Proper Orthogonal Decomposition-Based Modeling Framework for Improving Spatial Resolution of Measured Temperature Data

Rajat Ghosh and Yogendra Joshi

***Abstract*— This paper presents a proper orthogonal decomposition (POD)-based reduced-order modeling framework to improve spatial resolution of measured temperature data in an air-cooled data center. This data-driven approach is applied on transient air temperature data, acquired at the exhaust of a server simulator rack. Temperature data is collected by a distributed thermocouple network at 1 Hz sampling frequency following a step impulse in the rack heat load. The input data are organized in a 2-D array, comprising transient temperature signals measured at various spatial locations. Because its computational time scales logarithmically with the input size, the proposed POD-based approach is potentially useful as an efficient tool for handling large transient data sets. With spatial location being the parameter for the input data matrix, the proposed approach is suitable for rapid synthesis of transient temperature data at new spatial locations. The comparison between POD-based local air temperature predictions and corresponding data indicates a maximum prediction uncertainty of 3.2%, and root mean square prediction uncertainty of 1.9%.**

***Index Terms*— Computational sustainability, data center (DC), energy efficiency, live temperature streaming, proper orthogonal decomposition (POD), regression analysis, sensor placement.**

I. Introduction

SINCE data centers (DCs) consume as much as 2% of world electricity usage, improving DC energy efficiency is essential for the sustainable growth of the information technology (IT) industry [1]–[3]. The cost of DC cooling amounts to 35%–50% of a DC's total energy budget. Since cooling demand is strongly correlated to local air temperatures [4], a near-real-time characterization of the air temperature field is important for the energy auditing in a DC. Fig. 1 shows a typical alternating cold/hot aisles-based airflow scheme [5], in which the IT equipment inlets face cold aisles, and their outlets face hot aisles. Computer room air conditioning (CRAC) units drive pressurized cooling air into a raised-floor plenum. Unlike hot aisles, cold aisles have perforated tiles that allow cooling air to come up and get entrained into the servers via server fans. The hot exhaust air returns to CRAC units driven by the negative pressure gradient created by the CRAC blowers. Intermixing between cold and hot air and lack of cooling airflow can create localized hotspots. Unnecessarily conservative CRAC set-points for mitigating these hotspots and inappropriate server fan operations often lead to coolant bypass, in which cooling air directly returns back to the CRAC unit. The problem of cooling airflow management is further compounded by the introduction of hypervisor-based virtualization technologies [6] that facilitate dynamic server load migration.

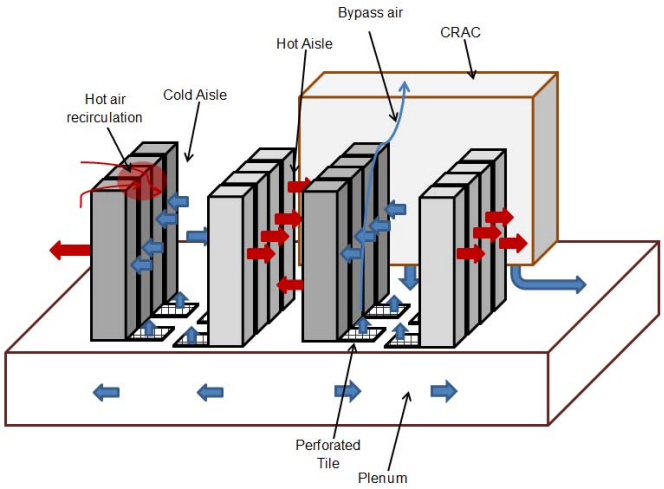

Fig. 1. Typical alternating cold/hot aisle-based forced convective cooling strategy in a raised-floor DC. Driven by a CRAC unit, cooling air comes into the room through perforated floor tiles and is then guided into IT racks by server fans. The potential problems for this cooling strategy are hot air recirculation in cold aisles and bypass air.

To avoid resource over- or under-provisioning, a real-time demand-aware cooling control system based on online temperature monitoring is required. A measurement-based monitoring framework needs to be supported by some modeling technology because temperature gradients in DCs can be quite large. For example, one might find air temperature at the corner of a server inlet differs by few degrees compared with the center of the server inlet. Therefore, it is imperative to measure temperature data by sensors deployed at multiple strategic locations and compute reliable temperature distributions to gain meaningful insight from the real-time measurements. The design of a measurement system involves resolution of tradeoffs between density of sensors, location of sensors, and their measurement frequency. The purpose of this paper is to develop a proper orthogonal decomposition (POD)-based algorithm that can improve the spatial resolution

Manuscript received May 26, 2013; revised October 1, 2013 and November 13, 2013; accepted November 14, 2013. Date of publication April 3, 2014; date of current version May 1, 2014. Recommended for publication by Associate Editor M. K. Iyengar upon evaluation of reviewers' comments.

The authors are with the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405 USA (e-mail: rajat.ghosh@gatech.edu; yogendra.joshi@me.gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCPMT.2013.2291791

of measured temperature data and ultimately reduce sensor requisition for DC temperature monitoring.

A widely used measurement technique for air temperatures near computing servers is to deploy a grid-based thermocouple (TC) network [7] and to mobilize it with a wheeled telescoping mechanism [8]. A key challenge in the design of such DC temperature measurement systems stems from the needed spatial resolution. The resolution of the turbulent forced-convective temperature field at the Kolmogorov length scale (where energy dissipation occurs) scales as $\mathrm{Re}^{-3/4}$. Therefore, for a rack of 2-m height, located in an air-cool DC with air velocity in order of 1 m/s, and the measurement resolution needed is 0.1 mm. On the other hand, the typical pore size of the perforated tiles is in the order of 10 mm. Therefore, an exhaustive resolution of DC temperature field demands measurement length scale in the order of few mm. Such a fine-grained temperature characterization is cost-prohibitive and unwarranted for thermal control. To reduce sensor density, data compression techniques can be utilized. One such technique is principal component analysis/POD, which transforms a data matrix into a product of a low-rank matrix (POD modes) and a coefficient matrix (POD coefficients) [9]. POD-based data compression algorithms are widely used in video surveillance [10], face recognition [11], and bioinformatics [12]. For characterizing turbulent flow, a POD model was introduced in [13] and extended in [14]. POD has been used as a parametric optimization tool for the DC infrastructure design problem. Typical parameters include rack heat load [15], CRAC flowrate [16], and time [17].

In this paper, a measurement-based parametric framework is developed to improve the spatial resolution of measured temperature data. Section II formalizes the optimization algorithm. The effectiveness of the proposed algorithm is verified in two measurement planes, located in a hot aisle. Section III describes the experimental setup and the measurement system. Section IV identifies a pertinent case study. Section V includes results and discussion. Finally, Section VI draws the conclusion.

II. Methodology

In this paper, a measurement-based reduced-order model of transient air temperature is developed with time as the independent variable and spatial location as the parametric variable. The measurement-based reduced-order air temperature model is developed via a POD-based statistical algorithm. Fig. 2 shows the proposed POD-based algorithm.

The intrinsic idea of a POD-model lies in the determination of an optimal basis space which maximizes the projection of input data. The general-purpose POD model is widely reported in [14] and [18]. In studies related to DC cooling, a detailed mathematical analysis of a domain-specific POD model is documented in [17]. As shown in the flowchart in Fig. 2, the data-driven algorithm consists of five major mathematical steps.

