Conference Papers

Dublin Energy Lab

2017

A Data Centre Air Flow Model for Predicting Computer Server Inlet Temperatures

Raymond Lloyd
IBM Research Dublin, Ireland, raylloyd@ie.ibm.com

Jer Hayes
IBM Research Dublin, Ireland, hayesjer@ie.ibm.ie

Marek Rebow
Technological University Dublin, marek.rebow@tudublin.ie

See next page for additional authors

Authors

Raymond Lloyd, Jer Hayes, Marek Rebow, and Brian Norton

A Data Centre Air Flow Model for predicting computer server inlet temperatures

Raymond Lloyd, Jer Hayes, Marek Rebow*, Brian Norton*

IBM, IBM Research Dublin, Ireland

*Dublin Institute of Technology, Dublin Energy Lab, Dublin, Ireland

Email: raylloyd@ie.ibm.com, hayesjer@ie.ibm.com, marek.rebow@dit.ie, president@dit.ie

ABSTRACT

Data centres account for approx. 1.3% of the world's electricity consumption, of which up to 50% of that power is dedicated to keeping the actual equipment cool. This represents a huge opportunity to reduce data centre energy consumption by tackling the cooling system operations with a focus on thermal management. This work presents a novel Data Centre Air Flow Model (DCAM) for temperature prediction of server inlet temperatures. The model is a physics-based model under-pinned by turbulent jet theory allowing a reduction in the solution domain size by using only local boundary conditions in front of the servers. Current physics-based modeling approaches require a solution domain of the entire data centre room which is expensive in terms of computation even if a small change occurs in a localised area. By limiting the solution domain and boundary conditions to a local level, the model focuses on the airflow mixing that affects temperatures while also simplifying the related computations. The DCAM model does not have the usual complexities of numerical computations, dependencies on computational grid size, meshing or the need to solve a full domain solution. The input boundary conditions required for the model can be supplied by the Building Management System (BMS), Power Distribution Units (PDU), sensors, or output from other modeling environments that only need updating when significant changes occur. Preliminary results validated on a real world data centre yield an overall prediction error of $1.2°C$ RMSE. The model can perform in real-time, giving way to applications for real-time monitoring, as input to optimise control of air conditioning units, and can complement sensor networks.

I. INTRODUCTION

Data centres are major consumers of electricity where according to the Uptime Institute's most recent survey [1], the average worldwide Power Utilisation Effectiveness (PUE) remained around 1.7 in recent years. PUE is the data centre total power / IT power and a value of 1.7 means over 40% of data centre energy is consumed by the Heating, Ventilating, and Air Conditioning (HVAC) systems. Ultimately, the goal is to maximise performance of Information and Communications Technology (ICT) equipment while minimising overall power consumption. To achieve this goal, new business practices and technologies are continuously being developed. Moreover, with the steep rise in energy costs, energy consumption has become a critical issue in all reaches of the world, with electricity use in data centres being about 1.3% worldwide in 2010, and 2% for the US, burning 198 Billion Kilowatts per year and it's expected to increase year on year [2]. The motivation for the model outlined in this paper stems from work performed in the area of thermal management and energy efficiency of data centres where the need for real-time temperature modeling often arose. Real-time monitoring of inlet temperatures of server equipment can help ensure energy efficient operations are continued, monitored and maintained.

The measuring of temperatures in data centres can of course be completed directly by placing wired or wireless sensors in appropriate locations. Measurement-based solutions give real-time sensor readings but generally have sensors sparsely placed within the data centre due to the installation and maintenance cost. Indirect modeling approaches of air flow and heat transfer provide alternatives to costly sensor deployments. Numerical methods such as Computational Fluid Dynamics (CFD) based on the Naiver-Stokes equations with heat transfer offer accurate representations of the physics but at a cost of computational complexity and long solution computation times. To simulate complex geometries, domain specific compact models are used to simplify the physics yielding significant improvements in computation time but inherently generate inaccuracies which are deemed acceptable. They are usually expensive and require trained personnel to operate. CFD models are not suitable for real-time applications where rapid solutions are required that may feed into a Computer Room Air Conditioner (CRAC) control system for example.

Reduced order physics models based on Potential Flow Theory (PFT) [3] offer faster solution times but at the expense of accuracy and still require a full domain solution. Accuracy was improved by the addition of buoyancy effects with Enhanced Potential Flow Models (EPFM) [4] but doubled the execution time. Another approach [5] [6] deals with buoyancy and recirculation physics by implementing a Rankine Vortex superposition with the core placed vertically at the top of the racks, halving the prediction error overall throughout the data centre room with significant improvement at server inlets but still require a full domain solution. Measurement-based physical modeling [7] [8] approaches which leverage real-time sensor data as the boundary conditions, offer a good compromise between sensor placement density and fitting the missing data with physics-based prediction.

Statistical solutions are very computationally efficient and model non-linearity very well but require training data which is not always to hand or easily obtainable for actual measurements, so simulated data is optionally used,e.g. Proper Orthogonal Decomposition (POD) [9][10]. Statistical models are susceptible to changes in physical conditions, e.g. layout change, and lose accuracy rapidly when they occur. They do provide rapid predictions which would be suitable for optimising CRAC control systems. However, in an ever changing data centre, the requirements for re-training may out-weigh the adoptability of this approach.

The need to obtain temperature predictions quickly is evident when we consider how fast conditions change, e.g. variations in perforated tile flow volume, air supply temperature or server power loads. The prediction data may be useful in many ways, from input into the HVAC control system, to alerting and reporting. Such a solution could be used standalone to provide temperature prediction, however it could also be used alongside or integrated with other modeling techniques and sensor readings to provide complete flow and temperature fields.

Our approach to reduce the complexity of obtaining temperature predictions is to reduce the solution domain. Solving local changes rather than the solving the complete solution domain would be far more efficient requiring only local boundary conditions. Thus we propose a model that will focus on the prediction of server inlet temperatures. The prediction method is fast enough to be real-time acceptable and accurate enough to be useful. The DCAM model is underpinned by turbulent jet theory but adapted for use in data centres.

We validate the model with real-world temperature data collected from a data centre which has undergone 25 step changes to the CRAC flow rates. The 25 scenarios provide a range of different input boundary conditions that test adaptability of the model to other data centres.

II. Assumptions and Limitations

The research is focused on data centres utilising a raised-floor cooling arrangement where the cold air is supplied by perforated floor tiles placed directly in front of the server racks. The model predicts inlet temperatures in front of servers and is limited to server racks which are not placed at the end of aisles. Power consumption by server racks are assumed to be uniformly distributed vertically from bottom to top of the rack. Possible increases in server fan speeds due to increased inlet temperatures are not currently accounted for.

