Forensically Discovering Simulation Feedback Knowledge from a Campus Energy Information System

Clayton Miller and Arno Schlueter
Architecture & Building Systems (A/S), Institute of Technology in Architecture (ITA), ETH Zürich
Zürich, Switzerland
{miller.clayton, schlueter}@arch.ethz.ch

ABSTRACT
Simulation model calibration has been long identified as a key means of reconciling the consumption and efficiency characteristics of buildings. A key step in this process is the creation of the actual diversity factor profiles for occupancy and various energy end uses such as lighting, plug-loads, and HVAC. Creation of these model inputs is conventionally a tedious process of site surveys, interviews or temporary sensor installation. Sometimes measured energy data can be used to create these schedules, however there are many challenges, especially when the sensor network available is large or unorganized. This paper describes a process applying a series of knowledge discovery filters to screen data quality, weather sensitivity, and temporal breakouts from large nonresidential building performance datasets collected by building management and energy information systems (BMS/EIS). These screening techniques are used to qualify the desirability for calibrated model diversity schedule creation from a forensic perspective. A diurnal pattern filtering technique is then applied that automatically extracts frequent daily performance profiles, which can then be normalized and used as model inputs according to conventional industry techniques. The process is applied on a raw dataset of 389 power meter data streams collected for eight years from the EIS of a campus of 32 higher education buildings. The results are discussed in the context of time and effort savings for creating urban and building scale simulation model inputs.

Author Keywords
Knowledge discovery; Measured building performance; Simulation feedback; Diversity schedules; Temporal data mining; Visual analytics

ACM Classification Keywords
J.5 [Arts and Humanities]: Architecture.; H.4.2 [Information Systems Applications]: Types of Systems Decision Support.; H.5.m [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous.

SimAUD 2015, April 12-15, 2015, Alexandria, VA

INTRODUCTION
Modern commercial buildings contain an ever-increasing amount and complexity of sensor systems designed to control and monitor the performance of energy consuming systems [7]. Building owners are starting to see the value in the storage and processing of this time-series data through the use of various analytics and monitoring techniques [6]. In addition, multiple commercial product and service providers are creating a new market on these technologies. One key focus in leveraging this data is design phase feedback through measurement and verification plans and calibration of the whole building energy models (BEM) [10]. Commissioning and operations experts are focusing the analysis of performance achievement as a means of getting the building off to a good start and maintaining optimal performance through the life of the building. A common scenario when leveraging the sensor data is when an analyst is forensically searching the stored dataset for general characteristics and anomalies as a basis for further investigation. The primary question of an analyst at this phase is: "What insight can we achieve at this point with the raw data available?"

Based on the authors' experience and through several industry and professional reviews, key challenges have been identified in the initial level of analysis of common BMS/EIS systems [5]:

- An unorganized data storage structure that impedes the analyst's ability to explore multiple data streams
- Indecipherable or forgotten point labeling schemes that make it difficult to semantically model the data
- Poor sensor accuracy or data storage quality
- Poor characterization of the influence of data streams on each other or from external factors such as weather, occupancy, etc.
- The sheer size of the database and number of sensors is often beyond the capabilities of most energy analysts' tools, which are often simply spreadsheets

The literature in commercial building model calibration treats temporal knowledge extraction from sensor data lightly and often focuses on the actual process of calibration once all of the inputs have been determined. One set of inputs that are crucial for model calibration are *diversity factor schedules* which inform the model of the approximate percentage of people, lighting or miscellaneous loads at a given point in time. Very often diversity factor schedules are only determined through labor-intensive site surveys, questionnaires or additional temporary sensors [4]. Another method of schedule creation is the use of measured energy performance of weather-independent end uses. One of the largest reviews of industry-standard diversity schedules creation relies on transforming measured energy performance data in this way [1]. Lighting and plug load meter data is transformed through manual analysis and normalization for the creation of diversity factor schedules. The techniques reviewed in this study rely on an in-depth understanding of the measured dataset and are restrained from implementation in other projects by the previously identified BMS/EMS challenges. The process of calibration on a campus or large portfolio EIS/BMS systems increases the challenge as the scale and complexity of the systems is larger. Recent research projects attempt to remedy these challenges such an example of automated utilization of measured data for *autotuning* simulations [9].

In this paper, a data screening process is defined that is designed for an analyst to dissect, characterize and acquire knowledge from BMS/EIS data repositories that are relevant to the diversity factor schedule creation procedure for non-residential buildings. Nonresidential buildings are targeted due to the similarity in the data collection capabilities and relatively systematic nature of load profiles as compared to residential. This process uses statistical techniques to filter relevant information as part of a *forensic* investigation of past data. The paper will stop short of the full transformation process of creating diversity schedules as these techniques are covered widely in the literature [1]. The forensic process is implemented on a large campus case study dataset that is in the process of analysis for calibrating a coupled urban-scale and whole building energy simulation.

