Real-time Model Predictive Control with Digital Twins and Edge Computing Technologies

Michael Berger [a], *Filippo* Bernardello [a], *Craig* Barry [a], *Pravesh* Badjoonauth [a], *Shruthi* Balaji [a], *Michael* Lakhdar [a]

[a] Conserve It, Melbourne, Australia, info@conserveitiot.com

Abstract. Buildings consume almost a quarter of Worlds Energy Consumption and hence are one of the major sources of emissions globally. In commercial buildings, HVAC is by far the most energy intensive system, accounting for close to half of the total energy consumption. For this reason every efficiency improvement in HVAC performance can significantly reduce the energy profile of the building, turning HVAC optimisation into a core requirement to deliver energy efficiency. Fundamental to optimising large energy consumers in today's modern buildings is the use of Machine Learning in order to dramatically improve the energy efficiency of modern central cooling and heating plants. This paper will demonstrate the techniques that have been implemented to deliver advanced Real-time Model Predictive Control on Edge Computing solutions that don't require Cloud connectivity or significant computing power. Through the use of deep domain knowledge and advances in Edge Computing, it is possible to 'learn' highly accurate models of how mechanical machines operate and apply those models to predict and then solve complex optimisation problems for advanced control and improvements in energy efficiency. The authors will show how, through the collection of real-time sensor data, our platform has successfully reduced energy consumption and electrical demand in real buildings without compromising space comfort in any way at all. The capability to generate self-adjusting control algorithms in an on-premises scenario not only delivers significant outcomes but lowers overall Total Cost of Ownership for the end client. The absence of ongoing subscription fees further improves the economic model and the case for on-premises, real-time, model predictive control. Furthermore, the paper will demonstrate how the same Digital Twins used for Model Predictive Control can be used for anomaly detection algorithms or Fault Detection and Diagnosis as well as Predictive Maintenance and that this will create new service opportunities and business models for smart companies of the future whilst continuing to deliver optimal performance of mechanical systems.

Keywords. Energy, Design of Innovative HVAC systems for optimized operational performances, Digitization, Digitization in HVAC control & Health Monitoring.
DOI: https://doi.org/10.34641/clima.2022.368

1. Introduction

Chilled water plants are widely used globally for air-conditioning and refrigeration applications, and account for a large proportion of commercial buildings' energy usage. As was highlighted again recently at the COP26 conference in Glasgow, curtailing CO2 emissions-driven global warming is one the biggest challenges of our time [1]. Reducing chilled water plants energy usage would contribute to this goal. Equipment efficiency has improved over the years, and variable speed drives (VSD) have been introduced to improve the part-load efficiency of the chillers, pumps, and cooling towers, and are now almost ubiquitous in new plants deployed around the world. However, those have also introduced many new variables to adjust by the supervisory control systems, such as the condenser water flow at each chiller for instance, and conventional controls solutions are not equipped to select the optimal setpoints in real-time as conditions vary throughout the days and seasons, and therefore do not allow these efficient machines to run together to their highest potential. Offline simulations have been carried out before to recommend some "rules-of-thumbs" that can be implemented in conventional control systems [2, 3], but those may not generalize well to every equipment sizing, performance, and systems design.

In this study, a deployable on-premises real-time model predictive control approach to address these challenges and minimize the energy usage of the chilled water plant is presented and assessed with actual site data and operation. It is intended to operate on site on the edge on an embedded controls platform, as displayed on Fig. 1.

Fig. 1 – Embedded Edge hardware for real-time monitoring and controls of live equipment

2. Modelling approach

The setpoints of interest in this study are:

- The leaving chilled water temperature (LCHWT) at each chiller
- The condenser water (CW) flow at each chiller

The former will dictate the loading of each chiller, when keeping the overall plant LCHWT as per specifications, and allow to set each chiller at or near their most optimal part-load point or "sweet-spot". The latter will tackle the trade-off between CW pump power (lower flow will result in lower power) and chiller power (lower flow will result in higher refrigerant discharge pressure and therefore higher power).

It is important to note that neither of those would impact the chilled water production of the plant, as the chilled water flow and temperature leaving the plant are not affected, and therefore would not compromise comfort in the building.