III. Data Centre Airflow Model

We now describe the DCAM model where the goal of this physics-based approach is to limit the solution domain and boundary conditions to a local level. The local level is defined as the area in front of the server, the perforated tile supplying the cool air, and the server itself. This lends itself to the application of turbulent jet theory to model the air jet from the perforated tile. The model is adapted to take into account unique characteristics of data centre air flow such as the intake of air from the cooling jet by the servers.

Research carried out on plume and jet theory models show that the velocity flow field of the jet after the initial formation conditions express self-similarity characteristics when looking at the cross section of the jet. This is also true of the concentration field which is the mixing of the jet fluid and the ambient fluid together and is defined by the concentration of jet fluid in the mix. This characteristic is leveraged in the design of the DCAM to simplify computation.

The study of jets and plumes has a long history [11] and it has been applied to studies of industrial chimney stacks, volcanic eruptions and underwater plume activity on the sea bed and more recently to building ventilation systems. Raised floor or Under Floor Air Distribution (UFAD) was first introduced in the 1950's to cool computer rooms. More recently UFAD is emerging as leading ventilation design for modern office buildings. As such, much work has been carried out to model UFAD in the office environment in terms of providing user comfort. Of interest, plume theory modeling approaches have been used to explore the impact of UFAD on room temperature stratification by means of buoyant heat plume and fountain jet model for the inertial diffuser flow [12], [13], [14] [15]. These approaches model steady state layered stratification of displacement ventilation in a room. The effect of different configurations on heat sources and cooling sources have been studied theoretically and experimentally.

In terms of applicability to a data centre, the lower zone of the UFAD models do not represent the dynamic airflow activity between the supply air, server consumption and exhaust of heated air very well. Firstly, the inertial forces are far greater than the buoyancy influences (Richardardson number Ri in the order of 0.1). Secondly, the models are zone-based assuming a well mixed zone but in fact, in data centres it is the mixing within a zone which is of interest and this is highly variable. The upper layer, which is at the ceiling height of a data centre is relatively well mixed, as the warm air travels back to the CRAC units for cooling and recirculation. In free cooling systems, displacement ventilation occurs as the heated air is displaced through ceiling vents.

More appropriately representing the airflow from a data centre's perforated tile are studies into turbulent jets since the air exiting the perforated tiles exhibit a high Reynolds number in the region of 27,000 (Eqn. 1 and Eqn. 2 for a 25% open perforated tile at $0.142m^3$/s). Importantly, the high inertial forces generated by the pressurised under floor plenum lend themselves to fit the model of a turbulent rather than a buoyant plume.

$$Re = \frac{\rho v D_H}{\mu} \qquad (1)$$

where ρ is the density of air, μ is viscosity, v velocity, D_H is the characteristic length.

$$Re = \frac{1.204 * 3.1 * .142}{1.983 * 10^{-5}} = 26,727 \qquad (2)$$

Fig. 1 shows a round free turbulent jet. The air stream exits the orifice along the jet axis y and spreads as the jets moves away from the orifice. This angle of spread has been empirically shown as approximately 11.8^o from the y axis, with little effect from the fluid type (water, air), orifice diameter, shape or discharge velocity [16][17][18]. The intersection of the spread lines under the orifice reveals the virtual source of the jet where $y = 0$. The relationship between the increase in radial distance r of the spread to the distance along the y axis is $\tan(11.8) \approx 1/5$.

$$R(y) = \frac{1}{5}y \qquad (3)$$

The distance from the orifice to the virtual source is $\frac{d}{2} = \frac{1}{5}y$ therefore

$$y = \frac{5d}{2} \qquad (4)$$

The distribution of velocity of the air jet exhibits a Gaussian bell curve shape as the flow develops along y. Initially on exit of the orifice, a top hat profile of velocity can be seen with an average exit velocity of U_0. The transition from a top hat to a Gaussian shape has been studied experimentally [19]. The experimental measurements reveal that the Zone of Flow Establishment (ZFE) ends at $\approx 6.2d$ where d is the diameter of the orifice. Up to the Zone of Established Flow (ZEF) the change of maximum centreline velocity u_{max} is minimal and self similarity begins as the Gaussian profile now characterises the velocity distribution across the jet and u_{max} starts to reduce almost linearly. Fig. 1 shows the transition from ZFE to ZEF. The transition to full Gaussian occurs when the eddies from the side ambient entrainment reach the flow axis of the jet y.

The Gaussian shape or bell curve shape of the velocity profile across the jet cross-section can be represented by;

$$u(r) = u_{max} \exp\left(\frac{r^2}{2\sigma^2}\right) \qquad (5)$$

where u is the velocity, r is the radial distance from the centreline and σ is the standard deviation. Since we know that 4σ's make up approximately 95% of the diameter of the jet which is $2r$, from Eqn. 3 we can say $\sigma = \frac{y}{10}$ therefore we can write

$$\sigma = \frac{y}{10} \rightarrow \frac{1}{2\sigma^2} = \frac{50}{y^2} \rightarrow u(y,r) = u_{max} \exp\left(-\frac{50r^2}{y^2}\right) \qquad (6)$$

The curve characteristic studies show the self similarity of the velocity profile as the flow develops away from the source of flow [17] [20]. Self similarity, from a modeling perspective helps with the simplification of the data centre
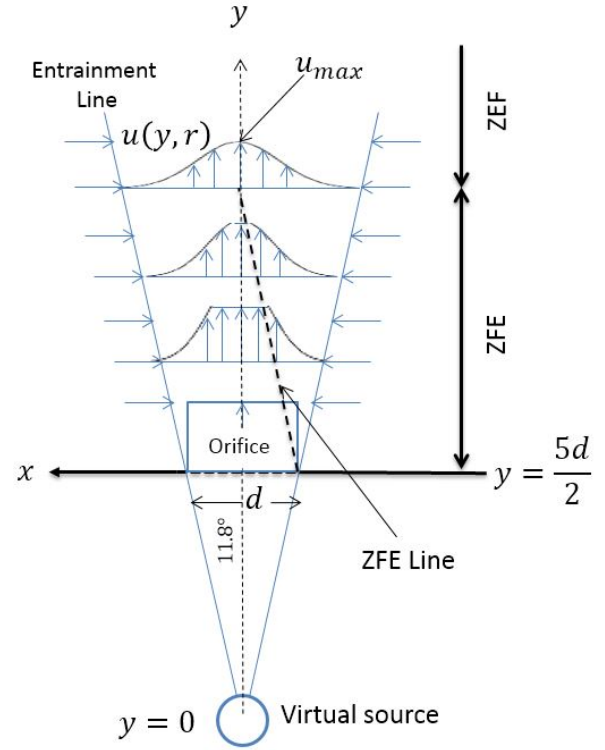

Fig. 1. Round Free Turbulent Jet with Zone of Flow Established Flow and Zone of Established Flow.

model due to the deterministic behavior of the flow with respect to distance from the source, initial velocity, and size of the orifice. This eliminates the complications of grid sizing, iterative processes and convergence as seen in numerical models. Another important observation from the jet flow is that velocity flow distribution across the jet is proportional to the concentration between the jet fluid and ambient fluid. This correlation between velocity and concentration has been found empirically[21].

