

Sensor-Based Predictive Modeling for Smart Lighting in Grid-Integrated Buildings

Chandrayee Basu, Julien J. Caubel, Kyunam Kim, Elizabeth Cheng, Aparna Dhinakaran, Alice M. Agogino, *Member, IEEE*, and Rodney A. Martin, *Senior Member, IEEE*

Abstract—Studies show that if we retrofit all the lighting systems in the buildings of California with dimming ballasts, then it would be possible to obtain a 450 MW of regulation, 2.5 GW of non-spinning reserve, and 380 MW of contingency reserve from participation of lighting loads in the energy market. However, in order to guarantee participation, it will be important to monitor and model lighting demand and supply in buildings. To this end, wireless sensor and actuator networks have proven to bear a great potential for personalized intelligent lighting with reduced energy use at 50%–70%. Closed-loop control of these lighting systems relies upon instantaneous and dense sensing. Such systems can be expensive to install and commission. In this paper, we present a sensor-based intelligent lighting system for future grid-integrated buildings. The system is intended to guarantee participation of lighting loads in the energy market, based on predictive models of indoor light distribution, developed using sparse sensing. We deployed ~60% fewer sensors compared with state-of-art systems using one photosensor per luminaire. The sensor modules contained small solar panels that were powered by ambient light. Reduction in sensor deployments is achieved using piecewise linear predictive models of indoor light, discretized by clustering for sky conditions and sun positions. Day-ahead daylight is predicted from forecasts of temperature, humidity, and cloud cover. With two weeks of daylight and artificial light training data acquired at the sustainability base at NASA Ames, our model was able to predict the illuminance at seven monitored workstations with 80%–95% accuracy. Moreover, our support vector regression model was able to predict day-ahead daylight at ~92% accuracy.

Index Terms—Clustering, daylight harvesting, inverse model, support vector regression, wireless sensor network.

Manuscript received March 2, 2014; revised August 13, 2014; accepted August 14, 2014. Date of publication August 26, 2014; date of current version October 21, 2014. This work was supported in part by the National Aeronautics and Space Administration, University of California at Berkeley, Berkeley, CA, USA, and in part by the California Energy Commission through the Energy Innovations Small Grant Program. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ashish Pandharipande.

C. Basu is with Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: basu.chandrayee@gmail.com).

J. J. Caubel, K. Kim, and A. M. Agogino are with the Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: jcaubel@berkeley.edu; knkim@berkeley.edu; agogino@berkeley.edu).

E. Cheng is with the Department of Computer Science and Applied Mathematics, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: elizabeth.y.cheng@gmail.com).

A. Dhinakaran is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: apamadhinak@gmail.com).

R. A. Martin is with the NASA Ames Research Center, Mountain View, CA 94035 USA (e-mail: rodney.martin@nasa.gov).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSEN.2014.2352331

I. INTRODUCTION

IBM's Instrumenting the Planet report [1] highlights the importance of wireless sensor-actuator networks and distributed analytics in the life cycle management of natural resources and technical infrastructures in agriculture, hydrological systems, land use, power grids, transportation systems, manufacturing and many more applications. The researchers introduced Real-World-Aware (RWA) systems, which extract information about the state of the real world from raw data aggregated from disparate sources and use it to complete the loop through automated and adaptive control. Cyber-physical systems are becoming pervasive in large infrastructures and are viewed as essential components of grid-connected buildings. Expert studies [2] show that if we retrofit all the lighting systems in the buildings of California with dimming ballasts, then it would be possible to obtain 450 MW of regulation, 2.5 GW of non-spinning reserve and 380 MW of contingency reserve from participation of lighting loads in the energy market. In some cities, such as Amsterdam, dimmable street LED's are integrated within their smart grid [3]. Ceriotti et al. [4] proposed wireless-enabled closed loop control for lighting in road tunnels. The advantage of controlling lighting loads is that they can be controlled to any intensity with dimming ballasts, unlike HVAC systems. Furthermore, low latency makes the dimmable lights competitive with generators, which have over one minute response time.

Wen et al. (2011) [5] found that closed loop control of building systems enabled by wireless sensor and actuator networks (WSANs) result in 28% cooling energy and 40% light energy savings in office buildings. Commercial lighting is one of the largest contributors to commercial energy consumption. Intelligent lighting forms an easy and low-cost avenue to energy conservation. According to the U.S. DOE Energy yearbook in 2010 [6] the maximum electricity consumption in commercial buildings (13.6%) is attributed to lighting. Our prior work has demonstrated that even without daylight harvesting (controlling artificial lights based on daylight availability), 50% of lighting energy can be saved from personalized control of wireless-enabled individually- dimmable luminaires. An additional 20% of energy savings could be achieved with daylight harvesting according to our simulation results [7]–[9]. Furthermore, there have been considerable improvements in lighting and shading controls [10] and in daylight harvesting systems [11], [12]. Singhvi et al. (2005) [13] developed a centralized lighting

system to increase user comfort and reduce energy costs by using a WSN. Lin *et al.* (2005) [14] proposed a decentralized algorithm for WSN-enabled optimal lighting control.

In spite of the growing impetus in lighting control research and some successful pilot projects, the actual adoption of intelligent lighting control systems in commercial buildings has been very limited. As of 2010, 70% of the US national stock of commercial buildings had no lighting controls for energy efficiency [15]. Some of the reasons include general lack of encouraging energy savings from expensive commissioning of lighting systems, particularly when usability was not considered appropriately. Rude [16] found that 50% of the intelligent lighting control systems they studied had been deactivated by the users and the remaining 50% operated at 50% of target performance. System usability problems include lack of interoperability between lighting, shading and building automation system drivers, software and databases.

II. CONTRIBUTIONS

In this paper we present an intelligent lighting system for future grid-integrated buildings with the following contributions:

1. Development of computationally inexpensive and data-driven predictive indoor light models for intelligent lighting control and smart grid integration.
2. Approximately 60% reduced sensing compared to state-of-art closed loop lighting control systems.
3. Design, development and testing of indoor light powered light sensor platform.

Our research is motivated by the driving need to increase adoption of wireless enabled intelligent lighting systems for building to grid integration and draws upon the expanding field of research in optimal sensing systems. The state-of-art commercial lighting control systems typically use one photo-sensor per luminaire or even 2-3 photo-sensors (to allow for redundancy), as shown in Wen & Agogino [7]. We demonstrated the proposed WSN platform could accurately estimate indoor light on work surfaces and perform day-ahead predictions for demand response with approximately 40% fewer sensors compared to these state-of-the-art commercial systems.

Reduced sensing is achieved by replacing many of the actual wireless sensor platforms by sensor inverse models, thereby reducing the cost of sensor deployments. These models are point estimates of indoor light in the form of clustered linear functions of measured daylight and artificial lights. Clustering captures the potential changes in spatial correlations in the light field, resulting from the physics of direct and diffuse light-distribution in space under varying sky conditions.

The advantage of mounting photo-sensors on the luminaires is that power supply to the sensing unit comes from the same circuit as the luminaire. However, this mounting position may lead to incorrect estimation of illuminance on the workplace due to large field of view of the photo-sensor. Our system circumvents this limitation by positioning photo-sensors on a few critical workstations supported by our adaptive regressor selection algorithm. The sensing units have miniature photovoltaic panels for harnessing power from indoor light. As part of our ongoing research on information-centric smart building control

systems, we deployed and tested the integrated hardware-software platform at the Sustainability Base at NASA Ames Research Center.

III. RELATED WORK

Maasoumy *et al.* (2013) [17] co-designed a coupled HVAC control algorithm and a temperature sensor system, optimized for energy and infrastructure cost, while meeting the occupant comfort needs. They observed that predictive control algorithms for optimal comfort and cost performances should be tailored differently to take into account sensor accuracy (represented by sensor position and number). In terms of temporal data density, Wen (2008) [9] and Singhvi *et al.* (2005) [13] demonstrated that sampling rates could be varied without compromising the control system performance, based on whether the light field is static or dynamic. Hence, reducing the number of sensors comes with an accuracy penalty. This can be mitigated by optimally selecting the spatial and temporal sampling frequency that adequately covers the indoor light field and maintains desired information accuracy. It is also important to define the desired information accuracy for user satisfaction and energy savings.

Many of distributed sensing applications, in particular for large infrastructures, face resource scarcity for which optimal sensor placement solutions have been proposed by researchers [18]–[20]. Most of these problems involve reverse engineering, where sensing parameters like position and sampling rate are changed based on feedback about the field. Such methods have been generalized for a wide range of applications. For example, near-optimal sensor placement algorithms using mutual information (MI) criteria assumes a Gaussian Process model of spatial distribution of environmental variables. This is essentially a sub-set selection problem (from all possible sensor locations) that maximizes the MI between the actual environmental variables (hidden variables) and the observed sensor readings. This method uses sub-modularity of MI criteria for obtaining at least a $\sim 63\%$ approximation of the optimal solution. One advantage of MI is that it can address non-linearity in spatial relationships of physical quantities. This algorithm was also validated for active sensing (e.g., changing sampling rates for battery life) as part of an intelligent lighting system.

Compressed sensing [21] is another alternative approach for reduced sensor deployment. It leverages the sparsity or redundancy of measured variables across the field, but requires prior knowledge of sparsity and randomized measurements. Compressed sensing has been mostly tested in audio and image acquisition. Sandhu *et al.* (2004) [22] proposed a Multi-Agent System (MAS) for distributed data processing and Influence Diagram (Bayes net)-based decision-making in closed loop lighting control. The main goal was to achieve flexibility of distributed computation. Sensor placement problems can be cast into the MAS framework, in which individual sensors are modeled as agents with a supervisory algorithm to minimize the average prediction error across the spatially distributed agents.

