Reduced-order Modeling Framework for Improving Spatial Resolution of Data Center Transient Air Temperatures

Rajat Ghosh, Yogendra Joshi
Georgia Institute of Technology
801 Ferst Drive
Atlanta, GA 30332-0405
rajat.ghosh@gatech.edu,
yogendra.joshi@me.gatech.edu

Levente Klein, Hendrik Hamann
IBM TJ Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598
kleinl@us.ibm.com,
hendrikh@us.ibm.com

Abstract

A proper orthogonal decomposition (POD)-based modeling framework is developed for improving the spatial resolution of transient rack air temperature data collected in a heterogeneous data center (DC) facility. Blocking cooling air inflow into racks periodically, three sets of transient temperature data are collected at the outlets of electronic equipment residing in three different racks. Using various combinations of initial discrete data as ensembles, the capability of the proposed POD/ interpolation framework for predicting new temperature data is demonstrated. The accuracy of POD-based temperature predictions is validated by comparing it to corresponding experimental data. The root mean square deviations between experimental data and POD-based predictions are found to be on the order of 5%.

Keywords

Data center, transient temperature measurement, proper orthogonal decomposition, dynamic events.

Nomenclature

POD	Proper orthogonal decomposition.
C.E.P.	Captured energy percentage.

1. Introduction

Data centers (DCs) are mission-critical information technology (IT) and telecommunications infrastructure, housing equipment in standard-size cabinets or racks. With facilities rapidly increasing cloud-computing-based e-commerce services supported by power-hungry electronic equipment, the energy consumption of data centers (DCs) has escalated in recent years [1, 2]. According to a DatacenterDynamics report in 2012, a staggering 20% rise in the DC energy consumption is projected [3] to represent 2% of the world's electricity consumption at an annual cost of $45 B. Furthermore, benchmarking studies reveal that the cooling systems in a typical data center consume as much as 30%-50% of the total facility power, and the life-cycle cost of cooling is fast becoming comparable to that of IT equipment for commodity computing [4]. Therefore to optimize cooling cost for a given IT load, real-time monitoring/ modeling of the cooling-related physical variables such as air velocity and temperature is highly desirable. Models for transient air temperature rise are particularly important during power loss [5, 6].

With increasing power density [7] inside a DC facility, the spatial and temporal temperature gradients have become too large to be resolved by a lumped system model. Therefore, one of the enabling features any DC modeling framework should have is the capability of high resolution data prediction under transient operating conditions. Currently, this is achieved through full-field numerical modeling [8], or detailed measurements [9]. Since DC thermal mapping involves large degrees of freedom, both of these techniques are time-intensive for real-time monitoring and feedback control. A lack of modeling algorithms for improving spatial resolution of thermal data stymies the development of an efficient monitoring technology and a model-based thermal control system [10] in data centers. In image processing, POD-based modal reduction is a prevalent technique for improving spatial resolution [11]. Recently in DC thermal management, the application of POD-based algorithms for improving parametric resolution of convective air temperature field has been made [12]. The literature cites a few studies that use time as the parameter [13]. To the best of authors' knowledge, no study has been conducted with spatial location as the parameter for a POD framework.

This paper develops a POD-based model for improving spatial resolution of experimentally-acquired transient temperature data inside a heterogeneous DC facility. Three different datasets are compiled by measuring air temperatures at multiple locations at the outlets of IT equipment in three different racks. The POD-based framework is applied on the transient dataset in order to generate spatially resolved predictions. The modeling accuracy is reported over a pertinent transient window of 1500 s.