1) *Compilation of Data Matrix:* The data matrix compiles the temperature data. For the data-driven algorithm, the data matrix forms the corresponding problem instance: each column of a data matrix includes a transient temperature signal collected at a particular sensor location. The sampling interval of the temperature signal is Δt over a domain $[0-t]$. The infinite domain problem of live streaming can potentially be reduced to a finite dimensional problem by estimating the signal settling time, t, when the temperature signal reaches its steady state. Several sensors are deployed in the measurement domain to capture temperature data matrix, $T_{m\times n}(t; x, y, z)$. The row rank, m, of the data matrix is equal to the length of the transient temperature signal while the column rank, n, of the data matrix is the number of sensors deployed.

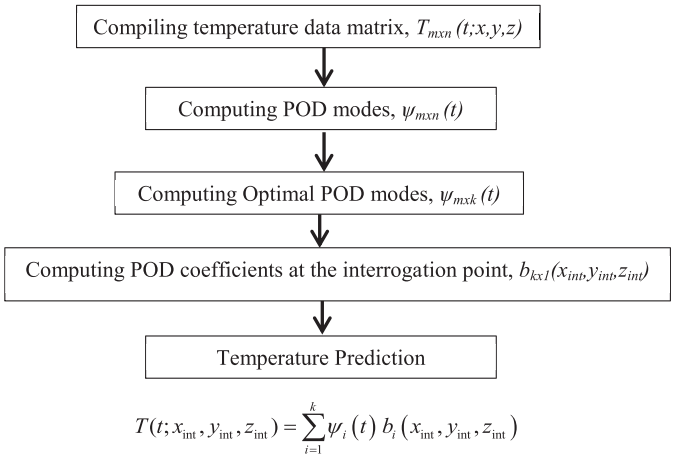

Fig. 2. POD-based reduced-order modeling algorithm with spatial location as the parameters. For a given time interval, the algorithm is applied on an ensemble of transient temperatures, $T(t; x_i^{\mathrm{en}}, y_i^{\mathrm{en}}, z_i^{\mathrm{en}})$. The temperature predictions are computed for the interrogation point, $(x^{\mathrm{int}}, y^{\mathrm{int}}, z^{\mathrm{int}})$.

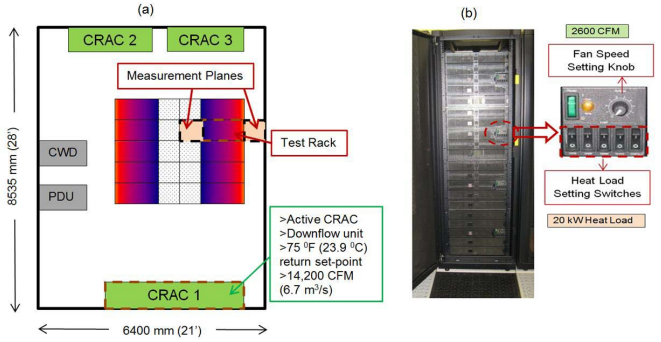

Fig. 3. Details of the experimental facility. (a) Plan view of the 672 ft^2 experimental facility, which has 10 racks arranged in a 5×2 architecture with alternating cold/hot aisles. (b) Photograph of the test rack, which is a server simulator rack, showing fan speed setting dial, and heat load control switches.

2) *Computation of POD Modes:* Sirovich [14] proposed the method of snapshots to compute POD modes (ψ), which are modeled as low-dimensional attractor spaces of the data matrix. The method involves following numerical procedure.
   a) *Computation of the correlation matrix:* The correlation matrix of a matrix is defined as

$$K_{n\times n} = \left(\frac{1}{n}\right) T^{*}_{n\times m} \otimes T_{m\times n} \tag{1}$$

   where superscript $*$ indicates the matrix transpose.

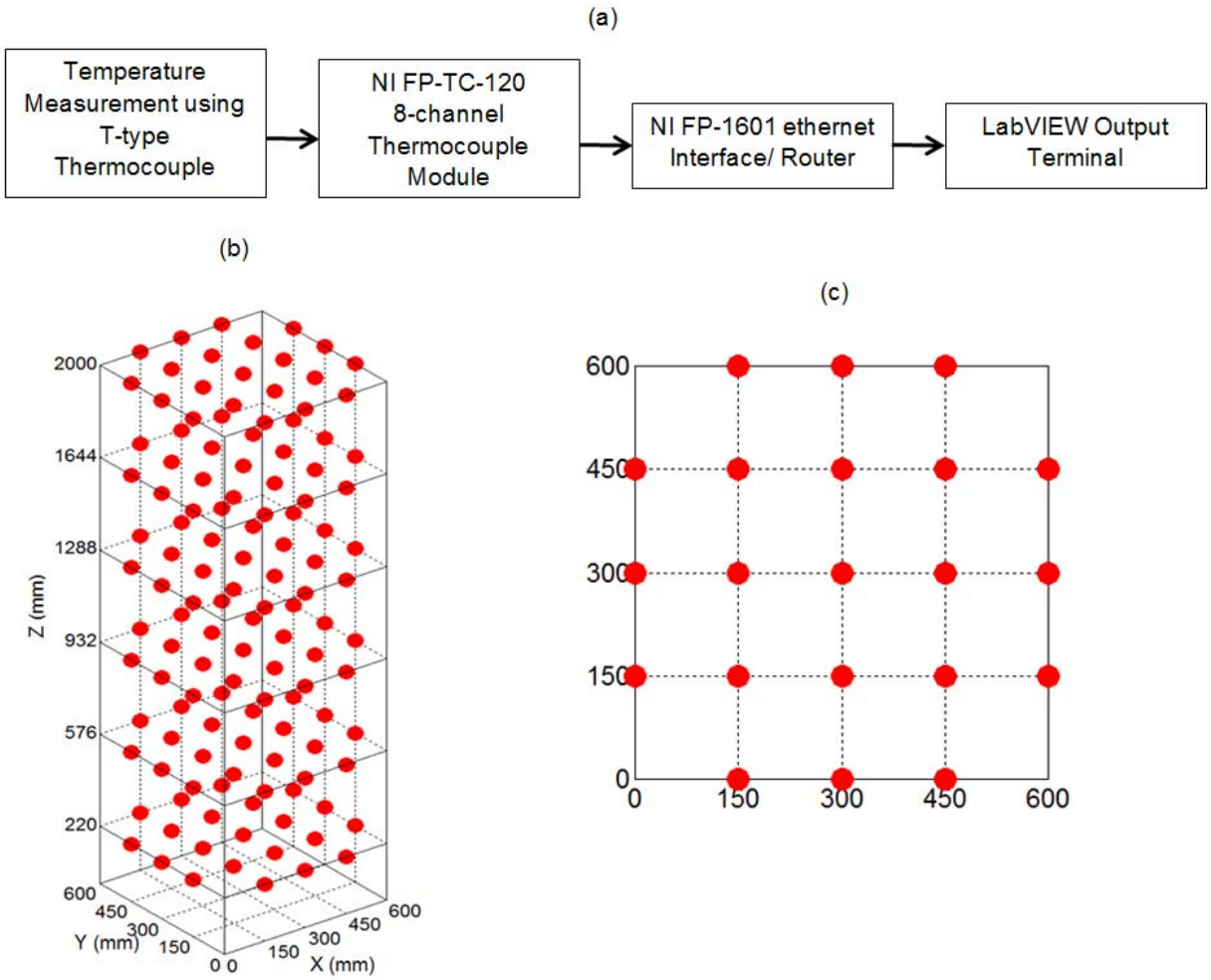

Fig. 4. Details of the temperature data acquisition system. (a) Measurement chain consists of generating TC-based temperature measurement data, processing at the TC module, processing at the network module, and transmitting processed data via a network router to the LabVIEW-based output terminal. (b) 3-D sensor arrangement at the rack exhaust. (c) Plan view of sensor network, located at 2000 mm. Each measurement plane consists of 21 T-type TC sensors.

