When Data Analytics Meet Site Operation: Benefits and Challenges

David H. Blum
Guanjing Lin
Michael Spears
Janie Page
Jessica Granderson

Lawrence Berkeley National Laboratory

Energy Technologies Area
August, 2018

For citation, please use 10.20357/B7KS39
Disclaimer:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
When Data Analytics Meet Site Operation: Benefits and Challenges

David H. Blum, Guanjing Lin, Michael Spears, Janie Page, and Jessica Granderson
Lawrence Berkeley National Laboratory

Abstract

Demand for using data analytics for energy management in buildings is rising. Such analytics are required for advanced measurement and verification, commissioning, automated fault-detection and diagnosis, and optimal control. While novel analytics algorithms continue to be developed, bottlenecks and challenges arise when deploying them for demonstration, for a number of reasons that do not necessarily have to do with the algorithms themselves. It is important for developers of new technologies to be aware of the challenges and potential solutions during demonstration. Therefore, this paper describes a recent deployment of an automated, physical model-based, FDD and optimal control tool, highlighting its design and as-operated benefits that the tool provides. Furthermore, the paper presents challenges faced during deployment and testing along with solutions used to overcome these challenges. The challenges have been grouped into four categories: Data Management, Physical Model Development and Integration, Software Development and Deployment, and Operator Use. The paper concludes by discussing how challenges with this project generalize to common cases, how they could compare to other projects in their severity, and how they may be addressed.

Introduction

A large focus has been put on the energy consumption of buildings, as they account for approximately 40% of primary energy consumption in the U.S. (EIA 2018). Methods to reduce this energy consumption during operation, improve system reliability, and provide other load management services, for instance for interactions with electric grids, are increasingly using data analytics and are included in what are known as advanced energy management and information systems (EMIS) (Granderson and Fernandes 2017). These methods include advanced measurement and verification, automated fault-detection and diagnosis (AFDD), and optimal control. Advanced measurement and verification, or M&V 2.0, refers to the use of automated analytics in combination with higher granularity data to quantify project energy savings (Granderson and Fernandes 2017). AFDD detects when operation diverges from intended operation and determines the cause for such divergence, such as in Granderson et al. (2017). Optimal control determines control setpoints or actions to meet desired performance objectives, such as energy or cost minimization, while providing an acceptable level of service, such as in Li et al. (2015) or De Coninck and Helsen (2016).

The methods can provide many benefits as outlined previously. However, they also commonly require the development and deployment of software and hardware to collect and store data, run analytics algorithms, and present data and results to owners, operators, and other users. In addition, they require the willingness of the owner and operator to implement, use, and aid in debugging and maintenance. Practical challenges associated with these tasks are often not discussed in the presentation of pre-commercial and demonstration work on EMIS with advanced data analytics in building applications. Doing so would enable future projects at a
similar stage to learn from others’ past experience and provide solutions in a more efficient manner.

Therefore, this paper will focus on the practical challenges and solutions utilized during the development, deployment, and field-testing of one such advanced EMIS, called PlantInsight. The PlantInsight project was chosen for this case study because it was a recently completed project that went through both development and real site demonstration, allowing the authors to discuss challenges that span both of these common EMIS research and development stages. First, an overview of the tool is given, highlighting the benefits it provides to operators, including data visualization, AFDD, and optimal control capabilities. This first section also discusses the realized benefits of the tool in terms of energy savings and operator satisfaction. Then, four main groups of challenges are presented, along with the solutions used by the project team to mitigate these challenges. These groups are Data Management, Physical Model Development and Integration, Software Development and Deployment, and Operator Use. Finally, the paper discusses how challenges on this project generalize to common cases, how they could compare to other projects in their severity, and how they may be addressed.

**Tool Overview**

**Site Description**

The PlantInsight tool was developed for campus chilled water plants and was field-tested in ASHRAE Climate Zone 4 as part of this project. The campus chilled water loop is served by two chiller plants: one which contains two 1,250-ton chillers, one 2,500-ton chiller, and four two-cell cooling towers and one which contains three 2,500-ton chillers and three two-cell cooling towers. The plants were built in 2006 and represent a state-of-the-art central chilled water plant, with variable frequency drives on cooling tower fans and secondary water pumps and a Johnson Metasys® Building Automation System (BAS) providing good system monitoring and data collection. The chillers are operated in stages to provide 42 °F +/- 2 °F chilled water and the cooling tower fans are modulated to maintain a 72 °F +/- 2 °F condenser water setpoint temperature.