Importantly, the mixing concentrations of the ambient fluid and the jet fluid represents the mixing of the hot exhausted air from the servers recirculating into the cold aisles and the cooling air supplied by the perforated tile. The inlet temperature is a function of the concentration proportions between the two i.e. $t = t(C_1, C_2)$ where C_1, C_2 are the jet and ambient fluid concentrations.

To determine the velocity of the centre line we consider the jet entering a body of fluid with the absence of external acceleration/deceleration. In this case the momentum flux in the jets cross-section remains constant, that is:

$$M = \int_0^\infty \rho u^2 2\pi dr = \rho U^2 \frac{\pi d^2}{4} \qquad (7)$$

where the momentum flux is $\rho u \times u$ with ρ the density and u the velocity accumulated over the jet cross-section with respect to the radius r. U and d are average exit velocity taken at the

orifice and the diameter. Therefore the max velocity along the jet axis can be expressed as

$$u_{max} = \frac{5d}{y}U \tag{8}$$

In the proceeding text, a theoretical two dimensional simplified model is proposed to model the behavior of the jet by placing a boundary wall consisting of a rack of server equipment to the side of the jet. The perforated tile airflow is modeled as a two dimensional round free turbulent jet Fig. 2. The area of the perforated tile is a 600mm square shape with an effective open area determined by the tile perforation percentage or damper setting (adjustable perforations).

In the DCAM, the velocity field is calculated on the premise of constant momentum flux, therefore we can use a concentrated jet placed in front of the rack with effective tile opening of A_{eff} where δ_{perf} is the percentage of perforated tile opening.

$$A_{eff} = \delta perf A_{tile} \tag{9}$$

The effective two dimensional cross-section value of d is therefore

$$d = \frac{A_{eff}}{2} \tag{10}$$

Typically, in a data centre, the perforated tiles are positioned directly in front of the server equipment. The jet is prohibited from expanding on one side with the server equipment restricting the flow of air from the perforated tile. The symmetrical velocity profile of the jet is now disturbed with a rack of servers on one side. The rack of servers prevents entrainment on this side, therefore we assume that mixing from entrainment on the server side of the flow axis y does not take place. The transition from top hat to Gaussian is schematically represented as a central flat spot on the top of the curve where u_{max} is constant as the flow progresses along y until the curve shape is reached. The simplified model assumes a smooth transition between ZFE and ZEF in terms of top hat to Gaussian profile.

The volumetric flux Q increases, due to entrainment of surrounding ambient air E as the jet travels along the jet axis y. The entrainment is the rate at which the volumetric flux grows and can be represented as

$$E = \frac{dQ}{dy} \tag{11}$$

where the volumetric flux Q is

$$Q = \int_0^\infty u 2\pi r dr = \frac{\pi}{50} u_{max} y^2 = \frac{\pi}{10} dU y \tag{12}$$

so E becomes

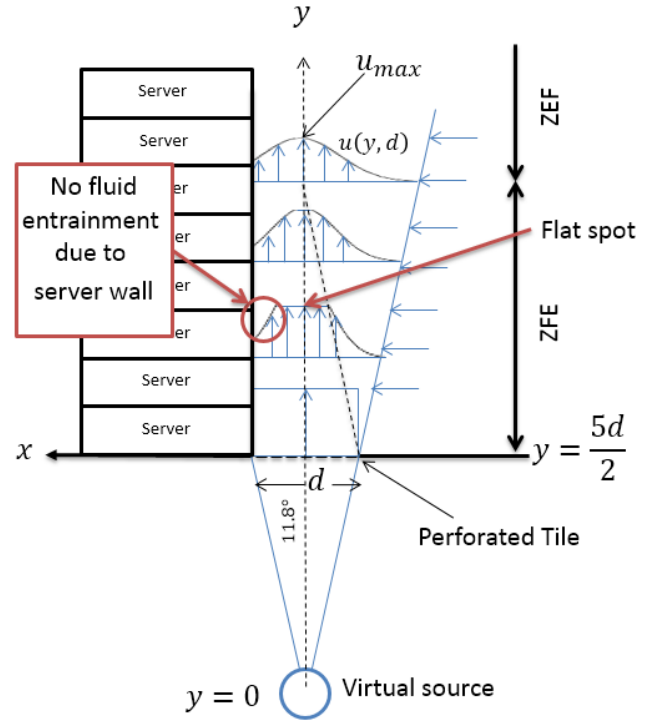$$E = \frac{dQ}{dy} = \frac{\pi dU}{10} \tag{13}$$

Fig. 2. Round Free Turbulent Jet with a computer server rack placed to the side of the jet

the radial entrainment transverse velocity v carries the entrainment into the jet so that

$$dQ = vdA \tag{14}$$

with $dA = 2\pi Rdy$ the lateral area of the jet section. Substitution of dA and further substitution of R from Eqn. 3 gives

$$\frac{dQ}{dy} = 2\pi Rv = \frac{2\pi yv}{5} \tag{15}$$

Equating this to the previous dQ/dy yields the value for transverse velocity of entrainment v in terms of average jet velocity $\bar{u}$

$$v = \frac{Ud}{4y} = \frac{u_{max}}{20} = 0.10\bar{u} \tag{16}$$

Fig. 3 is a two dimensional schematic showing the first step in the adapted model. In this schematic, the servers are not operational, therefore do not consume any of the jet flow and the server boundary interface is ignored. The sampling of jet cross-section is represented by the horizontal line at the bottom of the curve. The sampling distance along y can be as granular as required. Here it is set at 305mm intervals to match the experimental data resolution and sampling location.

