Energy-efficient data-based zonal control of temperature for data centers

Masoud Kheradmandi
Dept. of Computing and Software
McMaster University
Hamilton, Canada
kheradm@mcmaster.ca

Douglas G. Down
Dept. of Computing and Software
McMaster University
Hamilton, Canada
downd@mcmaster.ca

Hosein Moazamigoodarzi
Dept. of Mechanical Engineering
McMaster University
Hamilton, Canada
moazamih@mcmaster.ca

***Abstract*—In this work we address the problem of optimal economic thermal operation of data centers utilizing economic model predictive control. First, a data driven predictive zonal model is trained using subspace identification. This model is utilized in a model predictive controller to predict dynamic behavior of a data center. An economic model predictive control is designed to maintain temperature of each zone within an allowable range while minimizing operational costs of the cooling system. The effectiveness of the proposed method is illustrated through simulations on a mechanistic data center model.**

***Keywords*-Data centers; Thermal management; Zone-model predictive control; MPC; EMPC; Data driven predictive model.**

I. Introduction

Data centers (DC) are currently drawing up to 5% of the world's electricity [1], with this proportion growing because of increasing demand for cloud computing infrastructures. About 40% of this energy is usually provided for thermal management of IT equipment (ITE) [2]. The aim of DC cooling systems is to maintain server temperatures within a safe temperature region, where servers can be reliably operated. The safe temperature ranges for different types of DC hardware are available in guidelines such as those of ASHRAE [3]. Conventional cooling architectures for DCs are either room-based or row-based [4], [5]. Recently a new cooling unit infrastructure has been developed employing rack mountable cooling units. The idea in the new design is to reduce cooling unit power consumption by limiting the cooling area to the inside of a rack [6].

Violation of safe temperature guidelines, even by 1 or 2 degrees, can result in server failures [7], [8]. It can also result in poor performance [9]. Therefore a controller is required that can maintain safe temperatures while minimizing operating cost. In order to control DC temperature, we find that a control theoretic viewpoint is applicable. A number of different types of controllers can be utilized. The basic controllers, are ON/OFF [10] and proportional–integral–derivative (PID) controllers [11]. ON/OFF controllers trigger cooling when the temperature exceeds a certain threshold, and switch off when it falls below another threshold. However, this method cannot adjust the level of cooling and therefore may result in excessive power consumption. PID controllers can adjust cooling power according to errors between the measured and desired temperatures. However, a PID controller cannot predict the behavior of the system and the control action can result in abrupt changes in behavior if controller gains are poorly tuned. In addition, conventional PID is only appropriate for single input and single output systems. Finally, a PID controller does not directly yield the optimal solution.

A controller that can generate an optimum solution, for general multi input multi output systems is desired. Model Predictive Control (MPC) is an approach that can address all of these issues [12], [13]. MPC solves an optimization problem at each sampling instant over a finite time horizon, subject to a dynamic model of the system and general constraints, in order to calculate the control action. This objective function can be a combination of the temperature deviation from the setpoint and the input cost. Recently, a model predictive controller has been designed to stabilize DC temperature at a desired 'fixed' setpoint while reducing energy cost [14]. Therefore the objective function of the MPC should be designed in a manner that resolves the trade-off between output deviation from a fixed setpoint and input cost.

The models that are used in MPC to predict dynamic behavior of DCs are mostly mechanistic models [15], [16]. These models are not easy to simulate and their calculation is computationally expensive and hence not suitable for real-time operation. In addition, adopting a model with dynamic changes in structure may not be straight forward. Therefore in this work we will utilize a data-driven method to identify a predictive model. There are different machine learning modeling methods that can be utilized in an MPC framework. We will use a state space subspace identification method to model our system. This modeling technique is frequently utilized in system and control theory and is well developed for all aspects of signal processing. We demonstrate the effectiveness of our approach for a particular DC architecture, however the approach is given in sufficient generality to be appropriate for general architectures. With modern DCs instrumented to provide real-time monitoring data, our approach is an attractive means to perform thermal control without requiring an expensive mechanistic modeling

978-1-7281-5416-9/19 $31.00 © 2019 IEEE

exercise.

Motivated by the above considerations, in this work we address the problem of designing a model predictive control scheme to minimize the operating cost of a data center while ensuring that temperature is within a safe range. The rest of the manuscript is organized as follows: First, the general description for the data center considered in this work, a mechanistic model for DCs as a test bed, a subspace identification approach and a representative formulation for MPC are reviewed. Then the proposed zonal model predictive control approach is presented. The efficiency of the proposed zonal control is illustrated by simulating a DC with two rack mounted cooling units. The results are compared with a nominal MPC. Finally, concluding remarks are presented.

