The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0961-5539.htm

Reduced order thermal modeling of data centers via proper orthogonal decomposition: a review

Emad Samadiani and Yogendra Joshi
G.W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

Received 19 November 2009
Accepted 2 February 2010

Abstract

Purpose – The purpose of this paper is to review the available reduced order modeling approaches in the literature for predicting the flow and specially temperature fields inside data centers in terms of the involved design parameters.

Design/methodology/approach – This paper begins with a motivation for flow/thermal modeling needs for designing an energy-efficient thermal management system in data centers. Recent studies on air velocity and temperature field simulations in data centers through computational fluid dynamics/heat transfer (CFD/HT) are reviewed. Meta-modeling and reduced order modeling are tools to generate accurate and rapid surrogate models for a complex system. These tools, with a focus on low-dimensional models of turbulent flows are reviewed. Reduced order modeling techniques based on turbulent coherent structures identification, in particular the proper orthogonal decomposition (POD) are explained and reviewed in more details. Then, the available approaches for rapid thermal modeling of data centers are reviewed. Finally, recent studies on generating POD-based reduced order thermal models of data centers are reviewed and representative results are presented and compared for a case study.

Findings – It is concluded that low-dimensional models are needed in order to predict the multi-parameter dependent thermal behavior of data centers accurately and rapidly for design and control purposes. POD-based techniques have shown great approximation for multi-parameter thermal modeling of data centers. It is believed that wavelet-based techniques due to the their ability to separate between coherent and incoherent structures – something that POD cannot do – can be considered as new promising tools for reduced order thermal modeling of complex electronic systems such as data centers

Originality/value – The paper reviews different numerical methods and provides the reader with some insight for reduced order thermal modeling of complex convective systems such as data centers.

Keywords Modelling, Flow, Temperature measurement, Control systems, Data handling
Paper type Research paper

Nomenclature

- a, b — modal weight coefficient
- c_p — specific heat, J/kg K
- k_{eff} — effective thermal conductivity, W/mK
- n — number of observations
- m — number of retained modes
- P — pressure, Pa
- Q — heat generation, W
- q — volumetric heat generation, W/m^3
- T — temperature, K
- u — velocity, m/s

Greek symbols

- λ — eigenvalue
- ρ — density, kg/m^3
- ψ — POD mode

The authors acknowledge support of this research by the members of the Consortium for Energy-Efficient Thermal Management (CEETHERM).

International Journal of Numerical Methods for Heat & Fluid Flow
Vol. 20 No. 5, 2010
pp. 529-550

0961-5539
DOI 10.1108/09615531011048231

υ_{eff} effective viscosity, m^2/s

Subscripts

obs observation

Superscripts

$*$ Transpose

$+$ psuedo-inverse

1. Thermal management of data centers

Data centers, as shown in Figure 1, are computing infrastructure facilities that house arrays of electronic racks containing data processing and storage equipment, whose temperature must be maintained within allowable limits as it dissipates power. Data centers are utilized by a broad range of end-users including internet service providers, banks, stock exchanges, corporations, educational institutions, government installations, and research laboratories. The multi-scale nature of data centers spanning length scales from the chip to the facility level is shown in Figure 1. A common approach currently used for thermal management of air cooled data centers consists of computer room air conditioning (CRAC or AC) units that deliver cold air to the racks arranged in alternate cold/hot aisles through perforated tiles placed over an under-floor plenum, see Figure 2. The chip level determines the overall rate of the heat generation in the data center, while the CRAC units at the facility level are responsible in providing the cooling solution to maintain the chip temperatures in a safe range. Several alternate air-delivery and return configurations are employed, particularly when a raised floor arrangement is unavailable. Some of these are seen in Figure 3 (Rambo and Joshi, 2006).

The power consumption of data center facilities can be in the range of tens of MW, with an additional 30 percent or more needed for powering the cooling systems. Data center energy consumption is an increasingly important concern. In 2006 data centers in the USA consumed about 61 billion kWh, or 1.5 percent of total US electricity consumption, for a total electricity cost of about $4.5 billion (US Environmental Protection Agency, 2007). This estimated level of electricity consumption is estimated to

Figure 1. Data center and its multi-scale nature

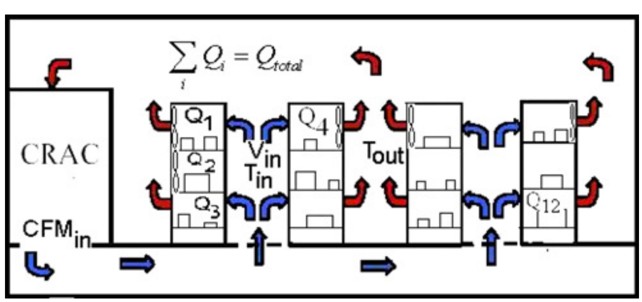
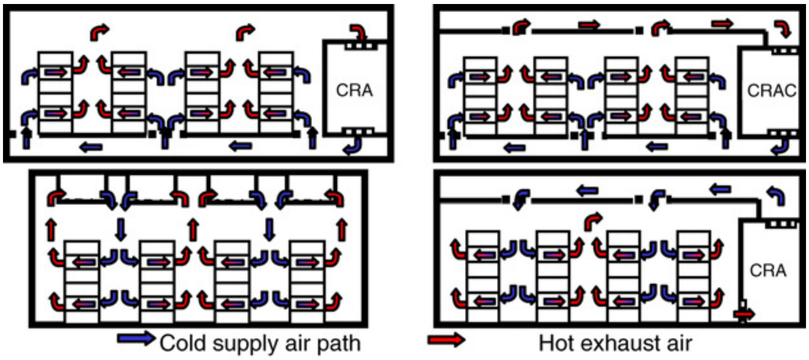


Figure 2. Typical air cooling system in data centers



Source: Rambo and Joshi (2006)

Figure 3. Different air-delivery and return configurations in data centers

be more than double that was consumed for this purpose in 2000 (US Environmental Protection Agency, 2007), and if this trend were to continue would require the construction of two new 500 MW power plants each year. Recent benchmarking studies by Lawrence Berkeley National Laboratories (Greenberg *et al.*, 2006) show an increase in data center floor heat loads per unit area over the past few years. This is consistent with the projected trend towards denser computing architectures, such as blade servers. The American Society of Heating Refrigeration and Air-conditioning (ASHRAE) projects significant increase in rack level powers (ASHRAE, 2005), as seen in Figure 4. Due to the relatively frequent upgrades in the computing equipment, both existing and new facilities are being subjected to these sharp increases in floor heat loading.

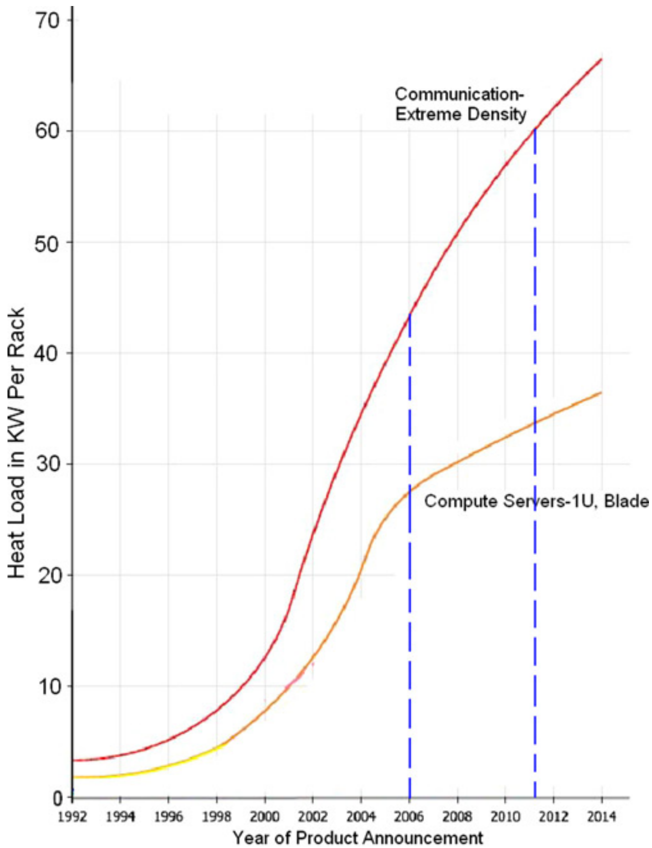
A significant fraction of the energy costs associated with the operation of a typical data center can be ascribed to the cooling hardware. In the recent benchmarking study of 11 existing facilities by the Lawrence Berkeley Laboratories (Greenberg *et al.*, 2006) the power consumption by the heating, ventilating and air-conditioning systems ranged from 22 to 54 percent of the overall supply. Some estimates suggest that the annual cost of powering and cooling servers already exceed their acquisition costs (Belady, 2007). Energy-efficient design of the cooling systems is essential for containing operating costs, and promoting sustainability. Through better design and preventing over-provisioning, it should be possible to reduce energy consumption by the cooling systems. Predicting the flow and specially temperature fields inside data centers in terms of the involved design parameters is necessary for an energy-efficient and reliable cooling system design.

