Data-driven prediction of flow parameters in a ventilated cavity using high-fidelity CFD simulations

Nina Morozova, F.Xavier Trias, Roser Capdevila, Assensi Oliva
Heat and Mass Transfer Technological Center (CTTC),
Universitat Politècnica de Catalunya (UPC), Terrassa (Barcelona), Spain

Abstract

In this study, we develop a machine-learning-based data-driven model, which predicts comfort-related flow parameters in a ventilated room. The model is based on the results of high-fidelity computational fluid dynamics (CFD) simulations with different geometrical configurations and boundary conditions. The developed model could be used as a cheaper alternative to CFD for applications where rapid predictions of complex flow configurations are required, such as model predictive control. Even though the developed model provides acceptable accuracy for most of the tested configurations, more input data is required to improve the model performance.

Key Innovations

- Developed model is capable of providing results at an accuracy comparable to CFD simulations.
- Once trained, the model provides the results almost instantly.

Practical Implications

The quality of the input data is crucial for the data-driven models. The input data should cover as many different working conditions as possible and come from a reliable source. Once the model is completed, the predictions are produced almost instantly.

Introduction

The outburst of airborne transmissible coronavirus disease (COVID-19) has highlighted the importance of ensuring adequate indoor air quality to reduce the risk of infection contamination in confined spaces (Morawska et al., 2020). One of the important aspects of ensuring indoor air quality is proper design and precise control of air parameters. Therefore, fast and accurate computation of indoor airflow is required for testing different design options or performing model predictive control (MPC) using real-time weather and occupant behavior data.

Nowadays, air distribution in buildings is usually evaluated by multizone models (Axley, 2007), zonal models (Megri and Fariborz, 2007), and CFD. Multizone models are the most popular choice due to the low computational cost, but they have limited applicability because each room is represented by only one node. Zonal models are considered intermediate between multizone and CFD, but they usually suffer from case dependency. In CFD, the solution provides a complete set of air parameters for each point of the physical domain. Nonetheless, CFD is always a compromise between computational cost and accuracy. Accurate CFD simulations require big computational resources while using ordinary office computers results in unreasonably long run-times (Morozova et al., 2018, 2019). Grid coarsening and RANS turbulence modeling are common ways to reduce the cost of CFD, but they can lead to overly inaccurate results. Moreover, according to Morozova et al. (2020), the growth of computational resources in the foreseeable future would not be enough to make CFD available for routine use in building applications. This means that new numerical models capable of proving the accuracy comparable with high-fidelity CFD, but at considerably lower computational cost, are needed.

Over the last years, several attempts to develop alternative to zonal and multizone reduced-order models for building simulations have been made. For example, Li et al. (2018) investigated a multiple model approach for predictive control of indoor thermal environment using proper orthogonal decomposition (POD). Phan and Lin (2017) developed a reduced-order model of a data center with multi-parameters using the POD method. Both developed models successfully reduce the simulation time while maintaining an acceptable level of accuracy. However, the models focus on temperature and thermal load predictions and do not consider motion-related flow parameters, which are usually more complex as they are described by nonlinear processes.

Data-driven models (DDMs) are gaining popularity in building modeling applications due to their accurate approximations of nonlinear processes. For example, Athavale et al. (2019) and Fang et al. (2019) compared different DDM approaches for temperature prediction in data centers, and concluded that predictions produced by the model are in good agreement with the reference CFD results. Warey et al. (2020) created a vehicle cabin thermal comfort model using machine learning (ML) and high-fidelity CFD simulation results with a test error of less than 5%. More-

over, DDMs are widely used in MPC. For instance, DDM-MPC for heating ventilation and air conditioning (HVAC) systems were developed for a university building (Ruano et al., 2016), an airport (Huang et al., 2015), and a residential building (Afram et al., 2017). All the aforementioned researchers note that DDMs produce accurate predictions at a low computational cost.

DDMs are based on using data analysis to find relations between system state variables without explicit knowledge of the physical behavior of the system. They can be developed relatively easy since they do not require an understanding of system physics. To train these models, a comprehensive set of the high-quality input-output dataset is needed for all possible working conditions. The accuracy of DDM decreases when training data deviates from testing data. Therefore, it is critical to use training data that covers all the operating conditions, which could be challenging. However, difficulties in obtaining high-fidelity training data are compensated by the high accuracy and the low computational cost of the resulting model, according to Afram et al. (2017). Therefore, DDMs could be used for complex indoor environments with stratification, natural and forced convection, where low order models cannot be relied on, and computationally expensive CFD simulations are required.