II. Preliminaries

A description of the data center with rack mounted cooling units is provided in this section, followed by the description of the test-bed model, and the relevant aspects of the system simulation employed to study practical application issues. A review of a subspace data-driven model approach for dynamic systems is then presented.

A. Data Center

As a test bed, a parameter-free transient zonal model to obtain the dynamic temperature profile inside a data center (DC) with rack mounted cooling unit (RMCU) with separated cold and hot chambers is reviewed. The schematic and zones for a DC IT rack within an enclosure that is cooled by two RMCUs are presented in Figure 1. This method of modeling was recently proposed in [6]. The model is based on mass and energy conservation rules for each zone within the DC.

A detailed explanation of the modeling details and formulations can be found in [6], and are omitted here for brevity. The system considered in this work consists of a single rack DC containing 20 servers with known power consumption, and two rack-mounted cooling units located at the top and bottom of the rack. The cooling unit consists of a heat exchanger and a set of five identical compact industrial fans.

1) Mechanistic Thermal Model of Data Center with Rack Mounted Cooling Unit: The mechanistic model for the test bed is adopted from a recently developed mechanistic model proposed in [6]. In this section we briefly present the mathematical formulation. The dynamic equation for the temperature is driven from the energy balance equation for an active server as follows:

$$\frac{M_T}{2}\left(\frac{dT_{e,i}}{dt}+\frac{dT_{c,i}}{dt}\right)=\rho_a c_{p,a} Q_{s,i}(T_{c,i}-T_{e,i})+\dot{P}_s, \quad (1)$$

where, M_T is the thermal mass of the server, $T_{e,i}$ and $T_{c,i}$ are the temperatures at the server outlet and cold chamber zone, ρ_a is air density, $c_{p,a}$ is the specific heat of air, and P_s is the power consumption of the server. $Q_{s,i}$ is the airflow through each active server and is calculated in m^3/s using the following equation:

$$Q_{s,i}=\begin{cases} 0.01415 & \text{if } T_{c,i}\leq 25 \\ 0.01415+0.00142(T_{c,i}-25) & \text{if } 25<T_{c,i}<35 \end{cases} \quad (2)$$

The energy balance for air and chilled water within the cooling unit can be written in the form of the following equations:

$$\rho_a c_{p,a} V_a\left(\frac{dT_c}{dt}+\frac{dT_h}{dt}\right)= \rho_a c_{p,a} Q_R(T_h-T_c)-\frac{UA}{2}(T_h+T_c-T_{i,w}-T_{o,w}), \quad (3)$$

$$\rho_w c_{p,w} V_w\left(\frac{dT_{i,w}}{dt}+\frac{dT_{o,w}}{dt}\right)= \rho_w c_{p,w} Q_w(T_{i,w}-T_{o,w})+\frac{UA}{2}(T_h+T_c-T_{i,w}-T_{o,w}), \quad (4)$$

where T_c and T_h denote the air temperatures at the cooling unit outlet and inlet, $T_{i,w}$ and $T_{o,w}$ the chilled water inlet and outlet temperatures, Q_w the water flow rate, and $c_{p,w}$ and ρ_w are the specific heat and the density of water. U and A denote the overall heat transfer coefficient and surface area inside the RMCU. In order to calculate the value of UA in the heat exchanger between two fluids, see [6]. Q_R in the equations in Table I is cold airflow and for each server i in a rack with n servers is calculated using the following equation:

$$Q_{R,i}=\left(\frac{n-i}{n(n+1)}+\frac{0.5}{n}\right)Q_R(1-\varphi) \quad \text{for } i=2,3,...n, \quad (5)$$

where φ is the portion of the cold airflow that exits the RMCU and enters the cold chamber zone in front of the first server.

For each chamber zone (cold and hot), using energy balance results in the following equations:

$$\rho_a c_{p,a} V_c \gamma \frac{dT_{c,i}}{dt}=\phi_R+\phi_i+\phi_o+\phi_l+\phi_s, \quad (6)$$

$$\rho_a c_{p,a} V_h \gamma \frac{dT_{h,i}}{dt}=\phi'_i+\phi'_o+\phi'_l+\phi'_s, \quad (7)$$

where V_c and V_h are the volumes of the cold and hot chambers. The parameter γ is a correction factor for the masses of the chamber zones. The quantities ϕ and ϕ' are the energy exchanges for the cold and hot chamber zones. The expressions for each energy exchange term in Equations 6 and 7 are presented in TableI, in terms of the temperature of the corresponding chamber zone.

