PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Affordable and personalized lighting using inverse modeling and virtual sensors

Chandrayee Basu
Benjamin Chen
Jacob Richards
Aparna Dhinakaran
Alice Agogino
Rodney Martin

Affordable and Personalized Lighting Using Inverse Modeling and Virtual Sensors

Chandrayee Basu[a], Benjamin Chen[a], Jacob Richards [a], Aparna Dhinakaran [a], Alice Agogino [a] and Rodney Martin [b]

[a]University of California Berkeley, Berkeley Energy and Sustainable Technologies Lab, Berkeley, United States;
[b]NASA Ames Research Center, Moffett Field, Mountain View, United States

ABSTRACT

Wireless sensor networks (WSN) have great potential to enable personalized intelligent lighting systems while reducing building energy use by 50%-70%. As a result WSN systems are being increasingly integrated in state-of-art intelligent lighting systems. In the future these systems will enable participation of lighting loads as ancillary services. However, such systems can be expensive to install and lack the plug-and-play quality necessary for user-friendly commissioning. In this paper we present an integrated system of wireless sensor platforms and modeling software to enable affordable and user-friendly intelligent lighting. It requires $\sim$ 60% fewer sensor deployments compared to current commercial systems. Reduction in sensor deployments has been achieved by optimally replacing the actual photo-sensors with real-time discrete predictive inverse models. Spatially sparse and clustered sub-hourly photo-sensor data captured by the WSN platforms are used to develop and validate a piece-wise linear regression of indoor light distribution. This deterministic data-driven model accounts for sky conditions and solar position. The optimal placement of photo-sensors is performed iteratively to achieve the best predictability of the light field desired for indoor lighting control. Using two weeks of daylight and artificial light training data acquired at the Sustainability Base at NASA Ames, the model was able to predict the light level at seven monitored workstations with 80%-95% accuracy. We estimate that 10% adoption of this intelligent wireless sensor system in commercial buildings could save 0.2-0.25 quads BTU of energy nationwide.

Keywords: wireless sensor network, daylight harvesting, inverse model, clustering, commissioning

1. INTRODUCTION

1.1 Motivation

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. With goals of improving energy efficienty, operations and power grid integration, new buildings are being instrumented with large-scale wireless sensor and actuator networks for research and development.[2,3] For example, Wen et al. (2011)[4] found that closed loop control of building systems enabled by wireless sensor and actuator network (WSAN) result in 28% cooling energy and 40% light energy savings in office buildings.

Commercial lighting contributes to one of the largest pieces of the commercial energy pie. Intelligent lighting offers an easy and low-cost avenue to energy conservation. According to the U.S. DOE Energy yearbook in 2010,[5] 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% energy savings could be achieved with daylight harvesting according to our simulation results.[6–8] Furthermore, there have been considerable improvements in lighting and shading controls,[9] and in daylight harvesting systems.[10,11] Singhvi et al. (2005)[12] developed a centralized lighting system to increase user comfort and reduce energy costs by using a WSN. Lin et al. (2005)[13] proposed a decentralized algorithm for WSAN-enabled optimal lighting control.

Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014,
edited by Jerome P. Lynch, Kon-Well Wang, Hoon Sohn, Proc. of SPIE Vol. 9061, 90612S

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.[14] Some of the reasons include general lack of encouraging energy savings from expensive commissioning of lighting systems. Moreover, system usability was not considered appropriately. Rude 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.[15] System usability problems include lack of interoperability between lighting, shading and building automation system drivers, software and database. Recent state-of-the-art commercial lighting systems use one photo-sensor and actuator per light fixture. A scenario of two to three wireless sensor platforms per workstation, including daylight sensors, amounts to one platform per 6.2 - 9.3 m^2. This is equivalent to two to three times the annual lighting energy cost per unit area of a medium-sized commercial building.[16] We project that even a 50% reduction in the cost of overall deployment of intelligent lighting system can result in improved market penetration by making these systems more affordable for smaller and medium commercial facilities. This represents a trade-off between density and accuracy of the lighting field data that may govern the energy cost savings and the infrastructure costs.

1.2 Contribution

In this paper we present a method of estimating an inverse model of an indoor light field. Our research is motivated by the driving need to increase adoption of state-of-the-art wireless enabled intelligent lighting systems and draws upon the expanding field of research in optimal sensing systems. We refer to this inverse light model at a point of interest as the *Clustering-Based Model*. The virtual sensors replace many of the actual wireless sensor platforms, thereby reducing the cost of sensor deployments in the operational phase of smart lighting systems. 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. 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 the NASA Ames Research Center. In the rest of the paper we will describe related work, research methods, experiment design, data acquisition, inverse model selection and validation followed by a summary of the results and a discussion.