2. Computational fluid dynamics/heat transfer (CFD/HT) modeling of data centers

Generally, the air flow inside data centers is turbulent. Also, buoyancy effects can usually be neglected (Rambo, 2006). The Reynolds-averaged Navier-Stokes equations are commonly used to simulate the turbulent mean flow in air-cooled data centers, by modeling the effect of turbulence on the mean flow as a spatially dependent effective viscosity:

$$\nabla u = 0 \tag{1}$$

$$u\nabla u - \nabla(\upsilon_{eff}\nabla u) + \frac{1}{\rho}\nabla p = 0. \tag{2}$$



Source: ASHRAE (2005)

Figure 4.
ASHRAE Heat load projections for communication and computing racks

Also, the mean energy equation with effective thermal conductivity can be used to compute the temperature field. The mean energy equation, ignoring viscous dissipation, is:

$$\rho c_p u \nabla T - \nabla(k_{eff} \nabla T) = q. \tag{3}$$

Several researchers have simulated the air flow and temperature fields in data centers (US Environmental Protection Agency, 2007; Greenberg *et al.*, 2006; Patel *et al.*, 2002; Rambo and Joshi, 2003a, b; Shrivastava *et al.*, 2005; Iyengar *et al.*, 2005; Schmidt *et al.*, 2004; VanGilder and Schmidt, 2005; Lawrence Berkeley National Laboratory and Rumsey Engineers, 2003; Samadiani *et al.*, 2007). Optimization (Shah *et al.*, 2005a, b; Bhopte *et al.*, 2005) and design (Schmit and Iyengar, 2005; Sharma *et al.*, 2002; Kang *et al.*, 2000; Boucher *et al.*, 2004; Rolander, 2005) incorporating different parameters involved in these systems have also been performed. CFD/HT is usually used to predict the air velocity and temperature fields inside the data center. Early application of CFD/HT modeling of data centers was done by Kang *et al.* (2000), Patel *et al.* (2001), Schmidt *et al.* (2001), and Rambo and Joshi (2003a). Schmidt *et al.* (2001) compared experimental measurements through raised floor data center perforated tiles with two-dimensional computational models. Their experimental validation shows fair overall agreement with mean tile flow rates, with large individual prediction errors. VanGilder and Schmidt (2005) parametrically studied plenum airflow for various data center footprints, tile arrangements, tile porosity and plenum depth. Initial studies to determine the air flow rates from the perforated tiles (Schmidt *et al.*, 2001, 2004; VanGilder and Schmidt, 2005; Radmehr *et al.*, 2005; Karki *et al.*, 2003) have modeled the plenum only and do not simulate the effect of the air flow inside the computer room on the perforated tile flow distribution. Samadiani *et al.* (2009) have shown that modeling the computer room, CRAC units, and/or the plenum pipes could change the tile flow distribution by up to 60 percent for the facility with 25 percent open perforated tiles and up to 135 percent for the facility with 56 percent open perforated tiles.

Numerical thermal modeling has been used for geometrical optimization of plenum depth, facility ceiling height, and cold aisle spacing for a single set of CRAC flow rates and uniform rack flow and power dissipation (Bhopte *et al.*, 2005). A unit cell architecture of a raised floor plenum data center is formulated in Rambo and Joshi (2003b) by considering the asymptotic flow distribution in the cold aisles with increasing number of racks in a row. The results indicated that for high flow rate racks, a "unit cell" containing four rows of seven racks adequately models the hot-aisle cold-aisle configuration in a "long" row of racks (Rambo and Joshi, 2003b). In Patel *et al.* (2002), Iyengar *et al.* (2005), Patel *et al.* (2001), and Schmidt and Cruz (2003), researchers have either modeled individual racks as black-boxes with prescribed flow rate and temperature rise, or with fixed flow rate and uniform heat generation. A procedure to model individual servers within each rack was developed in Rambo and Joshi (2003a). Rambo and Joshi (2003a) developed a multi-scale model of typical air-cooled data centers using commercial finite volume software. In their work, each rack is modeled as a series of sub-models designed to mimic the behavior of a server in a data center. Rambo and Joshi (2006) performed a parametric numerical study of various air supply and return schemes, coupled with various orientations of the racks and the CRAC units, to identify the causes of recirculation and non-uniformity in thermal performance throughout the data center.

The multi-scale nature of data centers needs to be considered in the numerical modeling. Also, as suggested in Samadiani *et al.* (2007), the future state-of-the-art of thermal management in data centers will include a combination of cooling solutions at different scales. This increases the need to have a multi-scale model for thermal phenomena happening at all important scales. The multi-scale model of a representative data center in Rambo and Joshi (2006, 2003a) consists of ~1,500,000 grid cells and needs more than 2,400 iterations to obtain a converged solution. This model took about 8 h to converge on a 2.8 GHz Xeon with 2 GB memory (Rambo and Joshi, 2006). Also, it should be noted that this model is still a significant departure from reality because it does not include finer details at the server and chip level. In light of this, a comprehensive CFD/HT multi-scale model of operational data centers, which may contain thousands of racks, seems infeasible due to limits on available computing. A compact or low-dimensional model which could run much faster, while including the influence of all important scale parameters with sufficient fidelity is essential. A comprehensive review of literature on data center numerical modeling with a study on the necessity of compact airflow/thermal modeling for data centers have been done in Rambo and Joshi (2007a).

3. Low-dimensional modeling approaches

Meta-modeling and reduced order modeling techniques can be used to extract the dominant characteristics of a system, trading a degree of accuracy for much greater computational speed. These techniques are briefly reviewed in sections 3.1 and 3.2.

3.1 Meta-modeling

Approaches such as linear response surfaces using design of experiments, krieging, multivariate adaptive regression splines, and other more advanced interpolation approaches offer approximations to generate a surrogate model of the system response in terms of the design variables (Simpson *et al.*, 2001). A literature review and comparison of different meta-modeling techniques with recommendations for computer-based engineering design has been done in Simpson *et al.* (2001).

Kriging, also called Gaussian process modeling, is a useful method for developing meta-models from expensive computer or experimental simulations for product design optimization (Sacks *et al.*, 1989; Santner *et al.*, 2003; Jin *et al.*, 2001). In kriging model, known as universal kriging, the true function of interest, Y(x), can be modeled as a combination of a known model plus departures (Wackernagel, 2002; Joseph *et al.*, 2008; Simpson *et al.*, 2001):

$$Y(\mathbf{x}) = g(\mathbf{x}) + Z(\mathbf{x}) \tag{4}$$

where $g(\mathbf{x}) = \sum_{i=0}^{r} g_i h_i(\mathbf{x})$ and $Z(x)$ is a weak stationary stochastic process with mean 0 and covariance function $\sigma^2\psi$. The h_is are some known functions and g_is are unknown parameters. The covariance function is defined as $\mathrm{cov}\{Y(\mathbf{x}+\mathbf{v}), Y(\mathbf{x})\} = \sigma^2\psi(\mathbf{v})$, where the correlation function $\psi(\mathbf{v})$is a positive semidefinite function with $\psi(0) = 1$ and $\psi(-\mathbf{h}) = \psi(\mathbf{h})$. In this formulation, $g(x)$ is used to capture the known trends, so that $Z(x)$ will be a stationary process. However, those trends are not usually known in reality. So, the following special case, known as ordinary kriging, is commonly used:

$$Y(\mathbf{x}) = g_0 + Z(\mathbf{x}). \tag{5}$$

The meta-model to predict the response function can be obtained as follows. If some function values $y = (y_1, \ldots, y_n)^*$ have been evaluated at corresponding n points $\{x_1, \ldots, x_n\}$, the ordinary kriging predictor is given by

$$\hat{y}(\mathbf{x}) = \hat{g}_0 + \psi(\mathbf{x})^* \Psi^{-1}(y - \hat{g}_0 \mathbf{I}) \tag{6}$$

where $\mathbf{I}$ is a column of 1s having length n, $\psi(\mathbf{x})^{*'} = (\psi(\mathbf{x}-\mathbf{x}_1), \ldots, \psi(\mathbf{x}-\mathbf{x}_n))$, and Ψ is an $n \times n$ matrix with elements $\psi(\mathbf{x}_i - \mathbf{x}_j)$, and $\hat{g}_0 = \mathbf{I}^*\Psi^{-1}\mathbf{y}/\mathbf{I}^{*'}\Psi^{-1}\mathbf{I}$. It is the best linear unbiased predictor, which minimizes the mean squared prediction error $E\{\hat{Y}(\mathbf{x}) - Y(\mathbf{x})\}^2$ under the model in Equation (5).