**Functionality**

There are four main functions of the tool: data collection and storage, graphical user interface (GUI) with data visualization, optimal control, and AFDD. Data is collected by the site BAS and stored in a Microsoft SQL Server located at the site. A custom-written Java 8 program pushes the data from the site server to a PostgreSQL database located on the PlantInsight server at LBNL. Finally, the PlantInsight tool processes all of the data, including cleaning, before restoring it back into the database.

The GUI with data visualization is implemented by a browser-based JavaScript, which interacts with backend scripts for running analyses through a REST API. The GUI front page displays for the time period specified by the user a plot of total load and load from each plant, the plant efficiency in kW/ton, a weather and load forecast for the next 24 hours, and the total cost of operations, total consumption, maximum load, and number of current faults. Additional pages provide more detailed data from the optimization and fault detection features of the tool,
including the current optimal setpoint, optimized vs. conventional power consumption estimates, time periods when faults are detected, and fault information about specific pieces of equipment.

The development of optimization and AFDD algorithms of the tool were previously described in (Granderson et al. 2016), along with challenges specific to the model development and calibration for physical model-based analytics. However, the algorithms are summarized here for clarity. Optimal control is achieved through model-predictive control (MPC) of the condenser water setpoint temperature. With this strategy, a model of the chiller plant operation is used to calculate the energy consumption of the plant for the following day of operation and an optimization algorithm is used to find the condenser water setpoint temperature that minimizes the energy consumption for the day. The chiller plant model is written in Modelica (Mattsson and Elmqvist 1997), an equation-based, declarative, open-source language specification that specializes in the simulation of physical systems, primarily using components from the Buildings Library (Wetter 2014). The optimization algorithm is invoked using GenOpt (Wetter 2001). The inputs to the model are forecasted ambient air temperature, relative humidity, and campus load. The campus load is predicted using a linear regression model with input variables of hour, day, month, and ambient air temperature. The plant model was calibrated using historic site performance data.

AFDD is achieved through both data-driven and physical model-based methods. Data-driven methods are used to detect chiller compressor and cooling tower fan cycling faults, while physical model-based methods are used to detect chiller efficiency faults. For cycling faults, compressor current and fan speed data are used to determine if the equipment is making the transition between on and off greater than a specified number per hour; eight for the chillers, ten for the fans. The chiller models used in the plant models described previously and written in Modelica are used to detect chiller efficiency faults by comparing the COP estimated from measured data with the COP calculated using these chiller models. The measured COP is estimated using a Bayesian nonlinear state estimation technique called Unscented Kalman Filtering (UKF) and is described in more detail in Bonvini et al. (2014a and 2014b). Finally, detected faults are reported as faulty periods if the persist for 90 minutes, and multiple faulty periods are aggregated into a single faulty period if they occur within two hours of each other.

**Operation**

The tool is run on a server located at LBNL, to which the plant operators have access through the browser-based GUI. Each morning, at 4:00 AM ET, the data collection, optimization, and AFDD scripts are run to update the data in the tool database, from which the GUI updates key performance indicators, AFDD results, and optimization results. At the conclusion of the optimization, an email is sent to the plant operators and tool developers indicating the determined optimal condenser water setpoints for each plant for the coming day. At the operator’s discretion, the optimal setpoint can be implemented manually. The choice of a single setpoint change per day was a request from the plant operators. Therefore, the optimization is set so that the condenser water setpoint is constant for the day.

**Realized Benefits**

The key objectives of the tool were to a) reduce energy consumption and associated greenhouse gas emissions of the plant and b) maintain or improve operator experience compared
to the existing BAS interface. Both of these objectives were met, though to a varying degree. Plant energy consumption and greenhouse gas emissions were reduced by an estimated 17% over a four-day period in April when the recommended optimal setpoints were actually implemented at the site. An annual simulation of the tool optimization indicated, however, that significant energy savings are only available in the winter and shoulder periods due to the humid climate. Such a climate prevents condenser water setpoint temperatures to be low enough to reduce total plant energy consumption. Therefore, the simulated annual energy and greenhouse gas emission savings were found to be less than 2%. However, the potential for savings using the tool would exist for drier climates. In addition, due to low implementation costs, the payback period of the tool was determined to be less than 1.5 years.