2.1 Optimization Formulation

The optimization problem to solve as part of the Model Predictive Control approach is to minimize the power usage of the chillers and the CW pumps, as defined in Eq. (1).

$$\begin{aligned}
&\underset{Load_i,\ Flow_{CW,i}}{\text{Minimize}} \sum_{i=0}^{n} P_{chiller,i}(Load_i, Flow_{CW,i}) \\
&\qquad\qquad + P_{CW\ pump,i}(Flow_{CW,i}) \\
&\text{s.t.:} \sum_{i=0}^{n} Load_i = Load_{Plant,Target} \\
&\quad Load_{Min,i} \le Load_i \le Load_{Max,i} \\
&\quad Flow_{CW,Min,i} \le Flow_{CW,i} \le \\
&\quad Flow_{CW,Max,i} \qquad (1)
\end{aligned}$$

The decisions variables here are the cooling load and the CW Flow of each chiller. The LCHWT of each chiller can be derived from their load and the overall plant LCHWT target. The problem is treated as steady-state as it only selects setpoints that are then fed to lower-level control loops tasked to reach and maintain them, and also due to the absence of other time-dependent state variables such as storage-related for instance. As such, the optimization is carried out on a single point short-term horizon, with the target overall plant load being computed based on the measured and recorded cooling load and regression-based prediction of its future value in the near future. The computed optimum decision variables can then be fed to the lower-level systems: The LCHWT setpoint command is communicated to the chillers via high-level or low-level communication, and the CW flow setpoint can be passed to a PID controller with the CW pump speed as output command. The optimization is repeated on a regular basis, every 1 minute or less, to account for continuous changes of conditions on site.

Additionally, this mathematical program can be repeated for each number of chillers in the plant and used to select the most efficient number of chillers. It is also flexible and can be simplified to only include one set of decision variables, in case of site limitations, for instance if the CW pump speed is not equipped with VSD.

2.2 Equipment models

The power usage of a chiller is modelled as a 2nd order multivariate polynomial function of its cooling load; the difference between the entering condenser water temperature (ECWT) and the LCHWT; and the CW Flow, as per Eq. (2).

$$P_{chiller} = f(Load, T_{ECW} - T_{LCHW}, Flow_{CW}) \qquad (2)$$

Chiller models have been proposed in [4, 5]. The selected model follows a similar direction with some modifications to further suit the requirements of live learning with potentially limited data, and of the optimization problem. Indeed, using the difference of water temperatures allow to capture the variation of power usage due to compression ratio without having to record significant data coverage of both chilled water and condenser water. Additionally, the use of the CW Flow and ECWT separately simplifies the optimization problem as they are both controllable variables, whereas LCWT is itself a function of the chiller power and would result in a recursive function that would add unnecessary complication to the downstream optimization algorithm. For air-cooled chillers, or water-cooled chillers with no CW Flow metering, the model can be simplified by omitting the CW Flow variable.

The efficiency of the chiller (and by extension the plant) is usually reported as Coefficient Of Performance (COP), following Eq. (3).

$$COP = \frac{Load}{Power} \quad (3)$$

The power usage of a CW pump is modelled as a 2nd order polynomial function of the CW Flow, as in Eq. (4)

$$P_{CW\ pump} = f(Flow_{\ CW}) \quad (4)$$

The above model is applicable to cases where there is a 1-to-1 relationship between chillers and pumps in the plant (often referred to as "dedicated" in the industry). Note that plants with 1-to-many chiller-pump relationship have been addressed by the authors separately but are not included in the present study.

2.3 Machine learning approach

The machine learning solution is intended to run autonomously on site and learn the performance of the site's equipment live, without human interaction. As such, one of the challenges is to handle cases where limited data is available, whether it is due to a small sample size or the dataset only spanning a small area of the operating range. After several previous site deployment reviews it was found that in numerous cases, the autonomously learnt model exhibited undesirable features, such as negative power values within the operating range in areas where little to no data had been available.

In addition to further data pre-processing, a solution that was investigated was to add constraints to the least-square learning of the model to further leverage the subject-matter experts' knowledge. Expected behaviours – for instance that the power usage of the equipment should always be positive in the operating range or that the power usage of a pump is expected to increase when the flow increases and all other values being equal – are embedded as linear constraints to the learning.

The resulting constrained least-square problem of a polynomial model with linear constraints is a standard convex Quadratic Program (QP) as defined in Eq. (5) and for the number of variables of the application (order of magnitude of 10 variables) can be solved in real-time well below 1 second on modern embedded hardware using state-of-the-art algorithms, such as an Interior Point method [6].

$$\begin{aligned} &\underset{x}{\text{Minimize}}\ \|Ax - b\|_2^2 \\ &\text{s.t.}: c\ .x\ \leq 0 \end{aligned} \quad (5)$$

In Eq. (5), the matrix A contains all the input data, the vector x the model coefficients, the vector b the output data (power usage in this case), and the vector c the linear constraints features.