Computer models are often deterministic and there is no random error in the output. So, kriging, providing an interpolating meta-model, is more suitable than other common alternatives such as quadratic response surface model. It has been used for the thermal design of wearable computers (Pacheco *et al.*, 2003) and a variable thickness piezoelectric bimorph actuator (Cappelleri *et al.*, 2002). Also, Giunta (1997) presents an investigation into the use of kriging for the multidisciplinary design optimization of a High Speed Civil Transport aircraft. See Simpson *et al.* (2001) for more examples of kriging applications.

Joseph *et al.* (2008) propose a modified kriging method, called blind kriging, which has an unknown mean model identified from experimental data using a Bayesian variable selection technique. In the blind kriging method, the functions h_i's are not assumed to be known. Instead, they are identified through some data-analytic procedures. So, the blind kriging model is given by:

$$Y(\mathbf{x}) = \mathbf{h}(\mathbf{x})'\mathbf{g}_r + Z(\mathbf{x}) \tag{7}$$

where $\mathbf{h}(x)' = (1, h_1, \ldots, h_r), \mathbf{g}_m = (g_0, g_1, \ldots, g_r)'$, and r are unknown. The most important step in blind kriging is to identify the unknown functions h_is. They can be chosen from a set of candidate functions, or variables, using variable selection techniques (Joseph *et al.*, 2008). The blind kriging has been applied for making a surrogate model of an engine block and head joint sealing assembly, piston slap noise, and for the flow rate through a borehole. Remarkable improvement is shown in prediction of the corresponding response functions using blind kriging over ordinary kriging. Also, it is concluded that a blind kriging predictor is simpler to interpret and is more robust against the mis-specification in the correlation parameters than an ordinary kriging predictor (Joseph *et al.*, 2008).

Qian *et al.* (2006) present a two-step approach for building low-cost surrogate models based on data from both detailed and approximate simulations. In their method, a Gaussian process model is first fitted using only approximate simulation data. Then, the fitted model is adjusted with detailed simulation data. They demonstrated the approach for the design of an electronic cooling heat exchanger involving linear cellular materials, using a detailed but slow simulation based on FLUENT finite volume analysis and an approximate but fast simulation using finite difference method. The approach is particularly suitable when both a physics-based model and an approximate model are available. Qian and Wu (2008) have extended the work in Qian *et al.* (2006) to carry out location and scale adjustments more flexibly and absorb uncertainty in the model parameters in the prediction.

3.2 Reduced order modeling through flow structure identification

The process of taking a model based either on detailed numerical simulations or full-field experimental measurements from a large number of degrees of freedom (DOF) to one involving significantly fewer DOF is termed model reduction; Shapiro, 2002). An approach to develop reduced order models of turbulent flows is based on the observation that many such flows are characterized by characteristic recurrent forms that are collectively called coherent structures. These are energetically dominant in many flows. So, it should be possible to build a relatively realistic, low-dimensional model of the flow by keeping only the dominant coherent structures, and simulating the effect of the smaller, less energetic, apparently incoherent part of the flow in some way (Holmes *et al.*, 1996). For this, one needs to identify the dynamically active structures, classify their elementary interactions, and define an averaging procedure to construct averaged quantities which would be the appropriate variables to describe turbulence, and then find the corresponding transport equations to compute the evolution of these new quantities (Farge *et al.*, 1996).

3.2.1 Turbulent coherent structure identification. Flow structure identification techniques can be used to capture the coherent structures of turbulent flows, using a time dependent data set obtained after refining some numerical or experimental velocity data. The approaches for physical description of the coherent structures fall into one of the two basic categories: Eulerian or Lagrangian. The Eulerian approach to coherent structures aims at partitioning the flow based on the instantaneous distribution of a scalar field, such as the vorticity, kinetic energy, enstrophy, or the strain (Haller and Yuan, 2000). For instance, Eulerian coherent vortices can be identified as regions with vorticity over a small threshold. On the other hand, the Lagrangian approach to the coherent structures is concerned with patterns emerging from the advection of passive tracers (Haller and Yuan, 2000). For instance, Lagrangian coherent vortices have been studied in terms of absolute and relative single particle dispersion.

Eulerian coherent structures can be obtained from, for example, Q-criterion (Hunt *et al.*, 1988) and the swirling strength criterion (Zhou *et al.*, 1999). These criteria are typically formulated in terms of the invariants of the velocity gradient tensor. Other Eulerian criteria have also been used for structure identification, and some of these have been compared to Lagrangian criteria (Haller, 2005). Lagrangian coherent structures can be obtained from the Okubo-Weiss criterion or finite-time Lyapunov exponents (Haller, 2002). Haller (2002) has examined the relevance of Lagrangian coherent structures for the true flow in two-dimensional domains. Green *et al.* (2007) have identified Lagrangian coherent structures for two three-dimensional flows in a plane channel, including an isolated hairpin vortex and a fully developed turbulent flow, by calculating the direct Lyapunov exponent (DLE). The Lagrangian method captures features of the flow that are familiar from flow visualization experiments, and are also described by various Eulerian criteria currently in use, but the DLE field yields greater detail than existing Eulerian criteria. This is partially because, unlike Eulerian criteria, the DLE may be evaluated on a finer grid than the original velocity data (Green *et al.*, 2007).

3.2.2 Procedures for coherent structure evolution calculation. 3.2.2.1 Fourier and wavelet-based techniques. Appropriate averaging procedures and corresponding transport equations are needed to compute the evolution of the coherent structures in turbulent flows. The fundamental principle in generating low-dimensional turbulence modeling based on the coherent structures is to find a representative set of modes or bases to project the governing equations onto, reducing the solution procedure to finding the appropriate weight coefficients that combine the modes into the desired approximate solution. All classical methods in turbulence rely on the Fourier representation. While the dissipation term is optimally represented in Fourier space because Fourier modes diagonalize the Laplacian operator (for periodic boundary conditions), the nonlinear convective term is very complicated in Fourier space where it becomes a convolution, i.e. all Fourier modes are involved (Farge *et al.*, 1996). Also, turbulent motions are nonseparable in the Fourier representation.

Wavelet transform-based techniques are alternative tools to identify the coherent/incoherent structures, and model and compute turbulent flows. The most useful property of the wavelet transform is its ability to detect and accurately measure the strength of individual singularities in a signal. So, wavelet-based techniques can be used to separately model the coherent and incoherent flow components, something that Fourier-based models cannot do. Farge *et al.* (1996) present a comprehensive review on the application of wavelet-based techniques for turbulent flows. They have shown numerous promising case studies in solving partial differential equations in wavelet space, including heat diffusion equation, Stokes flow in 2D, and Navier-Stokes equations in 2D. Also, wavelet-based techniques can be used to add detail to existing fluid flow simulations as a user-controlled post-process. Kim *et al.* (2008) have presented a novel wavelet method to enable large- and small-scale detail to be modeled separately. Instead of solving the Navier-Stokes equations over a highly refined mesh, they used the wavelet decomposition of a low-resolution simulation to determine the location and energy characteristics of missing high-frequency components. Then, these missing components were synthesized using a novel incompressible turbulence function (Kim *et al.*, 2008).

3.2.2.2 Proper orthogonal decomposition (POD)-based techniques. In addition to the Fourier and wavelet-based techniques, the POD can also be used to accomplish low-dimensional turbulence modeling (Rambo, 2006; Rolander, 2005; Holmes *et al.*, 1996; Rambo and Joshi, 2007b). The POD, also known as the Karhunen-Loeve decomposition, is a statistical technique and has several properties that make it well suited for turbulent flows. First, it has been shown experimentally that low-dimensional models using POD can well address the role of coherent structures in turbulence generation (Holmes *et al.*, 1996). Second, it captures more of the dominant dynamics for a given number of modes than any other linear decomposition (Holmes *et al.*, 1996). Finally, the empirical determination of the basis functions makes this method ideal for nonlinear problems. A review of the POD method and its application for turbulence modeling has been done in Holmes *et al.* (1996).