Table I
DEFINITIONS FOR THE TERMS IN EQUATIONS 6 AND (7)

Equation.(6)	Equation.(7)
$\phi_i = \begin{cases} \rho_a c_{p,a} Q_{in,i} T_{c,i-1} & Q_{in,i} \geq 0 \\ \rho_a c_{p,a} Q_{in,i} T_{c,i-1} & Q_{in,i} < 0 \end{cases}$	$\phi'_i = \begin{cases} \rho_a c_{p,a} Q'_{in,i} T_{h,i+1} & Q'_{in,i} \geq 0 \\ \rho_a c_{p,a} Q'_{in,i} T_{h,i} & Q'_{in,i} < 0 \end{cases}$
$\phi_o = \begin{cases} -\rho_a c_{p,a} Q_{o,i} T_{c,i} & Q_{in,i} \geq 0 \\ -\rho_a c_{p,a} Q_{o,i} T_{c,i+1} & Q_{in,i} < 0 \end{cases}$	$\phi'_o = \begin{cases} -\rho_a c_{p,a} Q'_{o,i} T_{h,i} & Q'_{o,i} \geq 0 \\ -\rho_a c_{p,a} Q'_{o,i} T_{c,i-1} & Q'_{o,i} < 0 \end{cases}$
$\phi_l = \begin{cases} \rho_a c_{p,a} Q_{l,i} T_{h,i} & \Delta P \geq 0 \\ \rho_a c_{p,a} Q_{l,i} T_{c,i} & \Delta P < 0 \end{cases}$	$\phi'_l = \begin{cases} \rho_a c_{p,a} Q_{l,i} T_{h,i} & \Delta P \geq 0 \\ \rho_a c_{p,a} Q_{l,i} T_{c,i} & \Delta P < 0 \end{cases}$
$\phi_R = \rho_a c_{p,a} Q_{R,i} T_c$	-

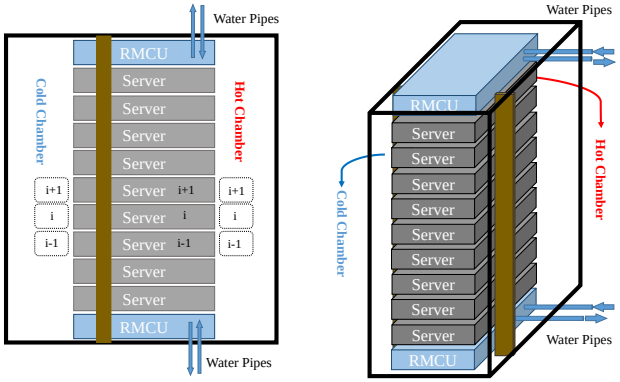

Figure 1. Schematic of the IT enclosure integrated with a single rack and two RMCUs with separated cold and hot chambers. The zones (control volumes) in the front and back chambers are shown.

III. DATA-DRIVEN ZONAL CONTROL IMPLEMENTATION

A. State-Space Based Subspace Identification

In this section we review the conventional subspace-based identification method which we will utilize to train a discrete state space model. The goal in model identification is to compute the linear time invariant (LTI) model parameters with the following form:

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{8}$$
$$y_k = Cx_k + Du_k + v_k \tag{9}$$

where $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$ denote the vectors of state variables and measured outputs, $w \in \mathbb{R}^{n_x}$ and $v \in \mathbb{R}^{n_y}$ are zero mean, white vectors of process noise and measurement noise with the following covariance matrices:

$$E[\begin{pmatrix} w_i \\ v_j \end{pmatrix} \begin{pmatrix} w_i^T & v_j^T \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{ij}$$

where $Q \in \mathbb{R}^{n_x \times n_x}$, $S \in \mathbb{R}^{n_x \times n_y}$ and $R \in \mathbb{R}^{n_y \times n_y}$ are covariance matrices, and, δ_{ij} is the Kronecker delta function. In order to compute the LTI model matrices, Hankel matrices are constructed by stacking the system variables as follows:

$$U_p = U_{1|i} = \begin{bmatrix} u_1 & u_2 & \dots & u_j \\ u_2 & u_3 & \dots & u_{j+1} \\ \vdots & \vdots & \vdots & \vdots \\ u_i & u_{i+1} & \dots & u_{i+j-1} \end{bmatrix} \tag{10}$$