2. 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) and Singhvi et al. (2005) 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 a 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.[18] 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 relationship of physical quantities. This algorithm was also validated for active sensing (changing sampling rates for battery life) as part of an intelligent lighting system.[12]

Compressive 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. Compressive sensing has been mostly tested in audio and image acquisition. Sandhu et al. (2004)[22] have 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. Lindelhf (2007)[24] have proposed and validated a predictive model of light, that assumed a linear relationship between vertical faade illuminance and indoor horizontal illuminance. In his work, this predictive model resulted in a standard deviation of 416 lux (close to standard light level in offices). The same authors also found that performance of the predictive model varied with exposure to direct sun. 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 solar altitude for daylight approximation in our prior research.[25]We refer to this algorithm as the *Sun Position-Based Model*.

3. ANALYSIS

3.1 Inverse problem theory

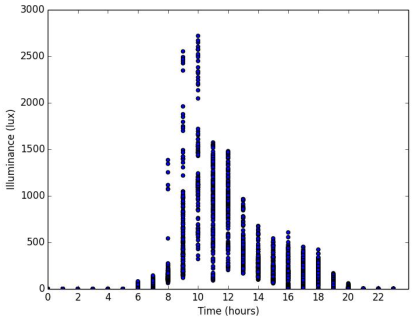
Ray-tracing light models can accurately approximate the indoor light distribution of buildings. These models, however, require accurate building and furniture dimensions and can be difficult to develop, requiring technicians and professional staff for calibration. An inverse model, by contrast, is a 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.[26]

3.2 Piecewise Multiple linear regression

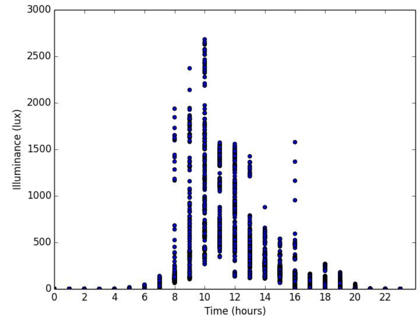
Multiple linear regression is an efficient and relatively simple procedure that can find a linear relationship between multiple regressors and a regressand. The ordinary least squares (OLS) method functions to create a best linear fit to a given data set by minimizing the sum of the squared residuals.[27] Based on this performance improvement achieved by the *Sun Position-Based Model* we assume a piecewise linear relationship between artificial and natural light sources and the illuminance measured at a workstation, with model parameters varying with solar altitude and half of the day. The time scale of each linear model is 30 minutes.

3.2.1 Clustering-based linear regression

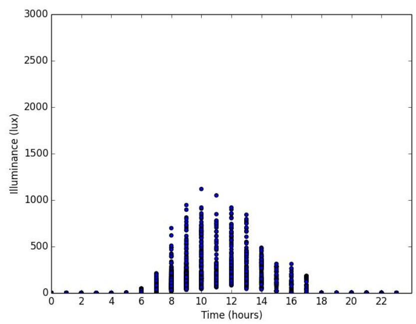
Considering the periodic nature of daylight, unpredictability of sky conditions and mismatch between onsite sky condition and weather station data, we clustered the daylight data into smaller sub-spaces using half-hourly means and standard deviations of light levels as features. Figure 1 shows a wide 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 precludes the use of regional sky conditions as a potential feature in our light models.



1 a). Hourly distribution of daylight values in test bed in scattered cloudy condition



1 b). Hourly distribution of daylight values in test bed in partly cloudy condition



1 c). Hourly disribution of daylight values in test bed in overcast sky condition

Figure 1. Hourly distribution of daylight level for different sky conditions. Sky conditions are obtained from the Wunderground repository for Berkeley

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 trade-offs between data requirements for convergence of clustering and model accuracy. We used clustering as a proxy for sky conditions, with a constant number of clusters. Then, we performed regression on the clustered data. This is different from clustering-based piecewise regression, which tries to find clusters such that the overall sum of squared error over the clusters is minimized.[28]

Clustering algorithms use unsupervised learning to discover natural groupings in unlabeled data. Clustering allowed us to account for the diverse onsite sky conditions without using actual weather data. We used K-means clustering algorithm for its simplicity[29] and availability of variants.[30] Let $x=\{x_i\}$ for $i=1,2,\ldots,n$ is a 2D dataset of mean and standard deviation of measured sunlight at the window for every half an hour during the day. Let $C=\{c_k\}$ for $k=1,2,\ldots,K$ be K clusters in each 30 minutes interval. In this case we used K=3 as a constant number, due to limited data points on one hand and empirical observations of the mean and standard deviation of daylight levels on the other. K-means partitions the data by minimizing the sum of squared distance between centroid of cluster c_k, identified as μ_k and $x_i \in c_k$. The squared error between each x_i and c_k is given by equation 1 and equation 2:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{1}$$

K-means minimizes the J function as:

$$J(C) = \sum_{k=1}^{3} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{2}$$

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 only one or two different final centroids, thereby obviating refined initializations. Some results of clustering are presented in the Results section. For this work, we used the K-means module of Scipy, Python with 20 initializations of cluster centroids and 100 iterations per model. The centroids were initialized randomly as subsets of x_i.

