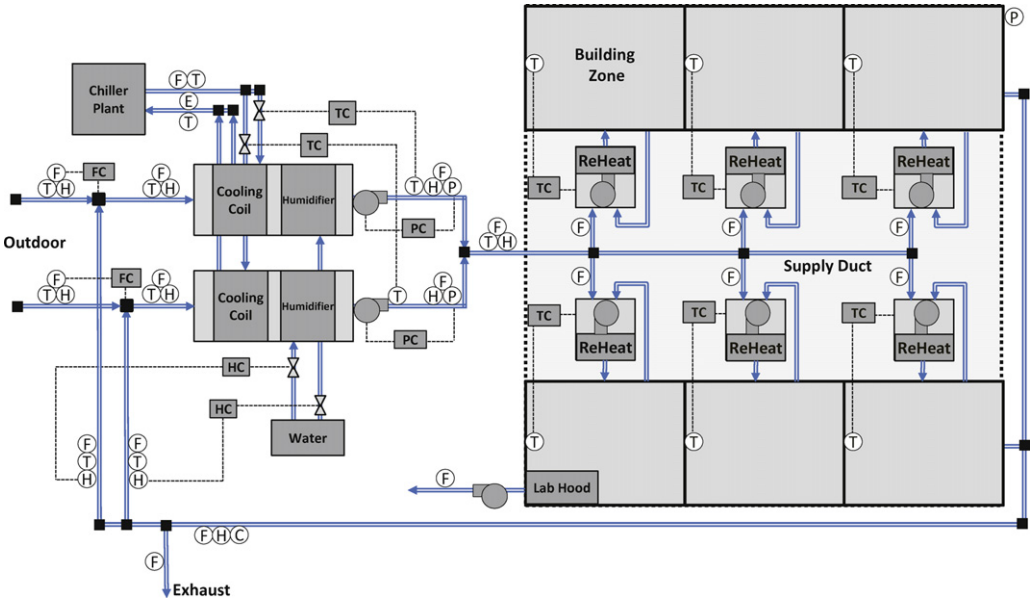

Fig. 5. Layout of HVAC system of Argonne's APS office building. Sensor labels are as follows: T = temperature, P = pressure, F = flow, C = CO_2 ppmV, and H = relative humidity. The control loops are indicated with boxes and dashed lines.

Table 2
Effect of local minima on GP model performance.

Explanatory variables	$\log p(\mathbf{y}\|\eta)$	ϵ_{test}^{SSE}	$\|\epsilon_{test}\|$	$\pm 2 \cdot \sigma_{test}$
$T_{amb} + OC + RH_{amb}$ (I)	6.92×10^2	5.1×10^5	31	±2290
$T_{amb} + OC + RH_{amb}$ (II)	7.67×10^2	5.8×10^5	1200	±586

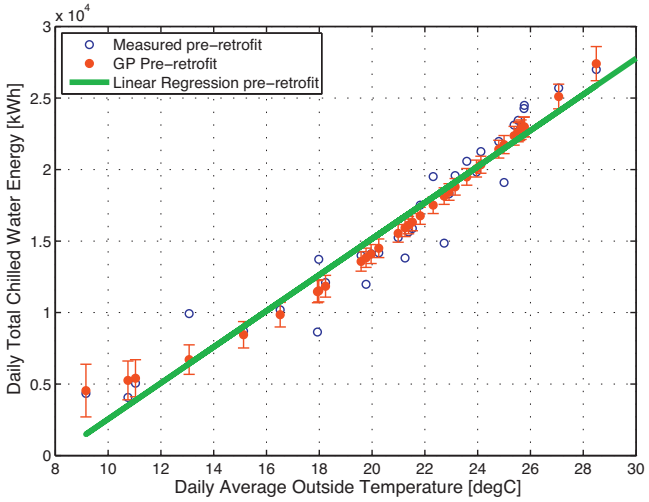

Fig. 6. Comparison of pre-retrofit models for chilled-water energy use.