In this work, we develop ML-based DDM, which uses the data from high-fidelity CFD simulations. It predicts comfort-related airflow parameters in a ventilated room with a heated floor. The main focus of our research is on investigating the capabilities and limitations of this model as a cheaper alternative to CFD, taking into account specific requirements for indoor environmental applications. We study how the input data affects the quality of prediction.

Governing equations and physical problem

Governing equations

The incompressible Navier-Stokes equations for a Newtonian fluid with constant physical properties are considered. The Boussinesq approximation is adopted to account for the density variations due to temperature difference. Thermal radiation is neglected. Under these assumptions, the governing equations are

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = \nu \nabla^2 \mathbf{u} - \frac{1}{\rho}\nabla p + \beta \mathbf{g} \Delta T, \tag{2}$$

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla)T = \alpha \nabla^2 T, \tag{3}$$

where $\mathbf{u} = (u, v, w)$ is the velocity vector in Cartesian coordinates $\mathbf{x} = (x, y, z)$, p the pressure, T the temperature, ν the kinematic viscosity, ρ the density, $\mathbf{g}$ the gravitational acceleration, β the thermal expansion coefficient and α the thermal diffusivity.

Hereafter, all the results are presented in dimensionless form. The reference values of time, velocity, temperature, and length are $t_{ref} = H/U_{ref}$, $U_{ref} = U_{in}$, ΔT, and H, respectively, where H is the cavity height, U_{in} - inlet bulk velocity, and $\Delta T = T_h - T_c$ - the temperature difference.

Physical problem

The physical set up used in the predictive model is a three-dimensional ventilated cavity with a heated floor. This configuration was studied experimentally by Blay et al. (1992). The geometry of the studied cavity is shown in Figure 1. The depth ratio of the cavity is $A_d = D/H = 0.3/1.04$. Cold air at $T_c = -0.5$ enters the cavity through the long thin inlet at the top of the left wall. The inlet velocity profile in the vertical (y) direction corresponds to a parabolic Poiseuille flow. The inlet slot has an aspect ratio $A_{in} = h_{in}/H = 0.018/1.04$. The air is discharged through the outlet with an aspect ratio $A_{out} = h_{out}/H = 0.024/1.04$ at the bottom of the right wall of the cavity. The bottom wall is maintained at a hot temperature of T_h, while the three other sidewalls are kept at the cold temperature of T_c. The cavity is filled with air ($Pr = \nu/\alpha = 0.71$) at Rayleigh number based on the cavity height $Ra = \rho g \beta \Delta T H^3/(\nu\alpha) = 2.4 \times 10^9$.

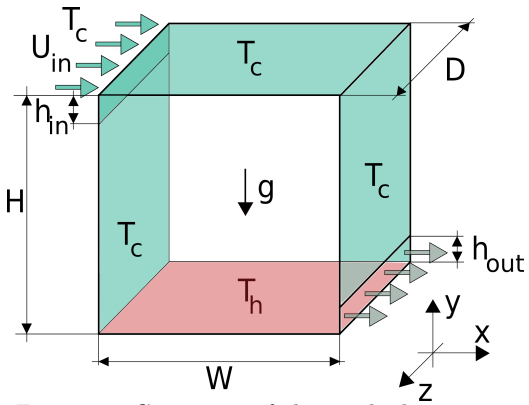

Figure 1: Geometry of the studied test case.

We build the input-output dataset by changing the height aspect ratio of the cavity ($A_h = H/W$) and the Froude number based on the inlet height ($Fr_{h_{in}} = U_{in}/\sqrt{\rho g \beta \Delta T h_{in}}$). We use 5 different height aspect ratios - $A_h = [0.25, 0.5, 1.0, 2.0, 4.0]$ and 20 different Froude numbers - $Fr_{h_{in}} = [1.10, 1.50, 2.00, ..., 5.00, 5.24, 5.50, ..., 10.00]$. The changes in the Froude number are produced by variations in the inlet bulk velocity (U_{in}), thus Pr and Ra numbers remain constant. Chosen combinations of $Fr_{h_{in}} - A_h$ are realistic and relevant for the indoor environmental applications.