In the POD-based model reduction technique, a set of data are expanded on empirically determined basis functions for modal decomposition. It can be used to numerically predict the temperature field more rapidly than full-field simulations. The temperature field is expanded into basis functions or POD modes:

$$T = T_0 + \sum_{i=1}^{m} b_i \psi_i. \tag{8}$$

The general algorithm to generate a POD-based reduced order thermal modeling in a system is illustrated in Figure 5 and is explained in the following:

(1) *Observation generation.* In the first step, the design variables of the system are changed n-times and the temperature field for the entire domain is obtained by

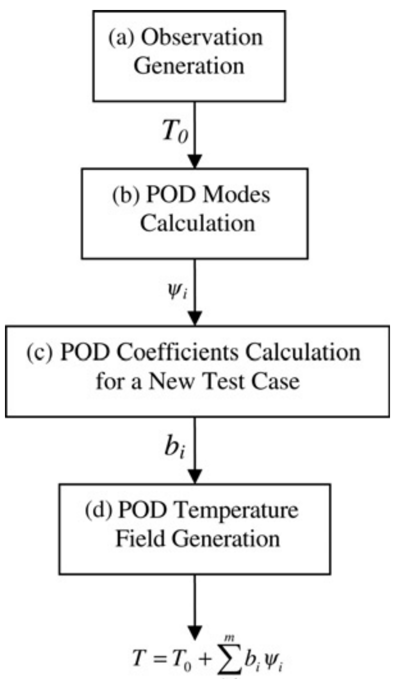


Figure 5.
General algorithm for POD temperature field generation

CFD/HT simulations, or detailed experimental measurements for each case. These thermal fields are called observations or snapshots. An element of the reference temperature field, T_0 in Equation (8), is typically considered as the average of the all observed data for a field point.

(2) *POD modes, ψ_i, calculation*. The POD modes of a thermal system, ψ_i, can be calculated from observations. In Equation (8), m is the number of retained POD modes in the decomposition which can be 1 up to n-1, where n is the number of observations. Using the method of snapshots, each POD mode can be expressed as a linear combination of the linearly independent observations (Holmes *et al.*, 1996):

$$\psi_i = \sum_{k=1}^{n} a_k (T_{obs,k} - T_0) \tag{9}$$

where T_{obs} is a matrix of which each column, $T_{obs,\ i}$, includes a complete temperature field data from an observation. The weight coefficients, a_k, in Equation (9) are obtained by solving the following n*n eigenvalue problem:

$$\sum_{k=1}^{n} R(i,k) a_k = \lambda a_i; \quad i = 1, \ldots, n \tag{10}$$

where $R = (T_{obs} - T_0)^* \otimes (T_{obs} - T_0)/n$ (Rambo, 2006; Rolander, 2005; Holmes *et al.*, 1996; Rambo and Joshi, 2007b). For a given set of observations, n eigenvalues, λ_i, and their relevant eigenvectors are obtained from Equation (10). Each eigenvector includes the weight coefficients, a_k, of the relative POD mode in Equation (9), so n POD modes are finally calculated. The energy captured by each POD mode in the system is proportional to the relevant eigenvalue. The eigenvalues are sorted in a descending order, so the first few POD modes in Equation (8) capture larger energy compared with the later modes.

(3) *POD coefficients, b_i, calculation for a new test case*. This key step is where the POD can be used to create a reduced order thermal/fluid model as a function of the system design variables. Generally, there are three methods to calculate the POD coefficients b_i for a new test case with a new set of design variables:

- *Galerkin projection of the system POD modes onto the governing equations*. This results in a set of coupled non-linear ordinary differential equations in time for transient systems, or a set of algebraic equations for steady state systems, to be solved for the POD coefficients. This method has been used to create reduced order models of transient temperature fields in terms of mostly one parameter, such as Reynolds/Raleigh number (Ravindran, 2002; Park and Cho, 1996a, b; Sirovich and Park, 1990a, b; Tarman and Sirovich, 1998; Park and Li, 2002; Ding *et al.*, 2008). The previous investigations have been either for prototypical flows (such as flow around a cylinder), or for simple geometries such as channel flow where inhomogeneous boundary conditions are easily homogenized by the inclusion of a source function in the decomposition.
- *Interpolation among modal coefficients*. In steady state, the POD coefficients at a new set of design variables can be obtained by an interpolation between the weight coefficients at the observed variables to match a desired new

variable value (Ding *et al.*, 2008; Ly and Tran, 2001). In this approach, the coefficients used to reconstruct an observed field $T_{obs,k}$ are found first by projecting each of the POD modes onto the observation in turn:

$$b_{i,obs} = (T_{obs,k} - T_0) \bullet \psi_i \quad i = 1, \ldots, m. \tag{11}$$

This can be computed for all observations within the ensemble T_{obs}. The complete coefficient matrix $B \in \Re^{m \times n}$, in which each column is the coefficient vector to reconstruct the corresponding observation from the ensemble T_{obs}, can be more efficiently computed as:

$$B = \psi^{+} \otimes (T_{obs} - T_0) \tag{12}$$

where $(.)^{+}$ is the Moore-Penrose pseudo-inverse giving the least squares solution (Strang, 1988). Once $b_{i,obs}$ has been found for all observations, each of which represents the solution under a specified combination of design variables, the POD coefficients b_i for a new set of design variables are calculated through the interpolation of the coefficients $b_{i,obs}$ between the corresponding observations. In other words, rather than directly interpolating between observations, interpolation is performed in the POD mode space using the coefficients $b_{i,obs}$. For systems with one design variable, this interpolation can be done through linear, or the slightly more accurate piecewise cubic spline interpolation between coefficients. This method has been applied only for a system with one parameter and simple geometry such as cavity flow (Ding *et al.*, 2008; Ly and Tran, 2001). However, the approach can be extended to more complex systems with multiple design variables using higher order multi-dimensional interpolation approaches, such as krieging or multivariate adaptive regression splines (MARS) (Rolander, 2005).

- *Flux matching process*. In the flux matching process (Rambo, 2006; Rambo and Joshi, 2007b), the coefficients b_i are obtained by applying Equation (8) to some locally specified region, such as system boundaries to match the known mass or heat fluxes. Although the flux matching process has been used to develop reduced order models of the flow behavior in complex steady state systems successfully (Rambo, 2006; Rolander, 2005; Rolander *et al.*, 2006; Nie and Joshi, 2008a), it has been applied only for thermal modeling of a simple 2D geometry of a channel with two iso-heat flux blocks (Rambo and Joshi, 2007b; Nie and Joshi, 2008a), with no consideration of complex 3D geometry. Nie and Joshi (2008b) have presented a POD-based reduced order modeling of steady turbulent convection in connected domains with the application for a 3D electronic rack. They developed a POD-based modeling for each component separately and then subsequently combined the models together using boundary profile based flux matching. Their method is only applicable to systems consisting of a series of nested sub-domains. Also, they applied their methods for a case study, where the thermal parameters, which are chip heat generation rates, existed only in one sub-domain, making the temperature distribution in other domains almost uniform. So, matching of the sub-domains' temperature fields was much easier than matching the flow and pressure fields in Nie and Joshi (2008a).

(4) *POD temperature field generation*. With calculated T_0, ψ_i, and b_i for a new set of design variables, the corresponding temperature field for the test case can be generated inside the entire domain from Equation (8) for different numbers of used POD modes, m.

4. Low-dimensional modeling of data centers

4.1 Heuristic methods

Aside from CFD/HT, simulation methods based on some heuristic approaches have also been explored (Moore *et al.*, 2005a, b; Karlsson *et al.*, 2004; Sharma *et al.*, 2003; Tang *et al.*, 2006; Nathuji *et al.*, 2008; Somani and Joshi, 2008; Qian, 2006) to predict the air temperature at discrete points, such as server inlets/outlets, for a new heat load distribution among the data center racks or servers. In Moore *et al.* (2005a, b), Karlsson *et al.* (2004) and Sharma *et al.* (2003), machine learning techniques based on the input from several deployed sensors are used to understand the relation between workload and internal and ambient temperatures. These methods require a large number of data points for interpolation and usually need a lengthy calibration for each data center of interest before they can be used for simulation. In Qian (2006), a three-fold latent variable model, using structural-equation method and errors-in-variables parameterization, is proposed to generate a surrogate model for maximum rack inlet temperatures in a non-raised floor data center in terms of nine design variables. The data center model in Somani and Joshi (2008) has four rows with six racks for the first two rows and four for the last two. They simulated the data center for 148 configuration runs by Flotherm. However, they just monitored the temperature at five points for each of the 20 rack positions, resulting in 100 points totally. The surrogate model has been used for determining practical values of the configuration variables of the data center to meet some physical and usage requirements.