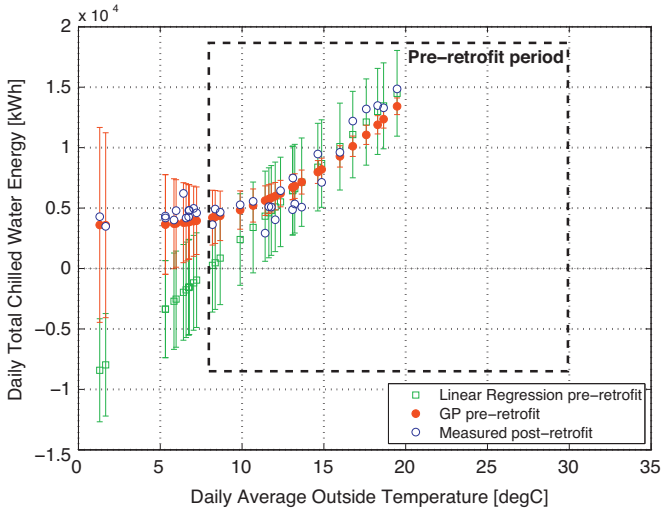

Fig. 7. Baseline model adjusted for the chilled-water energy during the post-retrofit period.

in the terminal units meet heating demands in individual spaces. A supervisory MPC control system has been recently installed to optimize supply air temperature and static pressure set-points in order to minimize energy demands while meeting thermal comfort criteria. We have submetered chilled-water energy consumed for the building and operation state variables in the AHUs and outdoor temperature conditions from the existing control system since August 2011. We consider 37 baselining days (weekdays) in a period spanning the months of July–August (previous to deployment) and 36 post-retrofit days (also weekdays) spanning the months of September–December (after the supervisory system was brought on-line). Given the sparse, limited data available, it is critical to understand the impact of data on uncertainty levels. This information will help us design suitable M&V strategies for the new control system and understand the effect of ambient conditions and operations on the savings uncertainty.

We aggregated 5-min-resolution data on chilled-water energy into daily total chilled-water energy. We developed a GP model that predicts daily total chilled-water energy as a function of the daily average outdoor temperature on the basis of the measured data during the pre-retrofit period. On the basis of the same data set, we also developed a linear regression model, following ASHRAE's Guideline 14. Fig. 6 compares the daily baseline energy use estimated by the two methods. We can see that the linear regression model captures the energy use in this temperature regime fairly well. The GP model, however, captures the slightly nonlinear trend at the extremes of the temperature domain more effectively. Furthermore, we can see that the GP model predicts narrower confidence intervals for the temperature range between 15 °C and 28 °C in which most of data points are available, and the confidence levels incrementally increase toward the tails. These results reflect consistency of the uncertainty levels and adaptivity to data availability, as opposed to the levels obtained with linear regression, which remain constant throughout the domain. One important caveat of the GP model, however, is that the uncertainty levels are underestimated at several points and thus several points are outside the main trend. We attribute this situation to variations in occupancy which are quite strong in this building. The first case study confirmed through simulations that the GP uncertainty levels can be underestimated because of the lack of occupancy information (see left graph in Fig. 3). Unfortunately, since occupancy estimates are not available, the only way to capture occupancy effects is to incorporate this as an input random variable. Our current GP framework is not capable of doing so, and this remains an important topic of future research.

The two pre-retrofit models were applied to predict the baseline energy outputs during the post-retrofit period. Fig. 7 presents the results. The dashed box indicates the temperature domain used to train the pre-retrofit model. As can be seen, the post-retrofit and pre-retrofit periods span different temperature domains, with an overlapping domain in the range of 8–20 °C. We see that the linear regression model does not capture the nonlinear behavior outside the pre-retrofit domain, and its predicted baseline is not reliable. With more data points below 15 °C, it would have been possible to identify a threshold point for different energy behavior and derive a piecewise linear model. Given the limited data points between 10 and 30 °C, however, it is difficult to observe the threshold point. We also see that the linear regression uncertainty estimates remain constant and wide in the post-retrofit domain. The GP model, on the other hand, yields more consistent behavior with wider confidence intervals for points outside the pre-retrofit domain. The M&V analysis based on the GP model suggests that the implemented control system shows statistically significant energy savings for a handful of days only. This result is expected because most of the chilled-water savings obtained with the control system occur on days with temperatures above 25 °C, a regime where the system has not yet been operated.