2. Modeling Algorithm

Figure 1 shows the flowchart for the POD-based algorithm. In this algorithm, the locations of temperature sensors are used as the parameters for the calculation of POD-coefficients. At first, an ensemble of data center air temperatures is constructed, as a $m \times n$ matrix, in which the row index, m is the dimensionality of the temperature data, and column index, n is the number of times the design parameter of the system is changed. A parameter-independent temperature $T_o(t)$ is computed by averaging all temperature data for a field point:

$$T_0(t) = \frac{T(x,y,z,t)}{n}. \quad \textbf{(1)}$$

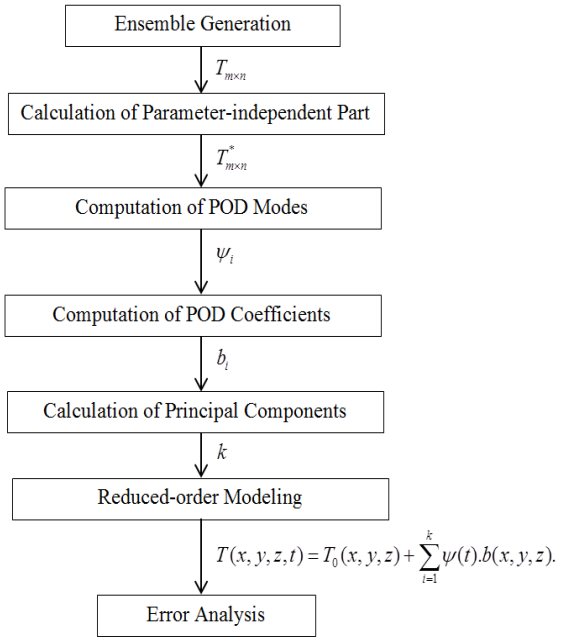

Figure 1: POD/interpolation-based framework for improving spatial resolution of experimental data. The POD modes are functions of time. POD coefficients are functions of spatial locations.

The parametric component of the temperature data, T^* is calculated by subtracting T_0 from each columns of $T_{m\times n}$.

$$T^*_{m\times n}(x,y,z,t)=T(x,y,z,t)-T_0(t). \quad \textbf{(2)}$$

The parametric component, $T^*_{m\times n}(x,y,z,t)$ is modeled as sum of products of POD modes, $\psi(t)$ and POD coefficients, $b(x,y,z)$. The POD modes are computed by a singular value decomposition-based technique [14] which ultimately leads to an eigenvalue equation for POD modes:

$$R\psi=\lambda\psi, \quad \textbf{(3)}$$

where,

$$R=\frac{1}{m}\left(T^*\right)^{tr}T. \quad \textbf{(4)}$$

The energy content captured by a POD mode is defined as the relative weightage of the corresponding eigenvalue, λ_i in the overall eigenvalue space, defined as:

$$E_i=\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}. \quad \textbf{(5)}$$

Overall, the governing equation for temperature is:

$$T(x,y,z,t)=T_0(t)+\sum_{i=1}^{k}\psi_i(t)b_i(x,y,z). \quad \textbf{(6)}$$

After POD modes are computed, POD coefficients are calculated by interpolation/ extrapolation. Unlike Galerkin projection-based technique [14], the interpolation/ extrapolation-based technique is computationally more efficient, particularly in working with discrete experimental data [15]. The upper bound for the sum of product terms in (6) is the principal component number, k which is calculated by:

$$\left(\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_i}>C.E.P\right)\bigcup\min(k). \quad \textbf{(7)}$$

$C.E.P.$ is defined as the captured energy percentage. For a high-fidelity model, it could be as high as 100%; whereas, for a crude model, it could be as low as 70% [13]. For this study, $C.E.P.$ is chosen to be 99%.For estimating the accuracy of the model, a prediction error is defined as:

$$E_{\text{Prediction}}=\left(T_{\text{Experiment}}-T_{\text{POD}}\right). \quad \textbf{(8)}$$

3. Case Study

The capability of the proposed modeling framework is demonstrated for three transient temperature datasets. The transients are introduced by periodic closing and opening of the rack inlet. For transient-1 and transient-2, temperature data are measured at the rack exits with the cabinet doors open. These racks have chilled water cooled rear-door heat exchangers. Hence, the closing of the rack is not possible with the experimental setup placed inside the cabinet. For transient-3, the rear door of the cabinet is kept closed because it does not have any rear door heat exchanger. The ensemble for the POD-based analysis is generated experimentally. Figure 2 shows the layout of the raised floor data center facility. Table 1 specifies the pertinent details of the facility and the experimental conditions.