$$U_f = U_{i+1|2i} = \begin{bmatrix} u_{i+1} & u_{i+2} & \dots & u_{i+j} \\ u_{i+2} & u_{i+3} & \dots & u_{i+j+1} \\ \vdots & \vdots & \vdots & \vdots \\ u_{2i} & u_{2i+1} & \dots & u_{2i+j-1} \end{bmatrix} \tag{11}$$

where U_p and U_f denote the past and future input Hankel matrices. The value i is a user-specified parameter that limits the order of the system (n) and i should be larger than n. Block-Hankel matrices are defined in a similar manner for y, v and w as $Y_p, Y_f, V_p, V_f \in \mathbb{R}^{in_y \times j}$ and $W_p, W_f \in \mathbb{R}^{in_x \times j}$. The state sequences are defined as follows:

$$X_p = \begin{bmatrix} x_1 & x_2 & \dots & x_j \end{bmatrix}, \tag{12}$$
$$X_f = \begin{bmatrix} x_{i+1} & x_{i+2} & \dots & x_{i+j} \end{bmatrix}. \tag{13}$$

We also define:

$$\Psi_p = \begin{bmatrix} Y_p \\ U_p \end{bmatrix} \tag{14}$$
$$\Psi_f = \begin{bmatrix} Y_f \\ U_f \end{bmatrix} \tag{15}$$

The orthogonal projection of the row space of matrix A onto the row space of matrix B, (A/B) is defined as:

$$A/B = AB^\dagger B, \tag{16}$$

where the superscript † denotes pseudo-inverse. By recursive substitution into the state space model, Equations (8) and (9), it is straightforward to show:

$$Y_f = \Gamma_i X_f + \Phi_i^d U_f + \Phi_i^s W_f + V_f \tag{17}$$
$$Y_p = \Gamma_i X_p + \Phi_i^d U_p + \Phi_i^s W_p + V_p \tag{18}$$
$$X_f = A^i X_p + \Delta_i^d U_p + \Delta_i^s W_p \tag{19}$$

where:

$$\Gamma_i = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix} \tag{20}$$

$$\Phi_i^d = \begin{bmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{bmatrix} \tag{21}$$

$$\Phi_i^s = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ C & 0 & 0 & \dots & 0 & 0 \\ CA & C & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ CA^{i-2} & CA^{i-3} & CA^{i-4} & \dots & C & 0 \end{bmatrix} \tag{22}$$

$$\Delta_i^d = \begin{bmatrix} A^{i-1}B & A^{i-2}B & \dots & AB & B \end{bmatrix} \tag{23}$$

$$\Delta_i^s = \begin{bmatrix} A^{i-1} & A^{i-2} & \dots & A & I \end{bmatrix} \tag{24}$$

Equation (17) can be rewritten in the following form to have the input and output data on the LHS [17]:

$$\begin{bmatrix} I & -\Phi_i^d \end{bmatrix} \begin{bmatrix} Y_f \\ U_f \end{bmatrix} = \Gamma_i X_f + \Phi_i^s W_f + V_f \tag{25}$$

By orthogonal projection of Equation (25) onto Ψ_p:

$$\begin{bmatrix} I & -\Phi_i^d \end{bmatrix} \Psi_f / \Psi_p = \Gamma_i X_f / \Psi_p + \Phi_i^s W_f / \Psi_p + V_f / \Psi_p \tag{26}$$

The last two terms in Equation (26) are the orthogonal projection of the future noise onto the row space of Ψ_p, and since the noise terms are independent, these two terms are equal to zero. Thus Equation (26) is simplified as follows:

$$\begin{bmatrix} I & -\Phi_i^d \end{bmatrix} \Psi_f / \Psi_p = \Gamma_i X_f / \Psi_p \tag{27}$$

Equation (27) indicates that the column space of Γ is equal to the column space of $\begin{bmatrix} I & -\Phi_i^d \end{bmatrix} \Psi_f / \Psi_p$, and the row space of X_f / Ψ_p is the same as the row space of $\begin{bmatrix} I & -\Phi_i^d \end{bmatrix} \Psi_f / \Psi_p$. This equation can be solved using singular value decomposition (SVD), and the system matrices can be calculated from the results. In existing results, the innovation form of the LTI model is used, as follows:

$$x_{k+1} = Ax_k + Bu_k + Ke_k \tag{28}$$

$$y_k = Cx_k + Du_k + e_k \tag{29}$$

where e_k is the innovation term, and, K is the filter gain. In these methods, after determining the A matrix, B and K are estimated using least squares. In contrast to the existing results, in this work we first estimate the noise terms, then calculate noise covariance matrices and based on these covariances the observer gain is calculated. The system identification procedures require the input signal to be quasi-stationary and persistently exciting of order $2i$. In the system identification step, only the observable part of a system is identified therefore the LTI model is always observable. Also note that, the order of the LTI system n is selected in a manner such that the identified system is controllable and the prediction of validation data is acceptable.