We developed a GP model that predicts hourly energy outputs to answer the following questions: Can GP capture hourly dynamic energy behavior? And, can finer-resolution data help reduce uncertainty in the analysis results? We used 862 hourly data points during the same pre-retrofit period and developed the hourly GP model as a function of outdoor temperatures at times t, $t-1$, and $t-2$ (lag of two time steps). To capture intraday occupancy variations, we regressed the energy output against a typical CO_2 profile at time steps t, $t-1$, and $t-2$. The average CO_2 profile is displayed in Fig. 8. We note that this is equivalent to using time-of-the-day as a surrogate input variable in order to capture occupancy variations.

Fig. 9 compares hourly energy predictions with measured values for one week of baseline operation. The results indicate that the GP model captures higher-resolution trends. We attribute the prediction errors to intraday occupancy variations from the average profile. Fig. 10 shows the daily total energy outputs aggregated from the hourly predicted energy outputs compared with those from the daily model for the post-retrofit period. The daily energy outputs aggregated from the hourly data have much smaller confidence intervals than those from the daily data. The comparison demonstrates that we can reduce uncertainty in predictions by fully

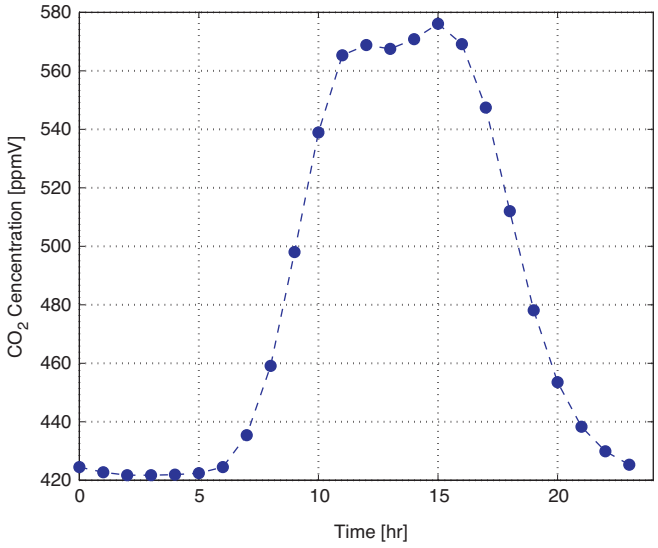

Fig. 8. Average CO_2 profile for weekdays.

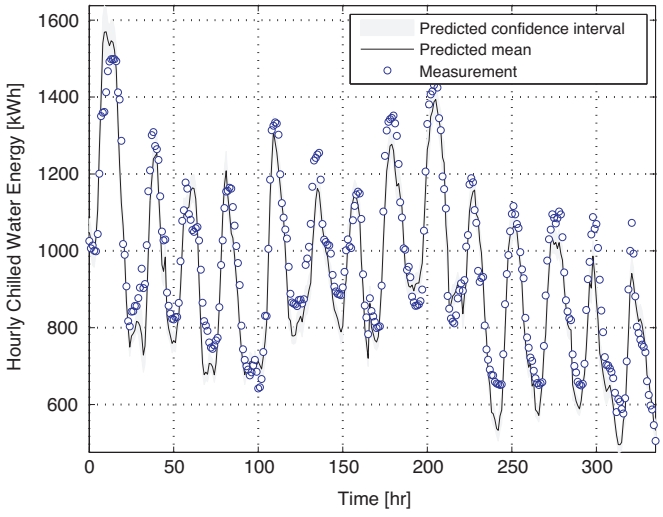

Fig. 9. Hourly pre-retrofit model predictions during the first week of pre-retrofit period.