A. Guillemin (2003) [23] and D. Lindelhof (2007) [24] have proposed and validated a predictive model of light

that assumed a linear relationship between vertical facade illuminance and indoor horizontal illuminance. Lindelhof [24] found that his linear predictive model resulted in a standard deviation of 416 lux (close to standard illuminance in offices). The same authors also found that performance of the predictive model varied with exposure to direct sunlight. Direct sunlight falling on a sensor is primarily responsible for the non-linear relationship between the sensed facade light and the sunlight distributed indoors. Ongoing research at the Lawrence Berkeley National Laboratory has shown that it is possible to predict the indoor light distribution in space as a linear function of one or two photo-sensor readings with reasonable accuracy in diffuse daylight conditions, for example, when blinds are drawn. But such correlations change rapidly when direct sun enters the space. Hence, it is important to train different models for direct and indirect sunlight conditions. In order to account for the temporal nature of the daylight distribution in space, we proposed a piecewise linear relationship between artificial and natural light sources and the illuminance measured at a workstation, discretized by one degree solar altitude for daylight approximation in our prior research [25]. We refer to this algorithm as the Sun Position-Based Model.

In terms of hardware related work on WSN enabled intelligent lighting system, Pandharipande et al. (2013) [26] proposed a wireless sensing system and a closed-loop illuminance feedback control algorithm for indoor lighting control. They showed that the wireless sensing system consisting of low-power, light energy harvesting sensor modules can be effectively used to provide illuminance measurements to the controller. However, the mounting locations and orientations of the sensor modules had to be limited in such a way that they are maximally exposed to ambient light to harvest sufficient energy for running the modules and charging the energy storages to be used in low-light condition. Our work proposes a wireless sensor network of a similar scheme, but with sensor modules that are not restricted in their placement and have extended operation time under low-light condition.

IV. SYSTEM ARCHITECTURE

Figure 1 provides a flowchart of our system architecture decomposed into the software (above) and WSN hardware (below) components. The WSN consists of two major components: remote light sensors and a base station with a central radio receiver and computer. Remote light sensors are stationed at selected workstations throughout the indoor space, and transmit local illuminance data to the base station using radio transceivers. The central base station receiver relays the data to the base computer through a serial port, and the data is stored locally in an SQLite database.

Once the data are collected from the WSN, the data processing modules are called for regularizing the matrix dimensions, eliminating zero illuminance readings during daytime, eliminating redundant data, and smoothing. The database stores the illuminance readings by mote number, Unix-time stamp (primary key), date and clock time, sky condition at the nearest weather station, solar altitude and azimuth, cluster ID. The software modules, written in Python, include

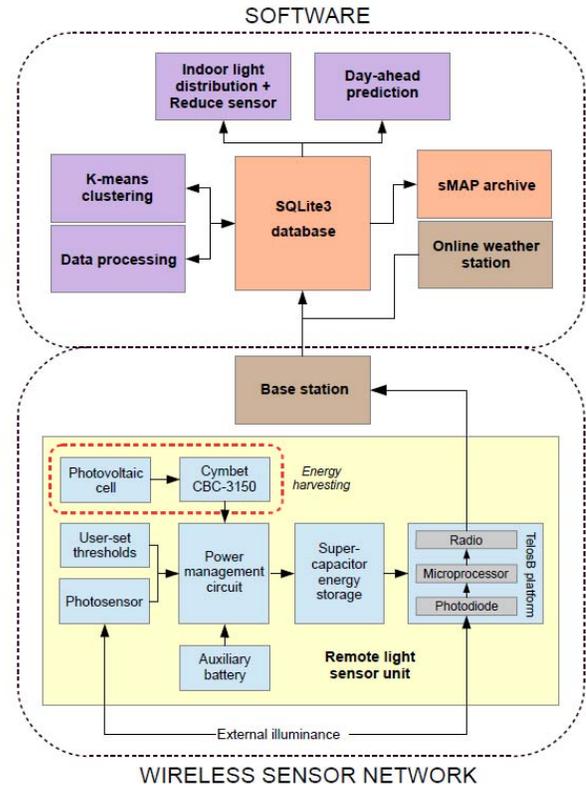


Fig. 1. System architecture showing hardware and software components.

a database driver, facade orientation prediction, a sun position calculator and programs for clustering, data processing, indoor light distribution and day-ahead prediction. The hourly sky conditions, temperature and relative humidity forecasts are called using Wunderground API. Solar altitude and azimuth are calculated using the Astronomer Almanacs solar position algorithm [27]. The same driver module is also used to forward the illuminance readings to an online database following a Simple Measurement and Actuation Profile (sMAP). sMAP was developed by UC Berkeley as a single web based platform for accessing large volumes of data from all possible sensor points from a multitude of disparate and distributed data sources such as building management systems [28], [29]. We will describe the light powered WSN platform in detail in section V, and discuss the components of the software in section VI

V. LIGHT POWERED WIRELESS SENSOR NETWORK (WSN)

The Wireless Sensor Network (WSN) is a deploy-and-forget illuminance data acquisition system optimized for low energy operation in indoor spaces. The WSN consists of remote light sensors that are primarily powered by small photovoltaic (PV) cells, harvesting ambient light energy in the indoor space to collect and transmit local light intensity readings.

A. Remote Light Sensor Unit

The WSN remote light sensors (see Figure 2) are centered on the TelosB platform, an open-source microprocessor-based



Fig. 2. Remote light sensor with Sanyo AM-1815 photovoltaic cell.

remote sensing platform developed at UC Berkeley. Illuminance data is collected by the TelosB’s onboard Hamamatsu S-1087 photodiode and transmitted to a base station receiver via the IEEE 802.15.4 layer over a five-minute duty cycle. Each remote sensor is fitted with a Sanyo AM-1815 photovoltaic (PV) cell to harvest ambient light energy in the indoor space. The energy harvesting system centers on the Cymbet CBC-3150 energy management module to regulate electrical power generated by the PV cell. In addition to the energy harvesting system, the units have an auxiliary battery to facilitate system start up and ensure reliable operation in low light conditions (see Figure 1).

The TelosB platform’s microprocessor is programmed using the open source TinyOS software. When using the default open-source code available from TelosB, it was found that the remote light sensor consumes 0.849 ± 0.003 mW of power when in the sleep state and 54 ± 3 mW during a 100 ms data transmission period. This equates to a total energy consumption of 260 ± 1 mJ over the nominal five-minute duty cycle.

The open source TinyOS code was modified to minimize the remote light sensor’s energy consumption. The data transmission period was reduced from 100 ms to 40 ms, the MCU clock speed was reduced from 4 MHz to 1 MHz, and an internal power-saving configuration was used to disable the TelosB microprocessor during the sleep portion of the duty cycle. Using these modifications, the TelosB platform consumes 0.2019 ± 0.0003 mW in the sleep state, and a maximum of 45 ± 3 mW during the 40 ms data transmission period. Given these performance characteristics, the platform uses 62.4 ± 0.2 mJ of energy over the five-minute duty cycle, roughly a 75% reduction from the original configuration, without noticeable effects on data transmission range or reliability.

B. Ambient Light Energy Harvesting

Studies have shown that ambient light energy harvesting can be suitably employed to power wireless sensor networks [26], [30]. For a light sensor platform, energy harvesting from ambient light is a natural choice. In order to assess the feasibility of this method, the power output of the Sanyo AM-1815 was characterized at several light intensity levels,

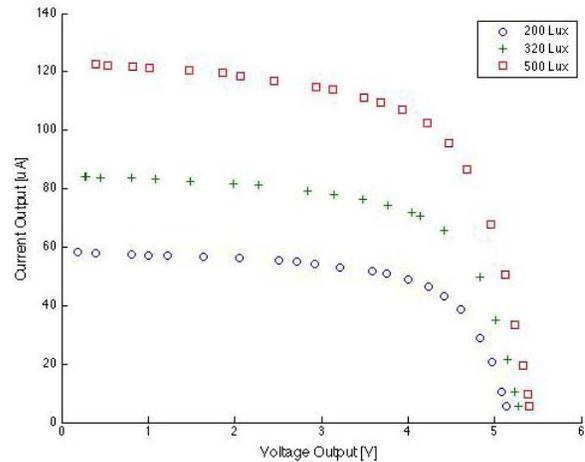


Fig. 3. PV cell power output at different controlled light levels.

TABLE I
MAXIMUM PV CELL POWER OUTPUT

Incident Illuminance (lux)	Maximum power output (mW)
200	0.196 ± 0.004
320	0.292 ± 0.004
500	0.433 ± 0.004

as shown in Figure 3. For each case, the PV cell was exposed to a constant illuminance under varying electrical loads. The electrical output of the cell was recorded at each load point, and the resulting performance curves are shown in Figure 3. Figure 3 shows that every illuminance level, a load point exists which maximizes the PV cell’s power output. Table I shows that the maximum output of the PV cell at an incident illuminance of 200 Lux is roughly equal to the TelosB’s power consumption in the sleep state. At the OSHA mandated minimum indoor workspace illuminance of 30 foot-candles (~ 320 Lux) [31], the maximum power output of the PV cell exceeds the TelosB’s power consumption in the sleep state by 0.090 ± 0.004 mW. This excess power is stored in the system’s super-capacitor, providing 1.8 ± 0.2 mJ of energy required during the TelosB’s data transmission period. At an incident illuminance of 320 Lux, the capacitor takes about 20 seconds to store the required transmission energy, and 166 seconds to charge to a maximum capacity of 14.9 ± 0.7 mJ at the nominal operating voltage of 3V. In this way, the energy generated by the PV cell and stored by the super-capacitor over the remote light sensor’s 5-minute duty cycle is well within the TelosB platform’s energy consumption requirements.

