Proceedings of the ASME 2011 Pacific Rim Technical Conference & Exposition on Packaging and Integration of Electronic and Photonic Systems
InterPACK2011
July 6-8, 2011, Portland, Oregon, USA

IPACK2011-52029

DYNAMIC REDUCED ORDER THERMAL MODELING OF DATA CENTER AIR TEMPERATURES

Rajat Ghosh, Yogendra Joshi
The George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract

We developed a Proper Orthogonal Decomposition (POD) based dynamic reduced order model that can predict transient temperature field in an air-cooled data center. A typical data center is modeled as a turbulent convective thermal system with multiple length scales. A representative case study is presented to validate the developed methodology. The model is observed to be capable of predicting the transient air temperature field accurately and rapidly. Comparing with the computational fluid mechanics/heat transfer (CFD/HT) based model, it is revealed that our model is 100x faster without compromising solution accuracy. The developed modeling framework is potentially useful for designing a control system that can regulate flow parameters in a transient data center.

Keywords: Transient data center, Proper Orthogonal Decomposition, Reduced order model

1. INTRODUCTION

As much as 30-50% of the total power consumed in a typical air-cooled data center is utilized for the supporting cooling systems – in fact, the cost for the cooling is becoming comparable to the IT equipments [1]. So, an optimal thermal design for a data center cooling is a major engineering challenge. The characterization of a commonly occurred transient scenario is critically important for the optimal cooling design – but limited literature exists in this area [2]. Shields et al.[3] have studied the transient thermal response to a sudden power failure in a data center. Other possible transient scenarios in data centers include IT equipment upgrades, facility level upgrades, and time-varying cooling power requirements. A real time control system that can monitor temperature fields, and adjusts the input parameters, such as cooling air velocity, would facilitate the parametric design of data center cooling. Such control systems need an enabling algorithm that can rapidly predict the temperature field in a data center with reasonable accuracy.

A typical air-cooled data center can be modeled as a multi-scale thermal system with turbulent, convective airflow condition. Since the heat dissipation could be as high as 40 kW/ rack in the multi-scale domain, a lumped capacitance type model usually does not have a sufficient level of predictability. Alternatively, full-scale CFD/HT simulation [4] would be too slow for dynamic modeling. To address this problem, we developed a proper orthogonal decomposition (POD) based reduced order model that captures the transient characteristics of a data center with reasonable accuracy and resolution. The turbulent air temperature field, inside a multi-scale data center, can be modeled as a system of large degree of freedom (DOF). Full-scale simulation of such large DOF system, typically with more than 10^4 data points, requires tremendously large computational resources. The proposed reduced order model utilizes POD-based modal decomposition to determine optimal basis functions that can capture dominant characteristics of the temperature field. Such model reduction has been utilized in predicting steady state temperature field in a data center [5]. The reduced order model for a transient temperature field involves determination of time-varying POD coefficients. The proposed model is validated against standard CFD/HT software. The model can generate new temperature data based on the existing temperature ensemble.

2. METHODOLOGY

To model air temperature field inside a data center, we consider incompressible, turbulent airflow with negligible radiation, viscous dissipation, and free convection. Air is also assumed to be a pure substance with invariant thermophysical properties. With these simplifying assumptions, the governing equations for the convective field inside a data center are given by

$$\partial_i V_i = 0 \qquad (1)$$

$$\partial_0 V_j + V_i \partial_i V_j = -\frac{1}{\rho}\partial_i P + \partial_i[(\gamma + \gamma_t)\partial_i V_j] \quad (2)$$

$$\partial_0 T + V_i \partial_i T = \partial_i[(\alpha + E_h)\partial_i T] + \frac{q'''}{\rho c_p} \quad (3)$$

These are solved numerically along with standard *k-ε* model. The corresponding model equations are given by

$$\frac{\partial k}{\partial t} + div(\rho V k) = div([\mu_{lam} + \frac{\rho\gamma_t}{\sigma_k}]gradk) + \rho\gamma_t G - \rho\varepsilon \quad (4)$$

$$\frac{\partial \varepsilon}{\partial t} + div(\rho V \varepsilon) = div([\mu_{lam} + \frac{\rho\gamma_t}{\sigma_\varepsilon}]grad\varepsilon) + C_{1\varepsilon}\rho\gamma_t G\frac{\varepsilon}{k} - C_{2\varepsilon}\rho\frac{\varepsilon^2}{k} \quad (5)$$

The corresponding parameters used in the calculation are

$$C_{1\varepsilon} = 1.44; C_{2\varepsilon} = 1.92; \sigma_k = 1; \sigma_\varepsilon = 1.3; \mathrm{Pr}_t = 0.85$$

We use commercially available Fluent 12.1 for the numerical simulation. First order upwind scheme is used to capture multi-scale and direction-sensitive flow information. For transient flow, first order implicit scheme is employed because it ensures solution stability. The coupled solver, together with SIMPLE scheme, is employed for pressure-velocity coupling [6].