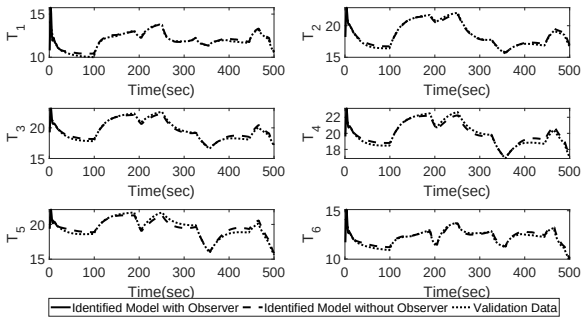

Figure 2. Model validation results (outputs)

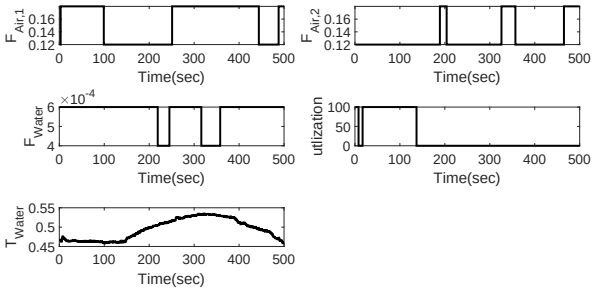

Figure 3. Model validation results (inputs and disturbances)

Remark 1: The order of the model or the number of states to be used for the model (the only model parameter choice) is decided from the training data fitting properties. This is done, at the singular value decomposition step, the choice of the number of states is equal to the number of dominant singular values. Note also that for the case that data are collected in closed-loop with noise or disturbance in the plant, open-loop identification methods are theoretically biased for data driven modeling, and closed-loop identification must be used, for example the closed-loop identification method utilized in [18].

B. Model Predictive control

In this section a conventional model predictive controller based on the identified model is designed. The goal of this controller is to keep the hottest point at the front of the servers at a desired temperature. The proposed controller is

found as a solution of the following optimization problem:

$$\min_{\tilde{u}(j)} \sum_{j=1}^{P} (\tilde{y}(j) - y_{sp})^T Q_y (\tilde{y}(j) - y_{sp}) + \tilde{u}(j)^T R_u \tilde{u}(j) \tag{30}$$

subject to: (31)

$$\tilde{x}(k+1) = A\tilde{x}(k) + B\tilde{u}(k) + B_d\tilde{d}(k) \tag{32}$$

$$\tilde{d}(k+1) = \tilde{d}(k) \tag{33}$$

$$\tilde{y}(k) = C\tilde{x}(k) + D\tilde{u}(k) + D_d\tilde{d}(k) \tag{34}$$

$$\tilde{u} \in \mathcal{U} \tag{35}$$

$$\tilde{x}(k) = \hat{x}_k, \quad \tilde{d}(k) = d_k, \tag{36}$$

where $\tilde{y}_k$ and $\tilde{u}_k$ are the predicted output trajectory and input at the kth sampling instant. $\tilde{x}$ and $\hat{x}$ are the predicted value and the estimate of the subspace state, obtained utilizing a state estimator. $\tilde{d}$ and d are the predicted value and the estimate of the measured disturbances, equal to the latest measurement. The matrices Q_y and R_u are output and input penalty matrices which are positive definite and pseudo-positive matrices. These are the means with which to tune the controller. The scalar y_{sp} is the desired output value for the system.

In order to compute the control action, the MPC optimization problem is solved at each sampling instant. The first input in the computed sequence is implemented on the system.

A Kalman filter is employed for state estimation in this study. The state estimator has the following form:

$$\begin{aligned}
\hat{x}_k^- &= A\hat{x}_{k-1} + Bu_k, \\
P_k^- &= AP_{k-1}A^T + Q, \\
K_k &= P_k^- C^T \left(CP_k^- C^T + R\right)^{-1}, \\
\hat{x}_k &= \hat{x}_k^- + K_k \left(y_k - C\hat{x}_k^-\right), \\
P_k &= \left(I - K_k C\right) P_k^-,
\end{aligned}$$

where $\hat{x}_k^-$ and P_k^- denote state and covariance matrix predictions at the kth sampling instant. Q and R are state and output covariance matrices, K_k is the Kalman filter gain at the kth sampling instant and I denotes the identity matrix.