The volume of cool air discharged from the perforated tile is represented in terms of its velocity profile with the initial volume of cool air exiting the perforated tile is represented as a two dimensional rectangle A_{cool} i.e. a top hat on exit. As

the jet progresses along y the area of A_{cool} remains constant but the shape changes to that of a Gaussian shape.

$$\frac{\partial A_{cool}}{\partial y} = 0 \qquad (17)$$

On the server side, the velocity profile is squared off as entrainment on that side is ignored.
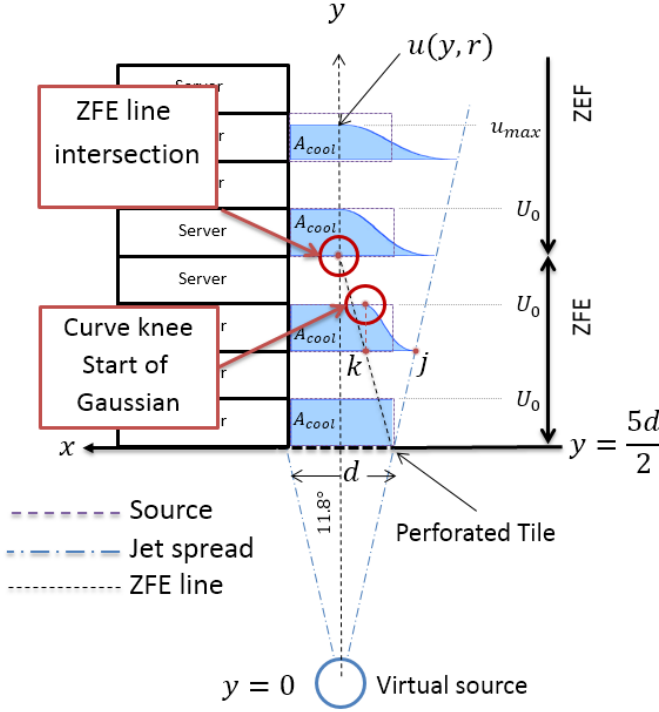

Fig. 3. Round Free Turbulent Jet - the velocity profile is simplified. A ZFE line is drawn from the edge of the perforated tile flow at U_0 to the centre of the curve at u_{max} at the point along y where the ZEF begins. The intersection of the ZFE line with the any of the sample lines determines the knee of the curve (start of the Gaussian curve).

To determine the shape of the curve along y in the ZFE, a line (ZFE line) is drawn from the edge of the perforated tile to the end of the ZFE where it intersects the y axis. Where this ZFE line intersects any horizontal sampling line at point k (within the ZFE) , this is represented as the "knee" of the curve, i.e. the start of the Gaussian curve, which is the point at which entrainment of ambient air has penetrated into the jet. The tail of the Gaussian curve is determined by the intersection of the entrainment line with the horizontal sampling line, point j. Recall that the slope of the entrainment line is a universal angle of 11.8^o so $y = 1/5x$ and that the start of the ZEF is $6.2d$, therefore points k and j can be calculated as follows

$$k = d - \frac{y}{12.4d} \qquad (18)$$

$$j = d + 5y \qquad (19)$$

where d is the width of the effective perforated tile.

In Fig 3, to the left of the Gaussian curve, the curve is flattened off. This represents the flat spot u_{max} the max velocity of the jet. This value of u_{max} is constant up to the ZFE where $u_{max} = U_0$. After this point the value of u_{max} can be determined by Eqn. 8.

After the ZFE, in the ZEF, the knee of the curve remains on the jet flow axis y and follows self similarity for values for $y > 6.2d$ the start of the ZFE. The DCAM model leverages the velocity/concentration profiles to determine mixing ratio's between cooling supply air from the perforated tile and warmer ambient air in the data centre.

The ratio's are captured as a two dimensional representation between the area of cooling air to area of ambient air, which of course represents the volume of cooling and volume of ambient air respectively.

The model requires scaling from units of length which determine the jet characteristics (diameter d in mm and height along the y axis in mm) to units of volumetric flow (m^3/s) on the x-axis and a dimensionless value on the y-axis $\frac{u_{max}}{U_0}$ for each calculation of mixing ratios. The y axis is conveniently scaled to this dimensionless unit whereby the areas representing the velocity profiles in the model can be calibrated directly to the volumetric flow rate. At jet exit from the perforated tile the top hat profile has vertical height of $\frac{u_{max}}{U_0} = \frac{U_0}{U_0} = 1$ so the x-axis can be scaled to the exit volumetric flow rate in m^3/s. The initial unit height is denoted as h. As we sample further up the server rack, both the shape of the area and the vertical height change. Where the servers are non-operational, the area of cooling air supply remains constant, only the shape of the area changes.

The next step is to model the server consumption at the inlets of the servers. The server consumption is depicted in Fig. 4 as a red vector V_s, with the volume of air consumed depicted are the red rectangle $A_{V_s} = V_s * h$. The server consumption rectangle has its rectangle height scaled to the unit height $h = 1$ and the length of which is scaled to the volumetric flow rate demanded by the server fans V_s (see legend in Fig. 4). The available cooling air is shown as $Acool$.

The server air demand V_s is positioned horizontally on y axis at the server inlet layer l_i. The sampling mixture area is depicted as the dashed rectangle A_{cons} and is equal to the server consumption A_{V_s}. It is drawn from the position of $x = V_S$ away from the server inlet (where $x = 0$ at the servers). The area A_{cons} is removed from the jet stream. The mixing proportions of cooling air to ambient air are determined by the areas bounded by A_{cons}.

As sampling location moves up the server rack along the y axis to the next server l_{i+1}, the proportion of cooling air consumed by the server in the previous step l_i is deducted from the available supply air A_{cool} and is denoted as $A_{coolcons}$, shown as "Reduction due to previous Server cool air demand" on the schematic in orange 4. This is effectively the remaining silhouette of A_{cool}, if the servers were non-operational.

The process continues as the sample location moves up the y-axis from server to server as in Fig. 4. In our case the
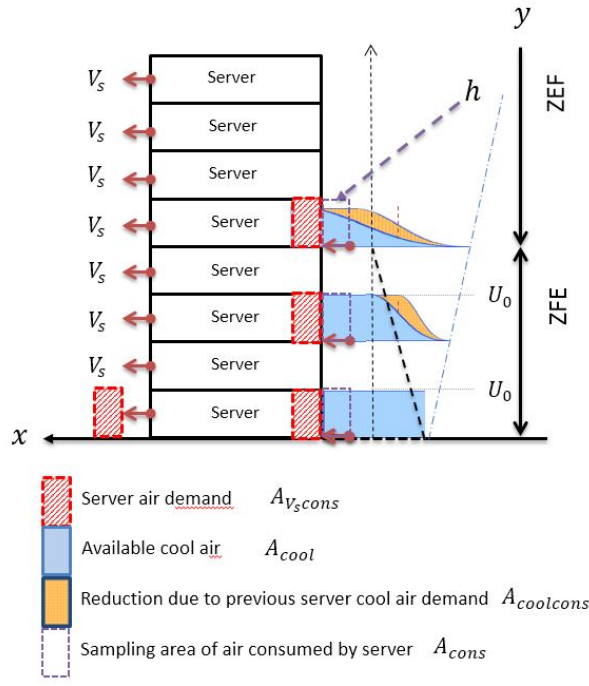
Fig. 4. Adapted Model depicting the demand of the computer servers on the jet as h increases