A Screening and Knowledge Extraction Process

In this paper, the development and implementation of a knowledge discovery process is discussed that addresses the context-specific challenges outlined. This process utilizes the following steps to explore a large, real-world case study dataset:

1. Data quality screening - Evaluation of the completeness and rough accuracy of the data in terms of gaps and extreme outliers

2. Weather sensitivity screening - Classification of each stream according to its influence from outdoor air temperature conditions

3. Breakout detection screening - Split each data stream according to subsets of continuously similar statistical behavior

4. Compared evaluation of metrics - Clustering and evaluation of the data streams available according to the combination of their metrics from the screening process

5. Typical daily profile filtering - Utilize a clustering process to extract and group similar performance profiles

Quality Metric	Description	Threshold
0	Missing Data	Data loaded as *NaN* as compared to other data stream availability
1	Flat-lined Data	Exact same value recorded for an entire day
2	Outliers	Individual data points as detected by Seasonal Hybrid ESD (S-H-ESD) algorithm
3	Normal	Data points that don't fit in the above criteria

Table 1. Data quality metric description and thresholds

METHODOLOGY

Data Quality Screening

General sensor data quality is defined in terms of the factors of temporal completeness, meaningful sensor output, and presence of significant anomalous behavior. Temporal completeness pertains to the absence of gaps in the dataset. These gaps are most often related to sensor failure, in which a sensor stops sending readings to the central repository. Meaningful sensor output pertains to the sensor stream's reading matching the expected behavior of the phenomenon. For example, a sensor that is constantly reading the same value for long periods of time is likely not giving meaningful information. This particular situation is often noted as "flat-lining" and it is defined as recording the exact same value for a 24 hour period for performance measurement data streams. Significant anomalous behavior is detected using a Seasonal Hybrid ESD (S-H-ESD) anomaly detection algorithm as implemented in the *AnomalyDetection* R library [12]. This library was developed by the web-based social media company, Twitter, as a means of finding interesting or anomalous behavior in the postings of their users online. This algorithm can be used to detect both global and local anomalies and it was developed with a focus on temporal datasets with potentially non-normal distributions and seasonal and other cyclical attributes.

In this screening process, four general classifications of data quality and completeness are created. The data set is first divided into these classifications based on various statistical thresholds. It should be noted that these thresholds were chosen for the dataset used in the implemented example and tuning may be required for implementation on other data sources. These levels with description and threshold definition are found in Table 1. Figure 1 illustrates the data quality screening process for 310 days taken from a commercial building. The initial 15 days have no data available and the next approximately 120 days have a flat-lined zero reading for the meter. This flat-lined behavior of this meter shows that the measured system is off or the meter is malfunctioning. Most of the remaining data are considered normal except for a peak towards the end that is classified as an anomaly. This anomaly is detected through time series decomposition and robust statistical metrics contained within the applied library.

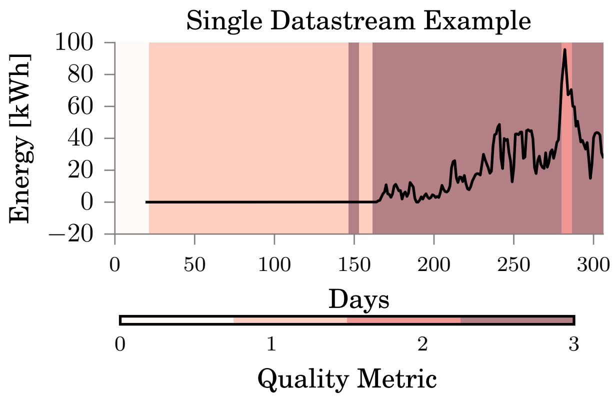

Figure 1. Data quality metrics of 310 days from a kWh meter

Weather Sensitivity Screening

Data stream influence characterization is the process of roughly classifying the dataset into streams and subsequences based on weather conditions sensitivity. This evaluation is important in understanding what measured performance is due to heating, cooling, and ventilation systems (HVAC) responses to outdoor conditions and what is due to schedule, occupancy, lighting, and miscellaneous loading conditions which are weather independent. Performance data that is influenced by weather can be used to better understand the HVAC system operation or be weather-normalized to understand occupant diversity schedules. Non-weather sensitive data streams are used with less pre-processing to create diversity schedules and to calculate miscellaneous and lighting load power densities.

In this filtering step, the Spearman Rank Order Correlation (ROC) is used to evaluate the positive or negative correlation between each performance measurement stream and the outdoor air dry bulb temperature. This technique has been previously used for weather sensitivity analysis [3]. The ROC coefficient, ρ, is calculated according to a comparison of two data streams, X and Y, in which the values at each time step, X_i and Y_i, are converted to a relative rank of magnitude, x_i and y_i, according to its respective dataset. These rankings are then used to calculate ρ that varies between +1 and -1 with each extreme corresponding to a perfect positive and negative correlation respectively. A value of 0 signifies no correlation between the datasets. This ρ value for a time-series is calculated according to Equation 1.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

The difference between the data stream rankings, x_i and y_i, is signified by a difference value, d_i, and the number of samples compared in each dataset is signified by n. Figure 2 illustrates the calculation of the ROC coefficient, ρ for three examples. The cooling sensitive data set shows a strong positive correlation between outside air temperature and energy consumption with a ρ value of 0.934. As the outside air temperature increases, the power consumption measured by this meter increases. The heating sensitive dataset shown has a strong negative correlation with a ρ of -0.68. A weather insensitive dataset is shown in the middle which has a ρ of 0.0, signifying no weather correlation, which is obvious due to the four levels of consumption which are independent from outdoor air conditions.