2.4 Solving the optimization problem

The simplest but most computationally costly way to solve an optimization problem is an exhaustive search, a method that cycles through all the possible combinations of all the decisions variables. In the present application, the computational complexity grows exponentially with the number of variables, and even with a small number of chillers in the plant it quickly becomes impractical for real-time controls. For instance, with a 3 chillers plant, which results in 6 decision variables (3 loads and 3 CW flows), and assuming 100 points being checked for each variable, it would result in 100^6, or a trillion, iterations, each of which consists of several operations to compute the total power to be minimized. Even assuming each iteration can be computed within 1 micro-second, which is a generous assumption, this would result in 1 million seconds of computational time, which is obviously not acceptable for real-time supervisory controls that may require sub-minute or even sub-second decisions in some cases.

The proposed approach is to use an Interior-point method for non-linear optimization [7]. The optimization problem is a QP, with a quadratic objective as a sum of quadratic functions, and linear constraints. Provided that the objective function is convex, which is expected for the pump power and chiller power based on their affinity law, an Interior-point method can typically find the global minimum of a QP with dozens of variables within milliseconds on modest modern hardware [6]. Note that even if the objective function is not convex, a sequential quadratic program method [7] was found to provide a good local optimum in similar time for all cases tested. Additionally, specific constraints can be added to the models' learning to enforce their convexity.

3. Modelling results

3.1 Model learning results

The chiller power model learning results are reported for 2 separate anonymised sites, with roughly 1 year of 15 minutes interval data each. Site A is equipped with 2 multi-centrifugal-compressors chillers rated at 1,700 kW of refrigeration, and site B with 2 centrifugal-compressors chillers of 1,800 kW of refrigeration. Chillers on Site A and B are from different manufacturers. As shown on Tab. 1 the proposed chiller power model captures with a good accuracy the actual measured trend for all tested chillers, and therefore proves to be a suitable predictor for the application, for several separate sites and chillers.

Tab. 1 – Chiller model learning results with site data.

	Site A		Site B	
Chiller n°	1	2	1	2
MAE (kW)	7.1	4.5	6.3	6.4
MAE/Mean (%)	5.17	4.98	5.49	4.32
R-squared	0.962	0.978	0.957	0.961

For chiller 1 of Site A, the fit of the predicted power in regard to the actual power is displayed on Fig. 2 and shows a good fit. The predicted power is also displayed on Fig. 3 in regard to the Load, the temperature difference, and the CW Flow.

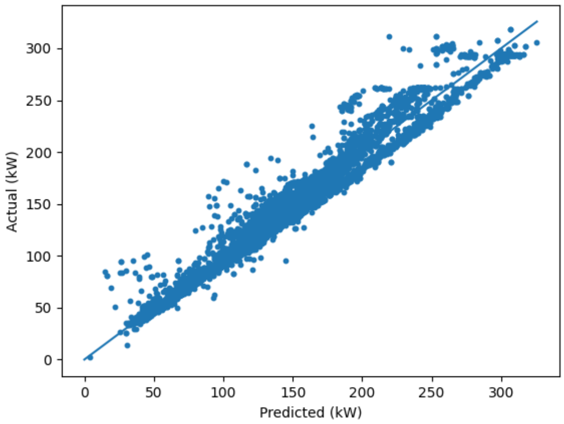

Fig. 2 – Site A, Chiller 1 predicted power compared to the actual measured power.

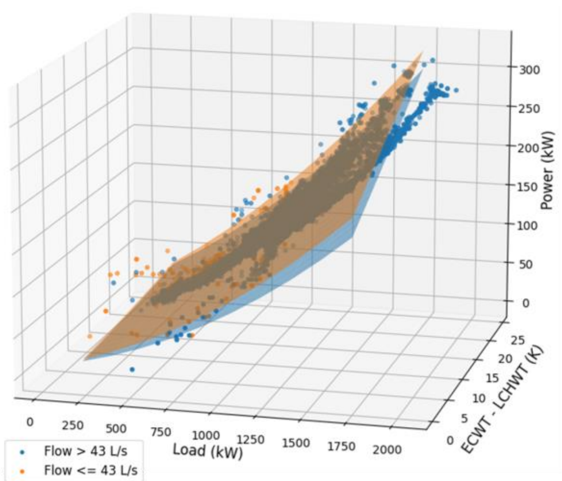

Fig. 3 – Site A, Chiller 1 predicted power in regard to the cooling load, difference of temperature and CW Flow. Surface plot: Predicted power. Scatter plot: Actual data.