Fig. 5. Mixing concentrations of cooling supply air and hot ambient air

sampling location is every 305mm starting at l_0 =152mm to l_7 =2286mm from the perforated tile to match the "layers" in the validation experimental data Section IV-B1. This sampling location could in fact be varied to suit the configuration of the servers installed in the rack. In Fig. 4 we can see the remaining A_{cool} reduces as the servers consume more and more of the original supply air from the perforated tile.

$$A_{cool}^{l_i} = A_{cool}^{l_0} - \sum_0^i A_{coolcons}^{l_i} \qquad (20)$$

Calculation of the temperature at the inlet T_{inlet} is a function of the volumetric mixing of air temperatures between the supply air T_{cool} and the surrounding ambient temperature T_{amb} contained in the sampling area $A_{cons}^{l_n}$. This can be seen in the area enclosed by A_{cons} in Fig. 5 which can be defined as;

$$T_{inlet}^{l_i} = \frac{T_{cool}*(A_{cons}^{l_i} \cap A_{cool}^{l_i}) + T_{amb}*(A_{cons}^{l_i} - (A_{cons}^{l_i} \cap A_{cool}^{l_i}))}{A_{cons}^{l_i}} \qquad (21)$$

where

$$A_{cons} = V_s * h \qquad (22)$$

IV. RESULTS

Validation of the DCAM model is performed with real-life data, however the model was first verified against a simplified CFD simulation to verify the the software implementation of the algorithm was correct.
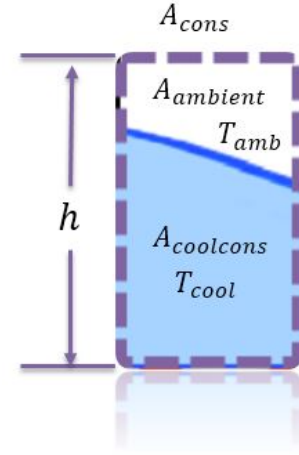
A. Comparison with simulation

The model is first verified using CFD software TileFlow [22]. The purpose of this testing is to ensure the model behaves and captures the implementation of the physics in a controlled simulated experiment with a well trusted simulation package specifically developed and optimised for data centres. In a live data centre, even under experimental conditions, some parameters are difficult to set and control and are often off-limits, e.g., server loads which in turn cause changes in power consumption. The CFD model is created to represent a simple data centre comprising of a single row of seven identical server racks with two CRAC units placed at each end and aligned with the hot aisles. The simulation results are taken from the centre server rack in the middle of the row of racks. Three simulations were created to model different degrees of recirculation from the top of the rack to the centre of the rack using different parameter settings for rack power and CRAC volumetric flow rate and temperature. The rack power is uniformly distributed from top to bottom of the rack.

The input parameters to the DCAM model are shown in Table. I. The parameters, which are local boundary conditions, are extracted from the CFD model. The T Ambient value is taken from the temperature at the top centre of the rack.

Boundary Conditions				
Test Case	**Rack Pwr**	**Tile Flow**	**Tile Temp**	**T Ambient**
#	kW	cfm	°C	°C
1	5	612	12.8	23.6
2	6	612	12.8	32.8
3	5	460	12.8	34.6

TABLE I
INPUT PARAMETERS FOR DCAM MODEL, VALUES EXTRACTED FROM CFD SIMULATION

The reference temperatures are taken from the CFD simulations directly in front of the server rack. The temperature is sampled at 305mm intervals from between 152mm to 2286mm. Using the extracted local boundary conditions from

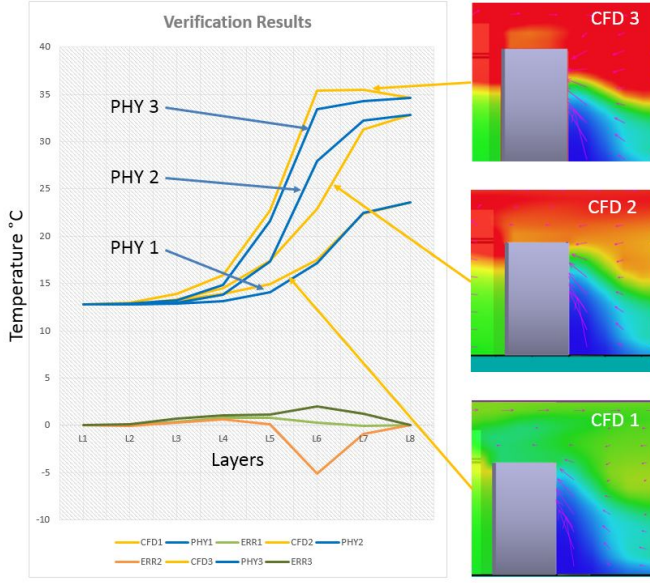the CFD simulation, the DCAM model is run and the results are shown in Fig. 6.

Fig. 6. Comparative verification results between DCAM model and CFD model in the simulated data centre

	L1	L2	L3	L4	L5	L6	L7	L8	
CFD1	12.8	12.8	13.2	13.9	14.9	17.5	22.4	23.6	
PHY1	12.8	12.81	12.86	13.1	14.1	17.19	22.45	23.6	
ERR1	0	-0.01	0.34	0.8	0.8	0.31	-0.05	0	0.432189
CFD2	12.8	12.8	13.3	14.5	17.4	22.9	31.3	32.8	
PHY2	12.8	12.83	12.99	13.84	17.3	27.98	32.24	32.8	
ERR2	0	-0.03	0.31	0.66	0.1	-5.08	-0.94	0	1.845014
CFD3	12.8	13	13.9	15.9	22.7	35.4	35.5	34.6	
PHY3	12.81	12.86	13.18	14.82	21.58	33.44	34.31	34.6	
ERR3	-0.01	0.14	0.72	1.08	1.12	1.96	1.19	0	1.013447
							Total RMSE Error		**1.096883**

TABLE II
MODEL VS SIMULATION: DCAM SHOWS REASONABLE AGREEMENT OF CFD RESULTS. HIGHLIGHTED ARE THE TWO MAX ERRORS.

Generally, the DCAM model agrees with the CFD model quite well except for one temperature location highlighted in Table II. The location of the error is at the inflection point of the curves, the most variable area of the inlet face of the servers. As both the DCAM and CFD models are simulations a definitive justification for the differences between the models can not be fully assessed, however the shapes and general profiles agree indicating that the DCAM model is executing the physics in the application as intended. To validate the model accuracy, real measured data is used and the results are presented below.