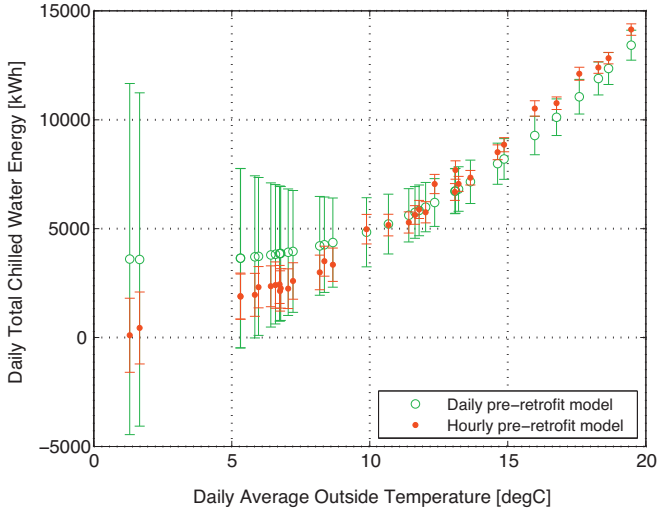

Fig. 10. Hourly pre-retrofit model predictions during post-retrofit period.

exploiting variations captured in the hourly data instead of the daily data when the measured data are limited.

6. Conclusions and future work

Measurement and verification are key to demonstrating whether guaranteed energy savings are achieved by deployed energy efficiency measures. Existing modeling practices used in M&V cannot accurately capture complex trends of energy behavior and cannot exploit limited and sparse data efficiently. Hence, M&V processes are lengthy and expensive and can lead to inaccurate uncertainty estimates that misguide decision-makers. To overcome these limitations, we have presented a Gaussian process (GP) modeling framework for M&V. We have demonstrated that GP models can capture complex behavior, including nonlinearities, multivariable interactions, and time correlations. In addition, because GP models are derived in a Bayesian setting, they can more naturally fuse different sources of uncertainty.

The GP modeling framework will be expanded to handle the following issues: (1) fusing input measurement uncertainty in explanatory variables (e.g., occupancy) and uncertainty due to model inadequacy, (2) handling multiple correlated outputs, and (3) managing large amounts of data. The GP strategy presented here estimates uncertainty levels by obtaining probability distributions for hyper-parameters of the covariance matrix while taking into account output measurement errors. It assumes, however, that explanatory variables are deterministic. The current framework handles a single output variable, so it can develop only an individual GP model per energy output measured at a system. This limitation, for instance, forces the use of ad hoc recursions for multistep predictions. Finally, the memory requirements and computational time involved in the solution of the maximum-likelihood problem increase cubically with the number of baselining points, since the training covariance matrix needs to be stored and factorized. To deal with this situation, we will investigate the use of matrix-free approaches to handle massive amounts of data.

Acknowledgment

This work was supported by the U.S. Department of Energy, under Contract No. DE-AC02-06CH11357.

References

[1] M. Anitescu, J. Chen, L. Wang, A Matrix-free Approach for Solving the Gaussian Process Maximum Likelihood Problem. Preprint ANL/MCS-P 1857-0311, Argonne National Laboratory, 2011.
[2] ASHRAE, ASHRAE Guideline 14: Measurement of Energy and Demand Savings, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc, Atlanta, GA, 2002.
[3] C.M. Bishop, Pattern Recognition and Machine Learning, New York, Springer, 2006.
[4] G.E.P. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[5] E.M. Constantinescu, M. Anitescu, Physics-based Covariance Models for Gaussian Processes with Multiple Outputs, Preprint ANL/MCS-P 1915-0711, Argonne National Laboratory, 2011.
[6] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley-Interscience, New York, 1998.
[7] Federal Energy Management Program. Energy Savings Performance Contracts. http://www1.eere.energy.gov/femp/financing/espcs.html, 2012.
[8] FEMP. M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 3.0, Technical report, U.S. Department of Energy Federal Energy Management Program, 2008.
[9] S.J. Hansen, J.W. Brown, Investment Grade Energy Audit: Making Smart Energy Choices, Fairmont Press, Lilburn, GA, 2004.
[10] S. Haykin, Kalman Filtering and Neural Networks, John Wiley & Sons, New York, 2001.
[11] Y. Heo, G. Augenbroe, R. Choudhary, Risk analysis of energy-efficiency projects based on Bayesian calibration of building energy models, in: Building Simulation 2011, Proceedings, 2011, pp. 2579–2586.