The CW pumps power model was tested on the same sites, with the dedicated pump of each chiller. As shown on Tab. 2 the proposed pump power model performs well overall with low error and high correlation fit in all cases.

Tab. 2 – Pump model learning results with site data.

	Site A		Site B	
Pump n°	1	2	1	2
MAE (kW)	0.56	0.31	0.38	0.41
MAE/Mean (%)	4.78	3.47	3.57	3.66
R-squared	0.978	0.988	0.958	0.947

For pump 1 of Site A, the fit of the predicted power and actual power in regard to the flow is displayed on Fig.4.

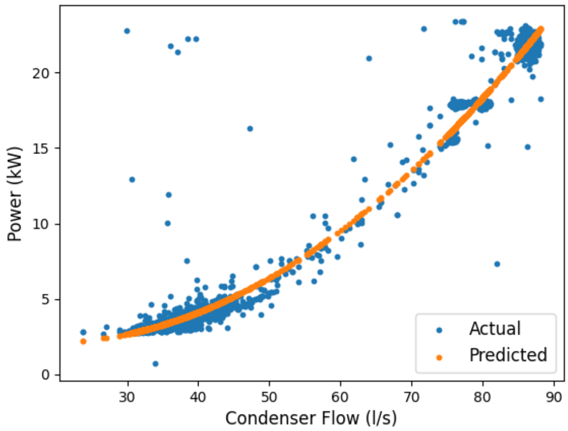

Fig. 4 – Site A, Pump 1 predicted and actual power in regard to the flow.

3.2 Optimization offline simulations

The optimization algorithms were initially run offline with models learnt from site data, to assess the validity of the premise as well as the savings potential.

As an example, the optimization of CW Flow setpoint was simulated for Site A's Chiller n°1 with its dedicated CW pump n°1. The rated data for this pair of equipment is a CW Flow of 86 L/s and a pump power of 22 kW. The power of each equipment individually and their sum is displayed on Fig. 5 and Fig. 6, for set conditions of LCHW and ECW temperatures, and respectively a low load of 30% and medium-high load of 70%. Those figures show the trade-off between pump power and chiller power and how at different conditions the optimal flow point is different. This illustrates the benefit of predictive modelling and continuous optimization to accurately assess the best CW Flow setpoint at all times under varying conditions.

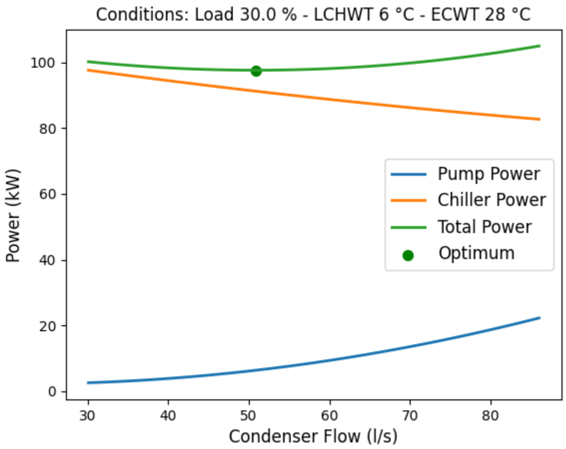

Fig. 5 – Site A, Chiller and Pump n°1 optimal CW Flow of 50 L/s (58% of design flow) under low load.

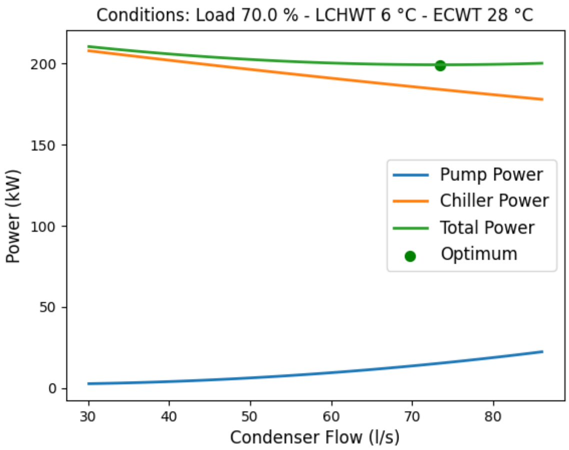

Fig. 6 – Site A, Chiller and Pump n°1 optimal CW Flow of 72 L/s (83% of design flow) under medium-high load.