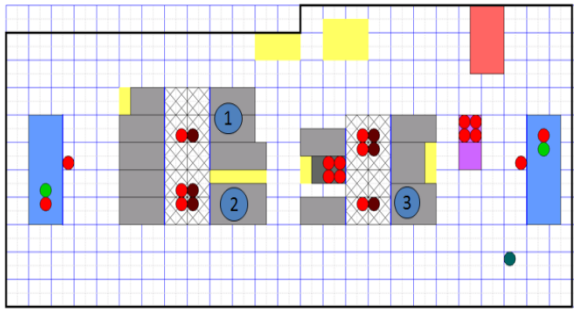

Figure 2: Layout of the data center test facility. Gray boxes represent IT equipment racks, yellow boxes storage units, blue boxes computer room air conditioning units (CRACs), red box power distribution units (PDUs). Hashed squares are perforated tiles. Red and green dots are the general-purpose sensors deployed in the facility. Blue circles with numbers 1, 2, 3 are the locations where transient measurements were conducted.

The temperature data are acquired by 10 calibrated chromel/alumel (K-type) thermocouples made from 24 gauge (0.02"/0.511 mm in diameter) wire with 25.4 mm (1") length sticking out in the air. The estimated uncertainty in the thermocouple measurement is 0.1 ^{0}C. Ten thermocouples are equally distributed on a pole over a 2,286 mm (7.5 ft.) height. Each thermocouple is fixed mechanically to the pole by wire wraps and extended until the data acquisition hardware. Table 2 includes the heights of different temperature sensors. The transient temperature data are acquired at every 1.5 s.

Table 1: Specifications of the data center facility

Part	Parameter	Details
Facility	Cross-section	186 sq. m. (2000 sq. ft.)
	Height	3 m. (10 ft.)
CRAC -1	Flowrate	5.85 m^3/s (12,400 cfm)
CRAC -2	Flowrate	5.85 m^3/s (12,400 cfm)
Perforated Tile	Flowrate	0.09 m^3/s (180 cfm)
Rack-1	Power	2013 W.
	Dimension	610 mm x610 mm x2000 mm.
Rack-2	Power	2075 W.
	Dimension	610mm. X549mm. x2000 mm.
Rack-3	Power	2753 W.
	Dimension	610 mm. x762 mm. x2000 mm.

For POD-based modeling, the choice of temperature ensemble is critical. In the present case study, the dimensionality of the transient temperature data or the number of thermocouples is equal to 10, i.e., $m = 10$. The column space dimension of the temperature ensemble is chosen as, $n = 2$. That means two sensor data constitute the ensemble. This choice is justified because the generation of temperature ensemble requires minimum number of temperature sensors.

Table 2: Locations of temperature sensors

Sensor #	*Height (mm (ft.))*
1	2286 (7.5)
2	2012 (6.6)
3	1768 (5.8)
4	1524 (5)
5	1280 (4.2)
6	1006 (3.3)
7	762 (2.5)
8	488 (1.6)
9	244 (0.8)
10	0(0)

As observed in Figure 3Figure 5, all transients show a similar general pattern: when the air intake is blocked, the exhaust temperature rises, and vice versa. However, for some locations, such as sensors at 2286 mm (7.5 ft.), 2012 mm (6.6 ft.), and 1768 mm (5.8 ft.) for the transients 1 and 2, the measured temperatures remain nearly unchanged. The sensor data at 2286 mm (7.5 ft.) for transient-3 indicate an opposing trend or a possible phase shift with respect to the rest of the sensor data: decreasing during blocking and increasing during unblocking. These anomalies may be attributed to possible cold air entrainment from the top of the aisle.