However, these experiments also showed that the PV cell’s power output is extremely susceptible to changes in both lighting conditions and electrical load, demonstrating the need for a management circuit to regulate and maximize this fluctuating output. The CBC-3150 module is equipped with an impedance matching function that varies the load on the PV cell to maximize the power output. This impedance matching function optimizes the PV cell’s power output with fluctuating incident illuminance. The CBC-3150 subsequently

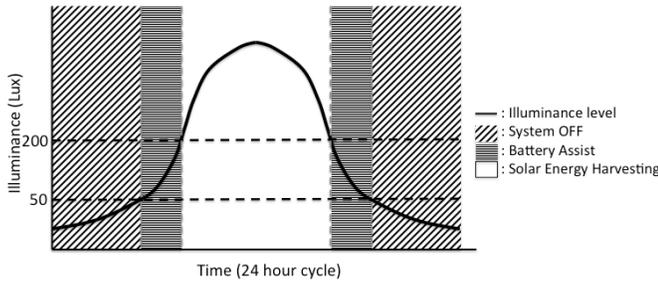


Fig. 4. Remote light sensor power management.

regulates this optimized PV power to maintain a maximum output voltage of 3.3 V to the TelosB platform.

C. Auxiliary Battery

A coin cell battery is added to the aforementioned energy harvesting system to address two major shortcomings, which we address next. Firstly, the system ceases to operate when exposed to illuminance levels below 200 Lux. At this incident illuminance, the power generated by the PV cell is roughly equal to the power consumption of the TelosB platform in the sleep state. Consequently, no excess energy can be stored in the system's super-capacitor over the sleep portion of the duty cycle to power the TelosB platform during the data transmission period. For the purpose of collecting data during building occupancy, it was deemed necessary to operate the remote sensors at a minimum illuminance of 50 Lux. Below this illuminance level, the space is deemed too dark for occupancy, and data collection is no longer required.

Secondly, the energy harvesting system was identified as having difficulties "waking up" following extended periods of complete shutdown, typically overnight. When the remote light sensor's TelosB platform and Cymbet CBC-3150 initially boot up, they require a surge in power to initialize various systems. It was found that the PV cell was typically unable to energize the super capacitor to the levels required to overcome this boot up surge until illuminances reached about 500 Lux. This often led to the remote light sensors remaining non-functional until late morning or early afternoon. These two observations led to the conclusion that an auxiliary battery was required to enhance the system's operational reliability.

Adhering to these requirements, auxiliary battery power should only be provided to the TelosB platform during periods when the incident illuminance is between roughly 50 and 200 lux, as illustrated in Figure 4. A window comparator enables auxiliary battery power to the platform when the voltage generated by a photo-resistor resides within a defined range. The lower and upper thresholds of this voltage range are initially calibrated to correspond to an incident illuminance of 50 and 200 Lux, respectively. In this manner, the auxiliary batteries both extend the effective data collection period, and provide the energy surge required to boot up the sensors in low light conditions. If needed, users can adjust the thresholds of the voltage range using potentiometers, to control the illuminance range during which auxiliary battery power is enabled. This feature allows users to control the sensor's

TABLE OF SYMBOLS

Symbols	Description
x_t	Raw illuminance reading at time step t
y_t	30-minutes moving average illuminance at time step t
w	Any workstation or illuminance sensor at the test bed
W	Number of workstation or illuminance sensors deployed
\mathcal{W}	Set of all illuminance sensors
J	Total number of 30-minutes interval in a day between 7 am and 7 pm
M	Total number of days during which illuminance data is collected
JK	Total number of clusters for which each of which a regression model is generated
z_j	2 D vector of mean and standard deviation of measured daylight at the window for the j^{th} 30-minutes interval of a day
y_w^{jk}	30-minutes moving average illuminance at any workstation w for the jk^{th} cluster
\hat{y}_w^{jk}	Predicted illuminance at workstation/ sensor at any workstation w for the jk^{th} cluster
C_j	Set of all clusters in the j^{th} 30-minutes interval of a day
c_{jk}	ID of each jk^{th} cluster
μ_{jk}	Centroid of each of the jk^{th} cluster
A	A set of all possible combinations of sensors $c \in \mathcal{W} \setminus w$
r	Number of regressor sensors at every iteration of the adaptive regressor selection algorithm varying between 1,2,3...0.5* W
e_s	Set of artificial light statuses used a features for regression
α	Set of weights for measured illuminance regressors
β	Set of weights for the artificial light statuses

boot up and shut down threshold, and modulate the period over which energy harvesting is enabled. Moreover, this flexible power management system enables users to easily configure the sensor units to operate efficiently in a wide variety of locations and incident illuminance levels.

Testing over two months showed that the sensor units were typically shut down for roughly 12 hours a day, using auxiliary battery power four hours a day, and harvesting light energy for eight hours a day. Given these performance characteristics, the system had a daily current consumption of 0.55 ± 0.09 mAh at a nominal voltage of 3V. The remote light sensor utilizes a CR2032 lithium battery, with a capacity of 240 mAh at 3V, allowing the sensor to operate over a year before requiring battery replacement. It should be noted, however, that the performance of the system is entirely reliant on ambient illuminance levels and the auxiliary battery management thresholds set by the user.

VI. SOFTWARE

A. Data Processing

The raw light data can be noisy due to dropped packets, redundant communication between the receiver and the sender nodes and low sensor accuracy. Other errors may stem from sensors that are shadowed or covered due to human activities or due to battery power drainage. Such errors must be handled

with sensor validation algorithms prior to basic data processing. The patterns in the data generated by each of these errors could be simulated and labeled for comparison with future data. Alternatively, the error patterns could be learned when the lighting system is running. We chose the latter to avoid intervention in real buildings.

For sensor validation we proposed a tolerance based on the 30-minutes moving average of light data. Empirical studies with the daylight data showed that a moving average over a 30-minute time window was able to capture the trend in daylight change due to changing sun position, without over-smoothing the data. Let x_t be the illuminance reading at current time step t . In place of raw illuminance data x_t we use 30-minutes moving average until time step t , y_t (see equation 1) as the input data for the regression models. Prior to the above we correct for erroneous raw data. If the difference between x_t and x_{t-1} is greater than the difference between y_{t-1} and x_{t-1} by a threshold percentage, chosen as a function of the illuminances, then x_t is assumed to be erroneous and replaced.

$$y_t = \frac{\sum_{n=0}^5 x_{t-n}}{6} \quad (1)$$

If $(x_{t-1} - x_t) > (f(x_{t-1} - y_t))(x_{t-1} - y_t)$, then replace (2)

$f(x_{t-1} - y_t)$ in rule (2) is a function of $(x_{t-1} - y_t)$, determined iteratively. The erroneous reading is replaced by illuminances from the same 30 minutes interval, averaged over the past seven most similar days. The distance metric used to compute similarity between one pair of 30-minute time spans is the day-to-day difference between averages of illuminance readings in that time span. Data points from various sensors with the nearest time stamps were also fused to avoid imputation as part of sensor data validation and processing.

B. Clustering Based Piecewise Linear Model

Ray-tracing light models can accurately approximate the indoor light distribution of buildings. These models, however, require accurate building and furniture dimensions and can be difficult to develop, requiring technicians and professional experts for calibration. An inverse model, by contrast, is a reduced-order model with only statistically significant inputs or features, and hence can be computationally inexpensive to perform simulations within a control loop. For these reasons, an inverse model is a promising choice for a predictive lighting control system designed for ease-of-use. Inverse problem theory describes methods by which a model of a system is developed by: (1) parameterizing the system in terms of a set of model parameters that adequately characterize the system in the desired point of view, (2) making predictions on the actual values based on relatively simple physical laws and given values of the model parameters, and (3) using actual results from measurements to determine the model parameters [32]. The ordinary least squares (OLS) method functions to create a best linear fit of a given dataset by minimizing the sum of the squared residuals. We used multiple linear regression models. Based on the performance improvement achieved by the Sun Position-Based Model we assume a piecewise linear

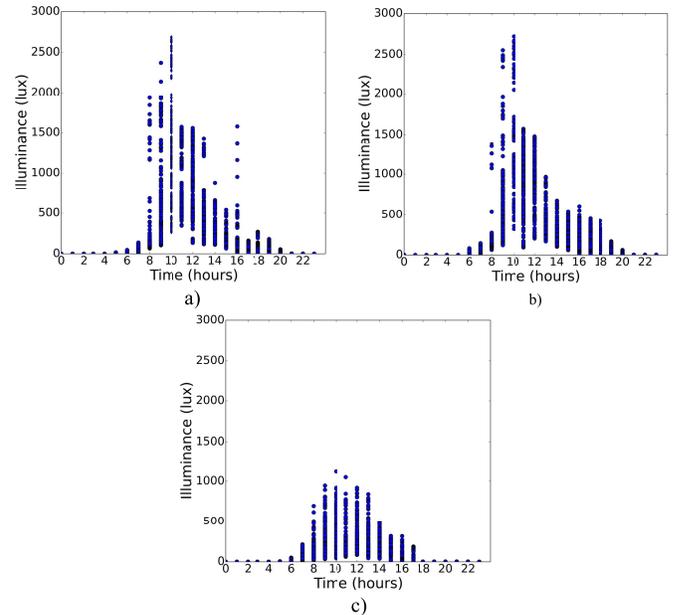


Fig. 5. Hourly daylight distribution under different clouded sky conditions. a) Partly cloudy conditions. b) Scattered clouds conditions. c) Overcast sky conditions.

relationship between the illuminances measured across the test bed at different workstations and between the artificial and natural light sources, with model parameters varying with solar altitude (time of the day) and sky conditions. The time scale of each linear model is 30 minutes i.e. we have one set of models for every 30 minutes-interval of the day during the daylight hours.