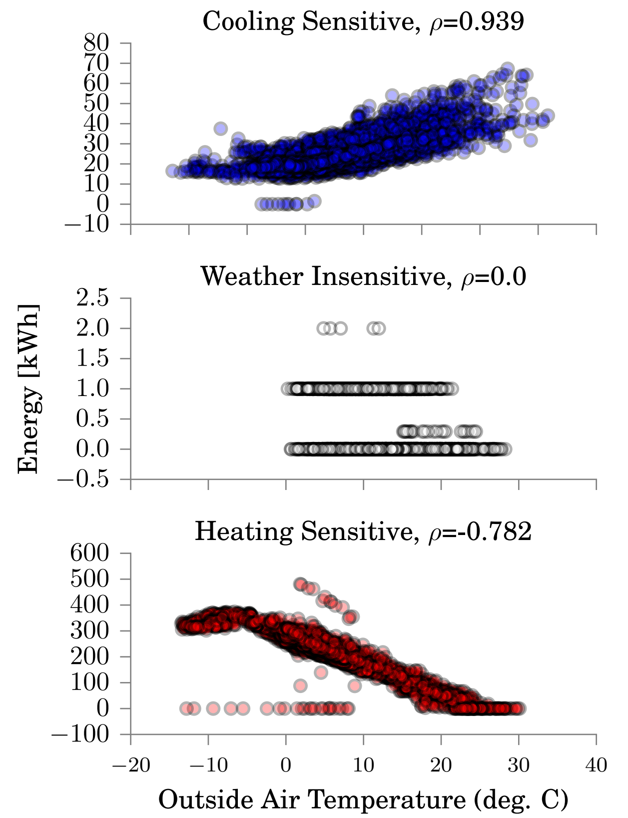

Figure 2. Weather sensitivity examples as energy vs. outdoor air temperature

The focus of the process is not just in the overall weather sensitivity of each data stream, but also the weather dependence over time. Therefore, in this filtering process, ρ is calculated for each month and plotted to show the changes over time. Figure 3 illustrates this process over the course of a year for two data stream examples. The weather sensitivity for these two streams changes slightly amongst the seasons possibly due to non-weather loads outside the typical season. For example, the cooling system will sometimes need to maintain a base cooling load due to internal people, lighting, and miscellaneous loads outside the cooling season and will thus be weather insensitive. Weather sensitivity analysis enables an analyst to divide the dataset into the classifications of heating and cooling sensitive and weather insensitive.

Breakout Detection Screening

Breakout detection screening is a process in which each data stream is analyzed according to the tendency to shift from one performance state to another with a transition period in between. Breakout detection is a type of change point detection

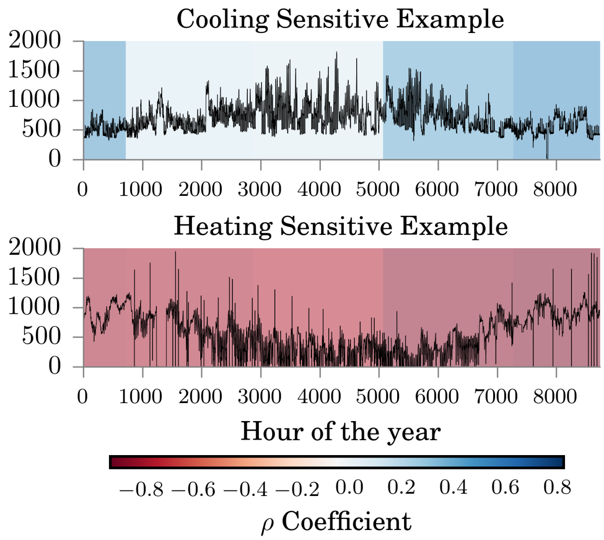

Figure 3. Weather sensitivity examples of one year of heating and cooling data

that determines whether a change has taken place in a time series dataset. Change detection enables the segmentation of the dataset to understand the nonstationarities caused by the underlying processes and is used in multiple disciplines involving time-series data such as quality control, navigation system monitoring, and linguistics [2]. Breakout detection is applied to temporal performance data to understand general, continuous areas of performance that are similar and the transition periods between them.

In this process, an R programming package, *BreakoutDetection*, is utilized, which is also developed by Twitter to process time-series data related to social media postings [13]. This package uses statistical techniques which calculate a divergence in mean and uses robust metrics to estimate the significance of a breakout through a permutation test. *BreakoutDetection* uses the E-Divisive with Medians (EDM) algorithm, which is robust amongst anomalies and is able to detect multiple breakouts per time series. It is able to detect the two types of breakouts, mean shift and ramp up. Mean shift is a sudden jump in the mean of a data stream and ramp up is a gradual change of the value of a metric from one steady state to another. The algorithm has parameter settings for the minimum number of samples between breakout events that allows the user to modulate the amount of temporal detail.