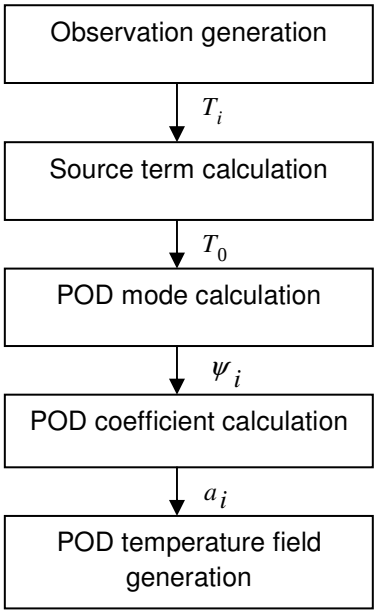

Figure 1: POD based algorithm for reduced order modeling

We use POD-based reduced order modeling framework - similar framework was previously used in analysis of steady transport processes with the large-DOF system [7]. The POD generates optimal basis functions for an ensemble of temperature observations from CFD/HT simulation. The pivotal feature of POD basis function or POD modes is its optimality: it needs fewer basis functions than the dimension of the function space to capture dominant characteristic of the temperature field. As shown in Figure 1, the algorithm involves application of linear algebra- similar to steady state POD modeling [5].A temperature ensemble with time as a parameter $T_i(x, y, z, t_i)$ is obtained by CFD simulation. $T_i \in R^{n\times m}$ is called data matrix with m observations of n dimensional temperature field. Since the temperature field is linear, it can be expressed as a linear expansion with suitable basis functions:

$$T(x, y, z, t) = T_0(x, y, z) + \sum_{i=1}^{l\le m} a_i(t)\psi_i(x, y, z) \quad (6)$$

The time-independent part of the temperature field is modeled by a source term, $T_0(x, y, z)$. The time varying part is modeled by a sum of products of time varying POD coefficients, $a_i(t)$ and time-independent POD modes, $\psi_i(x, y, z)$. The non-homogenous source term is calculated as:

$$T_0(x, y, z) = (\sum_{i=1}^{m} T_i(x, y, z, t_i))/m \quad (7)$$

At this point, we concentrate on finding optimal basis functions for the modified temperature ensemble:

$$T_i^*(x, y, z, t_i) = T_i(x, y, z, t_i) - T_0(x, y, z) \quad (8)$$

The POD is a stochastic tool that transforms a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible - it accounts for as much of the variability in the data as possible- and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components. These principal components can span entire set of observations so these are called optimal basis functions or POD modes. These basis functions ψ_i are optimal because they would have maximum projection of T_i^* onto ψ_i suitable normalized: $\max_{\phi\in L^2([0,1])} \frac{\left\langle \left|(T^*, \psi)\right|^2 \right\rangle}{\|\psi\|^2}$. The problem now is

to extremise $\left\langle \left| (T^*, \psi) \right|^2 \right\rangle$ subject to constraint $\|\psi\|^2 = 1$. The corresponding functional for this constrained variational problem is

$$J(\psi) = \left\langle \left| (T^*, \psi) \right|^2 \right\rangle - \lambda(\|\psi\|^2 - 1) \quad (9)$$

And a necessary condition for the variation is that the functional derivative vanishes for all variations of $\psi + \delta\theta \in L^2([0,1]), \delta \in \mathrm{R}$ is

$$\frac{d}{d\delta} J[\psi + \delta\theta]_{\delta=0} = 0 \quad (10)$$

On simplification of Equation (10) coupled with method of snapshots [8], we obtained the governing equation for POD modes:

$$R\psi = \lambda\psi \quad (11)$$

Where,

$$R = (1/m) T^{*t} T^* \in R^{m \times m} \quad (12)$$

POD modes are independent of time and the corresponding time-varying part of the temperature field is modeled by POD coefficients. The POD coefficients are determined via matrix inversion at a time instant when the temperature data $T(x, y, z, t)$ are available from CFD/HT simulation. Determination of temperature data at an intermediate time instant, when temperature is unknown, necessitates interpolation of POD coefficients at the time instants when temperature data are available. Alternatively, POD coefficients can be calculated by Galerkin projection [9]. In this technique, we project the governing equation in the space spanned by POD modes to obtain

$$\int_\Omega [\psi_i . \{ (\frac{\partial}{\partial t} + \vec{V}.\nabla)(T_0 + \sum_{i=1}^{l \le m} a_i \psi_i) - (\nabla.(\alpha + E_H)\nabla)(T_0 + \sum_{i=1}^{l \le m} a_i \psi_i) - \frac{q'''}{\rho c_p} \}] dxdydz = 0 \quad (13)$$

Also, POD modes are orthogonal basis function. Therefore

$$\psi_i . \psi_j = 0, \; i \neq j; \; \psi_i . \psi_j = |\psi_i|^2, \; i = j \quad (14)$$

We obtain a series of first order ordinary differential equations for POD coefficients from Equation (13) by suitable usage of orthogonality of POD modes:

$$A \frac{da_i}{dt} + B a_i + C = 0 \quad (15)$$

The system of ordinary differential equations represents evolution of the POD coefficients. The time-invariant constants A, B, C represent influence of the corresponding POD modes, gradients of the POD modes, the air velocity field and the corresponding spatial locations. We use finite difference method to obtain the gradients of discrete POD modes. Also, $a_i(t=0) = 0$, due to a specified initial temperature. The system of first order linear non-homogenous ordinary differential equations with constant coefficients is solved by the Runge-Kutta method based "ode 45" solver in commercial software MATLAB. POD coefficient determination by Galerkin projection is computationally involved and time consuming in comparison to interpolation techniques. However, unlike interpolation, Galerkin projection has no restriction on the number of variables. So for multivariable problems, such as transient air temperature with variable CRAC fan speed, Galerkin projection must be used.