The objective to improve operator experience was met by the tool based on an interview and survey conducted by LBNL with the lead plant operators and managers. Overall, the users indicated that the tool enabled them to more efficiently operate the plant by seeing load predictions and understanding how condenser water setpoint changes affect energy consumption. In the survey, the lead plant operator and manager rated the GUI, optimization and AFDD outputs, and overall tool design each at a four on a scale of one to five, with five being the highest level of user of satisfaction.

Deployment Challenges and Solutions

Data Management

Data management is a problem seen across projects implementing advanced building controls and diagnostics applications. First, data is often collected from a number of separate monitoring systems and not stored in a central location. This is especially true if additional sensors need to be added for the project and if new data is created by analytics programs themselves, such as key performance indicators or control signals. Such problems complicate the development, maintenance, and expansion of analytics and dashboard programs if multiple data reading and writing interfaces need to be used for each routine. Second, data needs to be interpreted and mapped to the appropriate routines. This can be particularly challenging if data is not tagged consistently or appropriately. Third, data quality may not be uniform across collected points, with points having different resolution and frequency of dropouts. These issues can make it difficult to make use of or visualize multiple data streams simultaneously and require methods for data filling, data replacement, and error handling. Data filling methods could include interpolation or padding. An example data replacement method is using power consumption in place of status signal. Error handling is especially needed to ensure critical software processes do not terminally fail when missing data is encountered. Finally, cybersecurity needs to be considered on all projects, though may vary depending on the application, and the relative stringency in requirements for organizations from the private sector, government agencies, or military installations. This includes protection of the server on which data is stored, encryption of data sent across communication networks, and accessibility to control systems. Not only can accommodation of security requirements slow development and deployment, varying requirements can impede seamless transfer of developed software across projects from one site to another.

The PlantInsight tool mitigated the data collection and storage challenge by establishing a central database in which system monitoring and analytics program data is stored and from...
which analytics and GUI data can be retrieved. The database was created separately from the database on which the site collected data to limit reliance on and interference with site operations. The general architecture is presented in Figure 1. The tool’s central data storage used a PostgreSQL (PostgreSQL 2018) database, an open-source, relational database (RDBMS) that is supported on Linux, macOS (and other UNIX-based platforms), and Windows. A number of different types of databases exist, where the selection of one versus another may depend on the volume of data, access speed, type, and structure. PostgreSQL was found to meet the needs of storing project data, both time series and json-like data structures, efficiently enough for the application and eliminated any need for maintaining multiple databases.

![Fig 1. - Relationship of site and PlantInsight software with arrows depicting communication. The tool utilized a local database to store a copy of site monitoring data and data generated from analytics algorithms. Site data was pushed to the tool database, while data from the tool database was accessible to the GUI and analytics algorithms.](image)

In terms of monitoring infrastructure, the PlantInsight project was fortunate to not have to install any new monitoring equipment at the site, and all monitoring data was already stored on an onsite database. Therefore, a single interface program had to be written to copy selected monitoring data from the site database to the PlantInsight database. One additional interface program was written for PlantInsight to collect weather forecast data for the optimization feature. Monitoring data did suffer from some quality issues, where data resolution varied across data streams and some data streams experienced both brief and significant outages. To mitigate these challenges, the PlantInsight tool preprocessed all monitoring data, resampling and interpolating missing values where necessary to a consistent five minute interval, and stored this version of all data in the database. It was this version of clean data that was used within the analytics and GUI software. In addition, error handling was added to PlantInsight software to ensure that missing data did not cause critical components of the software to terminate in failure. The project also benefited from prompt and consistent communication with site staff about missing data, which occurred due to site maintenance and network communication issues.

Finally, cybersecurity limited the ability for PlantInsight to directly control plant condenser water setpoints once the optimization was completed. Plant managers wanted to maintain human-in-the-loop control over all aspects of the plant to prevent harmful operation. Therefore, optimal setpoints were emailed to operators once per day for implementation.

**Physical Model Development and Integration**

For advanced EMIS using physical models to analyze performance or generate new control signals, the development of such models and integration of them into software design can be a challenge. Development involves identifying the relevant physics to be modeled, making
the appropriate assumptions to trade-off accuracy and complexity, writing the equations, and determining a method for solving the equations. Integration of the models into analytics software must consider in which language or on which platform the model is written, how the model is solved or simulated, and how this process is invoked by the analytics software.