As we indicated in Section III, the correlation between daylight and measured light distribution in space changes depending on whether the light sensor has direct sun in its field of view and/ or whether direct sunlight is entering the space. The presence of direct or indirect sunlight is also affected by sky conditions. One way to include sky conditions in the feature space would be to use satellite weather data. We compared the hourly sky conditions from online weather data with our onsite daylight measurements. We found that the online weather data did not reflect the site-microclimate adequately under partly cloudy conditions. Figure 5 shows a wide and comparable distribution of daylight under different clouded sky conditions obtained from the weather data. The lack of identifiable relationship between weather station data and onsite light distribution precluded the use of regional sky conditions as a potential feature in our light models. Instead we used clustering as a proxy for sky conditions, with a constant number of clusters (explained later in this sub-section). Dividing the data into half-hourly bins takes into account variations in the solar altitude at a lower resolution than our prior consideration [25]. The choice was made to accommodate tradeoffs between data requirements for convergence of clustering and model accuracy.

Clustering algorithms use unsupervised learning to discover natural groupings in unlabeled data. We used the K-means clustering algorithm for its simplicity [33] and availability of variants [34].

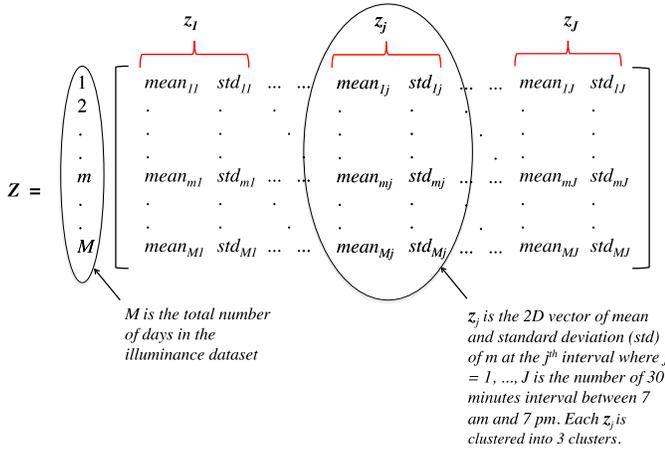


Fig. 6. Input matrix to clustering algorithm.

We describe the clustering method next. Clustering is performed on the mean and the standard deviation of the daylight data alone. This is based on the assumption that high mean daylight with low fluctuations characterize a clear sky, low mean daylight with low fluctuations correspond to overcast conditions, while under cloudy skies the daylight fluctuates the most but the mean light level is unpredictable. As mentioned earlier we divided our illuminance dataset into 30-minute bins, such that each bin for a day contains six data points at five minutes sampling interval. We then compute the mean and standard deviation of these six illuminances across all the days in our dataset.

Let $z_j \in \mathbb{R}^{M \times 2}$ be the data matrix of mean and standard deviation of measured daylight for the j^{th} 30-minute interval of a day, where $j = 1, 2, \dots, J$, and $J = 24$ between 7a.m. and 7p.m. M is number of days in the dataset (see Figure 6). We performed clustering on each of z_j . Each data point within z_j is assigned to one of the clusters c_{jk} in C_j for $k = 1, 2, 3$. K-means partitions the data by minimizing the sum of squared distance between a cluster centroid μ_{jk} and $z_j \in c_{jk}$. The resultant objective function for K-means is given by equation 3:

$$J(C_j) = \sum_{k=1}^3 \sum_{z_i \in c_k} \|z_i - \mu_{jk}\|^2 \quad (3)$$

Once clustering has been performed, the entire illuminance dataset is divided into $JK = 24 \times 3$ clusters. For the rest of the paper we refer to the processed illuminance at each workstation $w \in \mathbf{W}$ (set of workstations and sensors) for the jk^{th} cluster as y_w^{jk} . For this work, we used the K-means module of Scipy, Python with 20 initializations of cluster centroids and 100 iterations per model. One limitation of K-means is that the optimization problem presented in equation 2 can converge to local minima, which may differ with different random initializations of the centroids. However, most of our random centroid initializations resulted in similar final centroids, thereby obviating refined initializations. The results section provides more discussion of results and implications of clustering.

y_w^{jk} can be modeled as a linear combination of illuminances measured at other workstations $\{y_1^{jk}, \dots, y_{w-1}^{jk}$,

$y_{w+1}^{jk}, \dots, y_W^{jk}\}$ and artificial light statuses $\{e_s\}$ for S artificial lights in the influence zone as in equation 4.

$$\hat{y}_w^{jk} = \alpha_1^{jk} y_1^{jk} + \dots + \alpha_{w-1}^{jk} y_{w-1}^{jk} + \alpha_{w+1}^{jk} y_{w+1}^{jk} + \dots + \alpha_W^{jk} y_W^{jk} + \beta_1 e_1 + \dots + \beta_S e_S + \epsilon \quad (4)$$

$\alpha = \{\alpha_w\}$ and $\beta = \{\beta_s\}$ are model parameters and ϵ is random error. To solve this equation, the method of Ordinary Least Squares leads us to find the values of α and β that minimize the sum of the squared residuals. The above virtual sensor model is independent of the spatial layout of sensors for ease of implementation and scalability.

C. Adaptive Regressor Selection

One of the challenges in multivariate regression is the choice of an appropriate set of features or regressors, balancing a tradeoff between over-fitting and prediction accuracy of the inverse model.

In adaptive regressor selection, the optimization problem is to decide the location of the sensors for minimizing the prediction error across the workstations over the entire prediction period. Thus at the end of the training period we want to replace some of the light sensors by their inverse virtual models. The above problem is that of feature selection, in which we will retain the most informative features given a constraint on the number of features.

An adaptive regressor selection algorithm is a threshold-based heuristic feature selection process that minimizes the prediction error across all w . The algorithm selects a linear model for every sensor w and for every cluster c_{jk} , such that the number of regressors is utmost $0.5 * W$ in all. The training and the validation processes are executed in a single loop for each of the trained models y_w^{jk} . The prediction accuracy is measured in terms of Root Mean Square Error (RMSE) between the actual and the predicted illuminances of the sensor over the entire prediction period (independent) of the cluster.

$\mathbf{W} = \{1, 2, \dots, w, \dots, W\}$ is the set of all workstation sensors.

For sensor $w \in \mathbf{W}$,

For cluster jk ,

Train models $\{\alpha, \beta\}^A$ per equation 4, where $A = \{1, 2, \dots, \frac{(W-1)!}{(W-r-1)!r!}\}$ is a set of all possible combinations of sensors $\subset \mathbf{W} \setminus w$ and r is the number of regressors varying between $1, 2, 3, \dots, 0.5 * W$.

During testing period,

For sensor $w \in \mathbf{W}$,

Pick cluster jk

Predict $\{\hat{y}_w^{jk}\}^A \forall A$

Select A with the minimum root mean square error across all w as,

$$\operatorname{argmin}_A \sum_{w \in \mathbf{W} \setminus A} \sqrt{\sum_i (y_{wi}^{jk} - \hat{y}_{wi}^{jk})^2}$$

Pick the corresponding common set of sensors.

The above clustering-based model requires $\frac{(W-1)!}{(W-r-1)!r!}$ iterations per workstation and can be used to directly control the regressor numbers for reduced sensor deployment. In the results section, we compare this method with a commonly

used feature selection method called LASSO (Least Absolute Shrinkage and Selection Operator) [35]. LASSO is a one step regressor selection method with implicit set of iterations determined by regularization termination criteria.

D. Day Ahead Prediction of Indoor Light Distribution

The goal of day-ahead prediction of light distribution is to predict the available lighting load shedding from a building. Lighting loads could be reliable contingency reserve, spinning and non-spinning reserves. Most of these load participations require a short response time of 1 second to a few minutes and a total commitment of 1 to 2 hours. If we were to guarantee a minimum lighting load shed for 2 hours, we should know whether that continuous load shed would be comfortable to human eyes. Experiments shows that dimming artificial lights by even 80% is tolerable for most people in presence of sufficient daylight. Therefore, this reduces our problem to prediction of daylight availability in the next two hours. Such predictions will be important for spaces with low solar penetration.

Many researchers have focused on short-term predictions of daylight. For example, Lu *et al.* [12] proposed a short-term prediction of daylight using a weighted linear function of historical data, the weights being determined by a mean square error based similarity metric between current day and historical day. Day-ahead prediction is more challenging and would be necessary for long-term demand response. Therefore, in our current work, we focus on day-ahead prediction of indoor availability. To the best of our knowledge there has been very limited work on day-ahead prediction of indoor daylight availability. However, parallels exist in the context of solar radiation and PV power prediction. We found that, besides numerical weather simulation, neural networks are the most popular approaches to PV output prediction. We also turned to the literature on day-ahead prediction of building energy usage. Researchers have demonstrated day-ahead prediction of building energy use from smart meter data using Gaussian Process models as a function of temperature and time [36], as neural network support vector functions of forecasted temperature, humidity and solar radiation [37].