C. Zonal model predictive control

The primary objective of zonal control in a data center is to maintain the temperature at the front of servers within a desired range to ensure for safe operation. There is also a secondary control goal which is to reduce the operational cost of the cooling system. The manipulated values are the chilled water flow rate (or valve opening percentage) and air flow (or fan speed) at each sampling instant k, and are computed by solving the following optimization problem:

$$\min_{\tilde{u}(j),\epsilon_-(j),\epsilon_+(j)} \sum_{j=1}^{P} c_{\epsilon_-}^T \epsilon_-(j) + c_{\epsilon_+}^T \epsilon_+(j) + c_u^T \tilde{u}(j) \tag{37}$$

subject to: (38)

$$\tilde{x}(k+1) = A\tilde{x}(k) + B\tilde{u}(k) + B_d\tilde{d}(k) \tag{39}$$

$$\tilde{d}(k+1) = \tilde{d}(k) \tag{40}$$

$$\tilde{y}(k) = C\tilde{x}(k) + D\tilde{u}(k) + D_d\tilde{d}(k) \tag{41}$$

$$\tilde{u} \in \mathcal{U}, \quad \epsilon_- \geq 0, \quad \epsilon_+ \geq 0 \tag{42}$$

$$\tilde{x}(k) = \hat{x}_k, \quad \tilde{d}(k) = d_k \tag{43}$$

$$y_{min} - \epsilon_- \leq \tilde{y} \leq y_{max} - \epsilon_+, \tag{44}$$

where c_{ϵ_-} and c_{ϵ_+} denote penalties associated with slack variables. The vector c_u denotes the cost of the manipulated input variables. The slack variables ϵ_- and ϵ_+ are added to the lower and upper bounds on the outputs to ensure the feasibility of the optimization problem. The values y_{min} and y_{max} are lower and upper bounds on the output (temperatures). The schematic presentation of EMPC is presented in Figure (4).

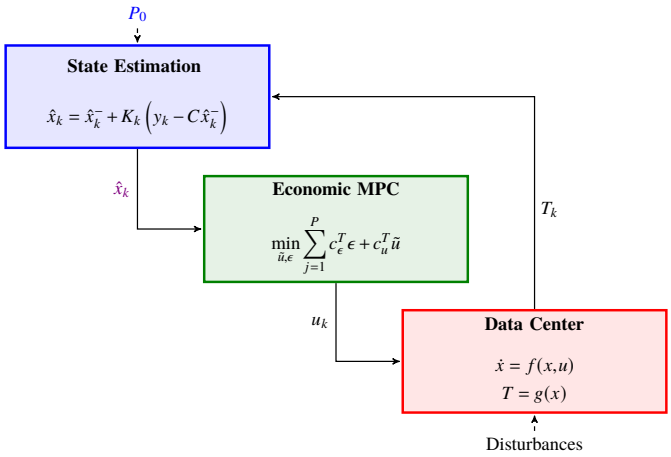

Figure 4. Schematic presentation of EMPC

Remark 2: The objective function serves two purposes. Firstly, it ensures that controlled outputs are within the defined range. Also, it minimizes operational costs based on the provided input costs (c_u). The main purpose is to ensure that temperatures are in the safe region, therefore ϵ_- and ϵ_+ are chosen in a way to ensure that the slack variables in the objective function are dominant.

Remark 3: The optimization problem defined is a linear programming (LP) problem. Therefore, using available solvers, the global optimum (if constraints are consistent) can be calculated. In order to implement it this formulation can be written in the standard form of an LP, which requires the addition of states and output variables to the decision

variables, which is a straightforward procedure. We shall not discuss this issue further in this paper.

Remark 4: Although a DC with rack mounted cooling unit is utilized as a testbed, the proposed framework can be used for other types of cooling unit and for any size of DC. We have not used any assumptions for the structure of the system and the model used in the predictive controller is data-driven.

Remark 5: There are two main differences in the two proposed frameworks. First, the optimization problem for MPC is a quadratic program, therefore it requires more computational effort than an LP. Secondly, in the MPC we have a setpoint for temperature inside a DC instead of a range for temperature. The problem of choosing the correct spots in a DC for being controlled may not be straightforward, as, the server workload distribution may change the location of hottest point(s). But, with EMPC we can monitor all the locations.