The relative energy content, E_k of a particular POD mode is an important parameter because it describes the degree of significance of the mode in overall temperature field. The relative energy content of a POD mode can be estimated by the fraction of the corresponding eigenvalue :

$$E_k = \frac{\lambda_k}{\sum_{j=1}^{n} \lambda_j} \quad (16)$$

The eigenvalue is obtained by taking square root of the corresponding singular value of the matrix R,

$$R = V \sum V^t \quad (17)$$

$$\lambda_k = \sqrt{\sum_{kk}} \quad (18)$$

Although enhancement of computation efficiency is the aim of the reduced order modeling framework, we must establish accuracy of the model by validating against detailed CFD/HT solution. The error of a POD predicted temperature field is

$$T_{error}(x, y, z, t) = T_{POD}(x, y, z, t) - T_{CFD/HT}(x, y, z, t) \quad (19)$$

3. RESULTS

3.1. *Case study definition*

We apply the POD based modeling framework to a data center cell, shown in Figure 2. There are eight racks in the cell of height *2000 mm*. In the present case study, we simulate 5 kW/ rack heat dissipation. Each rack includes six heat dissipating servers – simulated by thin-walled aluminum ducts with uniform volumetric generation within the enclosed space. The racks are arranged symmetrically in the hot aisles-cold aisles configuration – with perforated floor in the cold aisle for cooling air supply. The cooling airflow, supplied from the CRAC unit, is distributed via the raised floor plenum through perforated tiles in the cold aisles and driven into heat-dissipating servers by server fans. Hot air from the servers returns to CRAC via hot aisles. The initial conditions in the numerical solution are:

$$u = 0; v = 0; w = 0; T = 15^0 C; k = 1 m^2/s^2; \varepsilon = 1 m^2/s^3$$

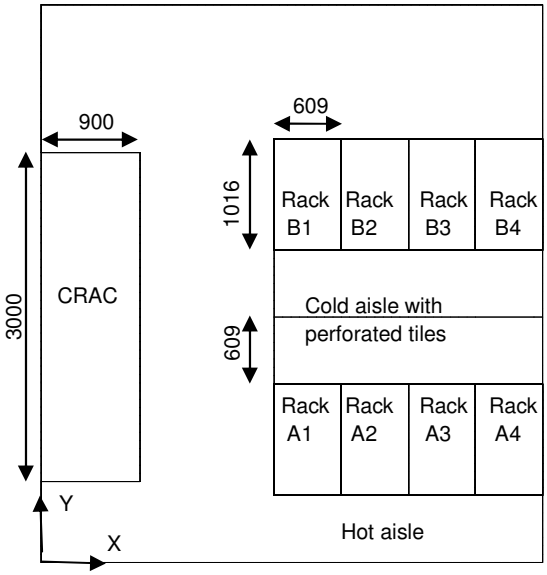

Figure 2: Case study data center cell top view: Dimensions are in mm

We use 182,000 grid cells. This represents a compromise between lack of accuracy with 100,000 grid cells, and high cost involved with 300,000 grid cells [10]. We use velocity inlet boundary condition for CRAC inlet. The inlet velocity is chosen to be 1 m/s in accordance with steady state airflow rate calculation for cooling 40 kW load. The air pressure drop, across 35mm thick and 20% porous plenum wall, is modeled by outlet vent boundary condition with loss coefficient of 0.5. The server inlets are modeled as pressure drop boundary conditions and server outlets are modeled by exhaust fan boundary condition [11].

3.2. *Characterization of the transient scenario*

The initial condition in the present cases study corresponds to a facility start-up scenario. Initially, we have a uniform temperature field of 15^0 C and uniform mean velocity field of 0 m/s. Then, the servers are subjected to heat loads and cooling air starts to flow into the plenum at 1 m/s. The resultant flow field is assumed turbulent and the heat transfer is assumed via forced convection. The air temperature rises due to heat dissipation from servers and ultimately reaches steady state at around 200 s, as shown in Figure 3. We place eight test points at the inlets of the highest servers of eight racks, at a height of 1830 mm for a 2000 mm rack. The test points are presumed to be representative of the entire domain; considering we are chiefly interested in the temperature field at the inlets of the racks. We analyze the transient data center temperature field from 0-200 s at 10 s time intervals to generate an ensemble, $T_i(x, y, z, t_i)$ of size $182{,}000 \times 20$.