At most sites, operators and conventional controls system typically maintain a constant CW Flow at all conditions, as the trade-off between the pump and chiller usage is not obvious. There is no clear rule-of-thumbs that can be followed. Indeed, as is shown on Fig.7, for a theoretical different pump sizing, at the same conditions and with the same chiller, the optimal CW Flow is very different.

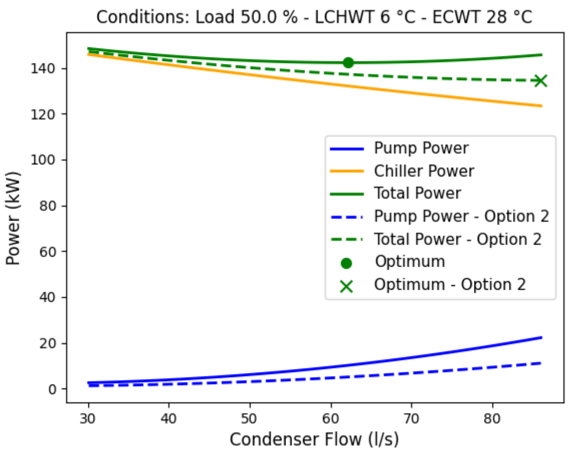

Fig. 7 – Optimal CW Flow comparison with a theoretical 2nd option of Pump sizing design.

Savings on Site A's Chiller and Pump n°1 power for the case of optimal CW Flow in comparison with the baseline case with constant CW Flow set at design value can be computed from the predictive models, as displayed on Fig. 8. The savings' opportunities are significant, at low and medium loads in particular, for this specific site.

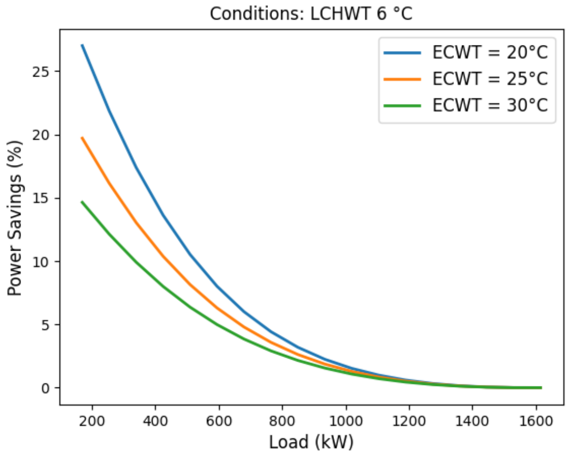

Fig. 8 – Site A, Chiller and Pump n°1 percentage savings for the optimal CW Flow setpoint, at different cooling loads and ECW temperatures.

On Site B, the combined impact of the optimal CW Flow and LCHWT is shown on Fig. 9, 10 and 11. The ECWT is configured to 28°C, the Plant LCHWT to 10°C, 2 chillers are running at the same time, and the simulation is run for plant cooling load points on their whole range of operation. The savings are computed by comparing with a case of constant CW Flows at each chiller and equal load distribution between the chillers.

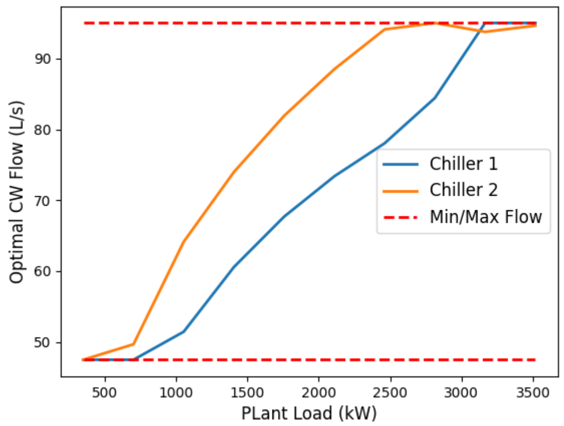

Fig. 9 – Site B, optimal CW Flow setpoints, at different plant cooling loads.

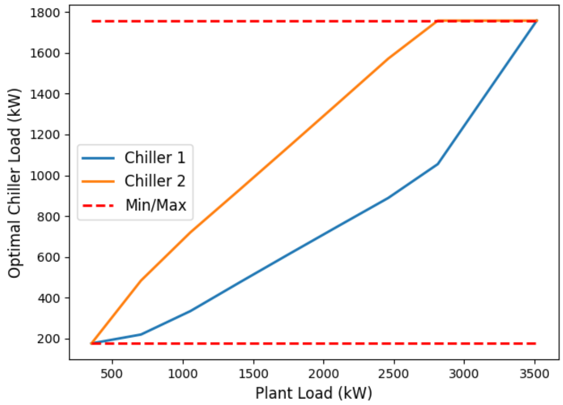

Fig. 10 – Site B, optimal chiller loads, at different plant cooling loads.