b) *Determination of eigenspace of K to compute POD modes:* The eigenspace of K is determined numerically by solving the eigenvalue problem (2) by the power iteration [19]. Then, POD modes are computed by taking dyadic product of T with V. V is the eigenvector for K, and λ is the corresponding eigenvalue. POD-modes describe the coherent structures [20] of the transient temperature data; hence, they depend on time

$$KV = \lambda V \tag{2}$$

$$\psi = T \otimes V. \tag{3}$$

3) *Formation of Optimal Basis Space from POD Modes:* The derived POD modes are optimal in the sense that N POD-modes capture more information/energy than any corresponding orthonormal modes derived from other comparable numerical procedures, such as fast Fourier transform [20]. The energy content of a POD mode is given by the corresponding eigenvalue. The degree of variability (energy) captured by an eigenvalue (λ_i) is given by a normalization factor

$$F_i = \frac{\lambda_i}{\sum_i \lambda_i}. \tag{4}$$

TABLE I
STANDARD DEVIATION OF HOT AISLE AIR TEMPERATURES AT DIFFERENT HEIGHTS

h	Standard Deviation
mm	^{0}C
2000	3.1
1644	3.1
1288	1.6
932	1.7
576	0.9
220	1.8

The optimal model order, k ($k \ll m; k < n$) can be derived by a preassigned tolerance criteria

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} < \text{tol.} \tag{5}$$

The ratio (k/n) indicates the degree of data compression provided by the POD model.

4) *Computation of POD Coefficients:* POD-coefficients capture location-dependent parametric components of the data matrix. The numerical procedure to compute POD-coefficients involves computation of the coefficient matrix. The coefficient matrix is computed by taking the dyadic product of the pseudoinverse of the POD modes and the data matrix

$$B = \psi^{+} \otimes T. \tag{6}$$

ψ^{+} is the pseudoinverse of POD-mode matrix, ψ. B is an n-by-n matrix. The rows of B indicate the weighing factors for the corresponding POD modes. The columns of B are the characteristics of the spatial locations of the sensors. It is assumed that the coefficient vector at a new spatial location, termed the POD-coefficient (b), lies in the column space of B. The literature reports various methods for the mapping: $B_{n\times n} \rightarrow b_{n\times 1}$. The most widely used method is Galerkin projection [18]. The Galerkin projection-based method is further simplified using the flux matching approach [21]. Nevertheless, the Galerkin projection is suitable for a simulation-generated highly resolved data matrix. On the other hand, the scarcity of experimental data dictates the application of statistical methods, such as spline-based interpolation [22] and kriging [23]. Given that this paper deals with experimentally acquired temperature data in a measurement domain with hot spots, a conditional procedure is proposed. At first, POD coefficients are generated using temperature sensors located at the boundary and the geometric center of the interrogation domain. If the resulting POD prediction uncertainty is more than a preassigned tolerance criterion (arbitrarily assumed to be equal to 5% in this paper), a new sensor arrangement is needed. The choice of new sensor locations is guided by a physics-based reduced-order model of forced convective local airflow field. The mathematical model for POD coefficients is

$$b_{k\times 1} = B_{k\times n} \otimes C_{n\times 1}. \tag{7}$$

Each element c_i ($\in C_{n\times 1}$) indicates the weighting factor for the POD coefficient. The determination of $C_{n\times 1}$ is based on identifying isothermal zones in the interrogation domain and then ascribing the influence of the neighboring sensors on an interrogation location. Therefore, the determination of $C_{n\times 1}$ is domain-dependent and is discussed in details in the results and discussion section. After $b_{n\times 1}$ is determined, an optimal POD coefficient, $b_{k\times 1}$ is extracted (5).

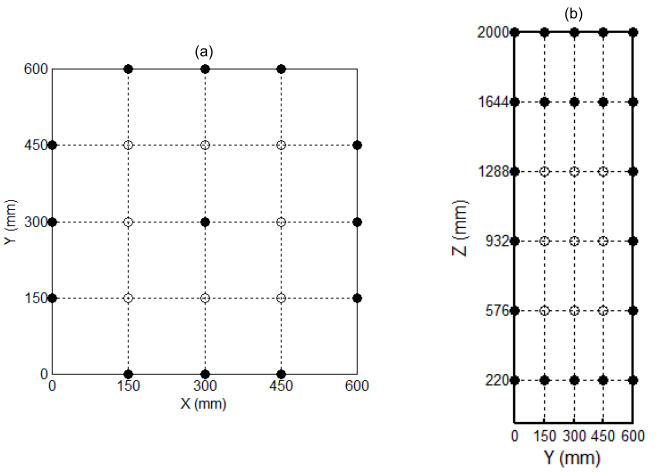

Fig. 5. Two planes are identified for validation purposes. (a) $z = 2000$ mm, a horizontal plane in the hot aisle located at near the top of the test rack. (b) $x = 150$ mm, a vertical plane in the hot aisle located parallel to the exhaust of the test rack. The data matrix is comprised of temperature data acquired by the sensors located at the positions marked by filled black circles. The open black circles represent locations where model predictions are validated with actual sensor data.

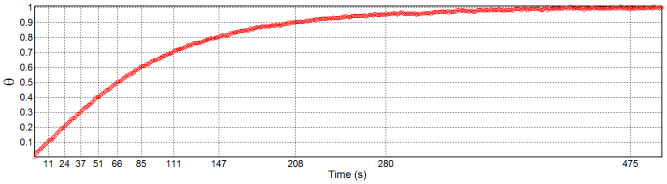

Fig. 6. Normalized air temperature acquired at (150, 150, 2000) mm in the hot aisle following sudden introduction of 20-kW test rack heat load. Different time instants are identified when the response reaches 10% 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, and 99% of steady state (s.s.).

5) *Temperature Prediction:* The temperature prediction at a new spatial location $(x_{\text{int}}, y_{\text{int}}, z_{\text{int}})$ is given by

$$T_{m\times 1}(t; x_{\text{int}}, y_{\text{int}}, z_{\text{int}}) = \psi_{m\times k}(t) \otimes b_{k\times 1}(x_{\text{int}}, y_{\text{int}}, z_{\text{int}}). \tag{8}$$

For a fidelity check, temperature data, acquired independently at the interrogation points, are compared with predictions. In this context, the prediction uncertainty is defined as the uncertainty in predicted local air temperatures. The comparison is quantified by rank correlation coefficients (ρ) and relative root mean square errors (rmse), as defined by the following equations:

$$\rho = \frac{\sum_i \left(T_i^{\text{data}} - \overline{T^{\text{data}}}\right)\left(T_i^{\text{prediction}} - \overline{T^{\text{prediction}}}\right)}{\sqrt{\sum_i \left(T_i^{\text{data}} - \overline{T^{\text{data}}}\right)^2 \sum_i \left(T_i^{\text{prediction}} - \overline{T^{\text{prediction}}}\right)^2}} \tag{9}$$

$$\text{rmse} = \sqrt{\frac{\sum_i \left(T_i^{\text{data}} - T_i^{\text{prediction}} \Big/ T_i^{\text{data}}\right)^2}{n}}. \tag{10}$$

On the basis of the mathematical procedure outlined in (1)–(10), a functional algorithm can be developed on efficient utilization of available temperature sensors. In the interest of simplicity, the algorithm is