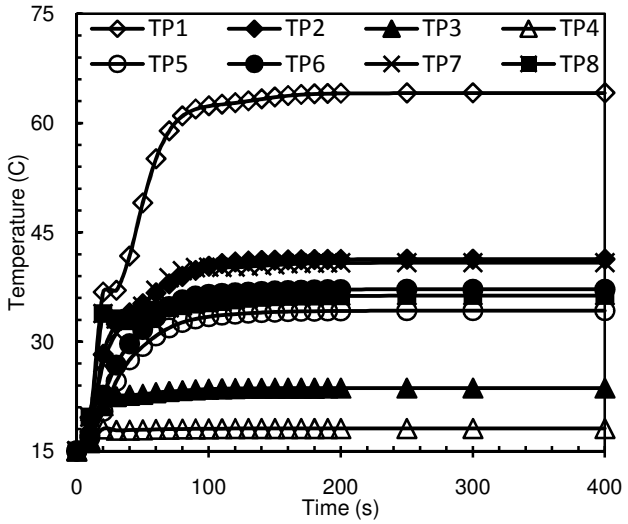

Figure 3: Transient response of different test points in the data center. The test points are at the inlets of the highest servers in the corresponding racks. The temperature profiles of multiple test points indicate that the system reaches steady state at 200 s

3.3. *Optimality of POD modes*

As demonstrated in Figure 4, first seven POD modes capture 99.9% of the total energy and hence, span the dominant component of the temperature field. So, 7 POD modes are sufficient to describe the temperature field. This modal reduction significantly improves POD based temperature computation.

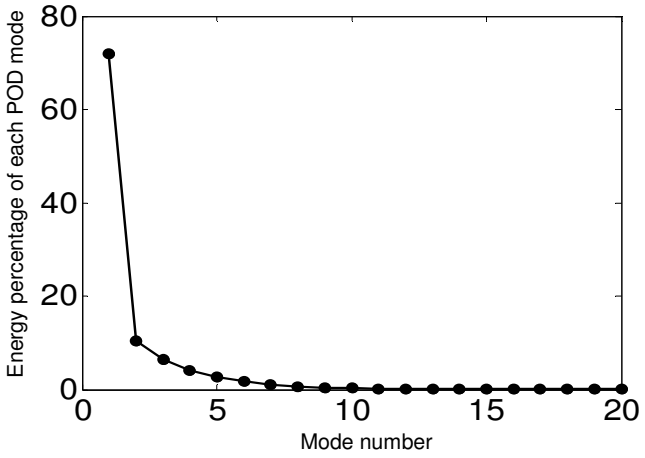

Figure 4: Energy percentage captured by each POD mode versus the mode number: First 7 modes capture 99.9% of total energy

3.4. *Determination of POD coefficients*

As shown in Figure 5, POD coefficients calculated by Galerkin projection or interpolation are almost identical. This is particularly true for the given case study where only variable is time. The interpolation based method, however, would not work

in multivariable situation. The interpolation is preferred method for single variable domain because it is faster than Galerkin projection.

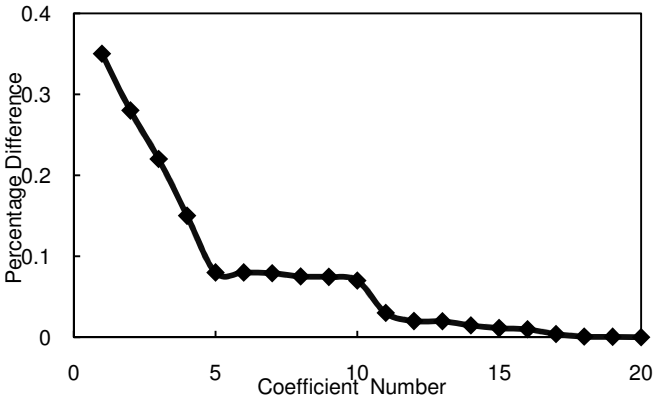

Figure 5: Percentage difference between POD coefficients calculated by Galerkin projection and interpolation. Negligible difference validates equivalence of two methods in a single variable case

3.5. *Model validation*

We focus at the inlets of the A and B racks. In Figure 6-9, we present the inlet temperature fields, the deviation between CFD/HT and POD temperature, and the velocity fields. The solid lines, parallel to Z axis, represent the boundary between two adjacent racks. Each rack is comprised of a stack of six servers, indicated by dotted line in Figure 6-9.

First we observe the temperature field at t=30 s, for both racks as shown in Figure 6 (a)-(b) and Figure 7 (a)-(b) respectively. As demonstrated in Figure 7 (c) and Figure 8 (c), the errors for all racks are within *~0.1%*. The spatial variations in the velocity fields shown in Figure 6 (d) and Figure 7 (d) can be correlated with the corresponding temperature field.