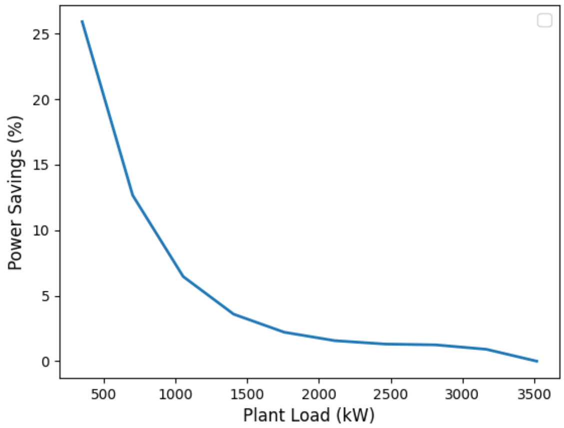

Fig. 11 – Site B, computed power savings at different plant cooling loads.

Due to the difference of part-load performance of the chillers, different optimal CW Flows and loads are found for each chiller, resulting in significant savings compared to the conventional case of maintaining constant CW Flow and equal loads on the chillers.

4. Real-time controls results

The approach has been implemented on site on embedded hardware for real-time direct controls of equipment via common high-level communication protocol such as Modbus or BACnet. This was integrated within the existing chiller plant controls hardware & software embedded solution PlantPRO®, which delivers a comprehensive on-premises controls solution that can fully manage the plant independently or in parallel with a Building Management System (BMS). The advantage of this implementation approach is that it does not rely on other control systems to manage the equipment, and contrarily to cloud-hosted solutions, it does not require a cloud subscription fee and removes downtime due to internet connection issues.

Two separate hardware platforms were used, initially a BeagleBone Black-based platform, with 1 core CPU and 500MB of RAM, and subsequently a Raspberry Pi 3+ compute module-based platform with a 4-core CPU and 1GB of RAM, as PlantPRO® transitioned to the latter. The solution was initially implemented as a prototype during the research phase with customizations on a site-by-site basis and has since then been implemented in the commercial product with full-suite of GUI and flexible configuration for various chiller plant configurations.

Resulting operational data further to deployments on Sites A, B & C were analysed. During operation on Site A, at a period of relatively constant conditions, the optimal CW Flow was overridden to the design CW Flow, to confirm that during actual operation the expected energy savings were indeed observed. As shown on Tab. 3, the optimal CW Flow brought savings of 10.4 kW of instantaneous power usage, or 22.6%. Albeit in a punctual scenario of low load, this site experiment does contribute to further validate the opportunities of optimizing over this variable in the plant.

Tab. 3 – Site A recorded site data comparison between optimal and standard CW Flow.

Case	Load (kW)	CW Flow (L/s)	Chiller Power (kW)	Pump Power (kW)	Total Power (kW)
Optimal	325	52.1	30.4	5.3	35.7
Standard	322	80.9	29.1	17.0	46.1

On Site C – which is equipped with 2 chillers of 3000 kW and 1 chiller of 1000 kW of refrigeration – the number of chillers and the loads were optimized but not the CW Flow due to sensor limitations at the condenser side. The overall plant COP was observed to increase noticeably at medium and high loads, which coincides with periods with more than 1 chiller enabled, when there are opportunities for varying the load proportion between chillers. This can be observed on Fig. 12. Note that periods with faulty equipment were removed.

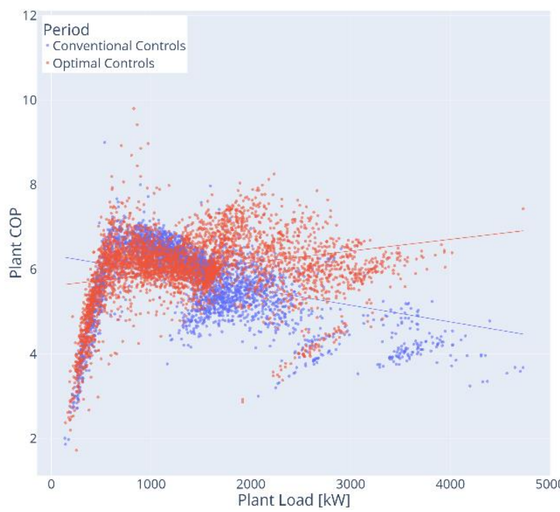

Fig. 12 – Instantaneous Plant COP at Site C with conventional controls period compared to optimal controls period, from measured data.