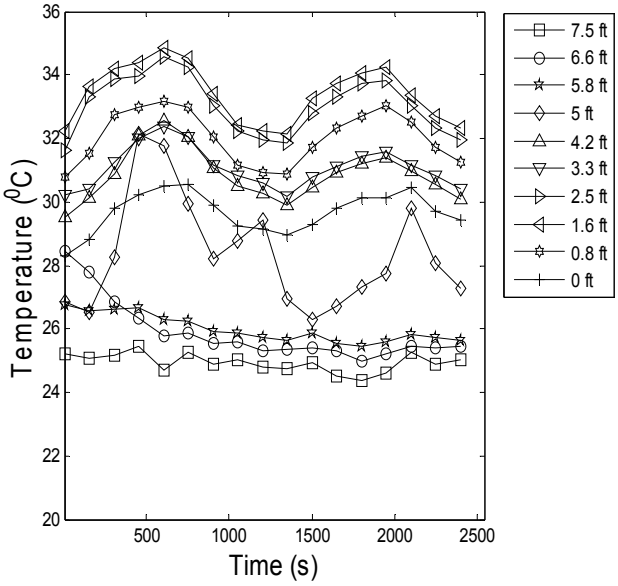

Figure 3: Experimentally-acquired temperature data showing the stratification during transient-1.

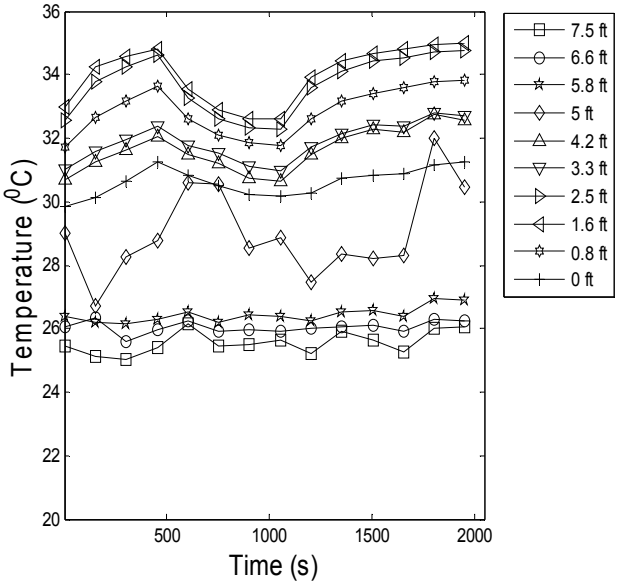

Figure 4: Experimentally-acquired temperature data showing stratification during transient-2.

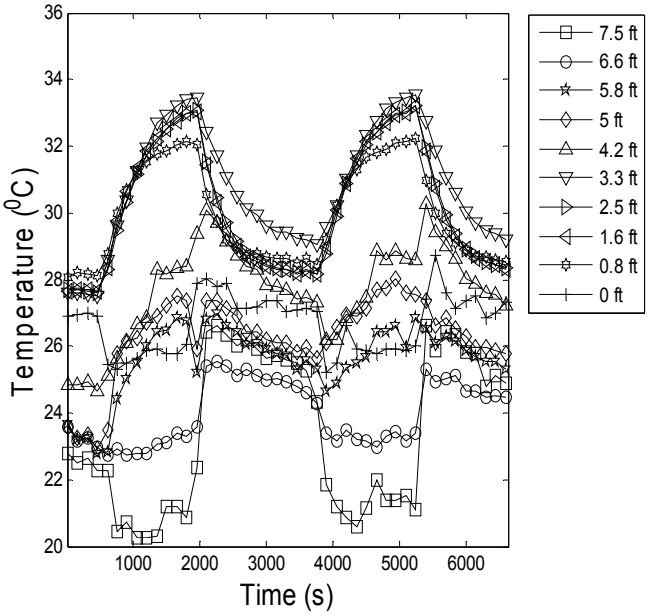

Figure 5: Experimentally-acquired temperature data showing stratification during transient-3.