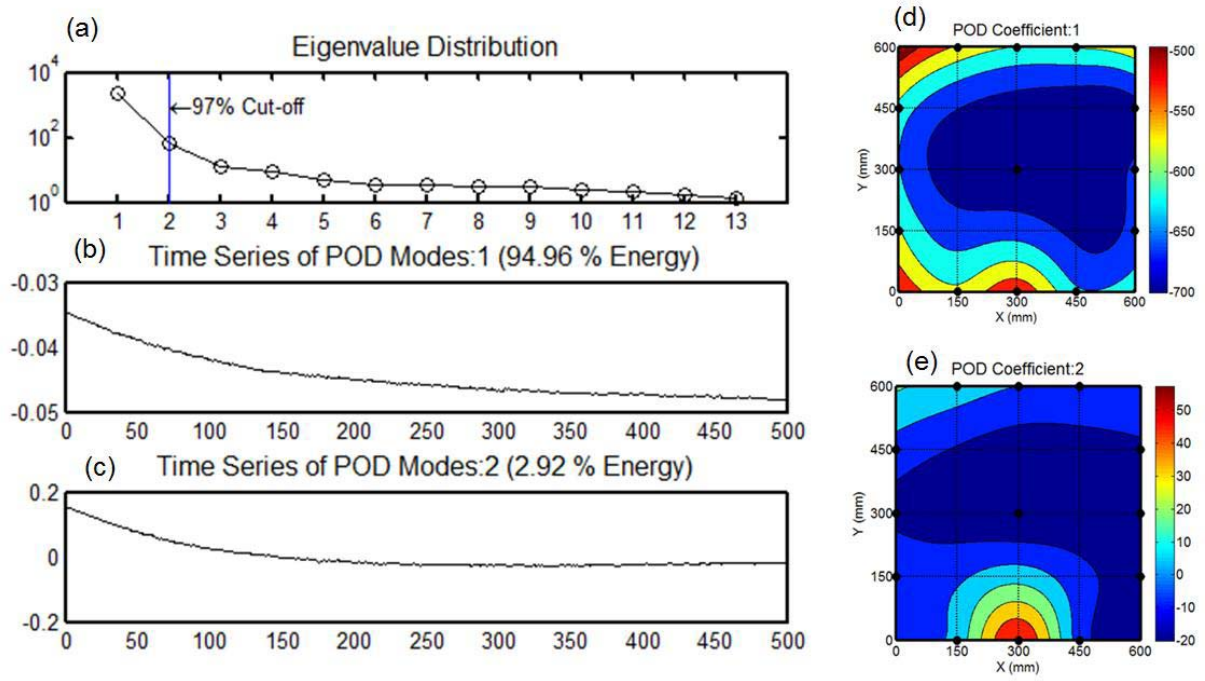

Fig. 7. POD-based model order reduction for $z = 2000$ mm. (a) Degree of data compression provided by POD. Two out of 19 POD-modes capture the coherent structure (97%) of the data sequence. (b) Time series for first POD mode, which captures 94.96% of energy. (c) Time series for second POD mode with 2.92% of energy. (d) POD coefficient for first POD mode. (e) POD coefficient for second POD mode.

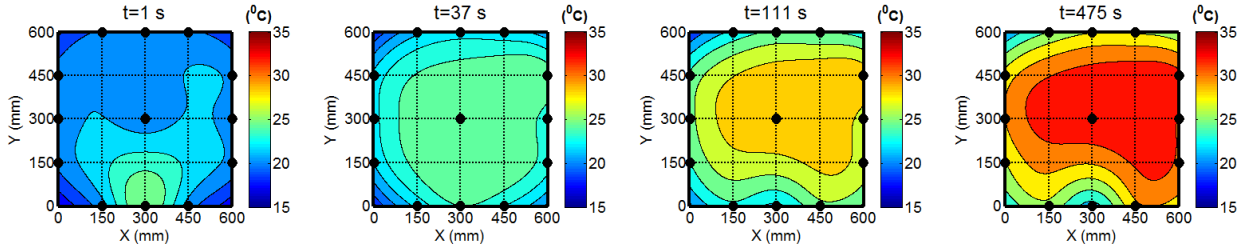

Fig. 8. Transient (initial, 30%, 70%, and 99% s.s.) air temperature contours at $z = 2000$ mm. The temperature contours are generated via the Delaunay triangulation technique using the temperature data acquired by the sensors located at the points marked by black filled points. The contours identify the influence of a sensor on various spatial locations.

applied on planar temperature data. This functional algorithm has two stages: the first one involves a geometry-based approach and the second one involves a physics-based approach. The physics-based approach is invoked only when the geometry-based approach fails to satisfy a preassigned tolerance criterion (5% relative deviation).

The steps for the geometry-based algorithm are the following.

1) *Data Acquisition:* It is assumed that the number of available sensors is equal to N. At first, one sensor is deployed at the geometric center. Remaining sensors are distributed equally on four edges of the interrogation plane. Therefore, each edge has $((N-1)/4)$ sensors. Of these $((N-1)/4)$ sensors, one sensor is placed at the center of the edge. Remaining sensors are placed symmetrically with respect to the edge center. Given that a corner point is shared by the two edges, no sensor is placed there. The measured temperature distribution is computed via statistical interpolation technique such as Delaunay triangulation [24].
2) *POD-Mode Computation:* Optimal POD basis space is computed using (1)–(5) with spatial location as the parameter. POD coefficients for the sensor locations are computed by (6).
3) *POD-Coefficient Computation for an Interrogation Location:* The relative location of an interrogation point is determined with respect to the sensor points in the measured temperature distribution. The POD coefficient for an interrogation point is determined by taking average of POD-coefficients corresponding to the sensor points lying in the same isothermal zone.
4) *POD-Based Temperature Computation:* The interrogation temperature is computed by (8). The temperature predictions are compared with the corresponding exper-

imentally measured data. If the percentage deviation [defined by (10)] is more than a preassigned tolerance criterion, the framework is considered to be unreliable.

In case the geometry-based algorithm fails to satisfy the tolerance criterion, it is recommended to follow a physics-based algorithm. The steps for the physics-based algorithm are the following.

1) *Data Acquisition:* At first, the forced-convective flow field is estimated either by an approximation model or by a coarse-grain computa-tional fluid dynamics (CFD) model. Depending on the directions of temperature gradients, the temperature field is segmented into different zones. In each temperature segment, the direction of the steepest temperature gradient is estimated. Based on that estimation, the temperature sensors are optimally distributed.

Remaining steps for the physics-based algorithm are identical to the steps 2–4 in the geometry-based algorithm.

III. Measurements

Fig. 3(a) shows the 672 ft^2 ($\sim$62.5 m^2) raised-floor experimental facility [located at Atlanta, GA; elevation 1027′ (313 m)] with three CRAC units. The height of the DC room is 9 ft ($\sim$2.75 m) with an under-floor plenum of height 3 ft ($\sim$0.9 m), and drop ceiling height of 5 ft ($\sim$1.52 m). The facility has 10 IT racks arranged in a 5×2 alternating cold/hot aisle architecture. The thermometry was conducted in the measurement planes, at locations indicated by the solid squares in Fig. 3(a). Fig. 3(b) shows a photograph of the test rack, which contains four vertically stacked 10-U (17.5 in $\sim$444.5 mm) server simulators. The heat load and the fan speed of a server simulator are controlled from the control unit [7] shown in Fig. 3(b). The heat load switches included are: 250, 500, 1000, 1000, and 2000 W. The fan airflow rate can be modulated to ten different levels via a dial knob. At the full capacity, a server simulator fan supplies 650 ft^3/min ($\sim$0.3068 m^3/s) airflow [7], simulating 2600 ft^3/min rack-flow.

The thermometry, as shown in Fig. 4(a), uses copper-constantan (T-type) TCs, made from 28 gauge (AWG) TC wire (0.32-mm diameter). A lumped capacitance analysis [25] in an air-driven convective environment ($h \sim 10$ W/m^2·K) indicates a response time of the order of 1 s. As shown in Fig. 4(b), each measurement domain has six grid-based TC units, each comprising 21 T-type exposed junction TCs. Six TC units are located at six different heights from the floor: 220, 576, 932, 1288, 1644, and 2000 mm. The rack exhaust is located at $x = 0$ mm.