On Fig. 13 and 14, respectively the actual cooling and LCHWT of each enabled chiller is displayed during a few hours of operation, illustrating how – as a result of the optimal controls approach – the load is not distributed equally between each chiller as it would typically be with conventional controls, while still targeting to meet the plant level production requirement.

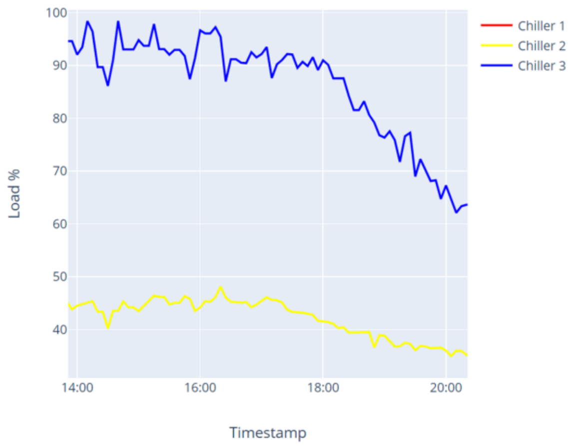

Fig. 13 – Cooling load distribution per chiller at Site C from measured data (with only chillers 2 and 3 running during that time period).

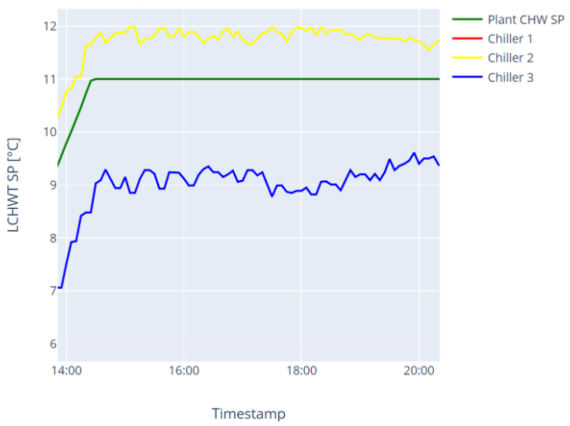

Fig. 14 – LCHWT distribution per chiller at Site C (with only chillers 2 and 3 running during that time period).

On Site B, both CW Flow and Loads were optimized. As shown on Fig. 15, the plant efficiency increased notably after deployment, compared to the previous conventional controls method.

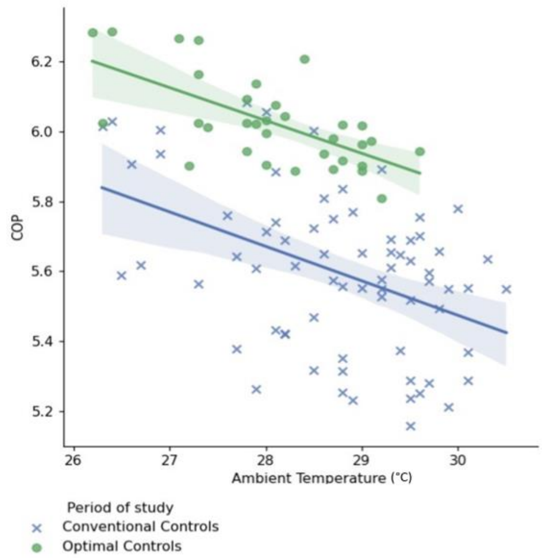

Fig. 15 – Daily Plant COP at Site B conventional controls period compared to optimal controls period, from measured data.

Measurement and verification (M&V) studies were carried out, following the guiding principles from the International Performance Measurement and Verification Protocol (IPMVP) [8], to assess the actual energy savings achieved while accounting for changes in conditions appropriately. For all 3 sites, a baseline was computed to represent the energy usage of the plant without optimal control and compared with the actual energy usage measured during the optimal controls period. For Site A, a calibrated simulation approach (IPMVP's option D) was used to simulate operation with constant CW flow, and the optimal controls period was of roughly 2 months. For Sites B and C, the measured energy usage of the plant prior to deployment was modelled as a function of the cooling load and outdoor air conditions (IPMVP's option B), and the optimal controls period was of roughly 1 month each. The savings are displayed on Tab. 4.

Tab. 4 – Energy savings.