Model definition

Input parameters

As input parameters of the model, we initially consider Froude number based on the inlet height ($Fr_{h_{in}}$), cavity height aspect ratio (A_h), temperature (T), and velocity magnitude (V) probes at 11 different locations on the mid-depth cavity plane ($z = D/2$). In total, we use 24 ($Fr_{h_{in}} + A_h + 11T + 11V$) input parameters. The positions of the probes are shown in the figure 2. In the results section, we analyze which probes have the smallest contribution to the accuracy of the prediction. As a result of this analysis, two probes are eliminated, and the total number of input parameters is reduced to 20.

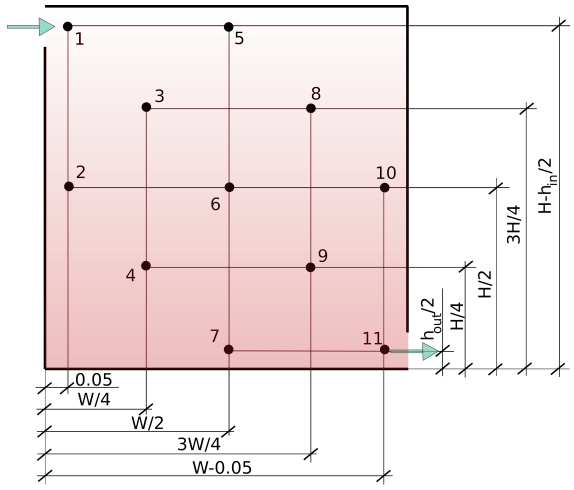

Figure 2: Locations of the input data probes on the mid-depth plane for the cavity with the height aspect ratio $A_h = 1$.

Output parameters

As the output parameters, we choose five global flow quantities for comparison: average Nusselt number on the hot wall - $<Nu>$, jet separation point - x_{sep}, average kinetic energy - $<E>$, average enstrophy - $<\Omega>$, and the average temperature of the cavity. The jet separation point is a point at the top cavity wall, where the wall-shear stress $<\tau_W>$ is equal to zero, as shown in the equation (6). Other flow parameters are calculated using equations (4) - (8).

$$<Nu>= -\frac{1}{A}\int_A \frac{\partial T}{\partial y} dA \quad \text{at} \quad y = 0 \tag{4}$$

$$<T_V>= \frac{1}{V}\int_V T dV \tag{5}$$

$$x_{sep} = x, \text{ at } <\tau_W>= \int \frac{\partial u}{\partial y} dz = 0, y = H, \tag{6}$$

$$<E>= \frac{1}{V}\int_V \frac{\mathbf{u}^2}{2} dV \tag{7}$$

$$<\Omega>= \frac{1}{V}\int_V \omega^2 dV, \tag{8}$$

where A is the surface of the hot wall, V is the volume of the cavity, $\mathbf{u} = (u, v, w)$ is the velocity vector in Cartesian coordinates $\mathbf{x} = (x, y, z)$, and $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ is the vorticity. All these quantities are time-dependent and averaged over time. Standard bracket "<>" notation is used for time-averaged values.

These quantities represent basic airflow properties and are relevant to the thermal comfort according to ISO (2005). Nusselt number and cavity temperature represent the thermal properties of the flow. Nusselt number is a measure of heat transfer. It is computed using the temperature gradient at the bottom hot wall. The average temperature is the operative room temperature. Kinetic energy measures the level of motion. Enstrophy corresponds to turbulence intensity. Kinetic energy, enstrophy, and jet separation point are used to measure draught and the level of velocity discomfort.

Methods

CFD simulations

To generate input and output data for the model, we use LES simulations on staggered grids with second-order symmetry-preserving spatial discretization (Verstappen and Veldman, 2003) and a one-parameter fully explicit second-order temporal discretization scheme developed by Trias and Lehmkuhl (2011). To perform the simulations, we use an in-house CFD code developed by Gorobets et al. (2010) and the LES-S3PQ turbulence model (Trias et al., 2015). This numerical configuration showed the best trade-off between the computational cost and accuracy for this flow configuration, according to to Morozova et al. (2020). The computational grid resolution for the cavity with the $A_h = 1$ is $N_x \times N_y \times N_z = 100 \times 160 \times 32 = 5.12 \times 10^5$ control volumes. The number of grid points in the horizontal (x) direction is rescaled accordingly to the value of A_h. All simulations run for 500 non-dimensional time units, which was found to be a long enough time-integration period to record the flow statistics for further averaging. In total, we carry out 100 CFD simulations, 15% of which are reserved for testing, and 85% are used in the model training.