[12] Y. Heo, R. Choudhary, G.A. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, Energy and Buildings 47 (2012) 550–560.
[13] IPMVP. International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings, vol. 1, Technical Report, Efficiency Valuation Organization, 2010.
[14] J. Jackson, Promoting energy efficiency investments with risk management decision tools, Energy Policy 38 (2010) 3865–3873.
[15] H.K.H. Lee, D.M. Higdon, C.A. Calder, C.H. Holloman, Efficient models for correlated data via convolutions of intrinsic processes, Statistical Modeling 5 (1) (2005) 53–74.
[16] P.A. Mathew, J.S. Kromer, O. Sezgen, S. Meyers, Actuarial pricing of energy efficiency projects: lessons foul and fair, Energy Policy 33 (10) (2005) 1319–1328.
[17] E. Mills, Risk transfer via energy-savings insurance, Energy Policy 31 (3) (2003) 273–281.
[18] E. Mills, S. Kromer, G. Weiss, P.A. Mathew, From volatility to value: analysing and managing financial and performance risk in energy savings projects, Energy Policy 34 (2) (2006) 188–199.
[19] C.J. Paciorek, M.J. Schervish, Nonstationary covariance functions for Gaussian process regression, in: Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference, Bradford Book, British Columbia, Canada, 2004, p. 273.
[20] C.E. Rasmussen, C.K. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, MA, 2006.
[21] T.A. Reddy, D.E. Claridge, Uncertainty of "measured" energy savings from statistical baseline models, HVAC&R Research 6 (1) (2000) 3–20.
[22] S.K. Sahu, S. Yip, D.M. Holland, Improved space-time forecasting of next day ozone concentrations in the eastern US, Atmospheric Environment 43 (3) (2009) 494–501.
[23] A. Satchwell, C. Goldman, P. Larsen, D. Gilligan, T. Singer, A Survey of the U.S. ESCO Industry: Market Growth and Development from 2008 to 2011. Technical report, Lawrence Berkeley National Laboratory, 2010.
[24] K. Soratana, J. Marriott, Increasing innovation in home energy efficiency: Monte carlo simulation of potential improvements, Energy and Buildings 42 (2010) 828–833.
[25] M. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York, 1999.
[26] M. Stein, J. Chen, M. Anitescu, Difference Filter Preconditioning for Large Covariance Matrices. Preprint ANL/MCS-1888-0511, Argonne National Laboratory, 2011.
[27] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[28] V. M. Zavala Real-time Optimization Strategies for Building Systems. Preprint ANL/MCS-P 1911-0611, Argonne National Laboratory, 2011.
[29] V.M. Zavala Real-time Resolution of Conflicting Objectives in Building Energy Management: An Utopia-tracking Approach. Preprint ANL/MCS-P 2056-0312, Argonne National Laboratory, 2011.
[30] V. M. Zavala, M. Anitescu, T. Krause, On the optimal on-line management of photovoltaic-hydrogen hybrid energy systems, in: R.M. de Brito Alves, C.A. Oller do Nascimento, E.C. Biscaia Jr., (Eds.), Proceedings of 10th International Symposium on Process Systems Engineering, pp. 2579–2586.
[31] V.M. Zavala, E.M. Constantinescu, M. Anitescu, T. Krause, On-line economic optimization of energy systems using weather forecast information, Journal of Process Control 19 (2009) 1725–1736.