While Artificial Neural Networks (ANNs) are state-of-the-art for prediction of solar irradiance for PV applications, they are prone to over-fitting and several local minima. We found that in addition to historical daylight measurements, we could leverage availability of the day-ahead weather forecasts in our prediction. However, as mentioned earlier, satellite weather data can only indicate the general daily trend of sky condition, but cannot reliably represent the onsite sky conditions. Under such circumstances over-fitting to online weather data, resulting from neural network models may generate erroneous predictions. Support Vector Regression, on the other hand, has been known to often outperform ANN and OLS regression by virtue of its generalizability. We, therefore, proposed a Support Vector Regression (SVR) day-ahead prediction model of indoor light.

We focused on modeling daylight at the windows, owing to its high hourly and daily variance. Barring the higher

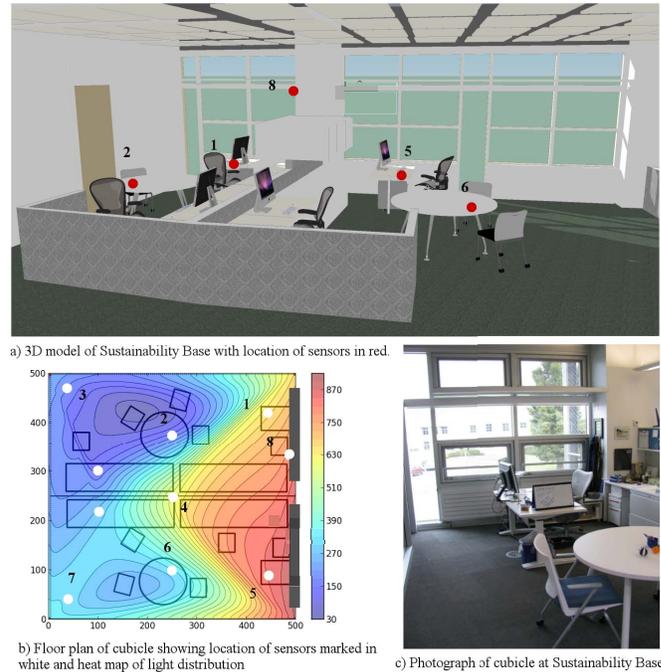


Fig. 7. Layout of cubicles and lighting conditions at Sustainability Base.

computational complexity of epsilon-SVR with non-linear kernels compared to OLS regression, epsilon-SVR has several advantages over OLS like flatness of function and error tolerance, besides the ability to handle non-linearity via kernels. The flatness of the function means SVR algorithm searches for small weights resulting in a more generalizable model. For introduction to SVR refer to Smola and Schölkopf [38] and LibSVM guide [39]. We used temperature, sky conditions and the hourly moving average of past three days of daylight as features. The main task of SVR is to set the hyper-parameters of the regression in order to get the most generalizable result. For selection of the hyper-parameters we refer to the recommendations of Cherkassky and Ma (2003) [40], as will be discussed later in Section VIII.

VII. DEPLOYMENT

Sensors were deployed across two cubicles in an open-plan office space in Sustainability Base (SB) at the NASA Ames Research Center. Sustainability Base is a 50,000 sq. ft. LEED Platinum certified high performance office building at NASA Ames Research Center. SB aims to redeploy innovations and technologies originally developed by NASA for aerospace missions to monitor and control building systems while reducing energy and water consumption. The ultimate vision of the SB is to provide a research test and demonstration site for different sustainable technologies and concepts. The three primary research objectives involved in this vision are to reduce building energy consumption and operating and maintenance costs, as well as to improve employee comfort levels.

Seven sensors were deployed on workstations (sensors 1-7 in Figures 7 a & b) and one sensor was placed on the wall near a window (sensor 8 in Figures 7 a & b). A 3D model of the

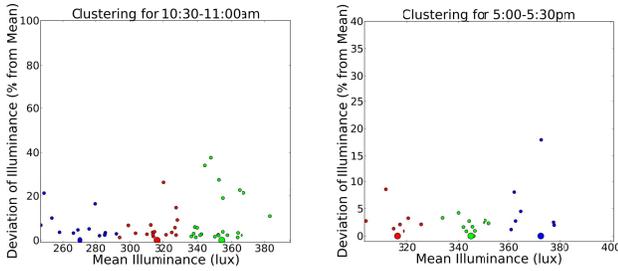


Fig. 8. Clustered 30 minutes light data showing 3 clusters, 10:30 AM - 11:00 AM (left) and 5:00 PM - 5:30 PM (right).

layout of the test bed cubicles with sensor locations (top), heat map of indoor light distribution and a photograph of the test bed are presented in Figure 7. Sensors 1, 2 and 3 were located at incremental distances from the window node 8, covering the work plane across the entire cubicle and sensors 5, 6 and 7 were replicated in the adjoining cubicle. Sensor 4 was located on top of a low height partition between the two cubicles. Sensors 1 through 7 will be referred to as workstation sensors in the rest of the paper. The goal is to use all of the above sensors for model training, but only deploy 50% or fewer of these eight sensors to predict the illuminance across all the workstations during the operational phase of intelligent lighting system. The sensors collected data for several weeks, reporting the data to a local server. Real-time trends could be accessed and viewed from sMAP (see Section IV: System Architecture) and a dedicated webpage. Artificial light statuses from four controllable luminaries were collected from lighting system data logs and were fed into the same database. Training and validation data were sampled from May 25 - June 5, 2012 and June 8 - June 20, 2012 respectively. During the training and validation period, the building was occupied and experienced normal operations.

VIII. RESULTS

A. Clustering Based Piecewise Linear Model

Figure 8 shows the results of clustering between 10:30 A.M. to 11:00 A.M. (left) and 5:00 P.M. to 5:30 P.M. (right). In Figure 8 (right), the mean illuminance has a narrow range towards the end of the day, 310-380 lux and a comparatively wider range in late morning 250-380 lux. However, in the morning the fluctuations in the light level are much higher (0-40% of mean), compared to late afternoon (0-20% of mean). The current dataset is taken from cloudy days. Therefore, while the mean illuminance does not change much throughout the day, the fluctuations vary due to generally higher light intensity in late morning as opposed to early afternoon.

The comparison of actual and predicted illuminances at workstations 2,3,6 and 7 are displayed in Figure 9. The two cubicles at SB are mirror images of each other, resulting in sensor positioning at identical locations with respect to the window. For example, workstations 2-6 and 3-7 have similar light profiles over the prediction period. Workstation 5 is a mirror image of workstation 1.

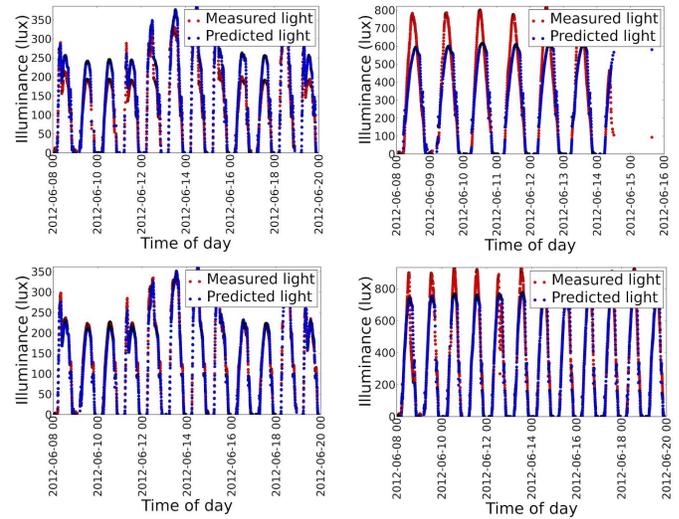


Fig. 9. Measured and predicted values at workstations 2 (top left), 3 (top right), 6 (bottom left) and 7 (bottom right).

TABLE II
ROOT MEAN-SQUARE ERROR FOR WORKSTATIONS 2,3,5,6,7
USING CLUSTERING-BASED MODEL

Workstation	2	3	5	6	7
RMSE (lux)	15.0	33.5	41.0	15.0	31.0
RMSE (%)	8.0	7.0	12.0	8.0	6.0
Regressors	8,1,4	8,4	8,4,a, b	8,1,4	8,d

The Root Mean Square Error (RMSE) of the prediction model (shown in both absolute value and as a percentage) calculated for the validation period (June 8 - June 20, 2012) is presented in Table II. Note that artificial lights have been identified by small letters a, b, c and d which are arranged in ascending order of distance from the window. The bottom row indicates the sensors used as the optimal set of regressors. Table II also lists the optimal set of regressors for best predictability of light distribution across the workstations. The RMSE is calculated across all the clusters for the entire validation period. Therefore only three physical sensors out of eight sensors deployed in the test bed were sufficient to predict the indoor light field with desirable accuracy. This amounts to 60% fewer sensors deployment compared to state-of-the-art intelligent lighting systems, which typically place a sensor in each luminary above each workstation. Results of the Sun Position-Based Model, applied to the same dataset and using the same set of regressors (as Table II), are presented in Table III. The average prediction error across the workstations, in our algorithm, has dropped to ~ 5 -15% (see Table II) with adequate data processing and clustering compared to 20-45% error using sun position-based data binning (see Table III). Moreover, the new Clustering-Based Model shows a more consistent prediction across the workstations with a narrower error range. The current RMSE is ~ 15 -40 lux as opposed to previous ~ 60 -250 lux across the workstations, reported in Paulson et al. [25]. As observed in Paulson et al., the prediction accuracy increases away from the window.