B. Validation via empirical data

1) Collecting Real World Empirical Data: Temperature data for validation was collected in a data centre using IBM Measurement and Management Technology (MMT) which was developed in IBM and is widely used in data centres around the world [23]. MMT uses a specially developed cart to collect temperatures from the data centre room. The cart consists of layers spaced vertically at 305mm intervals with

nine thermocouples per layer and one humidity sensor. The cart has motion encoding wheels and hardware to capture the data relative to the location in the data centre. The MMT cart fits on a standard 600mm tile and is traversed tile by tile throughout the data centre to digitise a high resolution three dimensional snapshot of the thermal environment.

Simultaneously, information is recorded from the CRAC units, power distribution units and perforated tiles. The data was gathered from a real operational data centre which is 102.2m^2 in size (25 x 11 tiles). The CRAC unit flow rates are varied with 25 different settings and the MMT scan conducted each time. The settings are shown in the in Table. III

The layout of the data centre is shown in Fig. 7. The racks are arranged in a hot aisle, cold aisle configuration with two CRAC units placed at each end of the data centre. The total IT power is 75KW. Racks A5 and C2 are fitted with rear door heat exchangers accounting for the removal of approx. 25KW of heat load. The data centre and measurements have been used previously in other modelling work, namely Measurement-based modelling [8] and Reduced-order modelling via Proper Orthogonal Decomposition [9] and thus has been held to be a valid data set by other researchers.

	CRAC1		CRAC2		
Case#	CFM/RPM Percentage (%)	Estimated CFM	CFM/RPM Percentage (%)	Estimated CFM	Estimated Total CFM
1	100	12400	0	0	12400
2	96	11904	0	0	11904
3	92	11408	0	0	11408
4	88	10912	0	0	10912
5	84	10416	0	0	10416
6	80	9920	0	0	9920
7	76	9424	0	0	9424
8	72	8928	0	0	8928
9	68	8432	0	0	8432
10	65	8060	0	0	8060
11	100	12400	100	12400	24800
12	85	10540	100	12400	22940
13	90	11160	90	11160	22320
14	100	12400	75	9300	21700
15	90	11160	80	9920	21080
16	70	8680	95	11780	20460
17	95	11780	65	8060	19840
18	80	9920	80	9920	19840
19	75	9300	80	9920	19220
20	65	8060	85	10540	18600
21	70	8680	75	9300	17980
22	70	8680	70	8680	17360
23	65	8060	70	8680	16740
24	70	8680	65	8060	16740
25	65	8060	65	8060	16120

TABLE III
FLOW RATES FOR CRAC 1 AND CRAC 2 FOR 25 SETTINGS

2) Validation: The validation of the model is completed using temperature data as gathered from the MMT cart outlined in Section IV-B1. The combined overall prediction error is measured by different metrics, Root Mean Square Error (RMSE) , Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) and results are show in Table IV. Overall the model preforms very well with an acceptable 1.20°C RMSE with standard deviation of 1.13°C and a mean absolute error of 0.86°C and standard deviation of 0.85°C.
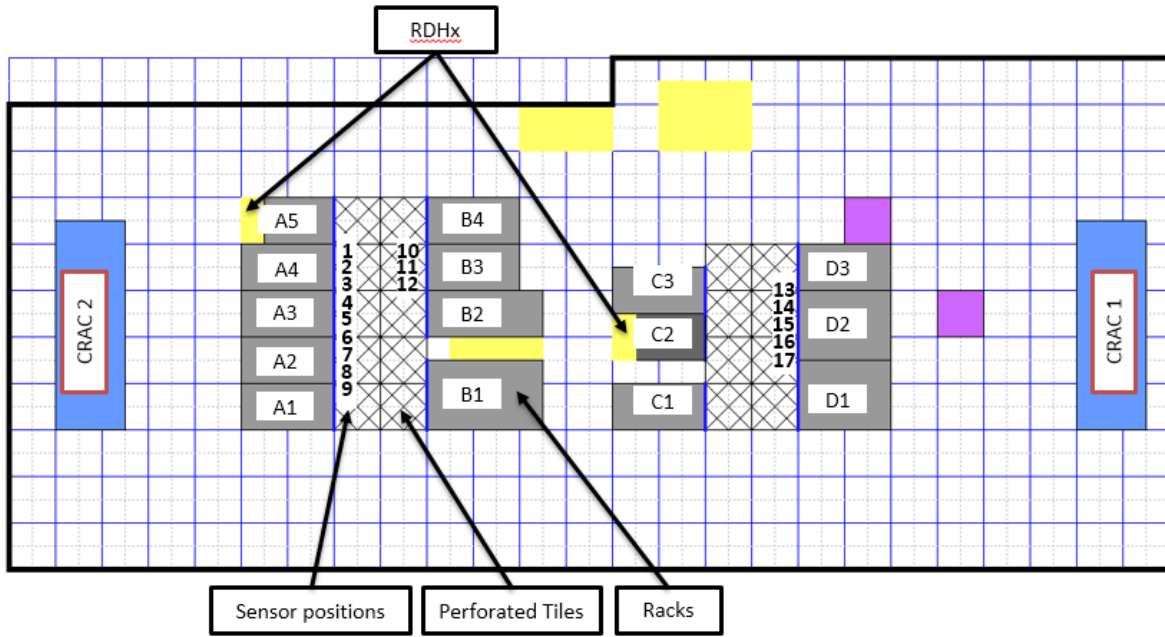
Fig. 7. Experimental Data Centre Layout shows locations of server racks, perforated tiles and CRAC units

This is comparable to results from CFD with an RMS error approximately 2°C [24] in a small controlled data centre. Reduced order methods such as measurement-based models [8] which have an inlet temperature RMSE ranging from 0.627°C to 2.044°C with the lower value requiring a quite large number of prescribed sensors nodes to supply the input boundary conditions, provide another example of error rates.

On a macro level, the error distribution at each layer for all sensors is shown in the chart Fig. 8. The chart plots the root squared error between each of the predicted and measured temperatures for each layer. We can see that maximum error for 50% of the temperature data for all layers is under 1.04°C with a range of approximately 0.60°C. This extends to approximately 1.68°C with a range of 1.17°C for 75% of the data and approximately 2.5°C for 90% of the temperature data. The extreme outliers occur beyond 95% of the data.