The results, presented in Figure 6 and Figure 7, indicate that POD can be utilized to replicate the CFD data with high fidelity. We have temperature field from $0-200\,s$ with $10\,s$ interval. We use POD based reduced order model to compute the unknown temperature fields at an intermediate time instant, $t=15\,s$, as shown by Figure 8 (a)-Figure 9 (a). We computed the temperatures at the same locations and time instant by CFD/HT simulations shown in Figure 8(b)-Figure 9(b). As evident from Figure 8(c)-Figure 9(c), two sets of results are in close agreement. Remarkably, the computational time for POD temperature field is approximately 4 s, whereas the corresponding time for the CFD/HT computation is approximately 8 min. This is a significant improvement in computational efficiency, without compromising solution accuracy.

4. CONCLUSION

We developed a POD-based reduced order modeling framework that can efficiently capture the transient air temperature field inside a data center. The reduced order model is validated against standard CFD/HT predictions and observed to be nearly two decades faster without compromising accuracy. The modeling framework is suitable for a data center control system that can regulate flow variables, and thereby deliver higher energy efficiency. Although the reduced order model is promising, it has some limitations. A multi-variable situation, such as a transient temperature field with variable cooling airflow, cannot be analyzed with the existing model. Then, the heat dissipation from the servers can be modeled more exhaustively with individual compact models.

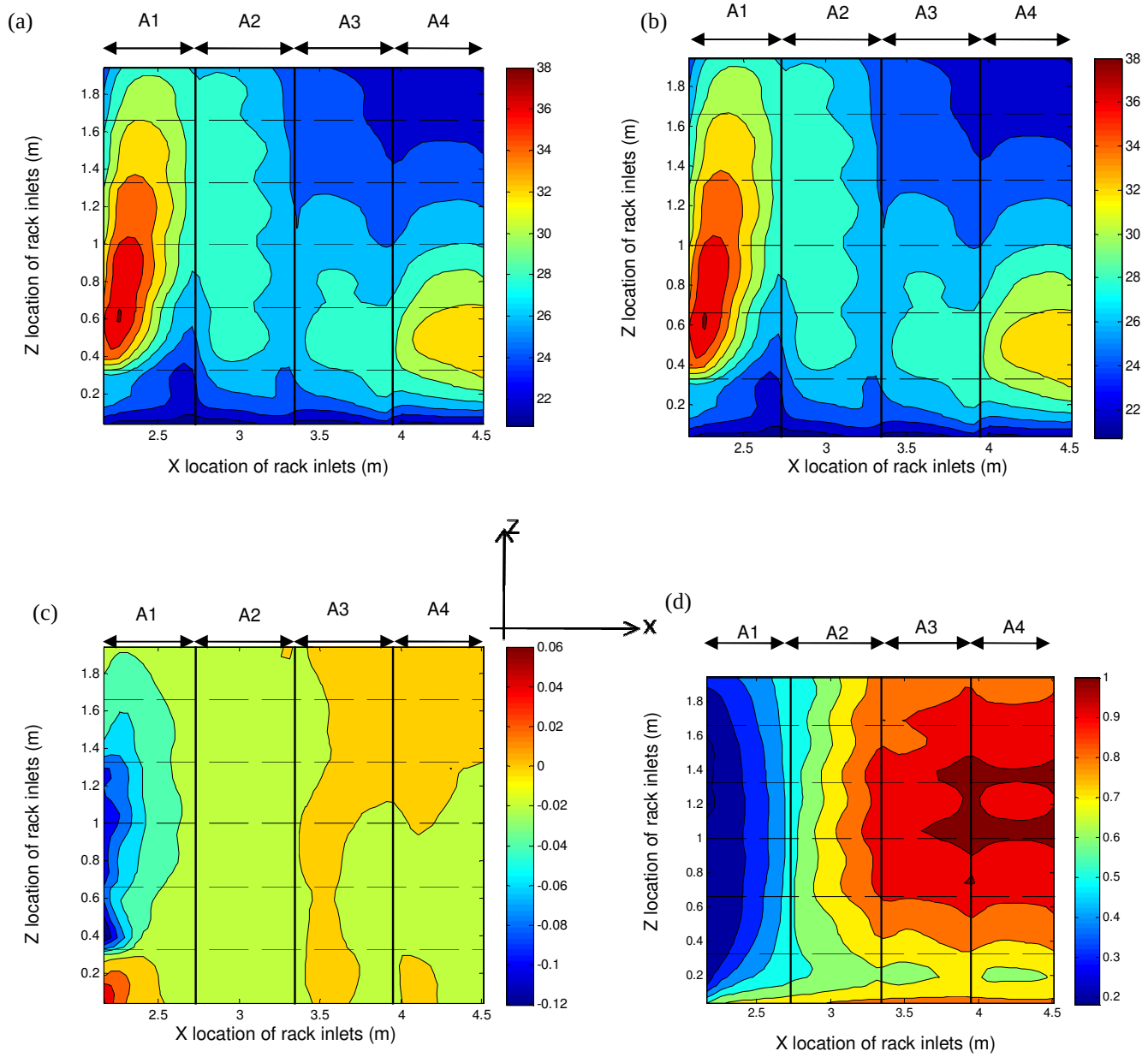

Figure 6: Validation of POD modeling framework at rack A inlets at 30 s: (a) POD based temperature (^{0}C) contour; (b) CFD based temperature (^{0}C) contour; (c) Error in temperature field; (d) CFD based velocity field (m/s).