3.2.2 Adaptive regressor selection

One of the challenges in multivariate regression using large-scale distributed sensor data is the choice of appropriate regressors, to minimize over-fitting and improve prediction accuracy of the inverse model. Our goal is to use sensor data as regressors and the optimization problem is to decide the location of the sensors for optimal prediction across the workstations. Spatial or geometric information about the distribution of photo-sensors and artificial lights would be useful in initial screening of regressors. For example, the physics of light attenuation as a quadratic distance relation can be used in regressor selection. We used precise location information of the sensors at the windows, which are likely to be places with highest variance in the light field. The rest the algorithm is automated to iteratively pick the best set of regressors to maximize prediction accuracy based on the percentage root mean square error of prediction. The algorithm selects those sensors that carries the maximum information about the rest of the light field. For example, the initial set of optimal sensors, in our algorithm, only contains a daylight sensor located near the window. The rest of the regressors are iteratively added from a set of n sensors, resulting in optimal sensor subset. By restricting the total sensor deployment to 50%,the number of iterations per workstation amounts to $\frac{(n)!}{(n-r)!r!}$ for $r = 1, 2, 3...0.5n$. The optimal sensor sets for all the workstations are saved in a matrix. We pick a set of size $\leq 0.5n$ sensors with the highest occurrence in the optimal sensor matrix. This method is sufficient for the small number of sensors that we deployed ($n = 6$). The above approach would be computationally expensive without at least low-resolution spatial information, such as room or hall dimensions. In order to address the scalability of the regressor selection method we also tested an MI-based optimal sensor placement algorithm using actual covariance between sensor readings.

3.2.3 Model equations

The following multivariate regression model was trained on the clustered photo-sensor data.

$$E_w = \alpha_1 E_{a1} + \ldots + \beta_1 E_{n1} + \ldots + \epsilon \tag{3}$$

E_w, E_a, and E_n are illuminance readings at the workstation, an artificial light source, and a natural light source, respectively, while α and β are constants defined by the model and ϵ is random error. If we have m samples, the equation becomes:

$$\begin{pmatrix} E_{w_1} \\ \vdots \\ E_{w_m} \end{pmatrix} = \alpha_1 \begin{pmatrix} E_{a1_1} \\ \vdots \\ E_{a1_m} \end{pmatrix} + \ldots + \beta_1 \begin{pmatrix} E_{n1_1} \\ \vdots \\ E_{n1_m} \end{pmatrix} + \ldots + \epsilon \tag{4}$$

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. A simple way to do this is to first arrange the data into the form:

$$\begin{pmatrix} E_{w_1} \\ \vdots \\ E_{w_m} \end{pmatrix} = \begin{pmatrix} E_{a1_1} & \cdots & E_{n1_1} & \cdots \\ \vdots & \ddots & \vdots & \ddots \\ E_{a1_m} & \cdots & E_{n1_m} & \cdots \end{pmatrix} \begin{pmatrix} \alpha 1 \\ \vdots \\ \beta_1 \\ \vdots \end{pmatrix} + \epsilon \tag{5}$$

Simplifying Equation 5 in matrix formulation:

$$\mathbf{Y} = \mathbf{X}b + \epsilon \tag{6}$$

Solving for b, the equation can be rearranged to form:

$$b = (1/n \sum_{i=1}^{n} x_i x_i')^{-1} 1/n \sum_{i=1}^{n} x_i y_i \tag{7}$$

This equation is the Ordinary Least Squares Estimator, and gives us the best fit linear model for the data.

4. DESIGN AND IMPLEMENTATION

4.1 Hardware

Figure 2. Sensor platform with optional solar panel for energy harvesting

A wireless photo-sensor network of TelosB motes running on AA-batteries was used for illuminance data acquisition. The motes are programmed in TinyOS, an open-source platform developed at UC Berkeley.[31] The motes along with sensors (left) and a prototype of the same platform with a solar panel (right) are shown in Figure 2. The sensors were sampled at 5-minute intervals and communicated the readings over the 802.15.4 layer to a base computer. The platform operated at low power mode by calling the appropriate low power function. This puts the platform automatically to sleep when the following conditions are met, the radio is off, all high speed clock output compare interrupts are disabled, SPI (serial peripheral interface) interrupt is disabled and the task queue is empty. One of the problems of the low power operation mode is lossy communication resulting in dropped packets and redundant data. Redundant data are communicated in less than sampling intervals when the listening sensor platform has poor connectivity with the sender platform and dropped packets result in missing data points.