Sites	Variables	Savings (kWh)	Savings (%)
Site A	CW Flows	15,948	3.6
Site B	CW Flows and Loads	7,358	7.1
Site C	Loads	4,056	5.1

The savings achieved in most cases are significant given that no mechanical equipment was replaced, and they are solely due to control strategies variations, moreover, only affecting 1 or 2 sets of setpoints (CW Flows and Loads) amongst all the variables in the plant. The efficiency of the plants prior to the implementation of the optimal controls features were already good in all cases, with conventional control strategies following best practices in the field. Site A and C already had PlantPRO® installed and had had extensive manual tuning by experienced operators to extract as much savings as possible with conventional approaches. Note that for Site C, the LCHWT allowable range configuration for each chiller was extended to allow for the chillers' loads to be distributed as per the optimization live decisions. For Site B, a BMS was controlling the plant prior to deployment, and the plant COP was also reasonably high, as shown previously on Fig. 15, in particular given that it is located in a hot-and-humid climate. Note that for that site, the transition to optimal controls included the installation of PlantPRO®, and there could have been some minor additional changes in regard to other control strategies. Due to the solution being deployable within an existing controls solution providing additional benefits, on low-cost embedded hardware, the cost of adoption is expected to be relatively low, and the level of savings to provide adequate benefits to the stakeholders.

The computing time to solve the optimization problem on the embedded hardware was also measured for a few key cases, as displayed on Tab. 5, and were fast enough for real-time controls, as demonstrated by the successful site deployments.

Tab. 5 – Loads-only optimization algorithm computing time on Raspberry Pi 3+ based platform – Average of 1,000 runs.

2 Chillers	5 chillers	10 Chillers
46.1 ms	54.2 ms	64.4 ms

5. Further use: Predictive maintenance

The models to predict the power usage of the chillers and condenser pumps, that have been developed and validated in this study, can also be used to track and detect performance degradation of the equipment. Operators are often unable to properly assess this due to only limited data at full-load being available from the manufacturer. As seen above, chiller power usage and efficiency are affected by the cooling load, water temperatures, condenser flow, and it can be hard to discern if a lower efficiency than surmised is due to performance degradation or simply to the conditions. With the predictive models, performance of the equipment can be captured initially based on manufacturer data or initial operation. This baseline calibrated model can then be continuously compared in the plant management system with actual measured efficiency of the equipment based on the live conditions, to alert the operator if it has decreased over time.

This can be used as a predictive maintenance tool to address otherwise undetected issues that would increase the power usage of the plant and may be early signs of a potential future failure that may cost more to be fixed later on than if prevented.

6. Conclusions

A real-time model-based optimal controls approach was proposed, validated from site data, and deployed in actual chiller plants to demonstrate its resulting operation and benefits. The performance of the equipment was predicted using machine learning models that demonstrated generally high prediction accuracy. The optimization models were used to run offline simulations and showed notable potential energy savings. Finally, the solution was deployed on several sites, and demonstrated actual energy savings in the range of 3.6% to 7.1%, while maintaining both reliability of operation and comfort in the buildings. The solution was also designed to operate at low computing cost and be deployable on modern low-cost embedded controllers, eliminating the hurdles encountered with cloud-hosted approaches such as on-going maintenance fees and security and reliability concerns due to reliance on internet connection for equipment controls. The predictive models can also be applied to further use, such as predictive maintenance and fault detection, and deliver additional indirect benefits.

7. Acknowledgement

The authors would like to acknowledge John Wessel from Airmaster for sharing his invaluable advice and expert domain knowledge.

8. References

[1] UN, COP26 The Glasgow Climate Pact, 2021

[2] Hydeman M., Zhou G. Optimizing Chilled Water Plant Control. ASHRAE Journal. 2007;44-55.

[3] Taylor S. Optimizing Design & Control of Chilled Water Plants Part 5: Optimized Control Sequences. ASHRAE Journal. 2012;56-75.

[4] Supervisory control strategies and optimization. ASHRAE Handbook – HVAC Applications – Chapter 41. 2007

[5] Hydeman M., Gillespie K. Tools and Techniques to Calibrate Electric Chiller Component Models. ASHRAE Transactions. 2002

[6] Boyd S., Vandenberghe L. Convex Optimization. Cambridge University Press. 2004

[7] Wächter A., Biegler L. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming. 2006;106:25-57

[8] Core Concepts – International Performance Measurement and Verification Protocol. EVO 10000. 2016

Data Statement

The datasets generated and/or analysed during the current study are not available due to commercial reasons but the authors will make every reasonable effort to publish them in the future.