The goal in using breakout detection for building performance data is to simply find when macro changes occur in sensor data stream. This detection is particularly interesting in weather insensitive data to understand when modifications are made to the underlying system in which performance is being measured. Figure 4 shows eight years of data from a weather insensitive data stream. Each color represents a group of continuous, steady-state operation and each change in color is, thus, a breakout. These breakouts could be the result of schedule or control sequence modifications, systematic behavior changes, space use type changes, etc. Creation of diversity factor schedules should target data streams which have few breakouts and the data between breakouts is the most applicable for model input.

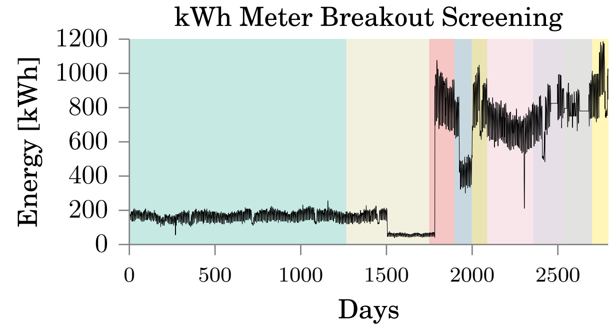

Figure 4. Breakout detection example of eight years of data

Compared Evaluation of Metrics

Comparison and evaluation of the individual metrics is necessary in order to understand which data streams are more applicable for extraction of diversity schedules related to the occupancy or load characterization of the building. In this step, the InCHlib (Interactive Cluster Heatmap library) is used to apply Wards hierarchical clustering with Euclidean distance to the metrics [11]. This clustering groups similar data streams according to similarities in the screening metrics. The data streams are qualified according to groups, or regions, of sensors with similar characteristics according to simulation data feedback.

Typical Daily Profile Filtering

The final step in the process is to characterize the data according to similar groups to understand the general modes of performance. This characterization divides the data into typical daily consumption profiles and further divides these diurnal sets into frequent patterns, or motifs, and infrequent patterns, or discords. This process has been implemented in previous case studies for performance characterization and is known as the *DayFilter* technique [8].

Up to this point, the previous steps are considered filters that are qualifying the data for inclusion or exclusion in the simulation data generation process. For example, an analyst may make decisions on whether to include certain types of outliers from the data quality test as they seem relevant to distinguishing discord patterns in this phase. Another decision might be to divide the dataset up into difference regions of analysis according to the breakout detection steps.

The *DayFilter* process starts by using Symbolic Aggregate approXimation (SAX) to distinguish patterns and then divide them by user-defined thresholds into motifs and discords. Discords are to be investigated further in order to understand infrequent behavior that may be caused by performance problems. The motifs are then clustered according to similarity using the k-means algorithm to create typical performance profiles.

IMPLEMENTATION

The filtering process is implemented to exemplify knowledge discovery for simulation according to a real-world case study. Typical model calibration processes of single buildings may only have a few submeters worth of data. These filtering and characterization techniques are applied to a relatively large dataset from a whole campus of buildings. This approach will show the value of filtering and visualizing large amounts of time-series data quickly and more accurately than more manual methods.

The case study chosen is a university campus with a large energy and building information system and a long log of data acquisition. The campus consists of 32 buildings located in a temperate climate. This campus has been modeled extensively in past projects using the EnergyPlus whole building simulation engine and the CitySim urban-scale modeling software. The intent at this point is to utilize the measured dataset from the campus EIS to tune and calibrate those models. Figure 5 illustrates the number and type of data streams available in the EIS of the campus. The key focus for simulation model calibration is the heating, cooling, and electricity consumption metrics, however the other streams can be processed to provide supporting evidence for model tuning decisions. Many of the older buildings on campus have very few automated measurement systems, while many of the newer or recently-renovated systems have a large number. Building 0 is home to the chiller and boiler plants for the campus and therefore it has the largest number of sensors. This project is within the scope of the goal of taking a large, relatively-unstructured dataset and *forensically* filtering the value and structure of the data out as it pertains to the model calibration objectives. Eight total years of data from January 1, 2006 to January 1, 2014 were collected from over 1200 sensors that were stored at 5-60 minute frequency intervals.