Plotting histograms in Fig. 9 we can see that layers 4,5,6,7 have greater variability and indeed standard deviation. This is as expected as the prediction of the inflection point in the curve is influenced by many factors in real data centres that are either not, difficult, or computationally prohibitive to capture in most modeling approaches. These factors include geometry of the server doors, individual server power loads (there can up to 42 individual servers in a single rack), layout of the data centre, recirculation, missing servers or filler panels etc. The experimental data only contains rack locations with no information on the servers contained therein, therefore a generalised uniform power distribution is implemented. Given the availability of such information, the algorithm makes provisions to vary the sampling height increments and distribute
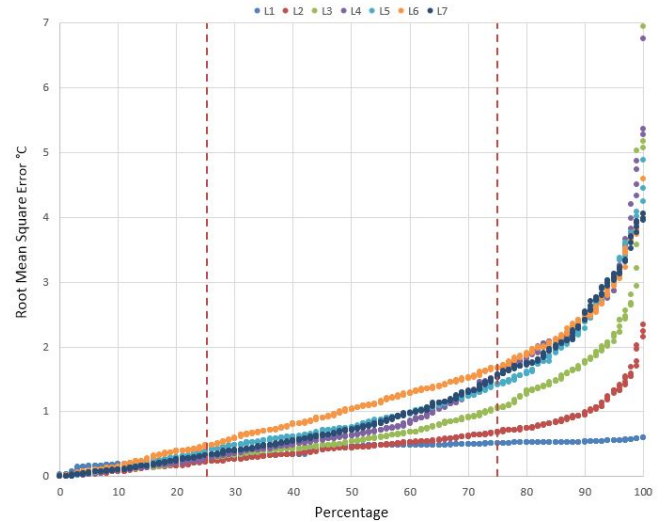

Fig. 8. The chart shows the root squared error of measured vs predicted temperatures for each layer. The x-axis is the percentage of the data and the y-axis is the RMSE.

the rack power to individual servers in the rack. Here the sampling height is dictated by the measurement data for direct comparison. The value at layer 8 is above the rack and used as the ceiling temperature, so is ignored.

Fig. 10 shows a color map representation of all results broken down over the 25 test cases. Each test case is a vertical column broken down into the layers. Within each case setting we predict 17 inlet temperature locations with 8 layers each. Each color in the column is the RMSE of all

RMSE and standard deviation on (predicted-measured)								
	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Overall**	**Range**
RMSE	0.42	0.64	1.12	1.46	1.40	1.50	1.40	1.20	-
Standard Deviation	0.14	0.51	0.92	1.37	1.28	1.38	1.40	1.13	-
Max Value	0.05	2.34	2.23	6.76	4.89	4.59	3.77	6.76	-
Min Value	-0.60	-1.39	-6.95	-4.74	-3.53	-3.63	-4.06	-6.95	-

MAE Statistics of SQRT((predicted-measured)^2)								
	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Overall**	**Range**
Mean Value	0.39	0.51	0.79	1.03	1.05	1.20	1.04	0.86	-
Standard deviation	0.14	0.39	0.80	1.04	0.93	0.91	0.94	0.85	-
Deviation Squared	8.58	65.46	271.27	459.60	362.94	347.99	373.70	-	
90% error under	0.54	0.98	1.78	2.49	2.39	2.46	2.53	1.88	1.99
75% error under	0.51	0.68	1.06	1.57	1.43	1.68	1.56	1.21	1.17
50% error under	0.47	0.44	0.54	0.63	0.76	1.04	0.72	0.66	0.60
Max Value	0.05	2.34	6.95	6.76	4.89	4.59	4.06	6.95	-

MAPE - Mean Absolute Percentage Error								
	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Overall**	**Range**
Mean Value	2.52	2.77	5.02	5.82	4.95	5.35	4.64	4.44	-
Standard deviation	1.18	2.08	5.63	5.74	4.18	3.87	4.20	4.31	-
Max Value	4.58	14.10	50.00	29.39	20.81	19.01	20.20	50.00	-

TABLE IV

MODEL VS EMPIRICAL DATA: STANDARD DEVIATION OF LAYERS L1-L7 OVER ALL 25 CASES.
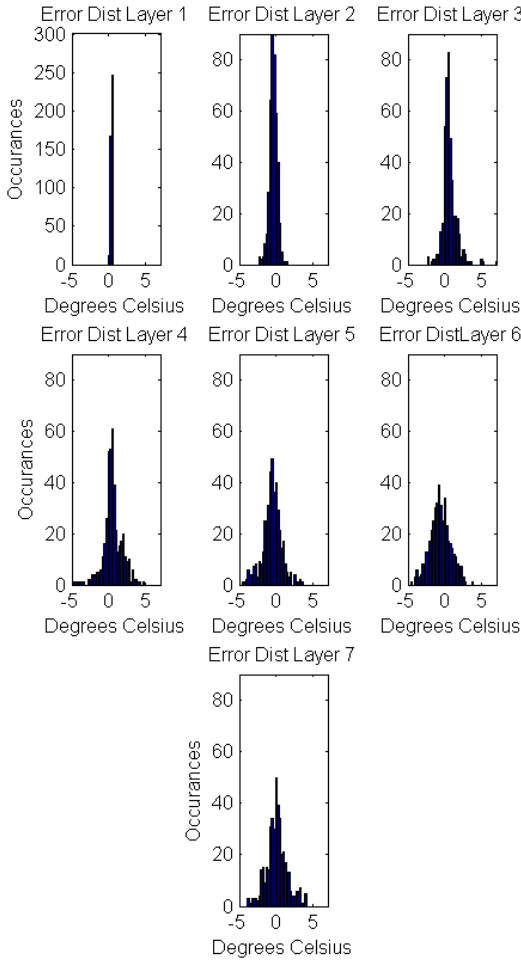

Fig. 9. Distribution of errors per layer binned by .25 degree c

deviates from the actual temperature.

Fig. 10. RMSE for all sensors per layer over the 25 case settings.

17 temperature predictions for that layer. As we can see, there is very good agreement between predicted values and actual values of temperature indicated by the blue color on the color map, however there is an area where the predicted temperature
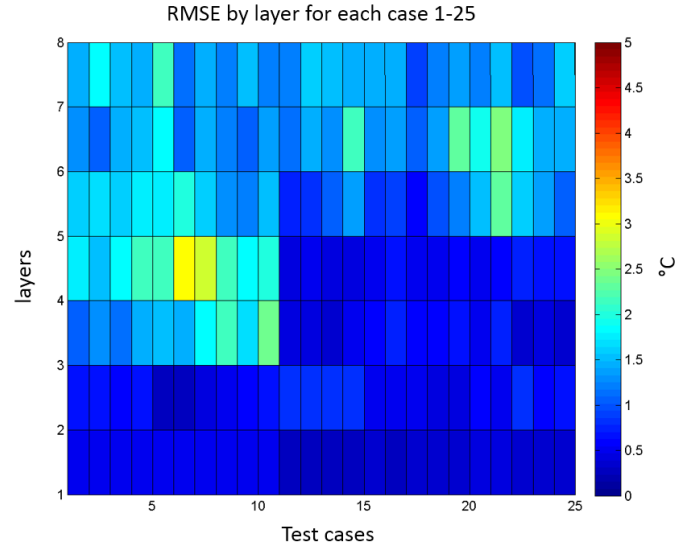
This can be explained when we examine the temperature distribution in the data centre in Fig. 11 which reveals a recirculation problem close to this location where the exhaust air from the servers is concentrated. This is due to CRAC 2 which is switched off in these cases. Additionally, there is an inherent bad layout of the data centre as the CRAC units and racks are parallel to each other. This prevents a direct return path for exhausted air, causing it to stagnate as the operational CRAC unit is too far to draw the hot air back for cooling. This encourages side recirculation at our inlet sensor location indicated by the red dot in Fig. 11. We account for only overhead recirculation in this preliminary version of the DCAM model but in future work, a 3 dimensional DCAM is proposed to take into account the effects of side recirculation.
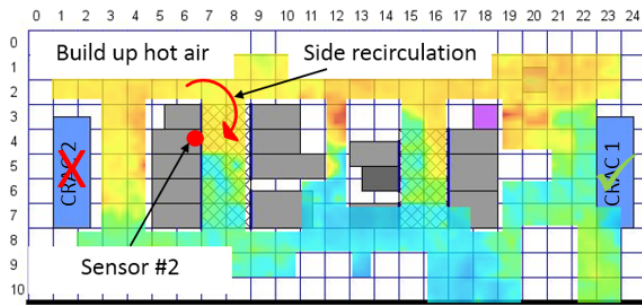
Fig. 11. Sensor location 2: build up of hot exhaust air in the location of sensor 2 due to CRAC 2 switched off and reduced air supply from CRAC 1.