TABLE III
ROOT MEAN-SQUARE ERROR FOR WORKSTATIONS 2,3,5,6,7
USING SUN POSITION-BASED MODEL

Workstation	2	3	5	6	7
RMSE (lux)	14.0	111.0	57.0	30.0	80.0
RMSE (%)	30.0	45.0	25.0	30.0	35.0

TABLE IV
NORMALIZED MEAN-SQUARE ERROR FOR WORKSTATIONS 2,3,5,6,7
USING CLUSTERING-BASED MODEL AND LASSO-BASED
REGRESSOR SELECTION

Workstation	2	3	5	6	7
Clustering	0.02	0.014	0.062	0.016	0.009
LASSO	0.019	0.036	0.127	0.007	0.122

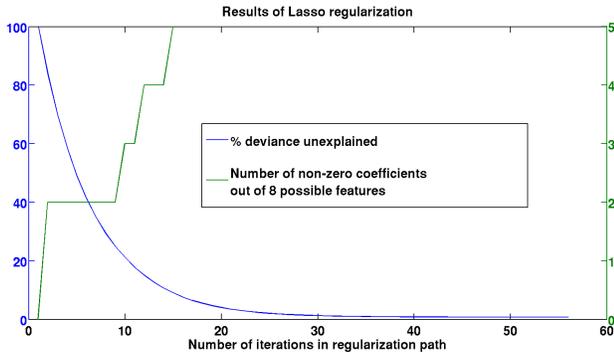


Fig. 10. Results of LASSO regularization showing five regressors essential for explaining percent deviance in prediction.

In order to test the long-term model performance we also trained and tested our inverse model on new data collected from the NASA Ames Sustainability Base in winter, during December 2, 2013 to February 4, 2014. The training times used on the test set were over the first ten days, from December 2, 2013 to December 12, 2013. Results of the root mean square percent error were within $\pm 1\%$ of prediction results obtained for summer months.

As mentioned in section VI, LASSO is a commonly used feature selection method and does not require exhaustive comparison as our method. Hence we compared the accuracy error of our regressor selection method with that of LASSO. We found that lowest prediction error was achieved using sensors 2, 3, 4, 5 and 8 as regressors, the normalized mean square error (NMSE) ranging from 0.007-0.127. The corresponding NMSE for clustering-based model using only sensors 1, 4 and 8 ranged from 0.009-0.06 (see Table IV). Figure 10 shows the regressor selection process, where all the five sensors are required to explain majority deviance of the readings.

B. Day Ahead Prediction of Indoor Light Distribution

Figure 11 illustrates the daily distribution of measured daylight level at the window on June 24-26, 2013. While light distributions on June 24 and 25 displays a similar trend

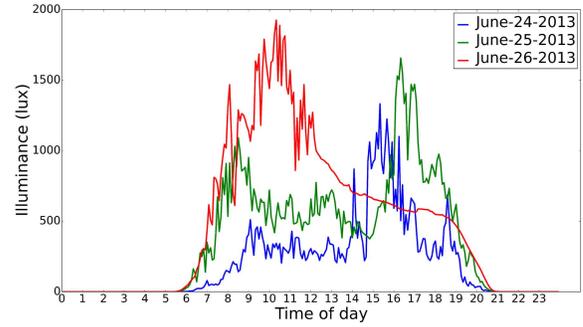


Fig. 11. Distribution of daylight level on three days of June in 2013, June 24, June 25 and June 26.

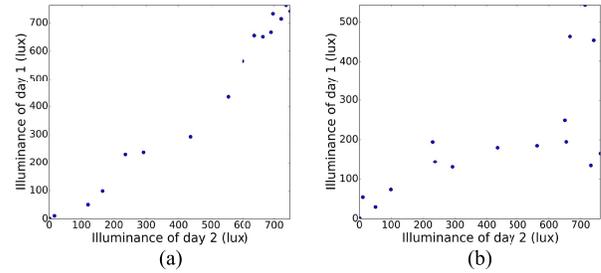


Fig. 12. (a) Scatter plot showing \sim linear relationship between hourly illuminances of two similar days (left). (b) Scatter plot showing deviation from linearity due to dissimilar sky conditions.

throughout the day with an offset between the two, light distribution on June 26 follows a similar pattern from 6:00 AM to 8:00 AM in the morning, with a sudden overshoot after that possibly due to clearer sky conditions, followed by a smoother profile in the second half of the day due to overcast sky. Therefore a simple regression model using historical values of hourly illuminances may give a good result when light data of June 24 is used to compute the day-ahead prediction of light distribution in June 25, but the same does not hold between June 25 and June 26.

We performed 5-fold cross validation to select features for SVR. We found that forecasted hourly outdoor temperature, hour of the day and hourly sky conditions are the most important features affecting the sunlight measured at the window. For similar days past light levels appeared to be a better predictor than any of the above features. Besides these, average hourly daylight levels over the past three days were considered as a feature for the SVR.

The scatter plot in Figure 12 a) shows an approximately linear relationship between hourly daylight levels measured on two similar days. Figure 11 b) on the other hand illustrates the deviation from linearity due to dissimilar sky conditions. A linear kernel produced the least mean square error of cross validation when data from historical days used in training and the test data have similar diurnal shape. On the other hand, Radial Basis Function (RBF) kernel is better able to handle occasional non-linearity as shown in Figure 12 b). We converted the sky conditions ‘clear’, ‘scattered clouds’, ‘partly cloudy’, ‘mostly cloudy’ and ‘overcast’ to numeric values from 1-5, for convenience of SVR. The similarity between the days was determined by the root mean square error between the sky conditions over 24 hours period. Depending on the similarity

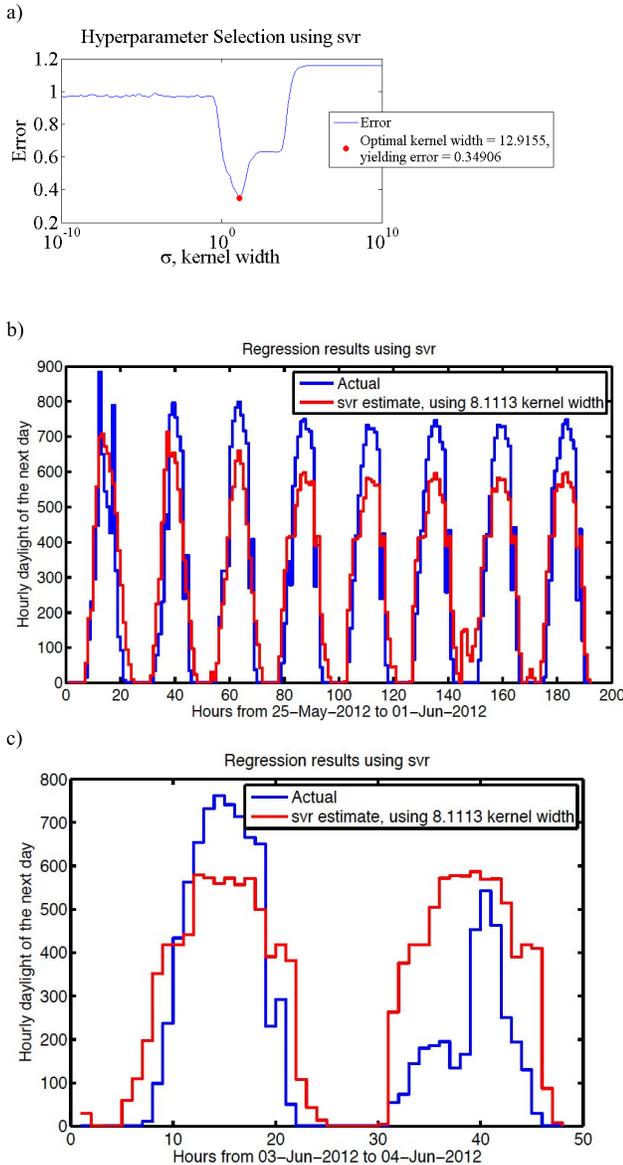


Fig. 13. a) Results for optimization of RBF kernel parameter γ . b) Training error of day-ahead prediction of daylight from May 25,2012 to Jun 1,2012. c) Day-ahead prediction of daylight level on June 3-4, 2012 from forecasted temperature, sky conditions and past 3 days hourly average measured daylight level.

between the forecasted sky condition of the prediction day and the previous three days, model cost function C , error tolerance ε and the RBF kernel parameter γ were adapted for improved prediction accuracy.

In SVR, C determines the trade-off between model complexity and error tolerance, while ε can affect the number of support vectors, which in turn governs SVR complexity. Higher error means fewer support vectors. As mentioned earlier C and ε are chosen according to the practical guidelines set forth in Cherkassky and Ma (2003) [40]. A univariate optimization of the RBF kernel parameter γ is then performed using a standard grid search. The results of the grid search and resulting fit to the training data are shown in Figure 13(a) and 13(b) respectively. The result of SVR based day-ahead prediction of daylight level on two consecutive days; June 3-4 2012 is illustrated in Figure 13(c). The training

data consisted of past six days of hourly temperature, sky conditions and hourly average daylight level of the past three days. The minimum RMSE was ~ 48 lux while the maximum error was 204 lux. The average accuracy of the SVR model over three days is $\sim 92\%$. The prediction error expressed as root mean square error (RMSE) was found to be 112 lux on an average with smaller error between similar testing and training light environment.