As shown in Fig. 4(c), 21 TCs are arranged symmetrically in a 2 ft $\times$ 2 ft (610 mm $\times$ 610 mm) square plane. Due to the symmetric arrangement, the distance between two neighboring TCs is equal to 150 mm. Overall, there are 126 TCs in each measurement domain. The sources of measurement uncertainty [26] are lack of data integrity, gain and offset, differential and integral nonlinearity, quantization, noise, cold junction compensation, networking, acoustic noise, and vibration. The measurement system is calibrated using an Omega CL122 TC calibrator (http://www.omega.com/pptst/CL120_134.html) and a NIST traceable calibrated thermometer. The measurement chain calibration is conducted in the 10 °C–35 °C temperature range. With 95% confidence interval, the average calibration error is estimated as (0.49 °C $\pm$ 0.19 °C).

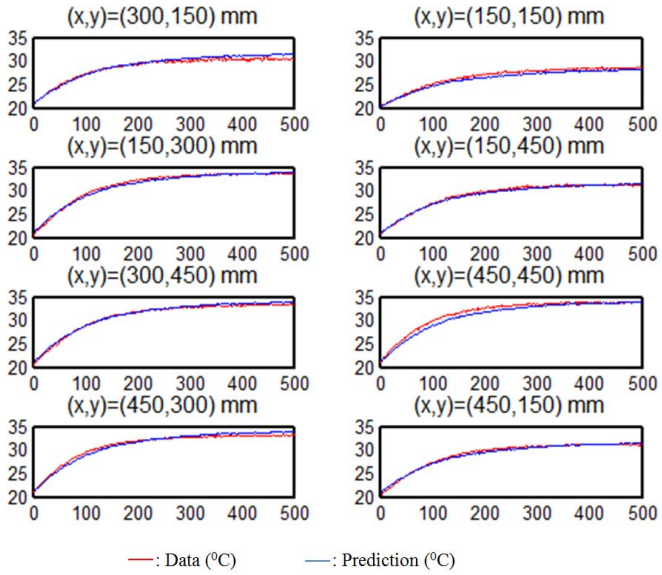

Fig. 9. Data versus POD-based predictions at different interrogation locations in $z = 2000$ mm plane.

TABLE II
Error Table Quantifying the Deviations Between Transient Data and Predictions in $z = 2000$ mm Plane

(x, y)	Correlation Coefficient	Relative Deviation
mm		(%)
(300, 150)	0.9928	1.8
(150, 150)	0.9975	1.9
(150, 300)	0.9970	1.0
(150, 450)	0.9970	0.8
(300, 450)	0.9974	1.0
(450, 450)	0.9941	2.2
(300, 150)	0.9935	1.6
(150, 150)	0.9950	1.1

IV. Case Study

In the present study, the test rack is suddenly switched to 20-kW power. This heat load is actuated by setting each server simulator to 5000 W by turning on 1000, 2000, and 2000 W heat load switches. Each server simulator fan is kept at a setting which corresponds to 650 ft^3/min ($\sim$0.3068 m^3/s) airflow [7]. The remaining racks in the facility were switched off during the experiment. The air temperature response is measured by 126 TCs deployed in the cold and hot aisles.

It is imperative that a sensor fusion algorithm be validated in the measurement planes that offer sufficiently large temperature gradients. Therefore, the remaining study focuses on the hot aisle. Table I shows the standard deviation of measured temperature data at six different heights in the hot aisle. In Table I, the plane at height 2000 mm has the highest

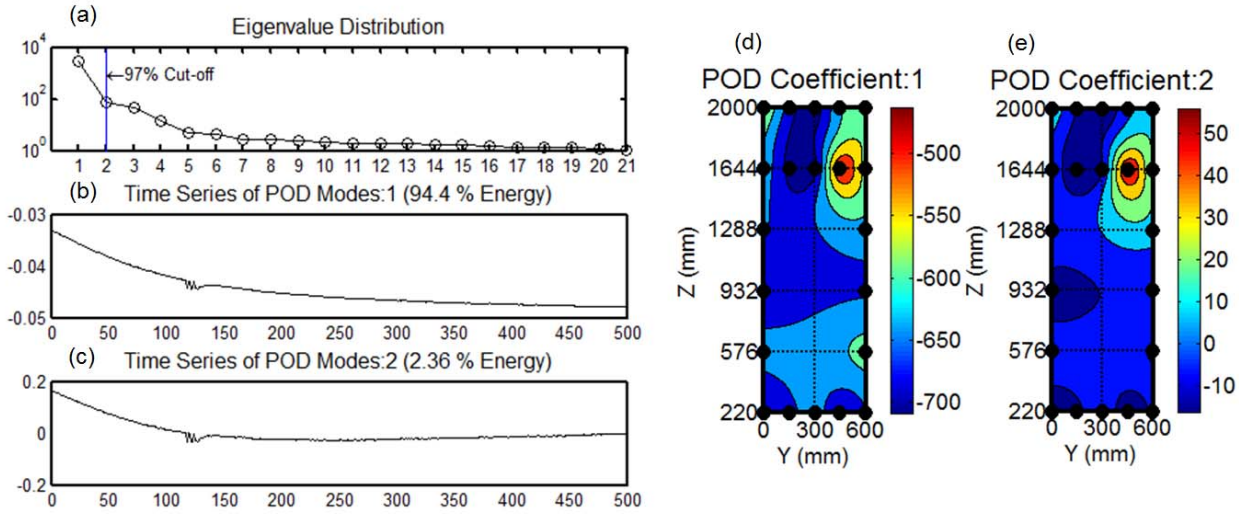

Fig. 10. POD-based model order reduction for $x = 150$ mm plane. (a) Degree of data compression provided by POD. Two out of 19 POD-modes capture the coherent structure (>97%) of the data sequence. (b) Time series for first POD, mode which captures 94.96% of energy. (c) Time series for second POD mode with 2.92% of energy. (d) POD coefficient for first POD mode. (e) POD coefficient for second POD mode.

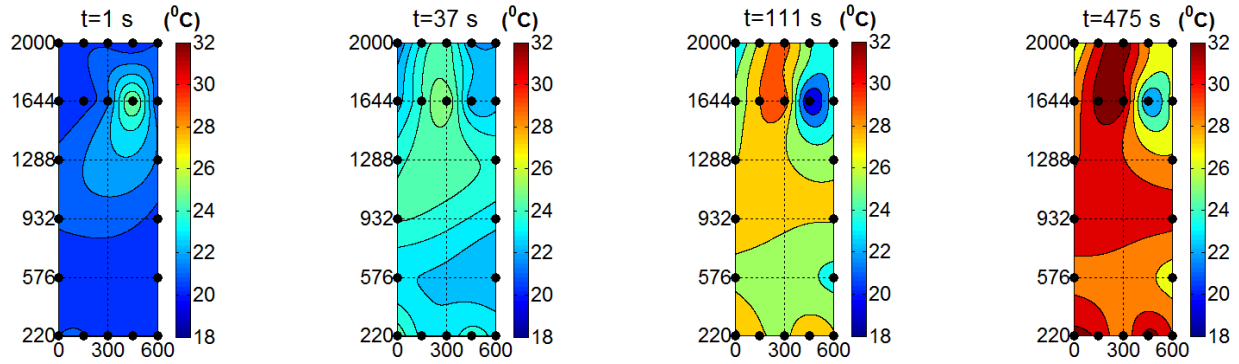

Fig. 11. Transient (initial, 30%, 70%, and 99% s.s.) evolution of air temperature in-plane parallel to the rack exhaust at $x = 150$ mm. The temperature contours are generated via Delaunay triangulation technique from the temperature data acquired by the sensors located at points marked by black filled points.

standard deviation of 3.1 °C. For prognostic-based thermal reliability modeling, this high-temperature plane is critically important because of its proximity to the fire suppression system. In addition, another measurement plane, located at 150-mm distance from the exhaust of the test rack, is chosen. The standard deviation of temperature data in this plane is equal to 3 °C. Due to its proximity to the server outlets, the temperature in this plane is very sensitive to the server IT workload variation. Therefore, rapid temperature assessment of this plane facilitates optimal dynamic cooling resource provisioning in DCs. Two temperature planes are chosen at $z = 2000$ mm (which includes 21 TCs) and at $x = 150$ mm (includes 30 TCs). After identifying these two measurement planes, two independent POD-based analyses are conducted, and their respective modeling fidelities are estimated.