To mitigate these challenges, PlantInsight utilized a state-of-the-art modeling language and standard. For model development, an open-source, declarative, equation-based, object-oriented, language called Modelica (Mattsson and Elmqvist 1997) was used. The open-source specification allows compilers to be written by any entity, commercial or also open-source. The declarative and equation-based characteristics of the language enable equations to be written in natural engineering form and for solution methods to be written separately from the equations, allowing engineers to focus on the modeling of the system, and experts in mathematics and computer science to focus on the simulation or optimization of such models. Finally, object-orientation allows for libraries of sub-components to be created and shared, which can then be used to build larger system models, preventing the modeler from having to start from scratch. The chiller, cooling tower, and whole-plant models developed for PlantInsight used many sub-components from the Modelica Standard Library (Modelica Association 2018) and Buildings Library (Wetter et al. 2014), as depicted in Figure 2.

![Modelica model for two parallel chillers with chilled water inlet and outlet ports and condenser water inlet and outlet ports, which can be connected to campus distribution and cooling tower models.](image)

For model integration, one standard that was utilized by PlanInsight was the Functional Mockup Interface (FMI) (FMI 2018). FMI standardizes the interface of physical models, including how to set inputs, retrieve outputs, package solvers, invoke simulation, and perform...
other necessary operations. Therefore, any software tool supporting the FMI standard has the potential to integrate any model that is packaged as a Functional Mockup Unit (FMU). This reduces the need for specialized and/or proprietary software to be integrated in the tool for modeling purposes. The chiller models used for efficiency fault detection in PlantInsight were developed using the commercial Modelica development environment Dymola (Dassault Systemes 2018), but were exported as FMUs and used within the open-source and free Python-based software package for UKF, EstimationPy (Bonvini et al. 2014c).

**Software Development and Deployment**

Software development can be a challenge when code needs to be written and maintained by multiple simultaneous contributors, or across the project lifetime. First, it requires avoidance of two people changing the same portion of code in incompatible ways and the ability for someone to efficiently update their version of code with edits made by someone else. It is also useful to know the bug fixes or new features that are being worked on by other contributors, along with the progress being made towards completion. Then, it requires keeping track of the versions of code which are considered stable and deployment-ready and which versions are considered in-progress and not yet completed. In addition, it requires traceability of who made specific changes so that issues later can be addressed by the appropriate people. Finally, it is important for contributors to understand the overall architecture of the software, the role each piece plays, and how it relates to other pieces. Software deployment can be a challenge when deployment is on servers or platforms that are different from where the software was developed. In this case, the computing hardware can be different, such as the speed, number, or cores of processor(s), amount of memory, size of the hard-drive, or network accessibility. In addition, the software environment can be different, such as the operating system or the presence of and version of support software.

To address the challenge of collaborative software development, the PlantInsight project utilized Git (https://git-scm.com/sfc), an open-source and free software versioning control system supported on Linux, Mac OS X, and Windows. Git keeps track of changes made to files within a particular directory, called a repository. The changes are tracked in a way that enables labeling of software versions, seamlessly transferring changes between versions, and repealing changes. In this way, a version of the software can be considered the deployment-ready version, typically called the “master”, with other development versions being worked on in parallel by any number of contributors, typically called “branches.” The changes “committed” to branches can be transferred to the master or to other branches, with Git checking for and aiding in resolution of portions of code simultaneously edited, called “conflicts.” This workflow is depicted in Figure 3. Repositories can be maintained online for free or for payment with varying degrees of privacy through sites such as GitHub (2018) or Bitbucket (Atlassian 2018). Often, these online maintainers also offer other collaborative development features such as issue tracking and discussion forums. For PlantInsight, Git was particularly useful for developing new features and bug fixes on separate branches from the master deployment version and transferring the required edits to this deployment version when ready.
To address the challenge of software deployment, the PlantInsight project utilized Docker (2018), a software that supports the development and deployment of software containers. Containers are lightweight, executable software packages that contain all requirements to run particular software, including the operating system, libraries, settings, and other support software, with specific versions. These requirements can be programmatically specified in Dockerfiles. In this way, a container can be created and/or deployed on any computing resource with Docker software, which is supported by Linux, Mac OS X, and Windows, as depicted in Figure 4. For the PlantInsight project, this was particularly useful for deploying the application on different servers from the deployment server for development, testing, and debugging and will also aid in transferring the software to other deployments in the future.