4.2 Data acquisition and 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.

A tolerance was proposed based on the moving average of 30-minute windows of light data. If the difference between the current light level and the immediate past light level is greater than the difference between the moving average and the past light level by a threshold percentage, chosen as a function of the light levels, then the current light level is assumed to be erroneous, as shown by equation 8 and 9.

$$m_i = \sum_{n=1}^{N} x_{i-n}/N \tag{8}$$

$$If\ x_{i-1} - x_i > \alpha(x_{i-1} - m_i)\ then\ replace \tag{9}$$

Here x_i is the light reading at current time step i, m_i is the moving average until time step $i-1$ and N is the averaging window. α in equation 9 is a function of $(x_{i-1} - m_i)$, determined iteratively. The erroneous reading is replaced by an average light level of same half hour from past 7 similar days. The similarity between each pair of 30-minute time spans is calculated based on the day-to-day difference between averages of light readings in that time span. The average light values over seven 30-minute time spans closest to the current 30- minute average were chosen to replace the erroneous light reading.

After the above processing we performed exponential smoothing of over 40 minutes windowed data and moving average over 1 hour windowed data, which were archived along with raw light readings. Imputation was not performed as data points from various sensors with nearest time stamps were fused for the inverse models.

4.3 System architecture

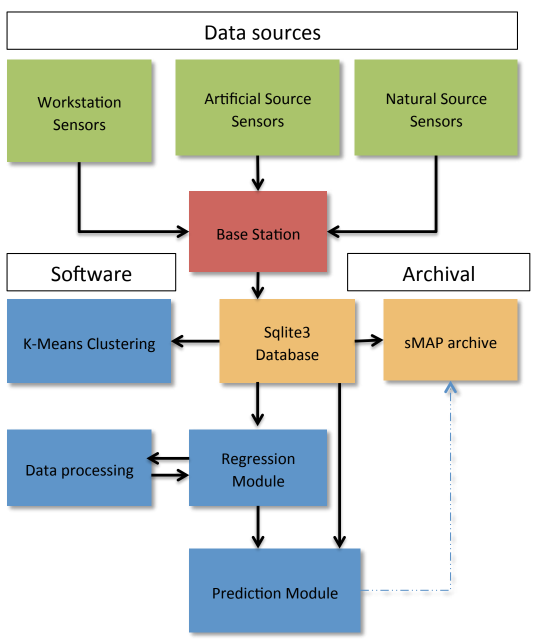
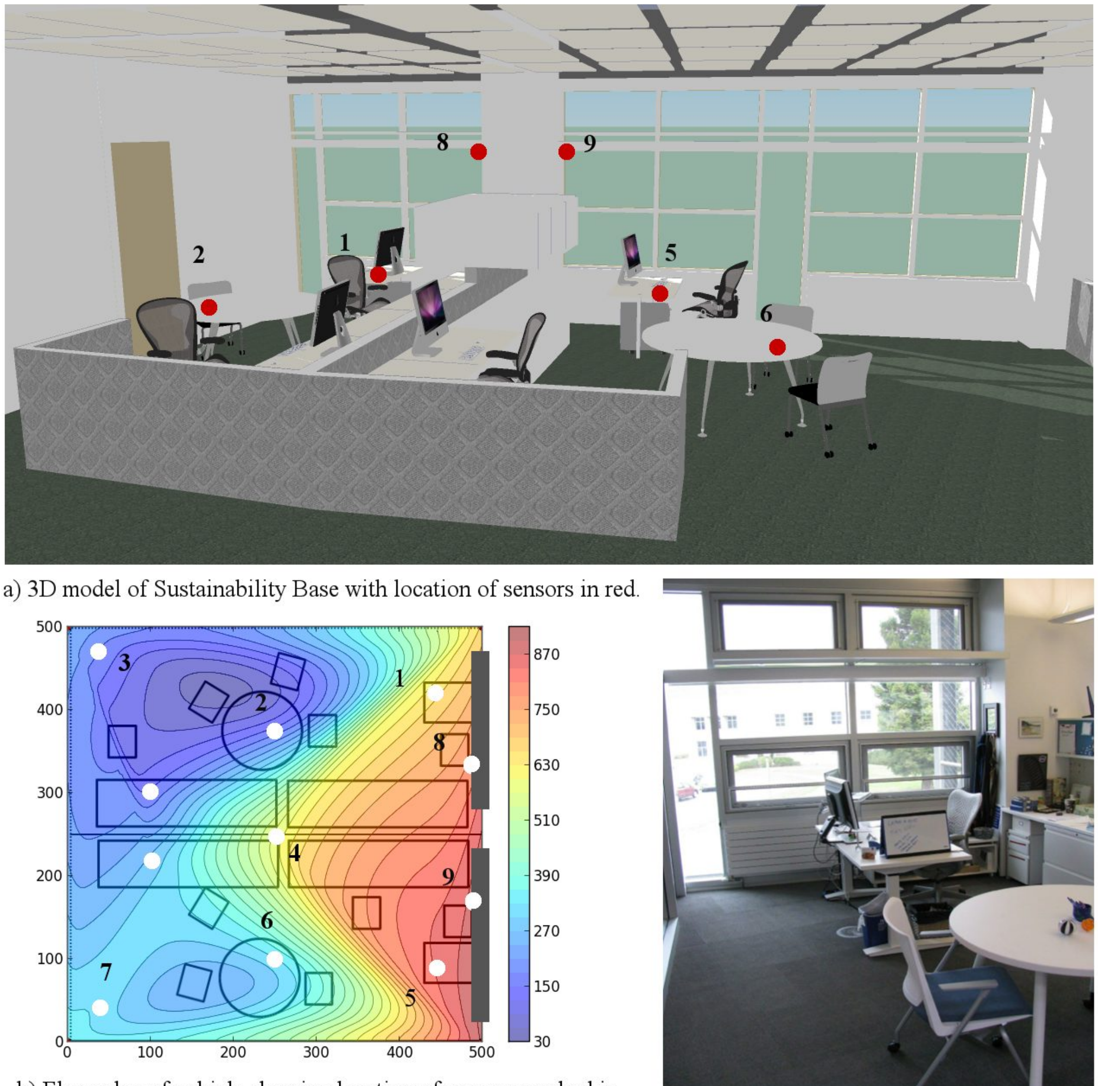