4. Results and Discussion

For the three case studies discussed, POD-based temperature predictions are computed using an ensemble of two temperature datasets. For a test point, data acquired by two neighboring sensors constitute the ensemble. For example, for a test point coincident with sensor#5, measurement data acquired by sensors 4-6 constitute the ensemble. For the temperature sensors 1 and 10 which are located at the two extremes, two nearest neighbors are sensors 2-3 and 8-9 respectively. For a given test point, the POD-based algorithm as shown in Figure 1 is applied on the corresponding ensemble, and new temperature data are computed. The computational time for generating new temperature data is 1 sec. on a desktop computer with an Intel® Core2Duo 2.66 GHz processor and 4.00 GB RAM. For three transients (transient-1-transient-3), POD-based transient predictions are shown in Figure 6-Figure 8. As evident from Figure 6-Figure 8, the POD-based predictions are qualitatively similar to the corresponding experimental data.

To estimate quantitative accuracy, the POD-based temperature predictions are compared with the corresponding experimental data. For transient-1, Figure 9 compares POD-based temperature predictions at $t = 1500$ s with corresponding experimental data. The deviations between experimental data and POD predictions vary non-uniformly with sensors. The maximum errors, observed at the locations coincident with sensors 5 and 10, are 3^0C (10%) and -3^0C (~11%), respectively. The root mean square (RMS) value of the prediction error at $t = 1500$ s for transient-1 is 1.7^0C (~5.9%). Similar data analyses are conducted for transient-2 and transient-3, and they are shown in Figure 10 and Figure 11, respectively. For transient-2, the maximum error of 2.9^0C (~9%) is observed at the spatial location coincident with sensor-5. The RMS value of the prediction error at $t = 1500$ s for transient-2 is 1.4^0C (~4.6%). On the other hand for transient-3, the maximum error of -6.4^0C (~23%) is observed at the location coincident with sensor-10. The RMS value of the prediction error at $t = 1500$ s for transient-3 is 2.3^0C (~8.6%). Relatively large errors in some locations indicate the possible presence of hotspots. The POD-based framework is ultimately a statistical model, unable to capture local thermal characteristics.

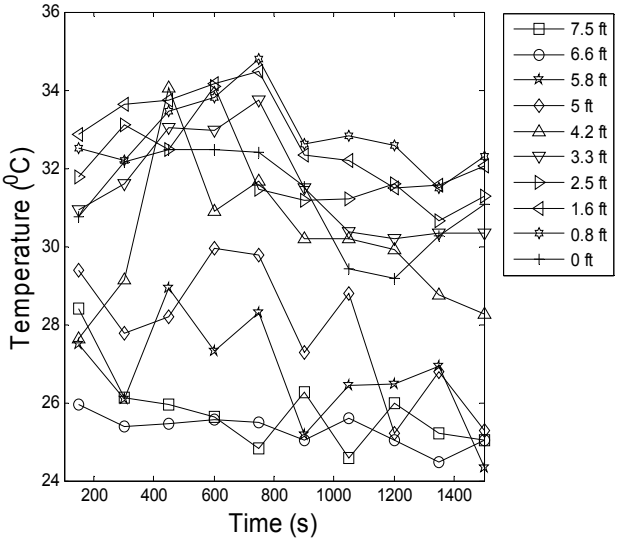

Figure 6: POD-based temperature data prediction for transient-1.

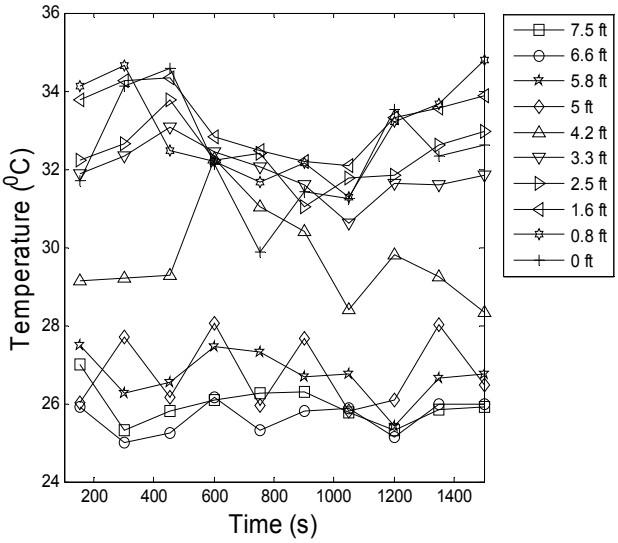

Figure 7: POD-based temperature data prediction for transient-2.