Data-driven model

According to the findings of Morozova et al. (2021) artificial neural network (ANN) ML algorithm shows the best performance for these applications. ANN is the most common ML-based data-driven modeling framework where the output is mapped to the input using a set of interconnected nodes or neurons. The neural network used in this work consists of one input layer, one hidden layer, and one output layer. The number of neurons in the input and output layer represents the number of model parameters and output values. The number of neurons in the hidden layer is equal to 16 and is chosen by trial and error (Figure 4).

Each neuron in the hidden layer is connected to every neuron in the preceding layer via links with specific weights W_{ij}. Equation (9) represents the output of a neuron (y_j) in a layer where n_i is the number of neurons in the preceding layer, i is the index for neurons in the preceding layer, f is the non-linear activation function, x_i is the input parameter, and b_j is the bias term associated with the neuron.

$$y_j = f\left(\sum_{i=1}^{n} W_{ij}x_i + bj\right) \tag{9}$$

Training the neural network involves determining the appropriate combination of the number of hidden layers, the number of neurons in each of the hidden layers, and the associated weight coefficients that minimize the prediction error.

In our work, we use open-source Keras (Chollet, 2015) ANN ML library. To improve the results of the prediction and avoid model overfitting (lack of generalization), we adopt k-fold cross-validation (Stone, 1974) and dropout regularisation (Srivastava et al., 2014) techniques. K-fold cross-validation is a resampling procedure based on the parameter k that refers to the number of groups a given data sample is to be split into. The value of k equal to 10 has shown the best performance for the developed model. Dropout is a regularization method, where during the model training procedure, some of the layer outputs are randomly ignored or "dropped out", increasing the sparsity of the system. In our work we adopted the dropout rate equal to 0.2, as it showed the best performance in the experiments.

Results

This section shows the results obtained by the created model and its assessment towards optimal performance. In order to quantify the accuracy of the model, we use local relative prediction error (RE), and mean relative error (MRE), which are calculated as detailed in the equations (10) and (11).

$$RE(\phi) = \frac{|\phi_d - \phi_p|}{|\phi_d|}, \tag{10}$$

$$MRE(\phi) = \frac{1}{N}\sum_{i=1}^{N} RE(\phi), \tag{11}$$

where N is the number of data points, ϕ_d stands for the data, and ϕ_p stands for the model prediction. We assume that the less than 15% RE is acceptable for this model.

Model assessment

In this subsection, we explain how some of the DDM parameters affect its accuracy. The important component of every DDM is the quantity and the quality of input data. To train our model, we used the results of 85 CFD simulations, but it is interesting to see how the prediction accuracy changes depending on the amount of input data. Results of this analysis are shown in Figure 3. The number of simulations in the input dataset has not yet converged and could be increased to improve the prediction accuracy. However, the computational cost of high-fidelity CFD should be taken into account. The model development is ongoing, and the results of this work will be taken into account when the decision of the configurations of the extra CFD simulations will be taken.

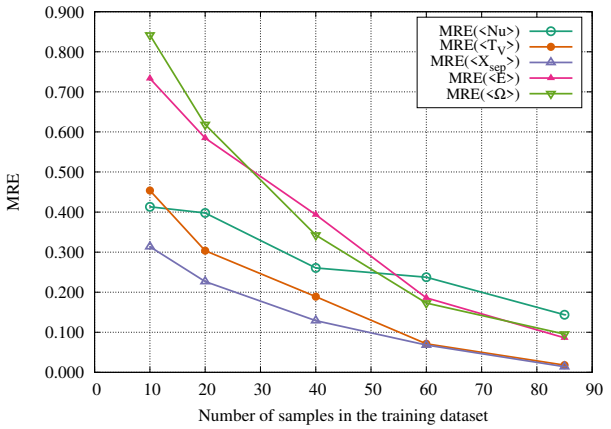

Figure 3: Mean relative error (MRE) of the studied flow parameters for different number of samples in the training dataset.