Figure 3. System architecture

A flowchart for the system architecture is given in Figure 3. The motes, each with their own unique ID, communicate data packages to a base station mote which forwards the data via a serial connection to a computer. The data are stored locally in an SQLite database.

The data processing modules are called from the regression algorithm for regularizing the matrix dimensions, eliminating zero illuminance readings during day-time, eliminating redundant data, smoothing. The database



a) 3D model of Sustainability Base with location of sensors in red.

b) Floor plan of cubicle showing location of sensors marked in white and heat map of light distribution

c) Photograph of cubicle at Sustainability Base

Figure 4. (a) 3 D CAD model of cubicles showing sensor locations at Sustainability Base, (b) Heat map of light distribution with sensor positions in plan of cubicles, (c) Photograph of test bed

stores the illuminance readings by sensor number, unixtime 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 a database driver, facade orientation prediction, a sun position calculator, sky condition calling, clustering, data processing, regression and prediction (see Figure 3). The hourly sky conditions are called using Wunderground API. Solar altitude and azimuth are calculated using the Astronomer Almanacs sun positioning algorithm.[32] The same driver module is also used to forward the illuminance readings to an online database following a Simple Measurement and Actuation Protocol (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.[33,34]

4.4 Deployment

Sensors were deployed across two cubicles in an open-plan office space in the Sustainability Base (SB) at NASA Ames Research Center.The Sustainability Base is a 50,000 sq. ft. LEED Platinum certified high performance office building at NASA Ames Research Center. The 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 and two sensors were placed on the walls near the windows. Layout of the testbed cubicles in 3D model with sensor locations (top), heat map of indoor light distribution and a photograph of the testbed are presented in Figure 4. Sensors 1, 2 and 3 were located at incremental distances from the window node 8, covering the workplane 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. The sensors collected data for several weeks, reporting the data to a local server. Real-time trends could be accessed and viewed from sMAP and a dedicated webpage. Artificial light statuses from 4 controllable luminaries were collected from lighting system data logs and were fed into the same database (snapshot of database would be good). The lights in the regression model are identified as a,b,c and d. 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.

5. RESULTS

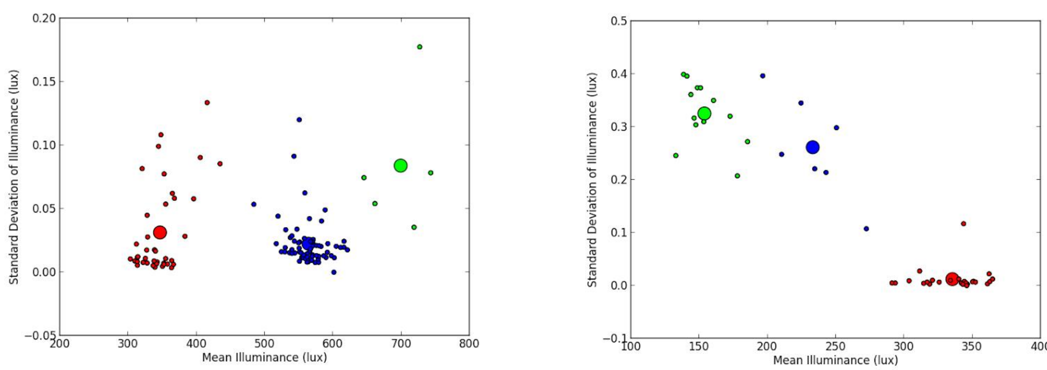


Figure 5. Clustered 30 minutes light data showing 3 clusters, 10:30 AM - 11:00 AM (left) and 5:00 PM - 5:30 PM (right)

Figure 5 shows the results of clustering between 10:30A.M. to 11:00 A.M. (left) and 5:00 P.M. to 5:30 P.M. (right). As per the results of clustering, the distribution of light level follows a diurnal pattern, justifying the use of 30-minute data bins. For example, in Figure 5 (right), the mean light level has a narrow range towards the end of the day, 100-370 lux and a wider range and a higher mean (300 to 750 lux) in late morning. The latter range, however, is much smaller than the natural fluctuation of daylight. The narrow range can be attributed to building geometry, lack of exposure to direct sun light or the limited data acquisition period. Furthermore, the clusters are distinct in Figure 5 (right) showing high light level with low standard and low light level with high standard deviation, as expected in clear and partially cloudy conditions respectively.