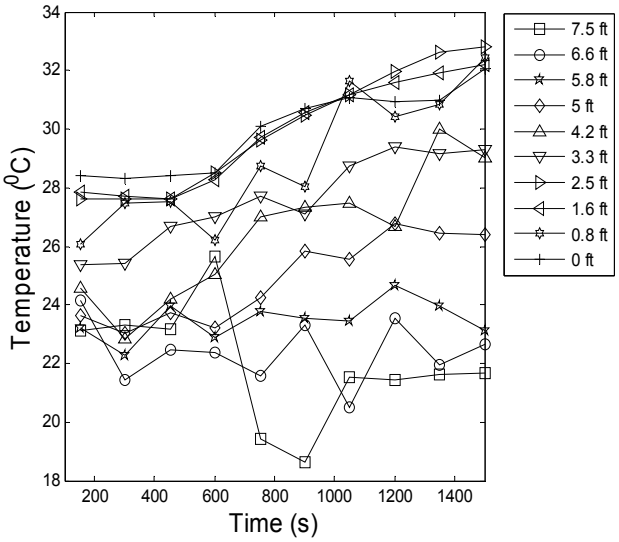

Figure 8: POD-based temperature data prediction for transient-3.

It is apparent from Figure 9-Figure 11 that the deviations between experimental data and corresponding POD predictions are varying non-uniformly along sensor heights. Therefore, it is imperative that transient data for different sensors are analyzed separately. For a finite time interval (0-1,500 s), the average errors and standard deviations are computed by finding the mean of the deviations between the experimental data and the corresponding POD predictions. For transient-1, transient-2, and transient-3, the results are presented in Figure 12-Figure 14, respectively. For the finite time interval (0-1,500 s), the overall RMS value of the average error for transient-1 is $1.2^0 C$. For transients-2 and 3, these values are $1^0 C$ and $1.6^0 C$, respectively.

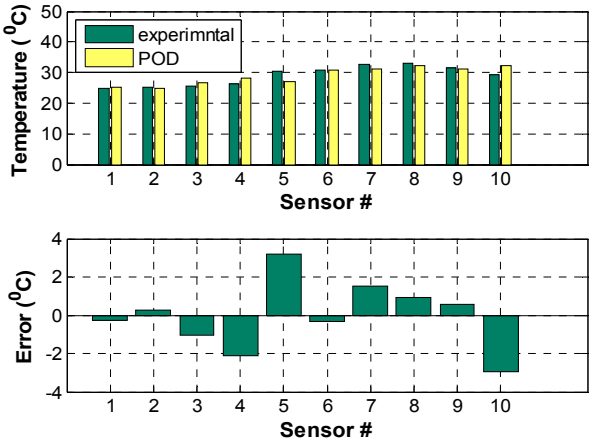

Figure 9: POD-based temperature prediction at $t = 1500$ s for transient-1. POD-based predictions are compared with corresponding experimental data.

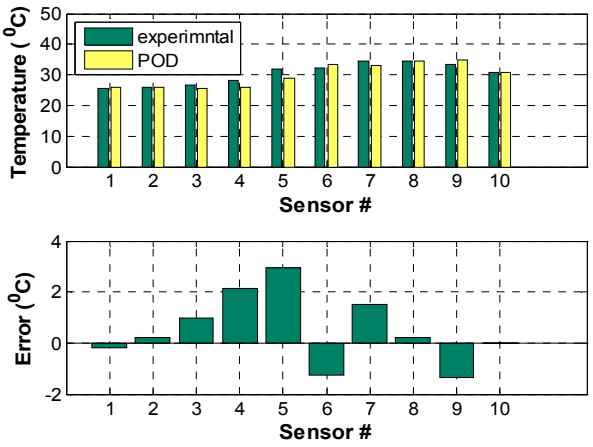

Figure 10: POD-based temperature prediction at $t = 1500$ s for transient-2. POD-based predictions are compared with corresponding experimental data.