In Tang *et al.* (2006), the rate of heat transferred by the airflow recirculation is described by a cross-interference coefficient matrix, which shows how much of the heat transferred by the air exiting from the outlet of each server contributes to the inlet of every other server. Having obtained this matrix through a calibration process for a specific data center, an abstract heat flow model is developed to predict the temperatures at the server inlets/outlets vs server power consumption. In Nathuji *et al.* (2008) and Somani and Joshi (2008), a coefficient matrix is assembled through a calibration process to provide an estimate of the sensitivity of each server inlet temperature to every other server heat load unit step change, for a given CRAC velocity. So an ambient intelligence-based load management approach is designed to determine the maximum possible heat loads of each server to meet the corresponding thermal constraint within a given air velocity.

The mentioned works above simulate the effects of the system parameters on the temperature field in data centers based on some heuristic approaches. These methods can predict the air temperatures only at some discrete points, such as server inlets/outlets. Recent work on reduced order thermal modeling of data centers using POD has extended the capabilities of the technique (Samadiani and Joshi, 2010a, b; Samadiani *et al.*, 2009). These extensions and representative results from Samadiani and Joshi (2010a, b) and Samadiani *et al.* (2009) are briefly reviewed in the following sections.

4.2 POD and Galerkin projection for data center modeling

Samadiani and Joshi (2010a) have presented an approach to handle the challenges of multi-parameter reduced order thermal modeling in complex multi-scale convective systems. The approach is centered on the integration of three constructs:

(1) POD;

(2) Galerkin projection; and

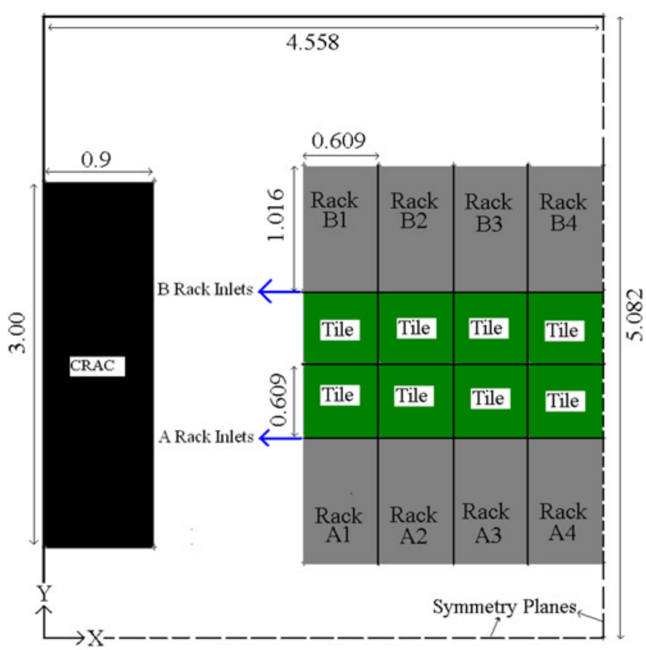
(3) energy balance/heat flux matching.

I is applicable to systems where the temperature field at selected scales or components drives the thermal design decision. The energy equation is solved only at these dominant components via system POD modes and Galerkin projection to obtain a more accurate zoomed prediction at these components, instead of the entire domain. The effects of the phenomena at other scales are modeled through simple energy balance equations and known heat flux and temperature matching, as well as appropriate matching conditions at the component interfaces. Unlike the previous work reviewed in section 4.1, the POD-based modeling is a deterministic approach which can predict the temperature at the whole data center domain in terms of multiple design variables. Also, a novel feature, compared with the previous POD related work reviewed in section Proper orthogonal decomposition (POD)-based techniques, is the use of POD modes and Galerkin projection for solving the governing turbulent convection equation in a complex multi-scale system. To the best of the authors' knowledge, this work is the first attempt to develop multi-parameter reduced order thermal modeling of complex multi-scale convective systems such as data centers.

The method has been applied to a representative data center, shown in Figure 6, to obtain a reduced order thermal modeling inside the data center with focus on the temperature filed at the rack scale. There are five design variables for the case study of Figure 6:

(1) inlet air velocity of CRAC unit, V_{in};

(2) heat load of racks A1 and B1, Q_1;



Note: Dimensions in m

Figure 6.
Case study data center cell top view

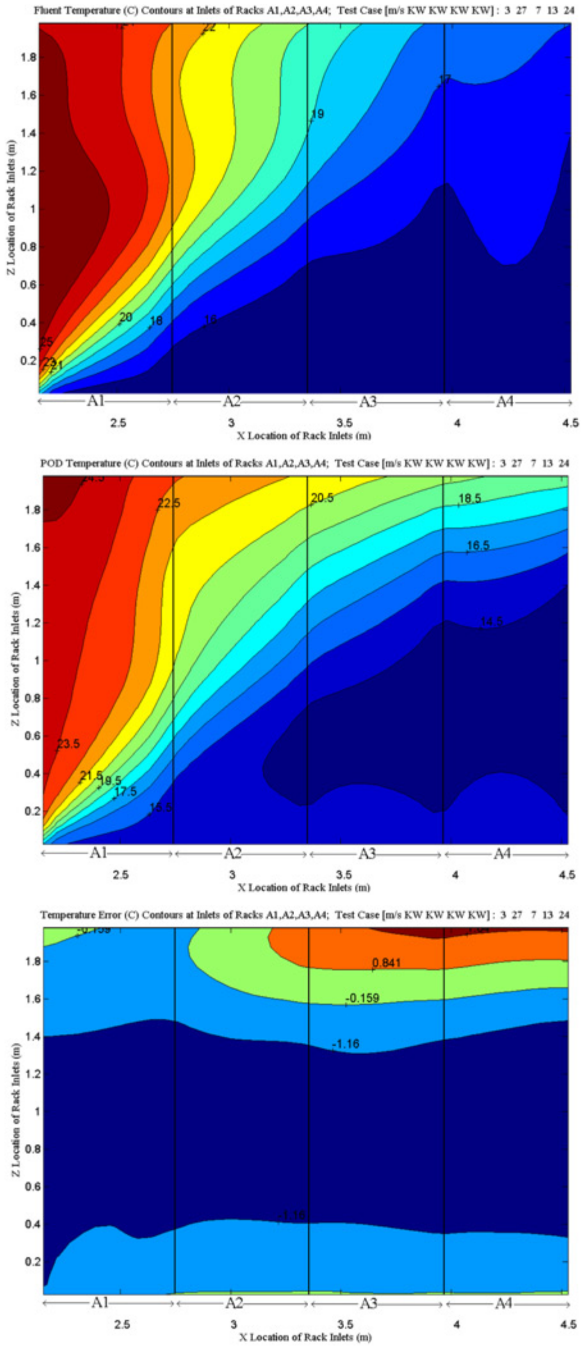
(3) heat load of racks A2 and B2, Q_2;
(4) heat load of racks A3 and B3, Q_3; and
(5) heat load of racks A4 and B4, Q_4.

The presented algorithm is used to generate temperature field for several new combinations of the design variables. It is shown that while the average temperature difference across the racks has converged after ~10 modes, the local temperatures need ~4 additional modes to converge for the same test case. To see if the POD method can predict the air temperatures at the rack inlets accurately for use in design decisions, the full field CFD/HT predictions using Fluent, POD simulations, and the POD temperature error from Samadiani and Joshi (2010a) are shown in Figure 7 for racks A1 through A4 for test case of [3 m/s, 27 kW, 7 kW, 13 kW, 24 kW] which is distinct from the original observations. It is shown that the average error at the rack inlets for different test cases is less than 1 °C, while the maximum local error is ~2.5 °C for some small regions (Samadiani and Joshi, 2010a). Also, the mean error, the standard deviation, and the Euclidean L2 norm of the POD temperature error at all 114,000 points of the rack scale for 15 test cases have been shown in Samadiani and Joshi (2010a). The mean error varies from 0.35 to 2.29 °C, while the average error is 1.36 °C, and the average standard deviation 1.12 °C. Also, the error norm changes from 1.8 to 10.1 percent, while the average is 6.2 percent (Samadiani and Joshi, 2010a). These values confirm that the presented POD method is accurate enough at the rack scale for design purposes.