The number of neurons in the hidden layer directly affects the performance of the ANN model. Nonetheless, the available literature does not provide any clear guidelines for the optimal number of neurons. In general, the number of neurons in the hidden layer should lie somewhere in between the number of neurons in the input layer and the number of neurons in the output layer. We performed a comparison of the model performance against the number of neurons in the hidden layer (Figure 4). The ANN with 16 neurons showed the smallest MREs for all of the studied flow parameters.

Optimal probe combinations

In order to create a reduced-order model, we initially chose 11 probes of velocity magnitude V and temperature T as model input parameters. Locations of the probes inside the cavity are chosen to take into account the possibility of using sensor readings as model input parameters. Furthermore, in order to reduce the number of input parameters, we analyze which probes are crucial for accurate predictions and which could be eliminated without losing the model accuracy. This is determined by the specific weights of the data and the mean relative prediction error, equation (11). Results of the analysis are detailed in Table 1.

Probes 3 and 4 have the smallest specific weights in the model, their elimination does not significantly de-

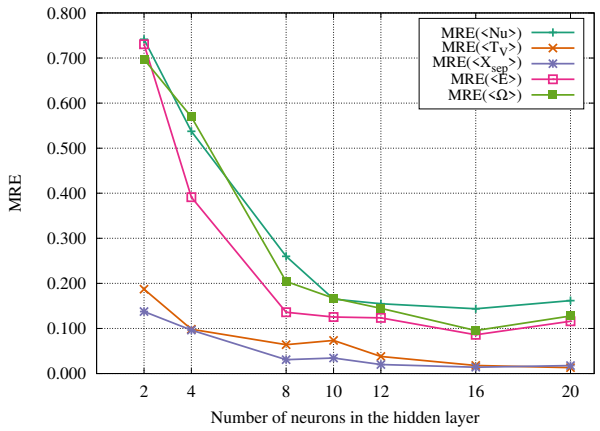

Figure 4: Mean relative error (MRE) of the studied flow parameters for different number of neurons in the hidden layer.

Table 1: Mean relative prediction error of the studied flow parameters for different combinations of probes.

Probes	MRE				
removed	$<Nu>$	$<T_V>$	x_{sep}	$<E>$	$<\Omega>$
none	0.144	0.018	0.014	0.086	0.095
3	0.146	0.023	0.067	0.082	0.102
3,4	0.154	0.051	0.068	0.093	0.118
3,4,6	0.197	0.076	0.084	0.145	0.144
3,4,6,7	0.343	0.165	0.115	0.195	0.228

crease the accuracy of the model. The specific weight of probe 6 is the next smallest after 3 and 4. After elimination of the probe 6, MRE of most the flow parameters does decrease much, however, the accuracy of the kinetic energy and enstrophy predictions has dropped significantly. Further elimination of the next probe with the smallest weight (probe 7) leads to big errors in the prediction of Nusselt number, kinetic energy, and enstrophy. For the further tests, we chose the model with 20 input parameters ($Fr_{h_{in}} + A_h + 9T + 9V$), using the probes 1, 2, and 5-11, showed in the Figure 2.

Average cavity temperature and the flow separation point are the parameters, which are predicted with the highest accuracy. Nusselt number is also predicted fairly well, however, the accuracy of its prediction mostly depends on the data from probe 7, located near the hot wall. Nonetheless, kinetic energy and enstrophy are difficult to predict accurately with the amount of data available. More simulations are required to improve the quality of the prediction for the kinetic energy and enstrophy.

Prediction quality across the test cases

The accuracy of the model is evaluated using two-dimensional plots, where the local relative prediction error (equation (10) is plotted for the tested combinations of Froude number ($Fr_{h_{in}}$) and height aspect ratios (A_h). Tested combinations are chosen randomly from the dataset and were not used in the model training process. The plots are shown in the Figures 5-9. The horizontal (x) axis of the figures is the Froude number, the spanwise (y) axis is the cavity height aspect ratio, and the vertical (z) axis is the relative prediction error $RE(\phi)$, detailed in the equation (10). The color scales from light to dark shows the values of errors from high to low, respectively.