The comparison of actual and predicted light levels at workstations 2,3,6 and 7 are displayed in Figure 6. 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

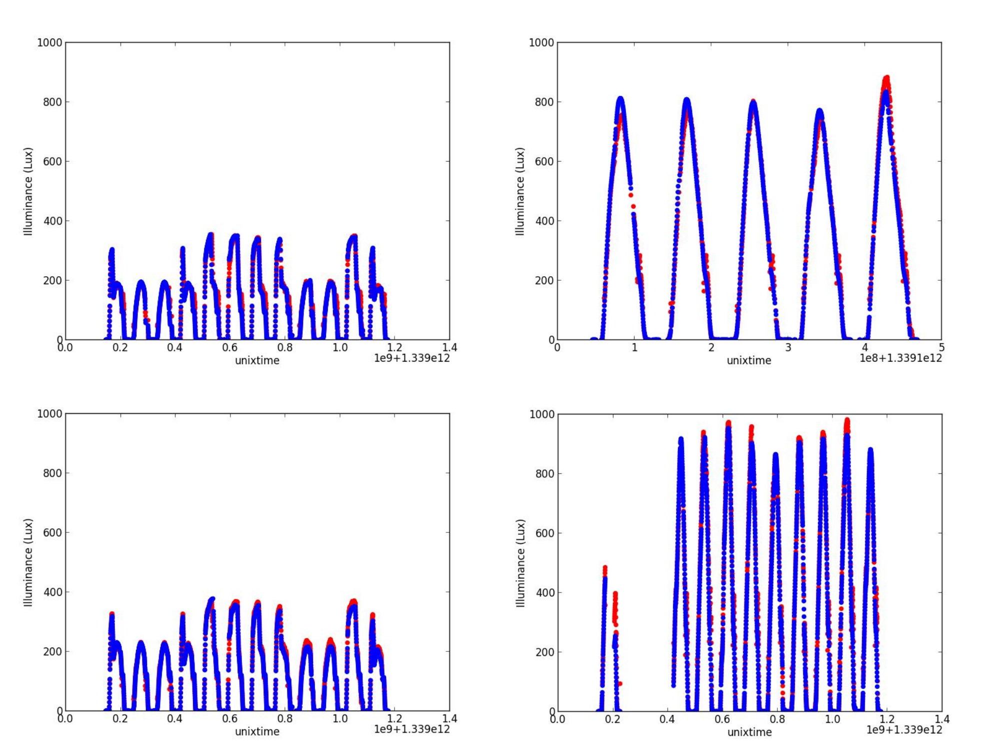


Figure 6. Measured and predicted values at workstations 2 (top left),3 (top right),6 (bottom left) and 7 (bottom right)

Table 1. 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

Table 2. 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

period. Workstation 5 is a mirror image of workstation 1. The Root Mean Square Error (RMSE) of the prediction model (shown in both absolute value and as a percentage) are presented in Table 1. The bottom row indicates the set of regressors used. The regressors 1, 4 and 8 were chosen both by the MI-based optimal sensor placement algorithm and by *Clustering-Based Model* with adaptive regressor selection. Table 1 also lists the optimal set of regressors for best predictability of light distribution across the workstations. Results of the *Sun Position-Based Model*, applied to the same dataset and using same set of regressors (as Table 1), are presented in Table 2.The average prediction error across the workstations, in our algorithm, has dropped to $\sim$ 5-15% (see Table 1) with adequate data processing and clustering compared to 20-45% error using sun position-based data binning (see Table 2). Moreover, the new *Clustering-Based Model* shows a more consistent prediction across the workstations with a narrower error range. The current RMSE is $\sim$ 15-40 lux as opposed to previous $\sim$ 60-250 lux across the workstations, reported in Paulson et al. 2012.[25] As observed in Paulson et al.,[25] the prediction accuracy increases away from the window.

6. DISCUSSION

6.1 Desired Prediction Accuracy

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 the above and settled on 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 light level is greater than 800 lux, inaction resulting from under-prediction may cause glare.

The recommended lux level for standard office work is 500 lux[35] 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[36] the human tolerance range at any light level is $\sim 50\%$, i.e. at 500 Lux the perceivable change is 250 Lux. This number was also adopted as the European standard.[37] IESNA Lighting Handbook[35] 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%.[38] 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 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.