The main goal of the suggested algorithm in Samadiani and Joshi (2010a) is to predict air temperatures at the rack inlet/outlets and inside the racks accurately and quickly for design purposes. So, the prediction error in the entire data center domain would be larger. It is shown that the mean error at 383,826 points of the entire data center cell changes from 1.64 °C up to 6.31 °C for the 15 test cases. The average of the mean errors and standard deviations are 4.08 and 3.67 °C, respectively. Also, the error norm changes from 10.7 percent up to 35.7 percent while the average is 21.1 percent. All these values confirm that the presented POD method is not accurate enough at the room scale (Samadiani and Joshi, 2010a).

Regarding the computational efficiency of the technique, the POD-based algorithm (Samadiani and Joshi, 2010a) generates the temperature field for a new test case with different CRAC velocity and rack heat loads in 12 min, while the CFD/HT simulation takes ~2 h for the same test case on the same computing platform (a desktop computer with Xeon™ CPU, 2.8 GHz and 2.75 GB of RAM). Also, the most time-consuming part of the method, integrating the velocity terms in the Galerkin projection over the domain, can be done once for all observed CRAC velocities, if the method is to be used for many simulations. It takes ~38 min to calculate these terms. After that, the algorithm is ready to obtain the POD temperature field for each new test case in only 4 s (Samadiani and Joshi, 2010a).

Although the presented method in Samadiani and Joshi (2010a) provides a quick and reasonably accurate thermal modeling of air-cooled data centers for design purposes, the approach is applicable only for systems where the temperature field at selected scales, called dominant scales, drives the thermal design decision. Accordingly, the generated temperature field based on this method at scales other than dominant scales is not very accurate. Also, the method requires the fluid flow solution at these dominant scales for integration of the energy equation via system POD modes and Galerkin projection.



Source: Samadiani and Joshi (2010a)

Figure 7.
Contours of CFD/HT temperature, POD temperature, and relative error (°C) at racks inlets for test case of [3 m/s, 27 kW, 7 kW, 13 kW, 24 kW]

4.3 POD and energy balance matching for data center modeling

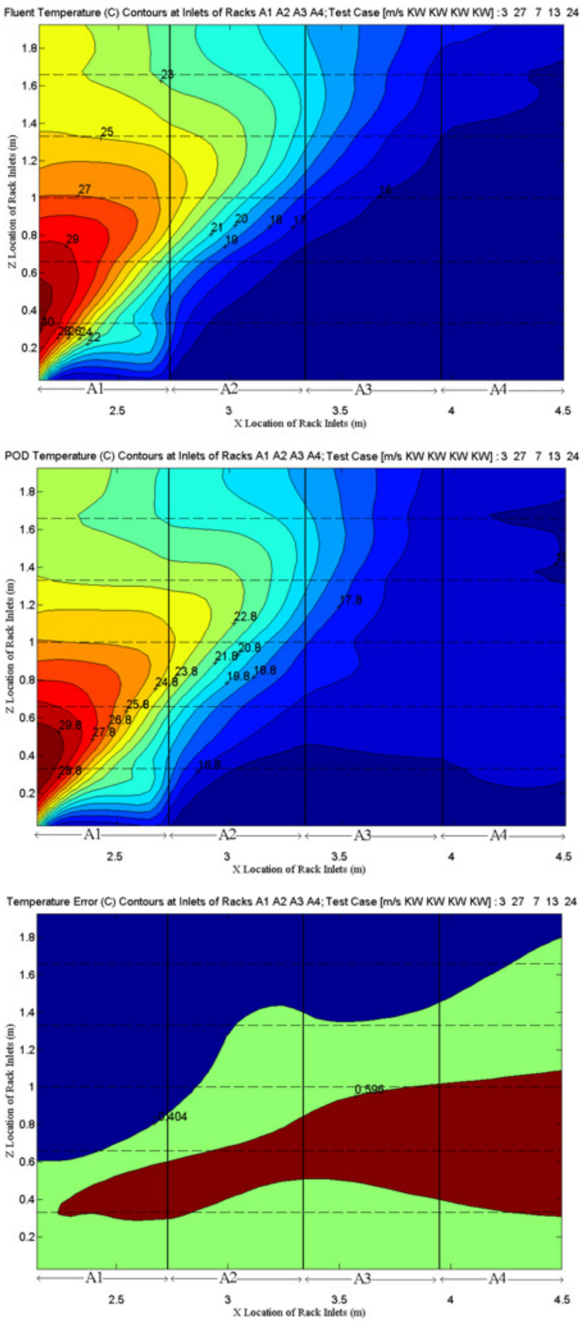
Samadiani and Joshi (2010b) have presented a simpler POD-based method to generate a reduced order thermal modeling of complex systems such as air cooled data centers. In this method, the algebraic equations to be solved for POD coefficients in Equation (4) are obtained simply through energy balance equations, heat flux matching (Rambo and Joshi, 2007b), and/or surface temperature matching for all convective components of the complex system. They applied this method to the case study shown in Figure 6 with the five design variables mentioned in section 4.2. The difference with the case study in Samadiani and Joshi (2010a) is that each server within a rack in Samadiani and Joshi (2010b) is modeled as a separate volumetric heat source. In Samadiani and Joshi (2010a), each rack was modeled as a uniform volumetric heat source.

It is shown that the temperature difference across the servers has converged after ~7 modes. Also, the local temperatures need ~3 additional modes to converge for the same test case. The full field CFD/HT predictions by Fluent, POD simulations, and the POD temperature error from Samadiani and Joshi (2010b) are shown in Figure 8 for racks A1 through A4 for test case of [3 m/s, 27 kW, 7 kW, 13 kW, 24 kW]. It is shown in Samadiani and Joshi (2010b) that the average error at the rack inlets for different test cases is less than 1.5 °C, while the maximum local error is ~2.5 °C for some small regions. Also, the temperature prediction error at all 431,120 points representing the entire data center cell have been studied for six test cases in Samadiani and Joshi (2010b). The mean error for the six test cases varies from 0.63 °C or 2.4 percent to 2.13 °C or 8.4 percent. The average of the mean absolute and relative error for all cases is 1.24 °C and 4.9 percent, while the average standard deviation is 1.46 °C. These values confirm that the presented POD method is accurate enough for the entire data center cell. Regarding the computational efficiency, it takes only ~48 s to obtain the POD temperature field by the method in Samadiani and Joshi (2010b), which is ~150 times faster than the full field CFD/HT simulation.

Each of the two POD-based methods explained in sections 4.2 and 4.3 has its own pros and cons. Unlike the POD-based method in Samadiani and Joshi (2010a), the presented method in Samadiani and Joshi (2010b) does not need fluid flow modeling and is accurate throughout the entire domain. Also, the method in Samadiani and Joshi (2010b) is much simpler and its application is easier for reduced order thermal modeling of operational data centers, where the observation data are gathered experimentally and thermal sensors are deployed at the inlet/outlet of the servers.

As a deficiency, the number of available algebraic equations to be solved for the POD coefficients in the presented method in Samadiani and Joshi (2010b) is limited by the number of convective components and available thermal information for the components in the system. This brings a limitation to the method whose effect on the results for the data center cell is studied in Samadiani and Joshi (2010b). It is concluded that the method can be used as a reliable and rapid predictor to obtain a new temperature field throughout the system, unless the number of components or available thermal information in the form of equations at the component boundaries is very close to or less than the number of dominant modes. This would not typically cause a problem in thermal model reduction of operational data centers with several housed servers if enough numbers of servers have thermal sensors at their inlet/outlet.

On the other hand, the POD technique based on Galerkin projection in Samadiani and Joshi (2010a) does not have any limitation regarding the number of components, since using Galerkin projection to obtain the algebraic equations results in m distinct algebraic equations for each component, if m POD modes are used.



Source: Samadiani and Joshi (2010b)

Figure 8.
Contours of CFD/HT temperature, POD temperature, and relative error (°C) at racks inlets for test case of [3 m/s, 27 kW, 7 kW, 13 kW, 24 kW]

5. Conclusions

Studies on the airflow/thermal modeling of air-cooled data centers through CFD/HT are reviewed and it is concluded that low-dimensional models are needed in order to predict the multi-parameter dependent thermal behavior of data centers accurately and rapidly for design and control purposes. Some studied meta-modeling techniques are reviewed. Reduced order modeling approaches based on coherent structures are explained and reviewed as available tools for low-dimensional turbulence modeling. While most of the studies in the literature on rapid thermal modeling of data centers are based on heuristic approaches, two recently developed POD-based reduced order thermal modeling methods for data centers are reviewed and compared. Also, the obtained results for a case study are presented. POD-based techniques have shown great approximation for multi-parameter thermal modeling of data centers. It is believed that wavelet-based techniques due to the their ability to separate between coherent and incoherent structures – something that POD cannot do – can be considered as new promising tools for reduced order thermal modeling of complex electronic systems such as data centers.