IX. DISCUSSION

One of the major goals of adaptive regressor selection is to ensure that the prediction accuracy demanded by the control system for occupant visual comfort and energy savings is not compromised. Therefore we analyzed the impact of prediction accuracy of the inverse model on occupant visual comfort and energy savings and determined an appropriate error threshold. The analysis assumes that unless the energy savings target is stringent and/or there is a Demand Response event, any under-estimation or over-estimation leading to prediction within 300 lux - 500 lux will lead to inaction. Any under-estimation below actual 300 lux will lead to energy wastage while an over-estimation $> 67\%$ above actual 300 lux is likely to cause visual discomfort due to inadequate light; whereas when the actual illuminance is greater than 800 lux, inaction resulting from under-prediction may cause glare.

The recommended lux level for standard office work is 500 lux [41] and, assuming a logarithmic sensitivity of the human eye, a momentary maximum error of 136 lux (as seen in our prediction) is hardly perceivable. According to experiments conducted by Luckiesh and Moss [42] the human tolerance range at any illuminance is $\sim 50\%$, i.e., at 500 lux the perceivable change threshold is 250 lux. This number was also adopted as the European standard [43]. IESNA Lighting Handbook [40] has a more conservative approach and assumes a tolerance of 20%. This number was, however, not experimentally validated.

The accuracy and predictive capability of first principle models of lighting, using sophisticated and computationally expensive ray tracing algorithms, vary widely depending on the expertise and the experience of the modelers, the average accuracy being 20% [44]. In comparison, $\sim 80\%$ -95% accuracy across the test bed, as obtained in our work indicates a model accuracy sufficient for occupant comfort. Moreover, the spatial distribution of the errors was found to be consistent except for workstation sensor 3. The temporal distribution of error is within 10% for most of the workstations in the test bed. Due to negligible under-estimation, we expect that the problem of energy wastage will not be encountered.

Furthermore we were able to reduce sensor deployment by 60% compared to the state-of-art intelligent lighting system, which use one photo-sensor and actuator per light fixture. A scenario of two to three wireless sensor platforms per occupant workstation, including daylight sensors, amounts to one platform/6.2 - 9.3 m^2 , assuming a standard occupancy of 18.6 m^2 /person as recommended by the ASHRAE standards for ventilation (ASHRAE, 2010) [45].

We compared our clustering-based method with a more standard feature selection approach, LASSO. One challenge

in this comparison was the inability to explicitly control the number of regressors in LASSO, unlike in the clustering-based method. We found that our method marginally outperforms LASSO in most cases with a smaller sensor deployment. However, LASSO is computationally more efficient than exhaustive regressor selection and can be considered as a competing method for further cross-seasonal testing and validation. Another approach would be to couple LASSO with the clustering-based method where the cost function is the aggregated error across all the clusters.

The average accuracy of the SVR day-ahead prediction model over three days was $\sim 92\%$, a significant improvement compared to prior literature. In related prior research on 24 hour prediction of solar irradiance, the researchers [46] found that ANN could predict day-ahead solar irradiance with 30%-50% RMSE on sunny days and 70% RMSE on cloudy days.

X. CONCLUSION

As part of our research endeavor to enable data-driven model-based predictive control of building systems with the Sustainability Base at the NASA Ames Research Center, we are developing a computationally inexpensive predictive model of indoor lighting. To this end we have deployed a low power wireless sensor network (with PV-energy harvesting) at this test bed and developed a piecewise linear regression model of clustered workstation illuminance, built on a month of data at seven workstations. In this work, clustering accounts for the complex nature of daylight resulting from unpredictable weather parameters such as sudden cloud cover and the relationship between building geometry and solar geometry. The clustering-based model was capable of predicting the illuminances with 80%-95% accuracy across the workstations. This was a significant improvement over our prior work using a sun position-based piecewise linear model. Clustering light data by mean and standard deviation revealed patterns in the data that could be utilized in refining the linear models. A support vector regression model was able to predict the day-ahead daylight availability with approximately 8% error. The predicted day ahead hourly daylight availability as function of forecasted hourly temperature, sky conditions and hourly average measured daylight of historical days is a potential valuable input to model predictive lighting control of grid-integrated buildings.

XI. FUTURE WORK

While our integrated WSN platform and software have demonstrated performance accuracy sufficient for intelligent lighting control and occupant comfort, further validation must be conducted for more generalizable results across larger test beds and for a year round performance evaluation. As we acquire more data from an operational test bed we will perform validation of the clustering-based model with randomly chosen training and validation sub-sets from a larger dataset. Our model has been developed using two weeks of training data, and therefore may not be extrapolated to all possible sky conditions or sun positions. Besides further training, deviations in indoor light distributions from training datasets can be accounted for in a robust control scheme through probabilistic

prediction, such as associating a confidence level with the virtual sensor predictions. The clustering-based model of indoor light will be extended to poll several explanatory variables as required by individual lighting scenarios and perform real time data fusion for reliability. Such a feature would be increasingly important for the platform reuse model. We will validate the day-ahead prediction model of daylight availability across all possible sky conditions and extend it to predict day-ahead spatial distribution of daylight.

ACKNOWLEDGMENT

The authors would like to thank their undergraduate students B. Chu and J. Richards, who were also coauthors in our former paper [47], for their support in the software development. This research was made possible with research funding from the Aligned Research Program through UARC/UCSC (a NASA Ames contractor) and the California Energy Commission's EISG programs.

REFERENCES

- [1] C.-H. Chen-Ritzo, C. Harrison, J. Paraszczak, and F. Parr, "Instrumenting the planet," *IBM J. Res. Develop.*, vol. 53, no. 3, pp. 1–16, May 2009.
- [2] F. Rubinstein, L. Xiaolei, and D. S. Watson. (Dec. 2010). Using dimmable lighting for regulation capacity and non-spinning reserves in the ancillary services market. A feasibility study. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [Online]. Available: <http://drcc.lbl.gov/sites/drcc.lbl.gov/files/LBNL-4190E.pdf>
- [3] Amsterdam Smart City. *Flexible Street Lighting*. [Online]. Available: <http://amsterdamsmartcity.com/projects/detail/id/62/slug/flexible-street-lighting>, accessed Sep. 13, 2014.
- [4] M. Ceriotti *et al.*, "Is there light at the ends of the tunnel? Wireless sensor networks for adaptive lighting in road tunnels," in Proc. 10th Int. Conf. Inf. Process. Sensor Netw. (IPSN), Chicago, IL, USA, Apr. 2011, pp. 187–198.
- [5] Y.-J. Wen, D. DiBartolomeo, and F. Rubinstein, "Co-simulation based building controls implementation with networked sensors and actuators," in Proc. 3rd ACM Workshop Embedded Sens. Syst. Energy-Efficiency Buildings, 2011, pp. 55–60.
- [6] Building Technologies Program, EERE, and U.S. Department of Energy, "Commercial Sector," in *Buildings Energy Data Book* (Department of Energy). Richland, WA, USA: PNNL, 2010.
- [7] Y.-J. Wen and A. M. Agogino, "Control of wireless-networked lighting in open-plan offices," *Lighting Res. Technol.*, vol. 43, no. 2, pp. 235–248, Jun. 2011.
- [8] Y.-J. Wen and A. M. Agogino, "Personalized dynamic design of networked lighting for energy-efficiency in open-plan offices," *Energy Buildings*, vol. 43, no. 8, pp. 1919–1924, Aug. 2011.
- [9] Y.-J. Wen, "Wireless sensor and actuator networks for lighting energy efficiency and user satisfaction." Ph.D. dissertation, Dept. Mech. Eng., Univ. California, Berkeley, CA, USA, 2008.
- [10] Y.-J. Wen, "Rapid-prototyping control implementation using the building controls virtual test bed," Philips Res. North Amer., Briarcliff Manor, NY, USA, Tech. Rep., 2011.
- [11] E. S. Lee and A. Tavil, "Energy and visual comfort performance of electrochromic windows with overhangs," *Building Environ.*, vol. 42, no. 6, pp. 2439–2449, Jun. 2007.
- [12] J. Lu and K. Whitehouse, "SunCast: Fine-grained prediction of natural sunlight levels for improved daylight harvesting," in Proc. 11th Int. Conf. Inf. Process. Sensor Netw., Beijing, China, 2012, pp. 245–256.
- [13] V. Singhvi, A. Krause, C. Guestrin, J. H. Garrett, Jr., and H. S. Matthews, "Intelligent light control using sensor networks," in Proc. 3rd Int. Conf. Embedded Netw. Sensor Syst., San Diego, CA, USA, 2005, pp. 218–229.
- [14] Y.-T. Lin and S. Megerian, "Low cost distributed actuation in large-scale ad hoc sensor-actuator networks," in Proc. Int. Conf. Wireless Netw., Commun. Mobile Comput., vol. 2. Jun. 2005, pp. 975–980.
- [15] C.-D. Ashe, M. C. de Monasterio, M. Gupta, and M. Pegors, "2010 U.S. lighting market characterization," U.S. Dept. Energy, Navigant Consulting, Inc., Washington, DC, USA, Tech. Rep., Jan. 2012.
- [16] D. Rude, "Why do daylight harvesting projects succeed or fail?" *Construct. Specifier*, vol. 59, no. 9, p. 108, 2006.