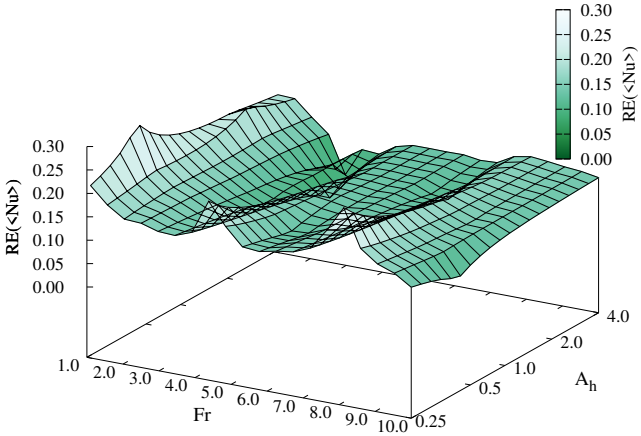

Figure 5: Relative error (RE) of the Nusselt number $<Nu>$ prediction plotted for different combinations of Froude number ($Fr_{h_{in}}$) and the cavity height aspect ratio (A_h).

Relative error for the Nusselt number is shown in the Figure 5. It shows low RE values for most of the combinations of $Fr_{h_{in}} - A_h$, except the area of small Froude numbers and the small cavity aspect ratio (tall cavity). For most of the configurations, the relative error lies within a 15% margin, however, it rises to 30% for some of the test points. The big values of errors could be explained by the dominance of the natural convection in the tall cavities with the small inlet velocities, which produce higher flow stratification. Flow stratification is difficult to resolve accurately.

The average cavity temperature (Figure 6) is accurately predicted for all of the Froude number - cavity width combinations. The relative error does not exceed 10%. However, the tendency is different - the errors are bigger in the area with the long cavities (big A_h).

Flow separation point is shown in the Figure 7. It is predicted well for most of the tested cases. The relative error value does not exceed 9% and is around 5% for most of the tested combinations of $Fr_{h_{in}} - A_h$. Nonetheless, the flow separation point is predicted more accurately for the small cavity A_h (short cavities), which are associated with lesser flow separation. In these geometrical configurations the flow separation happens closer to the right wall of the cavity or at the right wall itself, which makes its accurate evaluation easier.

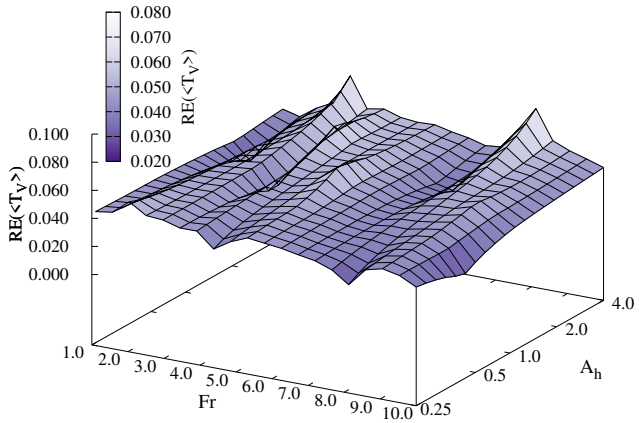

Figure 6: Relative error (RE) of the average temperature $<T_V>$ prediction plotted for different combinations of Froude number ($Fr_{h_{in}}$) and the cavity height aspect ratio (A_h).

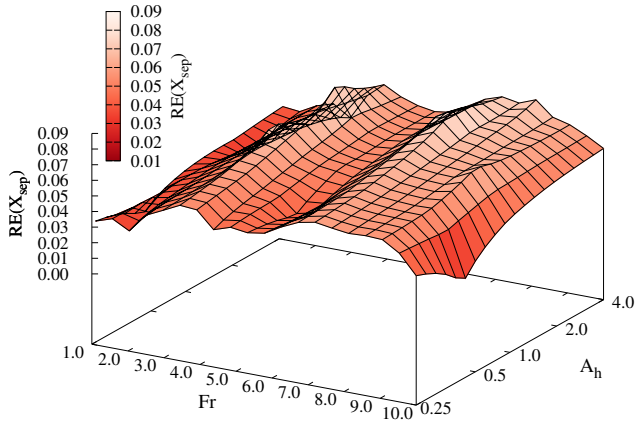

Figure 7: Relative error (RE) of the flow separation point x_{sep} prediction plotted for different combinations of Froude number ($Fr_{h_{in}}$) and the cavity height aspect ratio (A_h).