References

ASHRAE (2005), *Datacom Equipment Power Trends and Cooling Applications*, American Society of Heating, Refrigeration and Air-Conditioning Engineers, Atlanta, GA.

Belady, C.L. (2007), "In the data center, power and cooling costs more than the it equipment it supports", *Electronics Cooling*, Vol. 13 No. 1, available at: www.electronics-cooling.com/2007/02/in-the-data-center-power-and-cooling-costs-more-than-the-it-equipment-it-supports/

Bhopte, S., Agonafer, D., Schmidt, R. and Sammakia, B. (2005), "Optimization of data center room layout to minimize rack inlet air temperature", *Proceedings of the ASME InterPACK*, *San Francisco, CA*, ASME, IPACK2005-73027.

Boucher, T.D., Auslander, D.M., Bash, C.E., Federspiel, C.C. and Patel, C.D. (2004), "Viability of dynamic cooling control in a data center environment", *Proceedings of the International Society Conference on Thermal Phenomena*, pp. 593-600.

Cappelleri, D.J., Frecker, M.I., Simpson, T.W. and Snyder, A. (2002), "Design of a PZT bimorph actuator using a metamodel-based approach", *ASME Journal of Mechanical Design*, Vol. 124, pp. 354-7.

Ding, P., Wu, X.-H., He, Y.-L. and Tao, W.-Q. (2008), "A fast and efficient method for predicting fluid flow and heat transfer problems", *ASME Journal of Heat Transfer*, Vol. 130 No. 3, p. 032502.

Farge, M., Kevlahan, N., Perrier, V. and Goirand, E. (1996), "Wavelets and turbulence", *Proceedings of the IEEE*, Vol. 84 No. 4, pp. 639-69.

Giunta, A.A. (1997), "Aircraft multidisciplinary design optimization using design of experiments theory and response surface modeling", PhD thesis, Virginia Polytechnic and State University, Blacksburg, VA.

Green, M.A., Rowley, C.W. and Haller, G. (2007), "Detection of Lagrangian coherent structures in three-dimensional turbulence", *Journal of Fluid Mechanics*, Vol. 572, pp. 111-20.

Greenberg, S., Mills, E., Tschudi, W., Rumsey, P. and Myatt, B. (2006), "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, California*, pp. 76-87.

Haller, G. (2002), "Lagrangian coherent structures from approximate velocity data", *Physics of Fluids*, Vol. 14 No. 6, pp. 1851-61.

Haller, G. (2005), "An objective definition of a vortex", *Journal of Fluid Mechanics*, Vol. 525, pp. 1-26.

Haller, G. and Yuan, G. (2000), "Lagrangian coherent structures and mixing in two-dimensional turbulence", *Physica D*, Vol. 147, pp. 352-70.

Holmes, P., Lumley, J.L. and Berkooz, G. (1996), *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge.

Hunt, J.C.R., Wray, A.A. and Moin, P. (1988), "Eddies, stream, and convergence zones in turbulent flows", Center for Turbulence Research Report CTR-S88, Center for Turbulence Research, Stanford, CA.

Iyengar, M., Schmidt, R., Sharma, A., McVicker, G., Shrivastava, S., Sri-Jayantha, S., Amemiya, Y., Dang, H., Chainer, T. and Sammakia, B. (2005), "Thermal characterization of non-raised floor air cooled data centers using numerical modeling", *Proceedings of the ASME InterPACK05, San Francisco, CA*, ASME, IPACK2005-73387.

Jin, R., Chen, W. and Simpson, T. (2001), "Comparative studies of metamodeling techniques under multiple modeling criteria", *Structural and Multidisciplinary Optimization*, Vol. 23, pp. 1-13.

Joseph, V.R., Hung, Y. and Sudjianto, A. (2008), "Blind kriging: a new method for developing metamodels", *Journal of Mechanical Design*, Vol. 130, p. 031102 (8pp.).

Kang, S., Schmidt, R., Kelkar, K.M., Radmehr, A. and Patankar, S.V. (2000), "A methodology for the design of perforated tiles in a raised floor data center using computational flow analysis", *Proceedings of the ITHERM 2000 – Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, *Las Vegas, NV*.

Karki, K.C., Radmehr, A. and Patankar, S.V. (2003), "Use of computational fluid dynamics for calculating flow rates through perforated tiles in raised-floor data centers", *HVAC and R Research*, Vol. 9 No. 2, pp. 153-66.

Karlsson, M., Karamanolis, C. and Zhu, X. (2004), "Triage: performance isolation and differentiation for storage systems", *Proceedings of the Twelfth International Workshop on Quality of Service*, pp. 67-74.

Kim, T., Thürey, N., James, D. and Gross, M. (2008), "Wavelet turbulence for fluid simulation", *Proceedings of the International Conference on Computer Graphics and Interactive Techniques Archive (ACM SIGGRAPH)*, No. 502008.

Lawrence Berkeley National Laboratory and Rumsey Engineers (2003), "Data center energy benchmarking case study: facility 8", available at: http://hightech.lbl.gov/DCtraining/reading-room.html

Ly, H.V. and Tran, H.T. (2001), "Modeling and control of physical processes using proper orthogonal decomposition", *Mathematical and Computer Modeling*, Vol. 33, pp. 223-36.

Moore, J., Chase, J., Farkas, K. and Ranganathan, P. (2005a), "Data center workload monitoring, analysis, and emulation", *Proceedings of the Eighth Workshop on Computer Architecture Evaluation Using Commercial Workloads*.

Moore, J., Chase, J., Ranganathan, P. and Sharma, R. (2005b), "Making scheduling cool: temperature-aware workload placement in data centers", *Proceedings of the Usenix Technical Conference*.

Nathuji, R., Somani, A., Schwan, K. and Joshi, Y. (2008), "CoolIT: coordinating facility and IT management for efficient datacenters", *Proceedings of the HotPower'08; Workshop on Power Aware Computing and Systems*, *San Diego, CA*.

Nie, Q. and Joshi, Y. (2008a), "Multiscale thermal modeling methodology for thermoelectrically cooled electronic cabinets", *Numerical Heat Transfer, Part A: Applications*, Vol. 53 No. 3, pp. 225-48.

Nie, Q. and Joshi, Y. (2008b), "Reduced order modeling and experimental validation of steady turbulent convection in connected domains", *International Journal of Heat and Mass Transfer*, Vol. 51 Nos 25/26, pp. 6063-76.

Pacheco, J.E., Amon, C.H. and Finger, S. (2003), "Bayesian surrogates applied to conceptual stages of the engineering design process", *ASME Journal of Mechanical Design*, Vol. 125, pp. 664-72.

Park, H.M. and Cho, D.H. (1996a), "The use of the Karhunen-Loeve decomposition for the modeling of distributed parameter systems", *Chemical Engineering Science*, Vol. 51 No. 1, pp. 81-98.

Park, H.M. and Cho, D.H. (1996b), "Low dimensional modeling of flow reactors", *International Journal of Heat and Mass Transfer*, Vol. 39 No. 16, pp. 3311-23.

Park, H.M. and Li, W.J. (2002), "Boundary optimal control of natural convection by means of mode reduction", *Journal of Dynamic Systems Measurement and Control*, Vol. 124, pp. 47-54.

Patel, C.D., Sharma, R., Bash, C. and Beitelmal, M. (2002), "Thermal considerations in cooling of large scale high compute density data centers", *Proceedings of the ITHERM 2002 – Eight Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, *San Diego, CA*, pp. 767-76.

Patel, C., Bash, C., Belady, C., Stahl, L. and Sullivan, D. (2001), "Computational fluid dynamics modeling of high compute density data centers to assure system inlet air specifications", *Proceedings of the ASME IPACK'01*, *Kauai, HI*, ASME, IPACK2001-15622.

Qian, P.Z.G. and Wu, C.F.J. (2008), "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments", *Technometrics*, Vol. 50 No. 2, pp. 192-204.

Qian, Z. (2006), "Computer experiments: design, modeling, and integration", PhD thesis, Georgia Institute of Technology, Atlanta, GA.

Qian, Z., Seepersad, C.C., Joseph, V.R., Allen, J.K. and Wu, C.F.J. (2006), "Building surrogate models based on detailed and approximate simulations", *ASME Journal of Mechanical Design*, Vol. 128, pp. 668-77.