V. CONCLUSIONS

The data centre is a complex dynamic environment in terms of airflow and temperature distribution. Current physics-based modeling approaches require solving the full numeric calculations for the complete data centre domain which can be expensive in term of computation time and is usually done offline. Our approach, underpinned by turbulent jet theory is applied to only local boundary conditions and can be solved in isolation of the remaining data centre. The model is validated against a real-life data centre which was subject to 25 control change settings and returned and accuracy error of $1.2°C$ (RMSE). These results are in line with simulation software. The DCAM model can be used standalone or used to complement real-world sensors. Potential future applications include real-time monitoring of inlet temperatures to servers which adapts to changes in server workloads, cooling flow rates, supply temperatures and could be used in conjunction with variable frequency drives on air conditioners.

REFERENCES

[1] U. Institute, "2014 uptime institute data center industry," 2014, https://journal.uptimeinstitute.com/2014-data-center-industry-survey/ (accessed June 2016).

[2] J. G. Koomey, "Growth in data center electricity use 2005 to 2010, (a report by analytics press, completed at the request of the new york times)," August 2011, http://www.analyticspress.com/datacenters.html.

[3] M. Toulouse, G. Doljac, and C. Bash., "Exploration of a potential-flow-based compact model of air-flow transport in data centers," in *Proceedings of the 2009 International Mechanical Engineering Congress and Exposition*, November 2009.

[4] J. W. VanGilder, X. Zhang, and C. M. Healey, "Data center airflow prediction with an enhanced potential flow model." in *Thermal Management; Data Centers and Energy Efficient Electronic Systems.*, vol. doi:10.1115/ipack2013-7307, 2013.

[5] D. Lettieri., "Expeditious data center sustainability, flow, and temperature modeling: Life-cycle exergy consumption combined with a potential flow based, rankine vortex superposed, predictive method." vol. PhD Thesis. UC Berkeley (2012).

[6] M. M. Toulouse, D. J. Lettieri, V. P. Carey, C. E. Bash, and A. J. Shah., "Evaluation of a vortex model of buoyancy-driven recirculation in potential flow analysis of data center performance," in *13th Inter-Society Conference on Thermal and Thermomechanical Phenomena in Electronic Systems.*, vol. doi:10.1109/itherm.2012.6231413, 2012.

[7] H. F. Hamann, V. Lopez, and A. Stepanchuk, "Thermal zones for more efficient data center energy management," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on*, June 2010, pp. 1–6.

[8] V. Lopez and H. F. Hamann, "Measurement-based modeling for data centers," in *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on*, June 2010, pp. 1–8.

[9] E. Samadiani, Y. Joshi, H. F. Hamann, M. K. Iyengar, S. Kamalsy, and J. Lacey., "Reduced order thermal modeling of data centers via distributed sensor data," *In Proceedings of IPACK2009, 2009*, 2009.

[10] E. Samadiani and Y. Joshi, "Proper orthogonal decomposition for reduced order thermal modeling of air cooled data centers," *ASME Transactions J. Heat Transfer*, vol. 132, no. 7, pp. 071 402–071 402–14, 2010.

[11] O. Sutton, *Micrometeorology*. McGraw-Hill, 1953.

[12] Q. Liu, "Energy performance of underfloor air distribution (ufad) systems part iii: The fluid dynamics of a ufad system," 1996, phD Thesis, University of California.

[13] Q. Li and P. Linden, "The fluid mechanics of underfloor air distribution," *Journal of Fluid Mechanics*, vol. 554, pp. 323–341, May 2006.

[14] P. Linden and P. Cooper, "Multiple sources of buoyancy in a naturally ventilated enclosure," *Journal of Fluid Mechanics*, vol. 311, pp. 177–192, March 1996.

[15] P. Cooper, "The theory of plumes adapted to model air movement in naturally ventilated buildings," in *Building Simulation '93 Proceedings*, August 1993.

[16] S. B. Pope, *Turbulent Flows*. New York, NY: Cambridge University Press, 2000.

[17] H. J. Hussein, S. Capp, and W. K. George, "Velocity measurements in a high-reynolds-number, momentum-conserving, axisymmetric, turbulent jet," *J. Fluid Mech.*, vol. 258, pp. 31–75, 1994.

[18] B. Cushman-Roisin, *Environmental Fluid Mechanics*. John Wiley & Sons, Inc., New York, NY., 2010.

[19] M. L. Albertson, Y. Dai, R. Jensen, and H. Rouse, "Diffusion of submerged jets," *American Society of Civil Engineers: 637-677*, 1948.

[20] I. Wygnanski and H. Fiedler, "Some measurements in the self-preserving jet," *J. Fluid Mech*, vol. 38, pp. 577–612, 1969.

[21] N. Kotsovinos, "A study of entrainment and turbulence on a plane buoyant jet," 1975, report No. KH-R-32, W. M. Keck Laboratory of Hydraulics and Water Resources, California Institute of Technology, Pasadena, California.

[22] INRES, "Tileflow," http://inres.com/products/tileflow/overview.html.

[23] H. F. Hamann, M. Schappert, M. Iyengar, T. van Kessel, and A. Claassen, "Methods and techniques for measuring and improving data center best practices," in *ITHERM*, 2008.

[24] W. A. Abdelmaksoud, H. E. Khalifa, T. Q. Dang, R. R. Schmidt, , and M. Iyengar., "Improved cfd modeling of a small data center test cell." *12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, 2010.