V. Results and Discussion

This section demonstrates the effectiveness of the proposed framework at $z = 2000$ mm and $x = 150$ mm. Fig. 5 shows the proposed sensor arrangement strategies for $z = 2000$ mm [Fig. 5(a)] and $x = 150$ mm [Fig. 5(b)] planes. The filled circles are the locations of TCs which measure transient air temperatures to compile the data matrix for POD-based data analyses. For the fidelity verification of the framework, the POD-based framework generates predictions for the locations marked by the open circles. These predictions are subsequently compared with the corresponding experimental data, and the fidelity of the framework is assessed.

Before fidelity verification of the proposed framework in different measurement planes in the hot aisle, a representative transient temperature characteristic is analyzed by noting the transient temperature evolution at an arbitrarily chosen point in the hot aisle located at (150, 150, 2000) mm. Fig. 6 shows the transient characteristic curve. The normalized transient temperature is defined by θ:

$$\theta(t) = \frac{T(t) - T(t = 0\ \text{s})}{T(t = 500\ \text{s}) - T(t = 0\ \text{s})}. \tag{11}$$

The various time instants at which θ reaches the numerical value of 0.1 (11 s), 0.2 (24 s), 0.3 (37 s), …, 0.9 (208 s),

0.95 (280 s), 0.99 (475 s) are noted. These time instants will be used as the signposts in the ensuing discrete transient analysis.

A. Validation for Measurement Plane at $z = 2000$ mm

As shown in Fig. 5(a), there are 13 TCs in the measurement plane located at $z = 2000$ mm. Therefore, the data acquired by these 13 TCs constitute a data matrix of size 501×13. The row rank, 501 corresponds to a transient domain [0–500] s at the sampling interval of 1s. Subsequent POD-based analysis yields the eigenvalue spectrum [Fig. 7(a)], two optimal POD modes [Fig. 7(b) and (c)], and corresponding POD coefficients [Fig. 7(d) and (e)]. The optimality is characterized by the fact that first two POD-modes capture 97% of the information/energy of the temperature data. With the first POD-mode capturing 94.96% of the energy, this offers 84.6% data compression.

As proposed by (7), the POD coefficient at the interrogation location is mapped by the linear combination of two-optimal POD coefficients. Hence, the parametric variation of temperature prediction is governed by the weighting scalars of two optimal POD coefficients. The convective transport processes at the horizontal plane, $z = 2000$ mm are characterized by the upward airflow. Therefore, it can be concluded that the temperature variation in the $z = 2000$ mm plane is governed by local effects that can be analyzed by the geometry-based model alone. Fig. 8 shows Delaunay triangulation-based interpolation [24] of the temperature data captured by 13 sensors (shown by black filled circles). The interpolation shows different isothermal zones. It is proposed that the weighting vector, $C_{n\times 1}$ of a spatial location is governed by its position in the interpolated temperature mapping. All spatial locations in that isothermal zone have equal numerical impacts from the included sensors. Therefore, the proposed mathematical model is

$$C_{n\times 1}^{\text{zone}} = \left(\frac{1}{m}\right)\delta_{\text{zone,sensor}}$$
$$\delta_{\text{zone,sensor}} = 1, \quad \text{if the sensor lies in the zone}$$
$$\delta_{\text{zone,sensor}} = 0, \quad \text{otherwise} \tag{12}$$

where m is the number of sensors lying in the given zone. For example, the prediction at (450, 450) mm at $t = 475$ s depends on sensors located at (300, 300) mm, (600, 300) mm, and (600, 450) mm. Therefore, in this case, m will be equal to 3. It can be observed that the temperature distribution varies with time: initially the temperature distribution is relatively uniform varying between 15 °C and 20 °C with a warm air pocket at (300 mm, 0 mm). Thereafter, air temperature increases gradually, and the profile becomes more nonuniform. The air temperature seems to be highest around the central point (300 mm, 300 mm). This is because it is easier for hot air to pass through the center of the aisle. Eventually, air temperatures reach a steady-state around $t = 475$ s. It is observed that air temperature is most nonuniform at the steady state. Therefore, the value of m is determined based on the steady-state temperature profile.

With the proposed model, temperature signals are computed at the interrogation points and subsequently compared with the experimental data. There are eight validation points in this measurement plane.

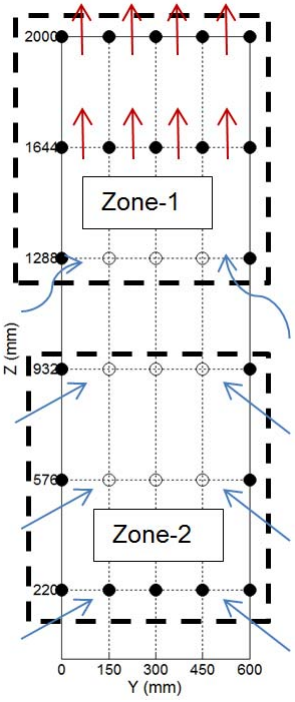

Fig. 12. Zonal abstraction of forced convective temperature field in the exhaust plane (parallel to the rack exhaust at $x = 150$ mm). Zone-1 is dominated by free shear flow directed upward to the ceiling. Zone-2 is dominated by flow entrainment from the two sides.

Fig. 9 compares the experimental temperature data at eight different interrogation locations to the corresponding POD-based predictions. The comparison suggests close similarity between data and predictions. The similarity is further quantified by the corresponding correlation coefficients and relative estimated errors shown in Table II.

Table II indicates correlation on the order of 99.5% and relative estimated error on the order 1% (maximum = 2.2%). The proposed POD model is capable of predicting temperature data with 99% relative accuracy, and reducing the sensor number from 21 to 13. That amounts to 38% sensor reduction.