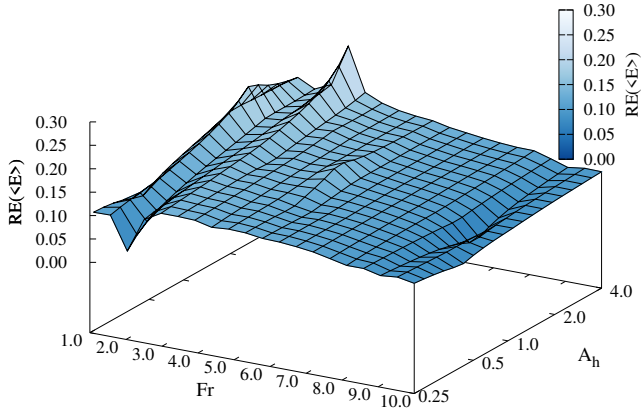

Figure 8: Relative error (RE) of the kinetic energy $<E>$ prediction plotted for different combinations of Froude number ($Fr_{h_{in}}$) and the cavity height aspect ratio (A_h).

Average kinetic energy (Figure 8) is more difficult to predict correctly. The specific geometry of the tested cases makes the accuracy of the model depend on the flow separation. In the areas with a higher $Fr_{h_{in}}$ number flow separation starts closer to the right wall of the cavity, thus, the secondary vortex is not formed. The presence of the secondary vortex makes the flow more complex, hence, the relative error increases. The secondary vortex is formed either when the cavity is long (big A_h) or when the inlet velocity is small (low $Fr_{h_{in}}$).

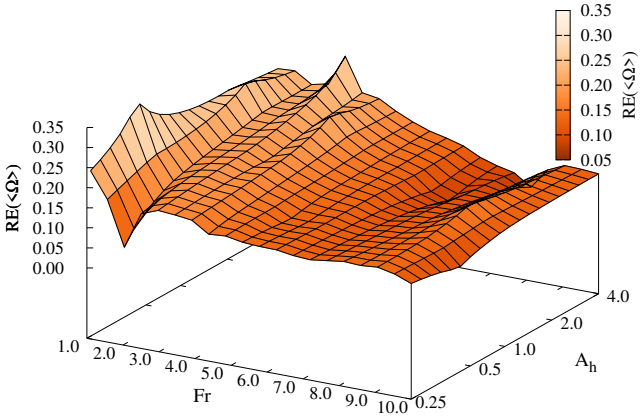

Figure 9: Relative error (RE) of the enstrophy $<\Omega>$ prediction plotted for different combinations of Froude number ($Fr_{h_{in}}$) and the cavity height aspect ratio (A_h).

Average enstrophy (Figure 9) is a characteristic of turbulence intensity, so it is the most difficult quantity to predict accurately. Moreover, the accuracy of the prediction, similar to the average kinetic energy, depends on the existence of the secondary flow vortex. For most of the tested combinations, the relative error is around 10%, however, it rises to 35% for the flow configurations with a secondary vortex.

As could be concluded from the results, an ML-based DDM is capable of providing predictions of airflow parameters at an acceptable level of accuracy for most of the tested ranges. However, the model struggles to provide reliable results for rather complex flows, where the flow separation phenomenon forms a secondary vortex. More data provided for these particularly complicated combinations of inlet velocity and cavity length could potentially improve the quality of the predictions.

Discussion

In this section, we discuss the computational cost of the simulations, problems, and limitations of the developed model.

The computational cost of the model development is a combination of the computational cost of CFD simulations and the model training cost. The latter is negligible - it takes approximately 3 minutes to train the model on a personal computer using only one CPU core. Moreover, after the training is completed, the

prediction is produced almost instantly. The main computational effort is used to produce high-fidelity CFD data for the model training. It takes approximately 215 CPU hours per simulation, thus, the whole dataset with 100 simulations uses around 21500 CPU hours. Taking into account, that the estimated electricity cost of one CPU hour using Amazon Web Services (2021) is 0.01€, the total electricity cost of the used dataset with 100 CFD simulations is 215€. This is reasonable computational and financial cost for the DDM development, taking into account that these CFD simulations are carried out only once, in order to form the input dataset, and will not be repeated.