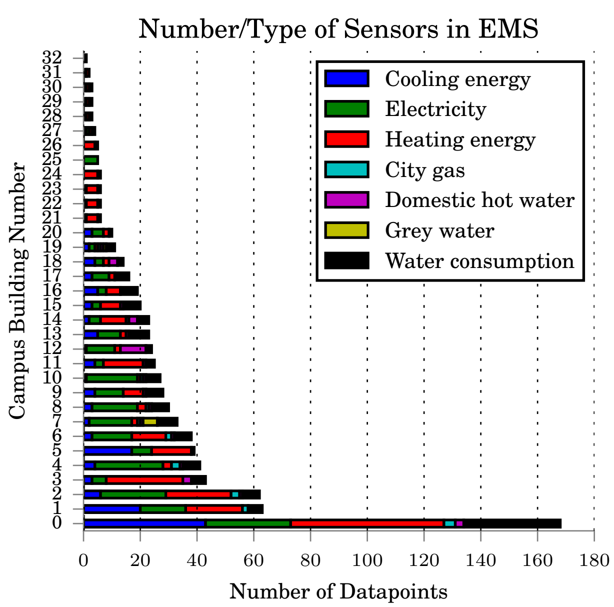

Figure 5. Quantity and type of data streams available from each building on the case study campus

In order to show the filtering and characterization steps visually, a subset of the EIS data is selected to apply the various filters and visualize them using a color mapping technique designed to show a high-level of information about the entire dataset as a whole without overwhelming an analyst with detail. The subset of data chosen for inclusion in this paper is 389 sensor streams measuring hourly energy readings in kilowatt-hours [kWh]. This dataset includes both electricity and thermal energy measurement devices from the heating and cooling systems. The intermediate screening steps are used to qualify the individual data streams for inclusion in the diversity profile generation step at the end of the process.

The first screening technique to be implemented is the data quality tests to calculate metrics according to data availability and general accuracy. Figure 6 shows an overview color mesh map of these tests applied to the 389 targeted energy measurement streams. The streams are sorted from bottom to top by increasing quality according to a summation of the metrics. Streams 0 to 130 contain a mixture of data streams with various types of gaps. The first obvious observation is that there are several streams at the bottom of the map in which a majority of the data is missing for much of the entire eight year time-span; insight that may be valuable to fix those particular sensors. Subsegments of the data can also be seen which have 'flat-lined' either due to sensor failure or to the measured phenomenon being turned off completely. Several groups of streams in this region can be observed which appear to have been installed at various stages in the last eight years.

The next screening process is the division of the streams according to weather sensitivity. This general application to the dataset can be seen in Figure 7 as a color map in which the Spearman Rank Correlation coefficient, ρ, is displayed for all data streams from -1, or highly negatively correlated, to +1, or highly positively correlated. The coefficient for each month has been calculated independently of the rest of each data stream and averaged across 6 months in order to visualize correlation across the entire year. This map has also been sorted from bottom to top according to increasing ρ to show the range of phenomenon occurring in the dataset. Approximately the bottom 70 data streams are highly correlated with heating; e.g., as the outdoor air drybulb temperature decreases, the energy consumption increases. The top 120 streams are positively correlated with cooling and seem to have more prominent seasonal correlation with weather. There are many non-weather sensitive streams in the middle which are not often as complete quality-wise.

The last screening process is the breakout detection. Figure 8 illustrates this step applied to the remaining data streams. The minimum span between breakouts is set to six months due to wide time range of eight years. This color map is sorted according to number of breakouts detected over the time range with the streams with more breakouts at the top and less at the bottom. This visualization shows that there is a wide range in terms of breakout behavior amongst the dataset from frequently changing sensors which show breakouts at every 6 month minimum to ones that are consistent across all eight years.

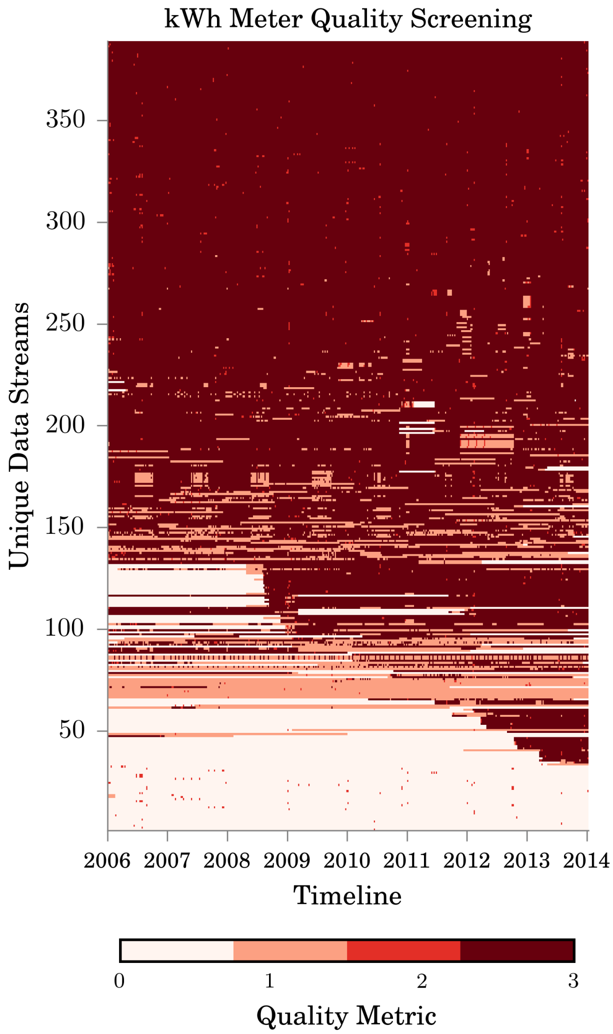

Figure 6. Data quality metrics map for campus sorted (bottom-to-top) according to increasing quality metric