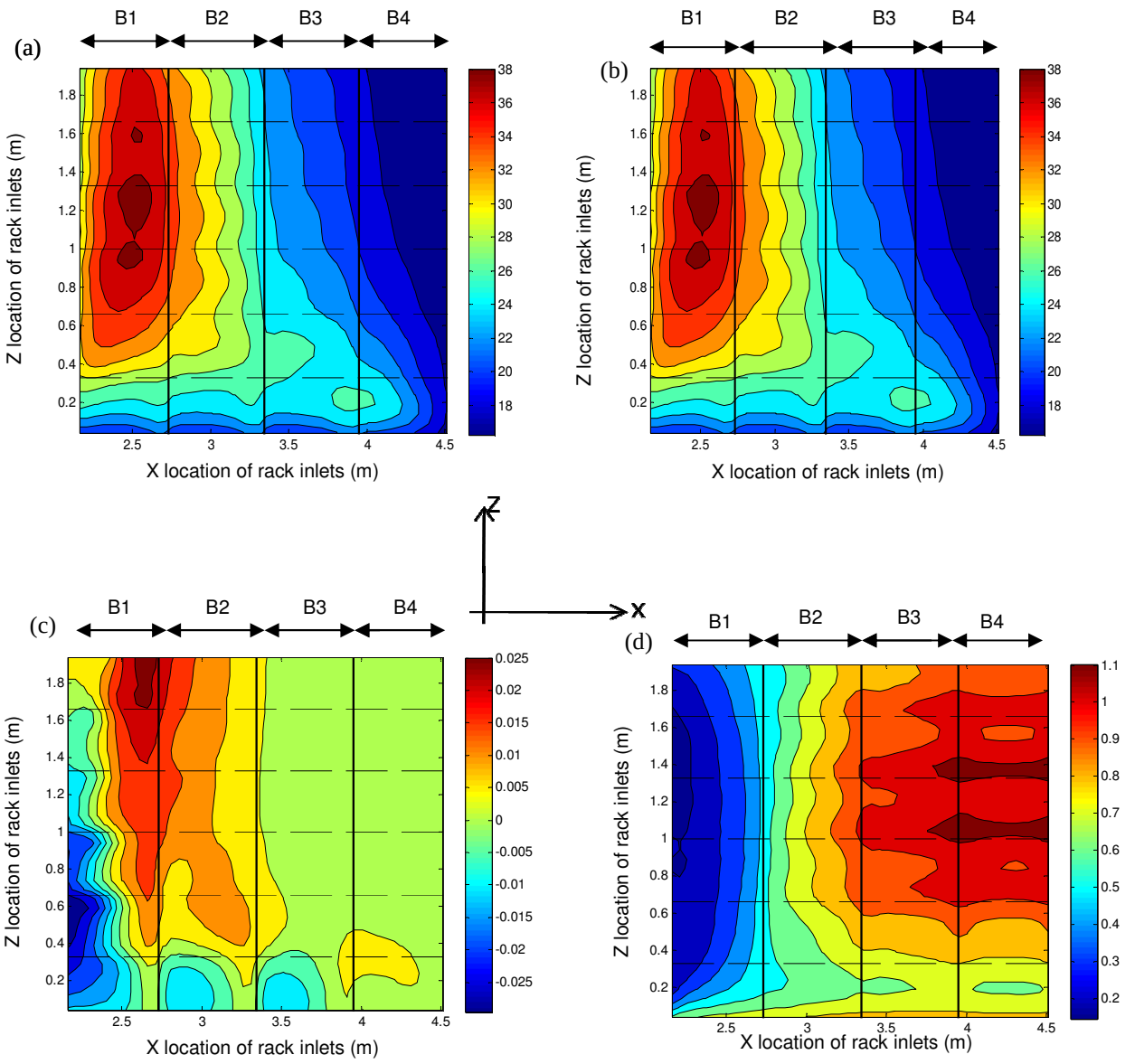

Figure 7: Validation of POD modeling framework at rack B inlets at 30 s: (a) POD based temperature (0C) contour; (b) CFD based temperature (0C) contour; (c) Error in temperature field; (d) CFD based velocity field (m/s).