IV. Illustrative Simulation Results

A. Data Driven Modeling

The model is computed using the proposed identification method. The manipulated variables are two air flow rates in front of the fans and overall chilled water flow rate. The measured disturbances are the server utilizations (assumed to be uniform among servers) and chilled water temperature. The controlled outputs are the six temperatures in front of the servers.

In order to identify a model, inputs and utilizations are perturbed in a pseudo-random binary sequence fashion, except chilled water temperature. The chilled water is typically taken from a utility service, and its temperature is subject to changes due to the time varying load. Here, we assume a sinusoidal signal. The identified model consists of twenty states, three manipulated input variables, two measured disturbances and six measured outputs.

The identified model has the following form:

$$x_{k+1} = Ax_k + Bu_k + B_d d_k + w_k, \tag{45}$$

$$y_k = Cx_k + Du_k + D_d d_k + v_k, \tag{46}$$

$$E[\begin{pmatrix} w_i \\ v_j \end{pmatrix} \begin{pmatrix} w_i^T & v_j^T \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{ij}, \tag{47}$$

where d is vector of measured variables. The matrices B_d and D_d are measured disturbances gains in the state and output equations. In order to use our identification approach, d is augmented with u to create the overall input, while the rest of the steps are as illustrated.

For model validation, a different training data batch was used. The model validation results are presented in Figures 2 and 3. Since the initial state of the state space model cannot be determined from the training data, a Kalman filter (a state estimator) is utilized in order to achieve output convergence, then open-loop prediction is used for model validation in order to evaluate model prediction performance. At time 300 seconds into the data batch, the output of the predictive model has converged to the plant output, and the identified LTI model in open-loop (without state update), along with the known input trajectory, is utilized for output prediction with the remainder of the data. The results show that after the convergence of the model states, the prediction performance of the model for simulating the process behavior is reasonable, and is appropriate for a predictive control implementation.

B. Economic Zone Model Predictive Control

The data center temperature is maintained in the desired range utilizing the controller presented in Equation 44. Since the test bed does not include a heating system we do not have a lower limit for temperature (or $T_{min} = -\infty$). The controller parameters are reported in Table II. Figure 7 shows the comparison of the zonal EMPC and conventional MPC. The simulation results are presented in Figures 5, 6. As expected, the EMPC achieves improved economic returns compared to the conventional MPC.

Table II
Controller Parameters

Variable	Value
c_{ϵ_-}	$\begin{bmatrix}10^6 & 10^6 & 10^6\end{bmatrix}^T$
c_{ϵ_+}	$\begin{bmatrix}10^6 & 10^6 & 10^6\end{bmatrix}^T$
c_u	$\begin{bmatrix}10^{-2} & 10^{-2} & 1\end{bmatrix}^T$
P	8
y_{max}	16

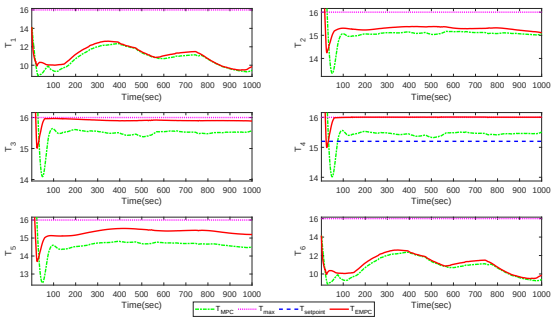

Figure 5. Closed-loop profile (output)

Remark 6: The issue of data center problems are amenable to control-theoretic techniques that are well established in other areas. We provide a convincing case that such techniques should be considered in this space.

V. Conclusions

In this study, a novel EMPC with a zonal control approach is developed that enables maintaining temperatures in a data

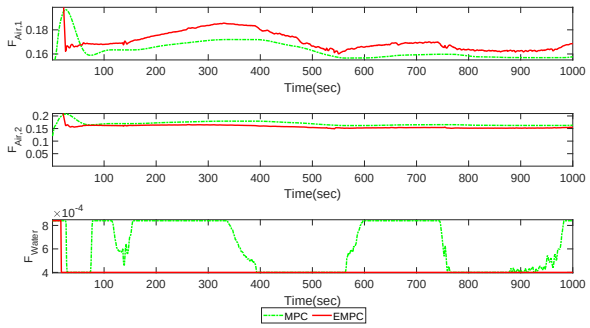

Figure 6. Closed-loop profile (input)