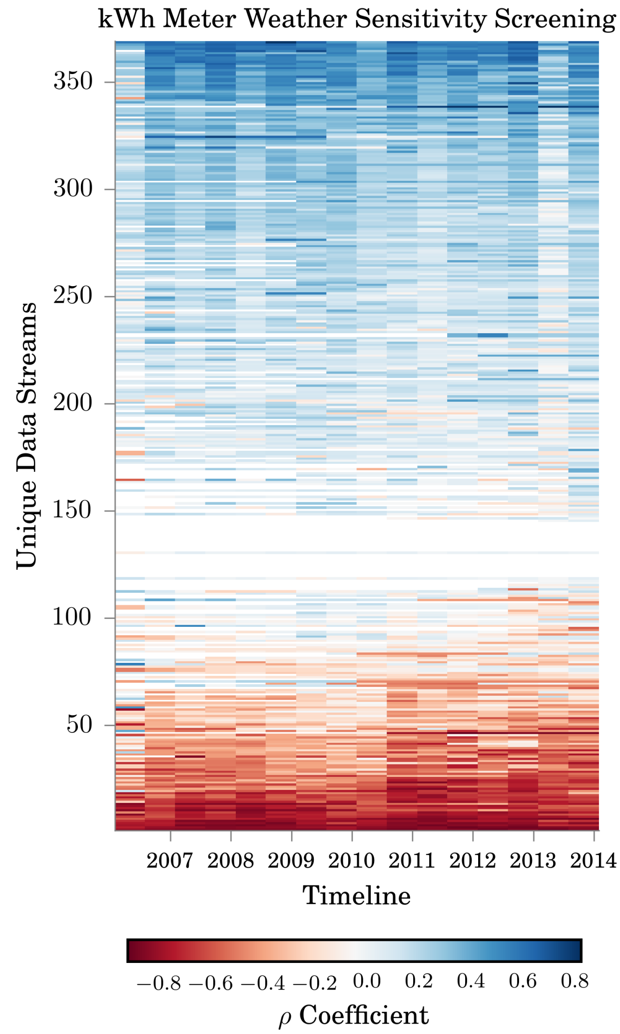

Figure 7. Weather sensitivity map sorted (bottom-to-top) from high negative to high positive ρ coefficient values

In order to analyze the strengths and weakness of the data streams based on the screening process, the metrics are combined and the values normalized to create a matrix of datastream attributes. Those attributes are then clustered to pinpoint subsets of data that are ready for typical schedule creation for simulation input. Figure 9 illustrates a hierarchical clustering of the data streams according to the mean of the quality metric, the mean weather sensitivity, and the max value of breakouts across the eight year time span. Prominent groups of sensors become apparent through this analysis as annotated on the figure. Region 1 includes data streams that are of low quality, mostly weather sensitive and have few breakouts. Region 2 are streams of high quality, varying weather sensitivity, but many breakouts. Region 3 has varying quality, is a majority weather insensitive, and has a moderate numbers of breakouts. Region 4 has varying quality, is mostly weather insensitive, and has a lower tendency for breakouts.

Region 3 and 4 are the most advantageous group of sensor data streams to investigate for this data are generally complete across many seasons (quality), not highly influenced by weather and have little need for normalization (weather insensitivity), and are relatively consistent over time (breakouts). These two sets of data streams are to be targeted in creation of diversity schedules for simulation feedback. The other regions have their strengths and weaknesses which need to be addressed through further analysis.

An example data stream from Region 4 is selected to illustrate the typical profile creation process. A set of optimal streams is queried from this region and a kWh meter is chosen that has a mean quality metric of 2.63, a mean weather sensitivity of 0.11, and 4 breakouts across the eight year timespan. This meter is also measuring a fairly large power load as its mean power consumption reading is 108 kWh. The process of SAX aggregation, discord and motif filtering is completed, and follow-up clustering is performed on the mo-