B. Validation for Measurement Plane at $x = 150$ mm

As shown in Fig. 5(b), for the temperature plane located at $x = 150$ mm, data acquired by 21 TCs constitutes the data matrix of size 501×21. The POD-based analysis yields the eigenvalue spectrum [Fig. 10(a)], two optimal POD modes [Fig. 10(b) and (c)], and corresponding POD coefficients [Fig. 10(d) and (e)]. The optimality is characterized by the fact that first two POD-modes capture 97% energy of the temperature data. The first POD-mode captures 94.4% of the energy. This offers 90.4% data compression. Fig. 11 shows Delaunay triangulation-based interpolation of the temperature data captured by 21 sensors (shown by black filled circles). The interpolation creates different temperature zones. However, unlike the plane at $z = 2000$ mm, the geometry-based algorithm fails to satisfy the tolerance criterion. Alternatively, a physics-based algorithm is proposed in light of the fact

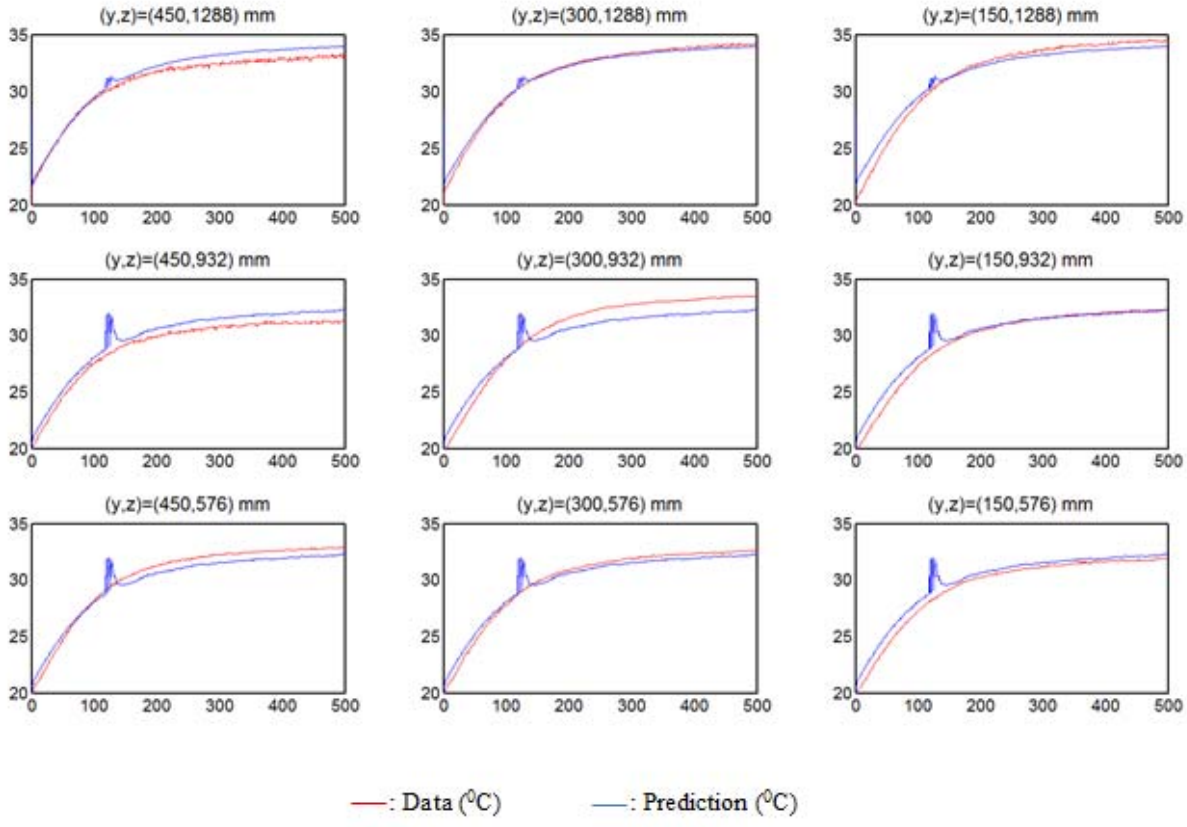

Fig. 13. Data versus POD-based predictions at different interrogation locations in $x = 150$ mm plane.

that $x = 150$ mm is a vertical plane parallel to the rack exhaust at $x = 0$. An estimation model is developed by identifying that it has two distinct convective environments: one near the top that is dominated by pressure gradient-driven upward airflow, and another near the bottom that is dominated by inertia-driven shear flow. Fig. 12 shows a schematic representation of such a flow pattern. Following this general notion, it is assumed that the predictions above 1288 mm are governed by the sensors near the top (Zone-1) and those below 932 mm are governed by the sensors at the two sides ($y = 0$ and 600 mm) of Zone-2.

This estimation of spatially dependent flow field directly influences the POD-coefficient computation which is modeled as

$$C_{n\times 1}^{\text{zone},i} = \left(\frac{1}{m_i}\right)\delta_{\text{zone,sensor}}, \quad i = 1, 2$$
$$\delta_{\text{zone,sensor}} = 1, \quad \text{if the sensor lies in the zone}$$
$$\delta_{\text{zone,sensor}} = 0, \quad \text{otherwise} \tag{13}$$

where m is the number of sensors lying in the given zone. For Zone-1 ($i = 1$), the number of data sensors is equal to 12; and that for Zone-2 ($i = 2$), it is equal to 9. While, three prediction points lie in Zone-1 (at height 1288 mm; marked by open circles), six prediction points lie in Zone-2 (three at height 932 mm and three at height 576 mm). The choice of two heights at 1288 mm and 932 mm is based on the standard deviation trend noted in Table I. It can be noted there is a sudden drop in temperature gradient from 1288 to 932 mm.

TABLE III
ERROR TABLE QUANTIFYING THE DEVIATIONS BETWEEN TRANSIENT DATA AND PREDICTIONS IN $x = 150$ mm PLANE

(y, z)	Correlation Coefficient	Relative Deviation
mm		(%)
(450, 1288)	0.9968	2.1
(300, 1288)	0.9983	1.2
(150, 1288)	0.9979	2.5
(450, 932)	0.9974	2.7
(300, 932)	0.9969	3.2
(150, 932)	0.9975	2.1
(450, 576)	0.9983	1.8
(300, 576)	0.9982	1.3
(150, 576)	0.9978	2.2

With the proposed model of optimal POD modes and corresponding POD coefficients, temperature signals are computed at the interrogation points and subsequently compared with the experimental data. Fig. 13 compares the experimental temperature data at nine different interrogation locations to the corresponding POD-based predictions. The comparison suggests close similarity between data and predictions. The similarity is further quantified by the corresponding correlation coefficients and relative estimated errors, as reported in Table III.

Table III indicates the correlation coefficient is in the order of 99.5% and relative estimated error of the order 1% (maximum = 3.2%). The proposed POD model is capable of predicting temperature data with 99% relative accuracy, and reducing the required sensor number from 30 to 21. That amounts to about 30% sensor reduction. On a related note, Fig. 13 shows few glitches in POD-based predictions between 115 s and 130 s. Similar glitches can be observed in transient POD modes. Given that POD modes are parameter independent, it can be inferred that these glitches are the property of this particular data matrix.

In summary, it can be claimed that a high-fidelity framework has been developed that is capable of improving spatial resolution of measured temperature data. The methodology can be further expanded with some formal optimization procedures to solve an optimization problem with the number of temperature sensors as the decision variable and the minimization of the sensor number as the objective function. The obvious main constraint is maintaining the temperature prediction error below certain preassigned error limit, such as the calibration error of the measurement system. The formal description of the optimization problem is

$$\begin{aligned} &\min \text{ (sensor number)} \\ &\text{s.t. deviation} < \text{error limit.} \end{aligned} \tag{14}$$

The proposed strategy is a measurement-based approach; therefore, the sensor pattern derived depends upon the temperature gradient, airflow pattern, rack power, and several other thermal variables. As shown in Fig. 12 for this case study, the number of convective environments needed for an arbitrary DC is contingent upon its airflow pattern. For the given case study, the airflow scheme is underfloor plenum supply and overhead ceiling return. Alternative airflow schemes could, for example, be underfloor plenum supply and room return, overhead ceiling supply and room return, and overhead ceiling supply and overhead ceiling return. The prediction of convective environments demands detailed CFD simulations or reduced order models. While CFD simulations are accurate, albeit computationally resource-intensive, and reduced-order models are efficient, but usually have a larger prediction uncertainty. The number of zones to be employed could be estimated based on exploratory experiments, or coarse grid CFD simulations. Indeed, an effective application of the proposed approach needs additional statistical analyses (as shown in Table I) or approximation models (as shown in Fig. 12) or CFD-based analyses.