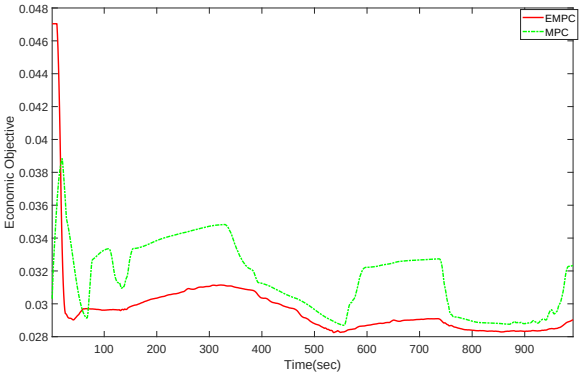

Figure 7. A comparison of the economic cost between the zone EMPC and MPC.

center within a safe region with cost efficient operation. The proposed approach is described and compared against a representative nominal MPC and shown to be able to provide improved closed-loop behavior through implementation on an example of a data center model.

VI. Acknowledgment

This research was supported grant CRDPI506142-16 and the Discovery grant program of the Natural Science and Engineering Research Council of Canada.

References

[1] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg *et al.*, "Us department of energy commercial reference building models of the national building stock," 2011.

[2] J. Dai, M. M. Ohadi, D. Das, and M. G. Pecht, *Optimum cooling of data centers*. Springer, 2016.

[3] R. K. Sharma, C. E. Bash, C. D. Patel, R. J. Friedrich, and J. S. Chase, "Balance of power: Dynamic thermal management for internet data centers," *IEEE Internet Computing*, vol. 9, no. 1, pp. 42–49, 2005.

[4] K. Dunlap and N. Rasmussen, "Choosing between room, row, and rack-based cooling for data centers," *APC White Paper*, vol. 130, 2012.

[5] G. Slessman, W. Slessman, K. Malik, J. Steffensen, and K. Holmgren, "Data center intelligent control and optimization," Apr. 9 2019, uS Patent App. 10/254,720.

[6] H. Moazamigoodarzi, S. Pal, S. Ghosh, and I. K. Puri, "Real-time temperature predictions in it server enclosures," *International Journal of Heat and Mass Transfer*, vol. 127, pp. 890–900, 2018.

[7] N. El-Sayed, I. A. Stefanovici, G. Amvrosiadis, A. A. Hwang, and B. Schroeder, "Temperature management in data centers: why some (might) like it hot," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 1, pp. 163–174, 2012.

[8] M. K. Patterson, "The effect of data center temperature on energy efficiency," in *2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*. IEEE, 2008, pp. 1167–1174.

[9] W. Torell, K. Brown, and V. Avelar, "The unexpected impact of raising data center temperatures," *Write paper 221, Revision*, 2015.

[10] D. H. Zervos, "On-off thermostat based modulating air flow controller," Dec. 3 1985, uS Patent 4,556,169.

[11] B. Durand-Estebe, C. Le Bot, J. N. Mancos, and E. Arquis, "Data center optimization using pid regulation in cfd simulations," *Energy and Buildings*, vol. 66, pp. 154–164, 2013.

[12] M. Kheradmandi and P. Mhaskar, "Prescribing closed-loop behavior using nonlinear model predictive control," *Industrial & Engineering Chemistry Research*, vol. 56, no. 51, pp. 15 083–15 093, 2017.

[13] B. Grosman, E. Dassau, H. C. Zisser, L. Jovanovič, and F. J. Doyle III, "Zone model predictive control: a strategy to minimize hyper-and hypoglycemic events," *Journal of diabetes science and technology*, vol. 4, no. 4, pp. 961–975, 2010.

[14] X. Zhao, Z. Xiong, L. Ding, X. Zhang, and F. Xu, "A smart coordinated temperature feedback controller for energy-efficient data centers," *Future Generation Computer Systems*, vol. 93, pp. 506–514, 2019.

[15] L. Parolini, B. Sinopoli, B. H. Krogh, and Z. Wang, "A cyber–physical systems approach to data center modeling and control for energy efficiency," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 254–268, 2012.

[16] C. J. Dawson, V. V. DiLuoffo, A. H. I. Rick, and M. D. Kendzierski, "System and method to control data center air handling systems," Jan. 3 2012, uS Patent 8,090,476.

[17] J. Wang and S. J. Qin, "A new subspace identification approach based on principal component analysis," *Journal of process control*, vol. 12, no. 8, pp. 841–855, 2002.

[18] M. Kheradmandi and P. Mhaskar, "Data driven economic model predictive control," *Mathematics*, vol. 6, no. 4, p. 51, 2018.