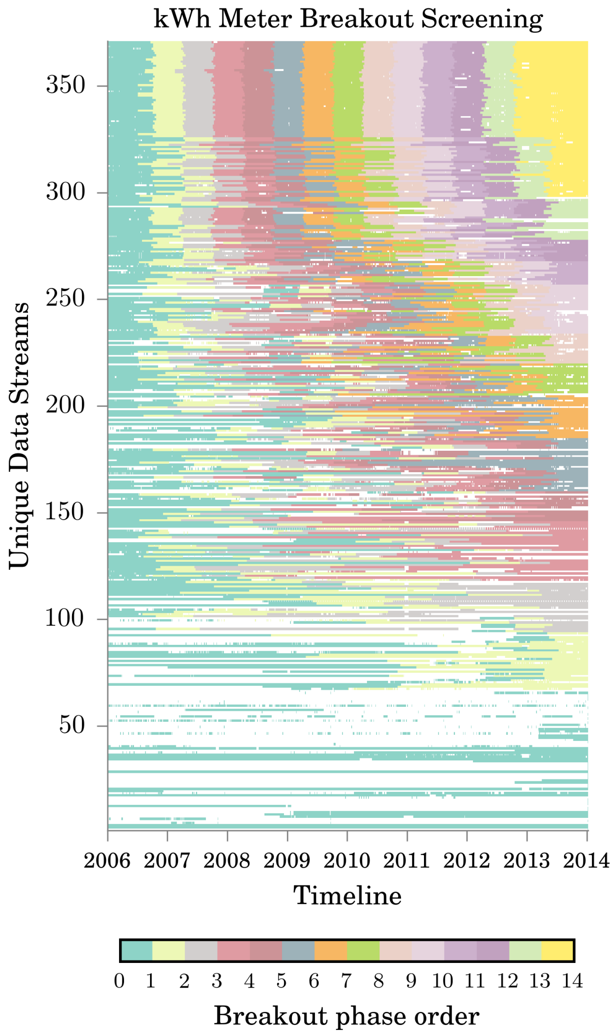

Figure 8. Breakout detection map sorted (bottom-to-top) according to increasing number of breakouts detected

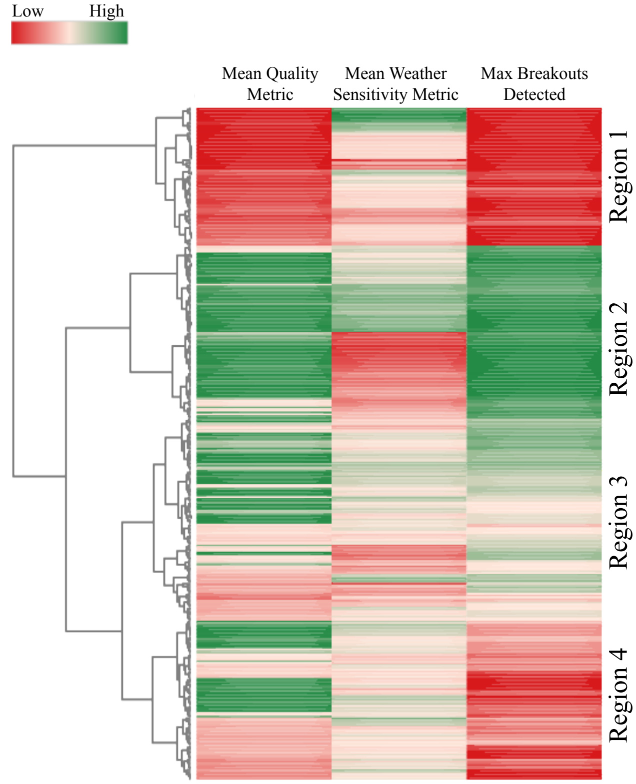

Figure 9. Clustering of energy measurement data streams according to screening metrics; each row represents a data stream

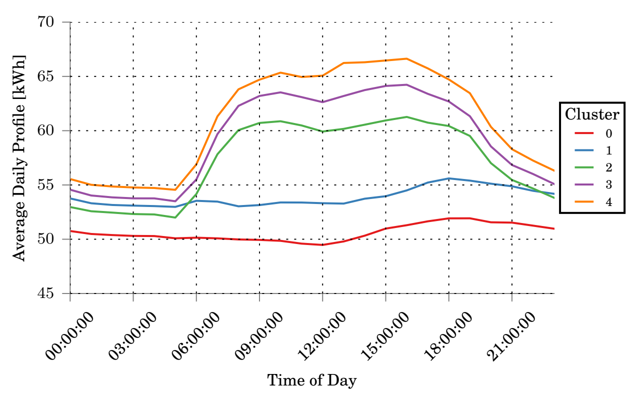

Figure 10. Example of a performance cluster profiles created by the typical profile creation process for a selected kWh meter from Region 4

tif candidates. A subset of the data is chosen from July 1, 2011 to January 1, 2014 as this data fits between the breakouts detected. Figure 10 illustrates the averaged daily profiles for each cluster across a 24 hour period. Figure 11 illustrates the resultant five typical profiles across the time range. The weather insensitive nature of this data stream is apparent as the daily totals modulate independent of the season. The profiles from this power meter follow a fairly conventional occupied versus unoccupied schedule. Figure 12 shows how the profiles are distributed across the days of the week. There are strong weekend clusters of type 0 and 1 and strong weekday clusters of types 2, 3, and 4. However, there are a few of these clusters that fall outside the conventions and are likely holidays. This example is now ready for further transformation into a simulation model input according to industry standards [1]. The revelation of these typical profiles is automatic and uninfluenced by infrequent daily patterns from the larger dataset.