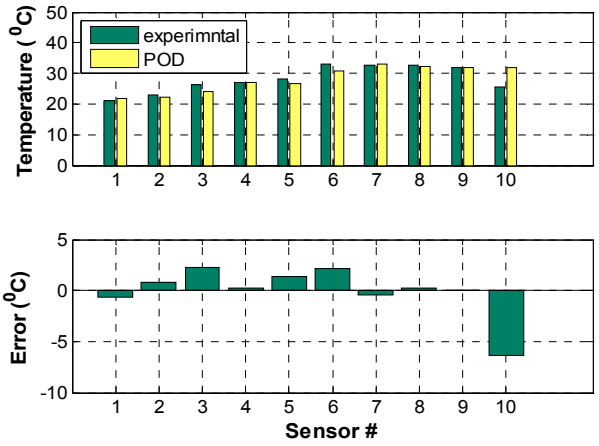

Figure 11: POD-based temperature prediction at $t = 1500$ s for transient-3. POD-based predictions are compared with corresponding experimental data.

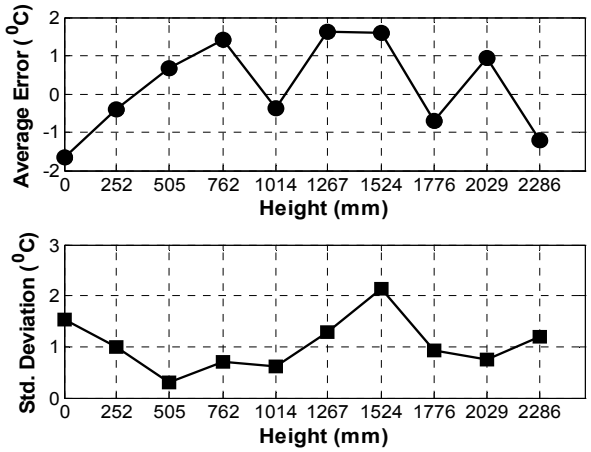

Figure 12: Average and standard deviation of the error for different sensors for transient-1 for the range: $t = 0 - 1500$ s.

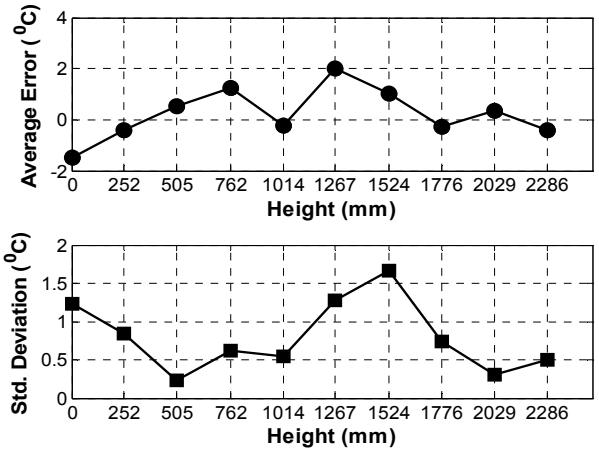

Figure 13: Average and standard deviation of the error for different sensors for transient-2 for the range: $t = 0 - 1500$ s.

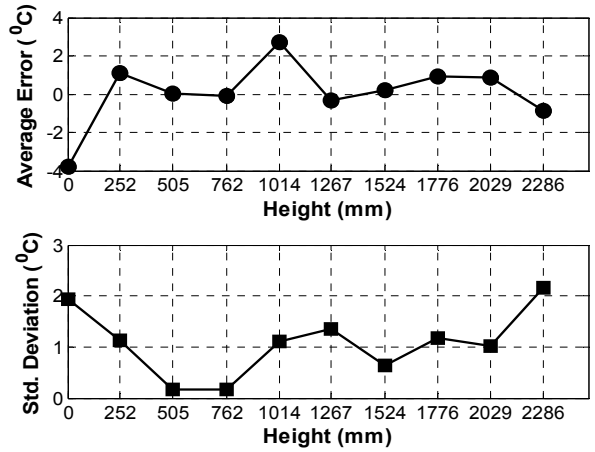

Figure 14: Average and standard deviation of the error for different sensors for transient-3 for the range: $t = 0 - 1500$ s.