Radmehr, A., Schmidt, R., Karki, K.C. and Patankar, S.V. (2005), "Distributed leakage flow in raised-floor data centers", *Proceedings of the ASME InterPACK*, *San Francisco, CA*, ASME, IPACK2005-73273.

Rambo, J. and Joshi, Y. (2003a), "Multi-scale modeling of high power density data centers", *Proceedings of ASME InterPACK03*, *Kauai, HI*, ASME, InterPack2003-35297.

Rambo, J. and Joshi, Y. (2003b), "Physical models in data center airflow simulations", *Proceedings of the IMECE-03 – ASME International Mechanical Engineering Congress and R&D Exposition*, *Washington, DC*, IMECE03-41381.

Rambo, J. and Joshi, Y. (2006), "Thermal modeling of technology infrastructure facilities: a case study of data centers", in Minkowycz, W.J., Sparrow, E.M. and Murthy, J.Y. (Eds), *The Handbook of Numerical Heat Transfer*, Taylor and Francis, New York, NY, pp. 821-50.

Rambo, J. and Joshi, Y. (2007a), "Modeling of data center airflow and heat transfer: state of the art and future trends", *Distributed and Parallel Databases*, Vol. 21, pp. 193-225.

Rambo, J. and Joshi, Y. (2007b), "Reduced-order modeling of turbulent forced convection with parametric conditions", *International Journal of Heat and Mass Transfer*, Vol. 50 Nos 3/4, pp. 539-51.

Rambo, J.D. (2006), "Reduced-order modeling of multiscale turbulent convection: application to data center thermal management", PhD thesis, Georgia Institute of Technology, Atlanta, GA.

Ravindran, S.S. (2002), "Adaptive reduced-order controllers for a thermal flow using proper orthogonal decomposition", *SIAM Journal of Scientific Computing*, Vol. 23 No. 6, pp. 1924-42.

Rolander, N. (2005), "An approach for the robust design of air cooled data center server cabinets", MS thesis, Georgia Institute of Technology, Atlanta, GA.

Rolander, N., Rambo, J., Joshi, Y., Mistree, F. and Allen, J.K. (2006), "Robust design of turbulent convective systems using the proper orthogonal decomposition", *Journal of Mechanical Design: Special Issue Robust and Risk Based Design*, Vol. 128, pp. 844-55.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989), "Design and analysis of computer experiments", *Statistical Science*, Vol. 4, pp. 409-23.

Samadiani, E. and Joshi, Y. (2010a), "Multi-parameter model reduction in multi-scale convective systems", *International Journal of Heat and Mass Transfer*, Vol. 53, pp. 2193-205.

Samadiani, E. and Joshi, Y. (2010b), "Proper orthogonal decomposition for reduced order thermal modeling of air cooled data centers", *Journal of Heat Transfer*, Vol. 132, No. 7, in press.

Samadiani, E., Joshi, Y. and Mistree, F. (2007), "The thermal design of a next generation data center: a conceptual exposition", *Proceedings of the Thermal Issues in Emerging Technologies, ThETA 1, Cairo*, pp. 93-102.

Samadiani, E., Rambo, J. and Joshi, Y. (2009.), "Numerical modeling of perforated tile flow distribution in a raised-floor data center", *Proceedings of the ASME InterPACK'07, Vancouver, British Columbia*, IPACK2007-334282007.

Samadiani, E., Joshi, Y., Hamann, H., Iyengar, M.K., Kamalsy, S. and Lacey, J. (2009), "Reduced order thermal modeling of data centers via distributed sensor data", *Proceedings of the ASME InterPACK'09, San Francisco, CA*.

Santner, T.J., Williams, B.J. and Notz, W.I. (2003), *The Design and Analysis of Computer Experiments*, Springer, New York, NY.

Schmidt, R. and Cruz, E. (2003), "Cluster of high powered racks within a raised floor computer data center: effects of perforated tiles flow distribution on rack inlet air temperature", *Proceedings of the IMECE-03 – ASME International Mechanical Engineering Congress and R&D Exposition, Washington, DC*.

Schmit, R. and Iyengar, M. (2005), "Effect of data center layout on rack inlet air temperatures", *Proceedings of the ASME InterPACK'05, San Francisco, CA*, ASME, IPACK2005-73385.

Schmidt, R., Karki, K.C. and Patankar, S.V. (2004), "Raised-floor data center: perforated tile flow rates for various tile layouts", *Proceedings of the ITHERM 2004 – Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, Las Vegas, NV*.

Schmidt, R., Karki, K.C., Kelkar, K.M., Radmehr, A. and Patankar, S.V. (2001), "Measurements and predictions of the flow distribution through perforated tiles in raised floor data centers", *Proceedings of the ASME InterPACK'01, Kauai, HI*, IPACK2001-15728.

Shah, A., Carey, V., Bash, C. and Patel, C. (2005a), "Exergy-based optimization strategies for multi-component data center thermal management: Part I, analysis", *Proceedings of the ASME InterPACK'05, San Francisco, CA*, ASME, IPACK2005-73137.

Shah, A., Carey, V., Bash, C. and Patel, C. (2005b), "Exergy-based optimization strategies for multi-component data center thermal management: Part II, application and validation", *Proceedings of the ASME InterPACK'05, San Francisco, CA*, ASME, IPACK2005-73138.

Shapiro, B. (2002), "Creating reduced-order models for electronic systems: an overview and suggested use of existing model reduction and experimental system identification tools", in Joshi, Y. and Garimella, S.V. (Eds), *Thermal Challenges in Next Generation Electronics Systems*, Millpress, Rotterdam, pp. 299-306.

Sharma, R.K., Bash, C.E. and Patel, C.D. (2002), "Dimensionless parameters for the evaluation of thermal design and performance of large-scale data centers", *Proceedings of the 8th AIAA/ASME Joint Thermophysics and Heat Transfer Conference, St. Louis*, AIAA-2002-3091.

Sharma, R.K., Bash, C., Patel, C.D., Friedrich, R.J. and Chase, J.S. (2003), "Balance of power: dynamic thermal management for internet data centers", Whitepaper issued by Hewlet Packard Laboratories, Tech. Rept. HPL-2003-5.

Shrivastava, S., Sammakia, B., Schmidt, R. and Iyengar, M. (2005), "Comparative analysis of different data center airflow management configurations", *Proceedings of the ASME InterPACK'05*, *San Francisco, CA*, ASME, IPACK2005-73234.

Simpson, T., Peplinski, J., Koch, P. and Allen, J. (2001), "Metamodels for computer-based engineering design: survey and recommendations", *Engineering with Computers*, Vol. 17, pp. 129-50.

Sirovich, L. and Park, H.M. (1990a), "Turbulent thermal convection in a finite domain: part I. Theory", *Physics of Fluids*, Vol. 2 No. 9, pp. 1649-57.

Sirovich, L. and Park, H.M. (1990b), "Turbulent thermal convection in a finite domain: part II. Numerical results", *Physics of Fluids*, Vol. 2 No. 9, pp. 1649-57.

Somani, A. and Joshi, Y. (2008), "Ambient intelligence based load management", *USPTO Provisional Patent, GTRC ID No. 4524*.

Strang, G. (1988), *Linear Algebra and Its Applications*, 3rd ed., Saunders, Philadelphia, PA.

Tang, Q., Mukherjee, T., Gupta, S.K.S. and Cayton, P. (2006), "Sensor-based fast thermal evaluation model for energy-efficient high-performance datacenters", *Proceedings of the Fourth International Conference on Intelligent Sensing and Information Processing (ICISIP)*, pp. 203-8.

Tarman, I.H. and Sirovich, L. (1998), "Extensions of Karhunen-Loeve based approximations of complicated phenomena", *Computer Methods in Applied Mechanics and Engineering*, Vol. 155, pp. 359-68.

US Environmental Protection Agency (2007), *Report to Congress on Server and Data Center Energy Efficiency*, Public Law 109-431, ENERGY STAR Program, US Environmental Protection Agency, Washington, DC.

VanGilder, J.W. and Schmidt, R. (2005), "Airflow uniformity through perforated tiles in a raised-floor data center", *Proceedings of the ASME InterPACK*, *San Francisco, CA*, ASME, IPACK2005-73375.

Wackernagel, H. (2002), *Multivariate Geostatistics*, Springer, New York, NY.

Zhou, J., Adrian, R.J., Balachandar, S. and Kendall, T.M. (1999), "Mechanisms for generating coherent packets of hairpin vortices in channel flow", *Journal of Fluid Mechanics*, Vol. 387, pp. 353-96.

Corresponding author
Emad Samadiani can be contacted at: samadianigatech.edu