On a related note, the authors concede that the switching from a geometry-based to a physics-based approach depends on the prediction fidelity of the geometry-based approach. That means the initial capital expenditure needs to be incurred for sensor deployment to validate the prediction fidelity of the geometry-based approach. In spite of this limitation, the framework is useful because one needs to establish the fidelity only during the equipment installations or major facility-level upgrades. The present approach is of high-fidelity during commonly occurring transient phenomena in DCs, such as rack heat load fluctuations, CRAC supply airflow changes and so on.

VI. Conclusion

This paper presents a reduced-order model to improve spatial resolution of experimentally measured air temperature data. The proposed framework uses a POD-based optimal model order reduction technique, which is a logarithmic time algorithm. The effectiveness of the framework is verified during a dynamic event triggered by sudden switching of a server simulator rack. It is observed that the proposed framework is capable of capturing temperature signals within 3.2% prediction uncertainty. The verification of the proposed approach is conducted in two interrogation planes: 1) one horizontal plane located near the top of the test rack in the hot aisle and 2) another vertical plane located near the exhaust of the test rack. In the horizontal plane, the POD coefficients are statistically computed. However, such a statistical approach fails in the vertical plane due to larger velocity gradient. Therefore, an approximation model for the velocity field is developed to compute the POD coefficients. The approximation model assumes the particular nature of the experimental setup, i.e., under-floor plenum supply, front-to-rear server flow, and drop ceiling return. Other airflow management strategies or rack arrangements might require different approximation models for determining regions with relatively low-temperature gradients. The cooling and IT equipment arrangement of a DC dictates the computation of the because it is a function of the spatial location of an interrogation point.

Unlike a method-driven approach like CFD/HT, the proposed approach is data-driven with spatial location as the parameter. Therefore, the presented solutions are specific for the given case studies. However, the proposed approach is general purpose. The presented framework can be easily adapted to monitor other physical quantities inside DCs, such as air velocity, air humidity, fan power, chiller flowrate, CPU utilization, and memory utilization. Such data-driven capability opens up the possibility of a rapid parametric decision-making platform to enable flexible thermal design of a DC. The decision-making platform would cooptimize various resources concurrently and identify the most efficient operating envelope. It could also serve as a prognostic-based thermal reliability modeling tool, useful during thermal runaway events, such as a power outage. By leveraging robustness and scalability of the POD-based algorithm, the measurement-based approach integrates online thermal monitoring with real-time analyses. Such thermal monalytics [27] will eventually offer following values to DC designers and operators: scalable local thermal analysis synchronized with virtual machine load migration, cost-effective dynamic sensor deployment, and data aggregation with minimal storage requirement. Overall, the framework facilitates stochastic decision making for DC facility operators by providing an effective alternative to first-order approximation-based sensor allocations.

References

[1] R. Brown, "Report to congress on server and data center energy efficiency: Public law 109-431," Tech. Rep., 2007.
[2] J. Koomey, *Growth in Data Center Electricity Use 2005 to 2010*, vol. 1. Oakland, CA, USA: Analytics Press, Aug. 2011.
[3] L. A. Barroso, "The price of performance," *Queue*, vol. 3, no. 7, pp. 48–53, 2005.

[4] M. K. Patterson and D. Fenwick. (2008). *The State of Datacenter Cooling*, Intel Corporation, Santa Clara, CA, USA [Online]. Available: http://download.intel.com/technology/eep/data-center-efficiency/stateof-date-center-cooling.pdf
[5] R. F. Sullivan, "Alternating cold and hot aisles provides more reliable cooling for server farms," Uptime Institute, New York, NY, USA, 2000.
[6] J. Stoess, C. Lang, and F. Bellosa, "Energy management for hypervisor-based virtual machines," in *Proc. USENIX Annu. Tech. Conf.*, 2007, pp. 1–14.
[7] G. M. Nelson, "Development of an experimentally-validated compact model of a server rack," Tech. Rep., 2007.
[8] 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," *IEEE Trans. Compon. Packag. Technol.*, vol. 31, no. 2, pp. 444–448, Jun. 2008.
[9] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, 2011.
[10] C. K. Chang and J. Huang, "Video surveillance for hazardous conditions using sensor networks," in *Proc. IEEE Int. Conf. Netw., Sens. Control*, Mar. 2004, pp. 1008–1013.
[11] H. Moon and P. J. Phillips, "Computational and performance aspects of PCA-based face-recognition algorithms," *Perception-London*, vol. 30, no. 3, pp. 303–322, 2001.
[12] M. Hubert and S. Engelen, "Robust PCA and classification in biosciences," *Bioinformatics*, vol. 20, no. 11, pp. 1728–1736, 2004.
[13] J. L. Lumley, "The structure of inhomogeneous turbulent flows," in *Atmospheric Turbulence and Radio Wave Propagation*. Moscow, Russia: Nauka, 1967, pp. 166–178.
[14] L. Sirovich, "Turbulence and the dynamics of coherent structures. I—Coherent structures. II—Symmetries and transformations. III—Dynamics and scaling," *Quart. Appl. Math.*, vol. 45, pp. 561–571, Oct. 1987.
[15] E. Samadiani and Y. Joshi, "Multi-parameter model reduction in multi-scale convective systems," *Int. J. Heat Mass Transf.*, vol. 53, nos. 9–10, pp. 2193–2205, 2010.
[16] N. Rolander, J. Rambo, Y. Joshi, J. K. Allen, and F. Mistree, "An approach to robust design of turbulent convective systems," *J. Mech. Design*, vol. 128, no. 4, pp. 844–855, 2006.
[17] R. Ghosh and Y. Joshi, "Error estimation in POD-based dynamic reduced-order thermal modeling of data centers," *Int. J. Heat Mass Transf.*, vol. 57, no. 2, pp. 698–707, 2013.
[18] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[19] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.
[20] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annu. Rev. Fluid Mech.*, vol. 25, pp. 539–575, Jan. 1993.
[21] E. Samadiani and Y. Joshi, "Proper orthogonal decomposition for reduced order thermal modeling of air cooled data centers," *J. Heat Transf.*, vol. 132, no. 7, pp. 071402-1–071402-14, 2010.
[22] P. Druault, P. Guibert, and F. Alizon, "Use of proper orthogonal decomposition for time interpolation from PIV data," *Experim. Fluids*, vol. 39, no. 6, pp. 1009–1023, 2005.
[23] N. Rolander, "An approach for the robust design of air cooled data center server cabinets," M.S. thesis, Dept. Mech. Eng., Georgia Institute of Technology, Atlanta, GA, USA, 2005.
[24] N. Dyn, D. Levin, and S. Rippa, "Data dependent triangulations for piecewise linear interpolation," *IMA J. Numer. Anal.*, vol. 10, no. 1, pp. 137–154, 1990.
[25] T. L. Bergman, F. P. Incropera, A. S. Lavine, and D. P. DeWitt, *Fundamentals of Heat and Mass Transfer*. New York, NY, USA: Wiley, 2011.
[26] R. J. Moffat, "Describing the uncertainties in experimental results," *Experim. Thermal Fluid Sci.*, vol. 1, no. 1, pp. 3–17, 1988.
[27] M. Kutare, G. Eisenhauer, C. Wang, K. Schwan, V. Talwar, and M. Wolf, "Monalytics: Online monitoring and analytics for managing large scale data centers," in *Proc. 7th Int. Conf. Auton. Comput.*, 2010, pp. 141–150.

Rajat Ghosh, photograph and biography not available at the time of publication.

Yogendra Joshi, photograph and biography not available at the time of publication.