6.2 Cost savings

Assuming one sensor per light fixture, the current best-marketed intelligent lighting product uses five to six photo-sensors in a 36 square meter space, while our sensor placement algorithm suggests that two photo-sensors are sufficient for estimating the light distribution in the same space. The current algorithm, capable of iteratively identifying the optimum combination of sensors with best predictive capability has shown significant improvement over our prior work. By replacing 60% of actual sensors with accurate predictive models, our WSN platform should be 50% less expensive than competitive technologies. With estimated market adoption, this affordable WSN platform could increase commercial lighting energy savings by 5-7% or by 0.2 to 0.25 quads BTU nationwide.

6.3 Error sources and corrections

Fluctuating weather patterns could affect the correlation of illuminance values between the motes at the workstations and those at the windows. Similar challenges in prediction accuracy might result if the major regressor sensor is giving erroneous readings from battery drainage or sudden movement or covering. An adaptive modeling algorithm was designed to appropriately deploy sensor data iteratively so that the current system could solve this problem partially. A more robust prediction system can take advantage of covariance across all the sensors for fusion and validation until a shared performance goal is reached. This algorithm could be extended to leverage other on-site sensing. For example, in our study, a preliminary comparison of window sensor 8 readings with three on-site roof-mounted radiometer data showed a good correlation between the two, but not for sensor 9. Results of further comparison with other reliable explanatory variables could eventually be used to weigh the sensor 9 readings based on data validity.

Training our model only on summer data may pose a scalability challenge. While light fluctuations might be similar (standard deviation 0%-25%) but mean light levels are likely to be lower in winter than in summer. In order to test the year round performance of the inverse model algorithm, we designed a quarter scale test bed, the size of a two-person office space and equipped it with scaled furniture, window and LED lights with realistic surface optical properties. This model provides a more controlled environment than the real building. Data were acquired from 16 Dec 2012 to 23 July, 2013 using similar data acquisition infrastructure as mentioned in subsection 4.2. We picked two weeks of training data from December 16 to December 26 and two weeks of validation data separated by 3 months from the training period, from March 16 to March 26. The RMSE was found to be 100 lux or 18%, still within the tolerance range by IESNA. We noted that the errors are high during mid-day when the maximum light level was 1200 lux, which is unlikely to be encountered in real buildings unless exposed to direct sunlight.

7. 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 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 light levels with 80%-95% accuracy across the workstations. This was a significant improvement over our prior work using 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.

8. FUTURE WORK

While our integrated WSN platform and software has 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 a real 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 data, can be accounted for in a robust control scheme through probabilistic prediction, such as associating a confidence level with the virtual sensor readings and online clustering. The online clustering-based model of indoor light will be extended to poll several explanatory variables as required by individual lighting scenarios and perform time to time data fusion for reliability. Such a feature would be increasingly important for the platform reuse model. In addition, we are incorporating day-ahead daylight prediction capability within the current model for better integration into a model predictive control framework for demand management.

9. ACKNOWLEDGEMENT

The authors wish to acknowledge the work of Andrew Sabelhaus, Julien Caubel and Kyunam Kim for helping mentor the undergraduate students in this research project and for their work in creating low power electronic solutions and solar panel energy harvesting. We would also like to thank our undergraduate student Elizabeth Cheng for her support in software development. This research was made possible with research funding from the NASA-UC 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 Journal of Research and Development* **53**(3), pp. 1–16, 2009.

[2] A. Aswani, H. Gonzalez, S. Shastri, and C. Tomlin, "Provably safe and robust learning-based model predictive control," *Automatica* **49**, pp. 1216–1226, 2013.

[3] A. Rowe, M. Berges, G. Bhatia, E. Goldman, R. Rajkumar, L. Soibelman, and J. Garret, "Sensor andrew: Large-scale campus-wide sensing and actuation," tech. rep., Carnegie Mellon University, 2008.

[4] Y.-J. Wen, D. D. Bartolomeo, and F. Rubinstein, "Co-simulation based building controls implementation with networked sensors and actuators," in *BuildSys, 3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency In Buildings*, ACM, 2011.

[5] PNNL, *Buildings Energy Data Book 2010.* Department of Energy, 2010.

[6] Y.-J. Wen and M. Agogino, "Control of wireless-networked lighting in an open-plan office," *Journal of Lighting Research and technology* **43**(2), pp. 235–248, 2011.

[7] Y.-J. Wen and M. Agogino, "Personalized dynamic design of networked lighting for energy-efficiency in open-plan offices," *Energy and Buildings* **43**(8), pp. 1919–1924, 2011.

[8] Y.-J. Wen, *Wireless Sensor and Actuator Networks for Lighting Energy Efficiency and User Satisfaction.* PhD thesis, University of California at Berkeley, 2008.

[9] Y.-J. Wen, "Rapid-prototyping control implementation using the building controls virtual test bed," tech. rep., Philips Research North America, 2011.

[10] E. Lee and A. Tavil, "Energy and visual comfort performance of electrochromic windows with overhangs," *Building and Environment* **42**(6), pp. 2439–2449, 2007.