CONCLUSION

A process of knowledge discovery is shown that is designed to automatically evaluate the applicability for simulation feedback for measured data streams within a large, raw EIS dataset. The analysis is shown of 389 power meters streams collected across eight years of data at a frequency of 60 minutes. The total data collected from these sensors is over 22 million measurements, an amount far beyond the capability of manual analysis. The novelty of this work is it showcases techniques which empower an analyst to screen this data before more manual approaches and before further data collection is planned. All of these screening steps can be imple-

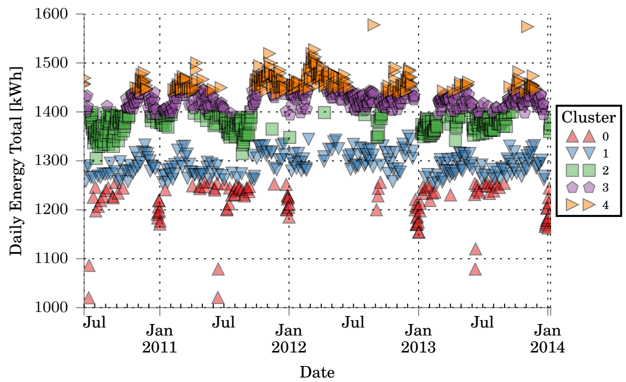

Figure 11. Performance clusters daily totals across time range

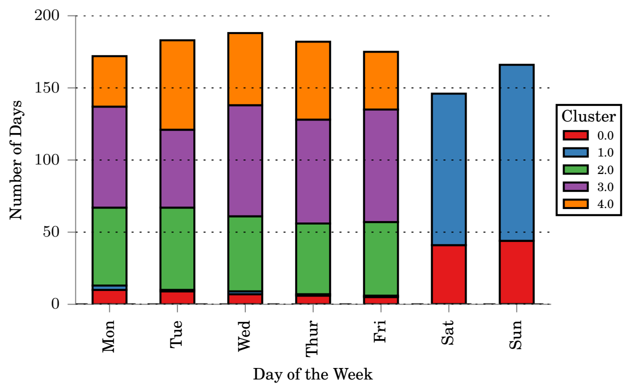

Figure 12. Number of instances of each performance cluster across the days of the week

mented by setting only a few parameters and the speed is only limited by the computational power in running the algorithms. While precise computation times were not kept for this study, it took approximately 30 minutes for all the screening steps to execute on a Mac OSX laptop computer for the 389 data streams. This process should save a lot of manual analysis and give more insight into the available dataset.

The dataset analyzed will continue to be developed as part of a larger study to model the campus with an urban scale modeling tool. Many of the Region 3 and 4 data streams are to be converted into diversity schedules for this simulation process and the comparison between manually generated simulation inputs and those created automatically in this process will be compared.

Process Replication

The data, code and additional explanation of these steps are available for download from http://www.datadrivenbuilding.org. Replication is possible using a series of IPython notebooks and a sample anonymized subset of the data.

ACKNOWLEDGMENTS

Funding for this project was provided by an ETH Zürich Institute of Technology in Architecture (ITA) Fellowship and the Swiss Competence Center Energy and Mobility (CCEM) through the Urban Multi-scale Energy Modeling (UMEM) project.

REFERENCES

1. Abushakra, B., Sreshthaputra, A., Haberl, J. S., and Claridge, D. Compilation of Diversity Factors and Schedules for Energy and Cooling Load Calculations, ASHRAE Research Project 1093-RP, Final Report. *Energy Systems Laboratory, Texas A&M University* (2001).
2. Basseville, M., and Nikiforov, I. V. Detection of abrupt changes: theory and application, 1993.
3. Coughlin, K., Piette, M. A., Goldman, C., and Kiliccote, S. Statistical analysis of baseline load models for non-residential buildings. *Energy and Buildings* (2009).
4. Davis III, J. A., and Nutter, D. W. Occupancy diversity factors for common university building types. *Energy and Buildings* (2010).
5. Effinger, M., Friedman, H., and Moser, D. Building Performance Tracking in Large Commercial Buildings: Tools and Strategies - Subtask 4.2 Research Report: Investigate Energy Performance Tracking Strategies in the Market. Tech. rep., Sept. 2010.
6. Granderson, J., Lin, G., and Piette, M. A. Energy information systems (EIS): Technology costs, benefit, and best practice uses. Tech. Rep. LBNL-6476E., Dec. 2013.
7. Khan, A., and Hornbæk, K. Big data from the built environment. In *Proceedings of the 2nd international workshop on Research in the large (LARGE '11)*, ACM (Beijing, China, 2011), 29–32.
8. Miller, C., Nagy, Z., and Schlueter, A. Automated daily pattern filtering of measured building performance data. *Automation in Construction* (2015).
9. New, J. R., Sanyal, J., Bhandari, M., and Shrestha, S. Autotune e+ building energy models. In *Proceedings of the 5th National SimBuild of IBPSA-USA*, IBPSA, Ed. (2012).
10. Raftery, P., Keane, M. M., and Costa, A. Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings* (2011).
11. Škuta, C., Bartůněk, P., and Svozil, D. InCHlib–interactive cluster heatmap for web applications. *Journal of cheminformatics* (2014).
12. Twitter. AnomalyDetection R Package. *https://github.com/twitter/AnomalyDetection*.
13. Twitter. BreakoutDetection R Package. *http://github.com/twitter/BreakoutDetection*.