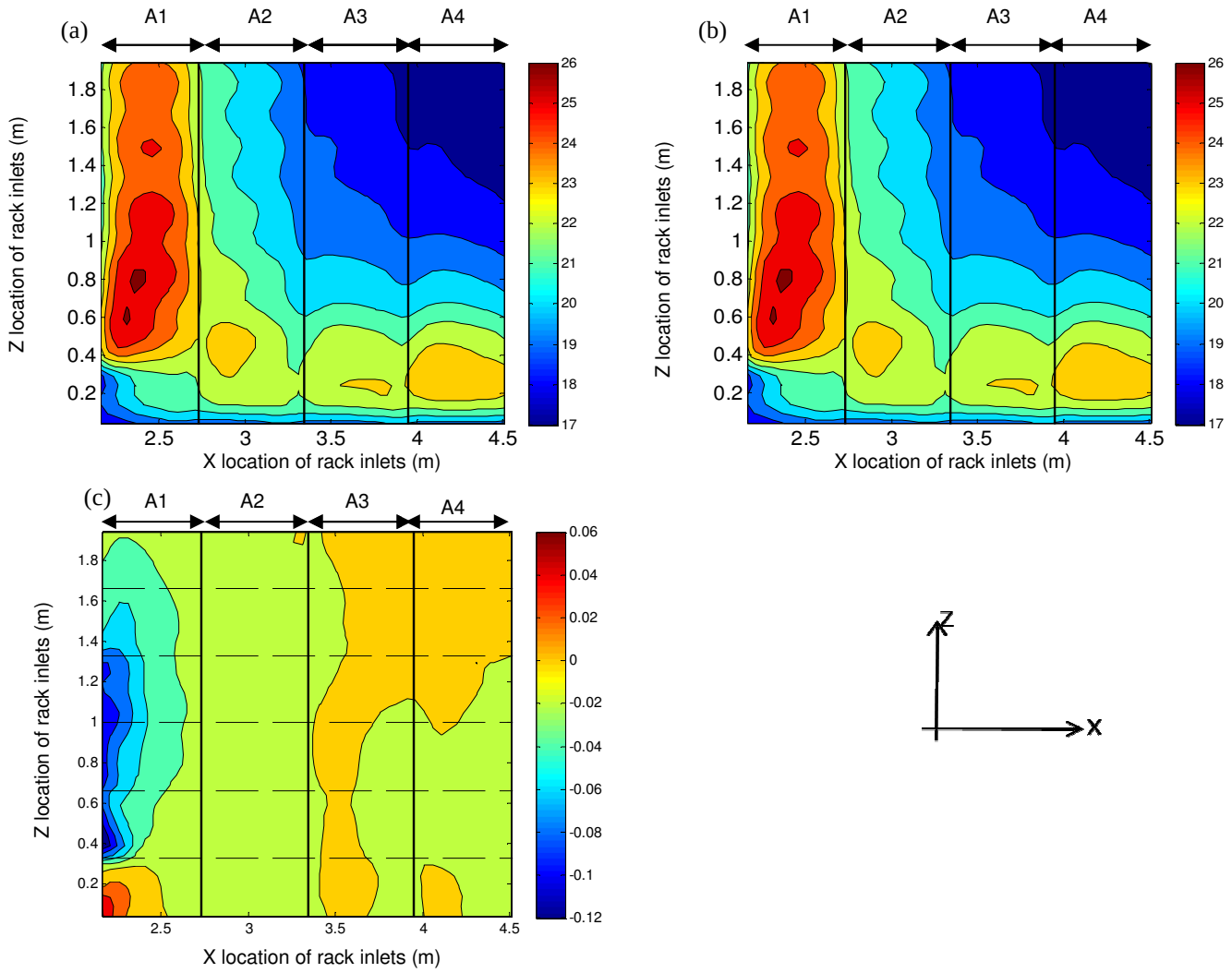

Figure 8: Validation of POD modeling framework at rack A inlets at 15 s: (a) POD based temperature (^{0}C) contour~4 s; (b) CFD based temperature (^{0}C) contour~8 min; (c) Error in temperature field.