5. Conclusions

This paper develops a POD-based framework for improving the spatial resolution of measured temperature data. A case study including three transients has been modeled. The accuracy of the model is estimated. It is shown that the RMS value of the average error at a given time instant is in the order of 5%. Taken over a finite time interval, this RMS value is in the order of 1%. A nominal modeling error establishes the fidelity of the POD-based modeling framework in improving the spatial resolution of transient air temperature data. Such a high-fidelity framework may be useful for developing high-resolution monitoring system for data centers.

Acknowledgments

The authors acknowledge support for this work from IBM Corporation, with Dr. Hendrik Hamann as the Technical Monitor. Acknowledgements are also due to the United States Department of Energy as the source of primary funds. The contributions from Stephan Barabasi and John Nelson, both from IBM, in setting up the data acquisition system are highly appreciated.

References

[1] J. Koomey, "Growth in data center electricity use 2005 to 2010," *Oakland, CA: Analytics Press. August,* vol. 1, p. 2010, 2011.

[2] S. Greenberg, E. Mills, B. Tschudi, P. Rumsey, and B. Myatt, "Best practices for data centers: Lessons learned from benchmarking 22 data centers," *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. ACEEE, August,* vol. 3, pp. 76-87, 2006.

[3] S. Campbell., *Data Center Power Consumption to Grow 20 Percent in 2012.* Available: http://it.tmcnet.com/channels/data-center-power/articles/239967-data-center-power-consumption-projected-grow-20-percent.htm, November 2012.

[4] J. G. Koomey, C. Belady, M. Patterson, A. Santos, and K. D. Lange, "Assessing trends over time in performance, costs, and energy use for servers," *Lawrence Berkeley National Laboratory, Stanford University, Microsoft Corporation, and Intel Corporation, Tech. Rep,* 2009.

[5] K. Khankari, "Thermal mass availability for cooling data centers during power shutdown," *ASHRAE Transactions,* vol. 116, 2010.

[6] D. Garday and J. Housley, "Thermal storage system provides emergency data center cooling," *Intel White Paper,* 2007.

[7] M. K. Patterson and D. Fenwick, "The state of datacenter cooling," Intel Corporation White Paper. Available at http://download.intel. com/technology/eep/data-center-efficiency/stateof-date-center-cooling. pdf, 2008.

[8] W. A. Abdelmaksoud, H. Ezzat Khalifa, T. Q. Dang, R. R. Schmidt, and M. Iyengar, "Improved CFD modeling of a small data center test cell," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on*, 2010, pp. 1-9, Las Vegas, NV.

[9] H. F. Hamann, J. A. Lacey, M. O'Boyle, R. R. Schmidt, and M. Iyengar, "Rapid Three-Dimensional Thermal Characterization of Large-Scale Computing Facilities," *Components and Packaging Technologies, IEEE Transactions on,* vol. 31, pp. 444-448, 2008.

[10] R. Zhou, Z. Wang, C. E. Bash, and A. McReynolds, "Data center cooling management and analysis-a model based approach," in *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2012 28th Annual IEEE*, 2012, pp. 98-103, Palo Alto, CA.

[11] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogrammetric Engineering & Remote Sensing,* vol. 66, pp. 49-61, 2000.

[12] E. Samadiani and Y. Joshi, "Reduced order thermal modeling of data centers via proper orthogonal decomposition: a review," *International Journal of Numerical Methods for Heat & Fluid Flow,* vol. 20, pp. 529-550, 2010.

[13] R. Ghosh and Y. Joshi, "Error estimation in POD-based dynamic reduced-order thermal modeling of data centers," *International Journal of Heat and Mass Transfer,* vol. 57, pp. 698-707, 2013.

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

[15] P. Druault, P. Guibert, and F. Alizon, "Use of proper orthogonal decomposition for time interpolation from PIV data," *Experiments in Fluids,* vol. 39, pp. 1009-1023, 2005.