[11] J. Lu and K. Whitehouse, "Suncast: Fine-grained prediction of natural sunlight levels for improved daylight harvesting," in *IPSN, 11th ACM Conference on Information Processing in Sensor Networks*, ACM, 2012.

[12] V. Singhvi, A. Krause, C. Guestrin, J. J. Garrett, and H. Matthews, "Intelligent light control using sensor networks," in *Proceedings of SenSys'05*, pp. 218–229, ACM, 2005.

[13] Y. Lin and S. Megerian, "Low cost distributed actuation in large-scale ad hoc sensor-actuator networks," in *Proceedings of 2005 International Conference on Wireless Networks, Communications and Mobile Computing*, pp. 975–980, ACM, 2005.

[14] C.-D. Ashe, M., C. de Monasterio, M. Gupta, and M. Pegors, "2010 u.s. lighting market characterization," tech. rep., U.S. Department of Energy and Navigant Consulting.

[15] D. Rude, "Why do daylight harvesting projects succeed or fail?," *Construction Specifier* **59**(9), 2006.

[16] CPUC, "Average rate by customer class years 2000-2011," tech. rep., California Public Utilities Commission, 2011.

[17] M. Maasoumy, Q. Zhu, C. Li, F. Meggers, and A. Sangiovanni-Vincentelli, "Co-design of control algorithm and embedded platform for building hvac systems," in *The 4th ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS13*, ACM, 2013.

[18] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research* **9**, pp. 235–284, 2008.

[19] A. Deshpande, C. Guestrin, S. Madden, J. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *VLDB '04 Proceedings of the Thirtieth international conference on Very large data bases*, **30**, pp. 588–599, ACM, 2004.

[20] A. Singh, R. Nowak, and P. Ramanathan, "Active learning for adaptive mobile sensing networks," in *ACM/IEEE Interntional Conference on Information Processing in Sensor Networks, IPSN*, ACM/IEEE, 2006.

[21] D. Donoho, "Compressive sensing," *IEEE Trans. Inform. Theory* **52**(4), pp. 1289–1306, 2006.

[22] J. Sandhu, M. Agogino, and A. Agogino, "Wireless sensor networks for commercial lighting control: Decision making with multi-agent systems," in *Proceedings of Working Notes of the AAAI-04 Sensor Networks Workshop*, pp. 88–92, 2004.

[23] A. Guellemin, *Using Genetic Algorithms to Take into Account User Wishes in an Advanced Building Control System.* PhD thesis, Ecole Polytechnique Federale de Lausanne, 2003.

[24] D. Lindelof, *Bayesian Optimization of Visual Comfort.* PhD thesis, Ecole Polytechnique Federale de Lausanne, 2007.

[25] R. Paulson, C. Basu, M. Agogino, and S. Poll, "Inverse modeling using a wireless sensor network (wsn) for personalized daylight harvesting," in *SENSORNETS 2013*, 2013.

[26] A. Tarantola, *Inverse model theory and methods for model parameter estimation.* SIAM, 2005.

[27] F. Hayashi, *Econometrics.* Princeton: Princeton University Press, 2000.

[28] B. Ari and H. Guevenir, "Clustered linear regression," *Knowledge-based Systems* **15**(3), pp. 167–173, 2002.

[29] X. Wu and V. Kumar, "Top 10 algorithms in datamining," *Knowledge Inf System* **14**, pp. 1–37, 2008.

[30] A. Jain, "Data clustering: 50 years beyond k-means," in *19th International Conference on Pattern Recognition (ICPR)*, 2008.

[31] P. Levis, S. Madden, J. Polastre, R. Szewczyk, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, *Tinyos: An operating system for sensor networks. Ambient Intelligence*, pp. 115–148. New York: Springer Berlin Heidelberg, 2005.

[32] J. Michalsky, "The astronomical almanacs algorithm for approximate solar position (1950-2050)," *Solar Energy* **40**(3), pp. 227–235, 1988.

[33] S. Dawson-Haggerty, X. Jiang, G. Tolle, J. Ortiz, and D. Culler, "smap a simple measurement and actuation profile for physical information," in *Proceedings of the Eighth ACM Conference on Embedded Networked Sensor Systems (SenSys 2010)*, 2010.

[34] S. Dawson-Haggerty, A. Krioukov, and D. Culler, "Experiences integrating building data with smap," tech. rep., UCB/EECS, 2012.

[35] *The Lighting Handbook*. Illuminating Engineering Society of North America, 9 ed., 2000.

[36] M. Luckiesh and F. Moss, "The visibility of various type faces," *Journal of the Franklin Institute* **223**(1), pp. 77–82, 1937.

[37] *EN 12464-1:2011:Light and lighting. Lighting of work places. Indoor work places.* EN, 2011.

[38] D. Ibarra and C. Reinhart, "Daylight factor simulation, how close do simulation beginners really get?," in *Building Simulation, 11th International IBPSA Conference*, 2009.