- [17] M. Maasoumy, Q. Zhu, C. Li, F. Meggers, and A. S. Vincentelli, "Co-design of control algorithm and embedded platform for building HVAC systems," in *Proc. ACM/IEEE Int. Conf. Cyber-Phys. Syst. (ICCPSS)*, Apr. 2013, pp. 61–70.
- [18] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, Jun. 2008.
- [19] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *Proc. 30th Int. Conf. Very Large Data Bases*, vol. 30. 2004, pp. 588–599.
- [20] A. Singh, R. Nowak, and P. Ramanathan, "Active learning for adaptive mobile sensing networks," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 60–68.
- [21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [22] J. S. Sandhu, A. M. Agogino, and A. K. Agogino, "Wireless sensor networks for commercial lighting control: Decision making with multi-agent systems," in *Proc. AAAI Workshop Sensor Netw.*, vol. 10. 2004, pp. 131–140.
- [23] A. Guellemine, "Using genetic algorithms to take into account user wishes in an advanced building control system," Ph.D. dissertation, Dept. Building Phys., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2003.
- [24] D. Lindelöf, "Bayesian optimization of visual comfort," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2007.
- [25] R. Paulson, C. Basu, A. M. Agogino, and S. Poll, "Inverse modeling using a wireless sensor network (WSN) for personalized daylight harvesting," in *Proc. SENSORNETS*, 2013, pp. 213–221.
- [26] A. Pandharipande and S. Li, "Light-harvesting wireless sensors for indoor lighting control," *IEEE Sensors J.*, vol. 13, no. 12, pp. 4599–4606, Dec. 2013.
- [27] J. J. Michalsky, "The *Astronomical Almanac's* algorithm for approximate solar position (1950–2050)," *Solar Energy*, vol. 40, no. 3, pp. 227–235, 1988.
- [28] S. Dawson-Haggerty, X. Jiang, G. Tolle, J. Ortiz, and D. Culler, "sMAP: A simple measurement and actuation profile for physical information," in *Proc. 8th ACM Conf. Embedded Netw. Sensor Syst.*, 2010, pp. 197–210.
- [29] S. Dawson-Haggerty, A. Krioukov, and D. E. Culler, "Experiences integrating building data with sMAP," Dept. Elect. Eng. Comput. Sci. Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2012-21, 2012.
- [30] H. Yua and Q. Yue, "Indoor light energy harvesting system for energy-aware wireless sensor node," *Energy Procedia*, vol. 16, Part B, pp. 1027–1032, 2012.
- [31] *Occupational Health and Environmental Controls: Illumination*, OSHA Standard 1926.56(a), 1926.
- [32] A. Tarantola, "The general discrete inverse problem," in *Inverse Model Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: SIAM, 2005.
- [33] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [34] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.
- [35] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.
- [36] H. Y. Nohand and R. Rajagopal, "Data-driven forecasting algorithms for building energy consumption," *Proc. SPIE Sensors Smart Struct. Technol. Civil. Mech., Aerosp. Syst.*, vol. 8692, Apr. 2013.
- [37] B. Donga, C. Caob, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy Buildings*, vol. 37, no. 5, pp. 545–553, 2005.
- [38] A. J. Smola and B. Schölkopf. *A Tutorial on Support Vector Regression*. [Online]. Available: <http://alex.smola.org/papers/2003/SmoSch03b.pdf>, accessed Sep. 13, 2014.
- [39] C. Hsu, C. Chang, and C. Lin. *A Practical Guide to Support Vector Classification*. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, accessed Sep. 13, 2014.
- [40] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, 2004.
- [41] *The Lighting Handbook*, 9th ed., Illuminating Engineering, D. DiLaura, K. Houser, R. Mistrick, and G. Steffy, Eds. New York, NY, USA, 2000.
- [42] M. Luckiesh and F. K. Moss, "The visibility of various type faces," *J. Franklin Inst.*, vol. 223, no. 1, pp. 77–82, 1937.
- [43] *Light and Lighting. Lighting of Work Places. Indoor Work Places*. EN Standard 12464-1:2011, 2011.
- [44] D. I. Ibarra and C. F. Reinhart, "Daylight factor simulations—How close do simulation beginners 'really' get?" in *Proc. 11th Int. IBPSA Conf.*, vol. 2009. 2009, pp. 196–203.
- [45] *ASHRAE Standard Ventilation for Acceptable Air Quality*, ASHRAE Standard 62.1-2010, 2010.
- [46] A. Mellit and A. M. Pavan, "A 24-hour forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy*, vol. 84, no. 5, pp. 807–821, 2010.
- [47] C. Basu, B. Chen, J. Richards, A. Dhinakaran, A. Agogino, and R. Martin, "Affordable and personalized lighting using inverse modeling and virtual sensors," *Proc. SPIE, Sensors Smart Struct. Technol. Civil, Mech., Aerosp. Syst.*, vol. 9061, Mar. 2014.

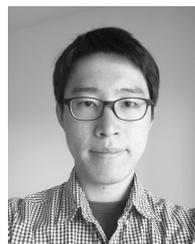


Chandrayee Basu is working towards the Ph.D. degree at Carnegie Mellon University, Pittsburgh, PA, USA. She is an intern with the Data Sciences Group, NASA Ames Research Center, Mountain View, CA, USA. She received the M.S. degree in architecture from the University of California at Berkeley, Berkeley, CA, USA, where she did her primary research project within the SmartLighting initiative of Berkeley Energy and Sustainable Technologies Laboratory, Department of Mechanical Engineering.

Her current research area encompasses distributed sensing and machine learning algorithms for smart infrastructure systems. She has participated in several projects related to wireless sensor networks and smart products like building energy monitoring, building vibration based occupancy detection and NanoSim, a simulator for NanoRK, an RTOS for wireless sensor networks, Carnegie Mellon University.

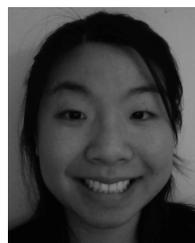
Julien J. Caubel received the B.Eng. degree in mechanical engineering from Cooper Union, New York, NY, USA, in 2012. He is currently pursuing the M.S./Ph.D. degree in mechanical engineering, with a concentration in design at the University of California at Berkeley, Berkeley, CA, USA.

He has been a Graduate Student Researcher with the Berkeley Energy and Sustainable Technologies Laboratory, University of California at Berkeley, since 2013. His research focuses on design and the sustainable development of emerging regions. From 2012 to 2013, he was a Fulbright Fellow in Ghana, developing and implementing sustainable biowaste to energy systems for isolated and agricultural communities. His current projects center on the development of sensor systems for building control and monitoring of airborne pollution. He received the awards, such as the Menschel Fellowship and the William G. Hunt Prize for Engineering Excellence.



Kyunam Kim received the B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Korea, in 2010, and the M.S. degree in mechanical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 2012, where he is currently pursuing the Ph.D. degree.

He has been a Graduate Researcher with the Berkeley Energy and Sustainable Technologies Laboratory, University of California at Berkeley, since 2013. His current research interests include development of energy-efficient sensor modules for wireless sensor network, design and dynamic analysis of robotic systems, and their control systems design.



Elizabeth Cheng is currently pursuing the B.A. degree in computer science and applied mathematics at the University of California at Berkeley, Berkeley, CA, USA, where she has been an Undergraduate Research Assistant with the Berkeley Energy Sustainable Technologies Laboratory, since 2013. Her research interests include big data processing and technology for sustainable development. She has been working as a Software Engineer with Cisco Meraki, San Francisco, CA, USA, since 2014.



Aparna Dhinakaran is currently pursuing the B.S. degree in electrical engineering and computer science at the University of California at Berkeley, Berkeley, CA, USA.

She has been an Undergraduate Research Assistant with the Berkeley Energy Sustainable Technologies Laboratory, University of California at Berkeley, since 2013. Her research interests include sensing technology, computer science education, and technology for sustainable development.



Alice M. Agogino (M'87) received the B.S. degree in mechanical engineering from the University of New Mexico, Albuquerque, NM, USA, in 1975, the M.S. degree in mechanical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 1978, and the Ph.D. degree in engineering-economic systems from Stanford University, Stanford, CA, USA, in 1984.

She is currently the Roscoe and Elizabeth Hughes Professor of Mechanical Engineering with the University of California at Berkeley. She worked in

industry for Dow Chemical, General Electric, and SRI International. She has supervised 106 M.S. projects/theses, 40 doctoral dissertations, and numerous undergraduate researchers. Her research interests include wireless sensor networks for diagnostics and monitoring, micro-electro-mechanical systems Computer Aided Design, artificial intelligence, sustainable design, smart products, and tensegrity robotics. She has received awards and honors, including the American Association for the Advancement of Science Lifetime Mentoring Award, the Pi Tau Sigma Professor of the Year Award, the NSF Director's Award for Distinguished Teaching Scholars, and ten best paper awards. She is a member of the National Academy of Engineering, and a fellow of ASME, AWIS, and AAAS.



Rodney A. Martin (SM'12) received the B.S. degree in mechanical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1992, followed by service as a Naval Officer. For most of his time on active duty, he was a Civil Engineer Corps Officer, serving as an Assistant Zone Manager and a Facility Systems Integrator for the Navy Public Works Center, San Diego, CA, USA. He gained experience with relational database integration issues for facilities maintenance management systems.

He received the M.S. and Ph.D. degrees in mechanical engineering from University of California at Berkeley, Berkeley, CA, USA, in 2000 and 2004, where he pursued research objectives influenced by his previous experience in the Civil Engineer Corps. He was with the NASA Ames Research Center, Mountain View, CA, USA, where he has worked in the application areas of robotics, data mining for aviation safety and space propulsion, and most recently computational sustainability for intelligent buildings. He is currently the Lead Researcher of Sustainability Base with the NASA Ames Research Center. His research interests include the intersection of mathematical statistics, extreme value analysis, optimal level-crossing prediction, control theory, machine learning and detection, and estimation theory. He has authored or co-authored over 30 publications. He is a Senior Member of AIAA and an Associate Member of ASHRAE.