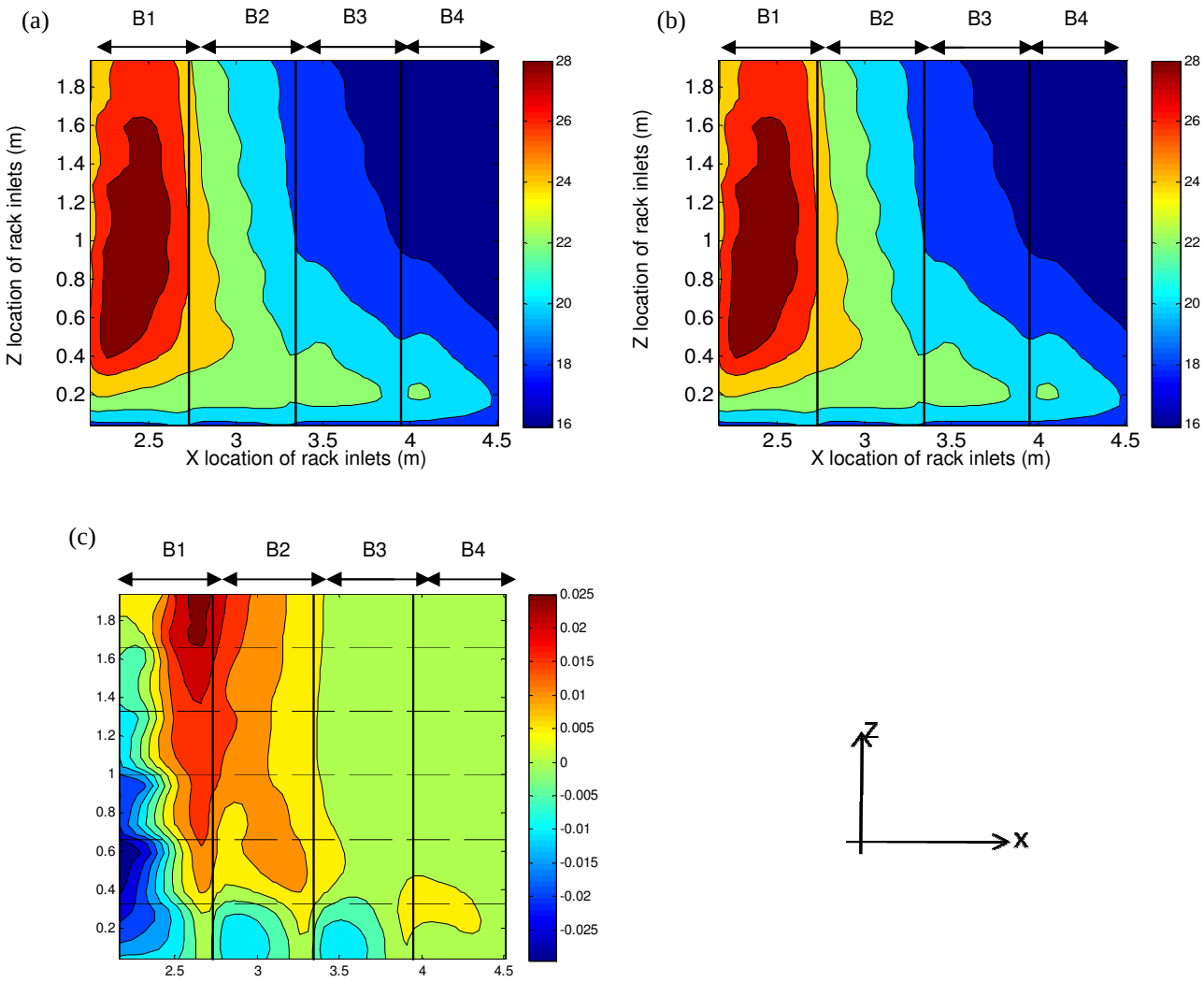

Figure 9: Validation of POD modeling framework at rack B inlets at 15 s: (a) POD based temperature (^{0}C) contour, t~4s; (b) CFD based temperature (^{0}C) contour, t~8 min; (c) Error in temperature field.

NOMENCLATURE

ψ	POD modes
a	POD coefficients
V	Average velocity
ρ	Density
γ	Kinematic viscosity
γ_t	Kinematic viscosity of turbulence
P	Pressure
α	Thermal diffusivity
E_h	Eddy diffusivity
k	Turbulent kinetic energy
ε	Dissipation
λ	Eigenvalue
Σ	Singular Value
l	Number of Retained POD modes
	Superscripts
t	Transpose

REFERENCES

[1] Christian L. Belady, P. E., 2007, "" In the data center, power and cooling costs more than the it equipment it supports", http://www.electronics-cooling.com/2007/02/".

[2] Somani, A., 2008, "Advanced thermal management strategies for energy-efficient data centers",Masters Thesis, Georgia Institute of Technology.

[3] Shields, S., 2009, "Dynamic thermal response of the data center to cooling loss during facility power failure",Masters Thesis, Georgia Institute of Technology.

[4] Rambo, J. D., 2006, "Reduced-order modeling of multiscale turbulent convection: Application to data center thermal management",Ph.D. Dissertation, Georgia Institute of Technology.

[5] Samadiani, E., and Joshi, Y., 2010, "Proper orthogonal decomposition for reduced order thermal modeling of air cooled data centers", Journal of Heat Transfer, 132(Compendex), pp. 1-14.

[6] Patankar, S., 1980, Numerical heat transfer and fluid flow, Hemisphere Pub.

[7] Berkooz, G., Holmes, P., and Lumley, J. L., 1993, "The proper orthogonal decomposition in the analysis of turbulent flows", Annual Review of Fluid Mechanics, 25(1), pp. 539-575.

[8] Sirovich, L., 1987, "Turbulent and the dynamics of coherent structures", Quarterly of Applied Mathematics, 45(3), pp. 583-590.

[9] P.Holmes, J. L. L., G. Berkooz, 1996, Turbulences, Coherent Structures, Dynamical Systems, and Symmetry, Cambridge University Press, Great Britain.

[10] Rambo, J., and Joshi, Y., 2006, "Convective transport processes in data centers", Numerical Heat Transfer; Part A: Applications, 49(Compendex), pp. 923-945.

[11] Crippen, M., Alo, R., Champion, D., Clemo, R., Grosser, C., Gruendler, N., Mansuria, M., Matteson, J., Miller, M., and Trumbo, B., 2010, "BladeCenter packaging, power, and cooling", IBM Journal of Research and Development, 49(6), pp. 887-904.