Operator Use

Final challenges that can be encountered during any project are those associated with use of the new software by site operators and their level of engagement during installation, debugging, and maintenance. Operators can be hesitant to trust the results of software aimed at determining new control strategies or too busy with other daily operational tasks to truly test the new capabilities offered by the software. Failure to report and help address issues encountered during installation and operation can quickly derail the deployment altogether.

The PlantInsight project benefited from a high level of interest, engagement, and communication between the project development team and site operations management and staff. Phone conferences, iterative needs assessments, and mockups that occurred continuously
throughout development between these parties enabled feedback from operators to the development team about which portions of the tool worked well, which needed improvement, and which new features should be added. For example, performance metrics that were not central to the optimization and AFDD were not displayed on the dashboard initially, but were requested by the operators and added by the development team. In addition, the timing, frequency, and boundaries of optimal setpoint calculations were determined through discussions about how the site operators were comfortable operating the system and piloting the tool. Finally, the site responded promptly to inquiries by the development team about missing performance data, checking their monitoring systems to determine if the issue was with the monitoring system itself or the transfer of data to the PlantInsight tool. Overall, the high level of engagement led to a high level of satisfaction with the tool by the site management and operations staff and general success of the project.

Discussion

This paper has described a number of challenges that were faced during the development and deployment of PlantInsight, as well as the solutions that were used to mitigate them. While these solutions helped lead to a successful demonstration, the circumstances of this project also helped avoid challenges that may be present on others. For example, the project benefited from the presence of a state-of-the-art monitoring and data collection system at the site before the project began relieved a large amount of effort that could go into specifying and installing such a system on other projects. Another circumstance of the project that enabled some of the solutions was the presence of Modelica modeling and software development experts on the project team. Modelica has not yet become commonplace in the U.S. and requires some time for a user to learn. The project team utilized in-house expertise in the use of Modelica for building energy modeling to avoid time required for a modeler to learn and challenges stemming from modelers’ lack of experience that may be present on other projects. In addition, not every project team in the building research or industry communities has expertise in software development to leverage. For example a project team could be composed only of mechanical engineers with expertise in the analytics algorithms being developed. Such software development expertise on the PlantInsight development team greatly aided in the choice of database, use of Git and Docker, and overall functionality of the software. Finally, the effect of the high level of interest and engagement by site staff on this project is something that may not be shared on others. Feedback on development, prompt responses to debugging requests, and a willingness to test the tool went a long way in alleviating additional challenges.

Conclusion

A focus on building energy consumption has prompted many pre-commercial and demonstration projects to develop advanced EMIS with analytics that can provide owners and operators advanced, new techniques for FDD and model-based optimization. While these systems have the potential to provide many benefits, there can be many practical challenges to their development and deployment. This paper presented the challenges faced during the development and field-testing of one such advanced EMIS called PlantInsight, which provides data visualization, AFDD, and optimal control for a campus chilled water plant. The challenges were grouped into four categories; Data Management, Physical Model Development and
Integration, Software Development and Deployment, and Operator Use. For each category, solutions utilized by the PlantInsight project team were described, such as the creation of a central database for the tool, the use of state-of-the-art physical modeling standards called Modelica and FMI, utilization of Git and Docker software for collaborative development and rapid deployment, and a high level of engagement with the site operations management and staff.

While some of these solutions were facilitated by project team expertise and site characteristics specific to the project, they may still generalizable and applicable to other projects. For example, the value that Git and Docker provide for software development, or that Modelica and FMI provide for model development and integration, may be worth the time investment even in cases where the development/demonstration team is less familiar with their use. In addition, some of the site characteristics that were found beneficial in this project, such as a high level of interest and engagement by operations staff as well as the presence of a state-of-the-art metering infrastructure, could be considered as prerequisites for site selection to alleviate challenges in later stages of a project.

Acknowledgements

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors would like to thank the Environmental Security Technology Certification Program for supporting the work in this study, with special thanks to Timothy Tetreault (Noblis), Christopher Crouse (NAVFAC HQ), Tim Rath (USNA), John Barton (USNA), Wangda Zuo (University of Colorado, Boulder), Michael Wetter, and Stephen Czarnecki (all of LBNL), and Daniel McQuillen (McQuillen Interactive).

References