One of the main bottlenecks of the DDMs is the necessity of big sets of high-fidelity data. In our work, we used a dataset of 100 CFD simulations. It is a good dataset to start the model development and identify critical points with weak predictions, but it should be significantly increased, in order to develop a reliable model.

Another important limitation of the developed modeling framework is its applicability. Even though a wide range of room geometries and inlet velocities was tested, the model is not universal and does not cover a full extension of possible indoor environmental configurations. The DDMs could be used for the flow configurations with complex physical phenomena like natural and forced convection or stratification and widely used building geometry. This makes DDMs applicable in the situations where other reduced-order models do not provide sufficient accuracy and reduces its case dependency. Unfortunately, these models are not suitable for complex geometries. Hence, DDMs could be used for applications where a combination of fast and accurate predictions is required, for example, for model predictive control. Data-driven models are capable of providing rapid predictions for common room geometries. And as they are based on high-fidelity data (CFD simulations), they take into account temperature and velocity distributions inside individual rooms. The combination of the aforementioned factors makes data-driven modeling frameworks an interesting option for the MPC.

Conclusions

In this work, we created a ML-based DDM for predicting the comfort-related flow parameters in a three-dimensional ventilated cavity with a heated floor. The developed DDM provides rapid and accurate predictions using an ordinary office computer. The modeling framework chosen is the ANN with one hidden layer. The model input parameters are the values of the temperature and velocity magnitude at 9 locations within the cavity domain (Figure 2), and the output parameters are the average Nusselt number on the hot wall, jet separation point, average kinetic energy, average enstrophy, and the average temperature of the cavity. The model is capable of predicting these parameters with acceptable accuracy (the mean relative error varies between 5% and 15%), however, the accuracy for some of the most complex flow configurations was insufficient. More high-fidelity data is required to construct a robust and reliable model. Extra CFD simulations will be carried out for these complex flow configurations.

The work on improving this model is ongoing. In the future, we would like to amplify the available training data, study how different input configurations affect the quality of the predictions, find a trade-off between the quantity and the quality of the training data, namely, explore different turbulence models and grid resolutions. Moreover, we plan to work on the extrapolating capabilities of the DDMs.

Acknowledgments

This work is supported by the Ministerio de Economía y Competitividad, Spain [ENE2017-88697-R]. N. Morozova is supported by the by the Ministerio de Economía y Competitividad, Spain [FPU16/06333 predoctoral contract]. Part of the calculations was performed on the MareNostrum 4 supercomputer at the Barcelona Supercomputing Center [RES project IM-2021-1-0015]. The authors thankfully acknowledge these institutions.

References

Afram, A., F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar (2017). Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy and Buildings 141*, 96–113.

Amazon Web Services (2021). Amazon EC2 On-Demand Pricing. `https://aws.amazon.com/ec2/pricing/on-demand/`.

Athavale, J., M. Yoda, and Y. Joshi (2019). Comparison of data driven modeling approaches for temperature prediction in data centers. *International Journal of Heat and Mass Transfer 135*, 1039–1052.

Axley, J. (2007). Multizone airflow modeling in buildings: History and theory. *HVAC&R Research 13*(6), 907–928.

Blay, D., S. Mergui, J. L. Tuhault, and F. Penot (1992). Experimental turbulent mixed convection created by confined buoyant wall jets. In *Proceedings of the First European Heat Transfer Conference*, pp. 821–828. Birmingham (UK), Sept 1992.

Chollet, F. (2015). keras. `https://github.com/fchollet/keras`.

Fang, Q., Z. Li, Y. Wang, M. Song, and J. Wang (2019). A neural-network enhanced modeling

method for real-time evaluation of the temperature distribution in a data center. *Neural Computing and Applications 31*(12), 8379–8391.

Gorobets, A., F. X. Trias, M. Soria, and A. Oliva (2010). A scalable parallel Poisson solver for three-dimensional problems with one periodic direction. *Computers & Fluids 39*(3), 525–538.

Huang, H., L. Chen, and E. Hu (2015). A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. *Energy and Buildings 97*, 86–97.

International Organization for Standardization (2005). *Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*.

Li, K., X. Wenping, C. Xu, and H. Mao (2018). A multiple model approach for predictive control of indoor thermal environment with high resolution. *Journal of Building Performance Simulation 11*(2), 164–178.

Megri, A. C. and H. Fariborz (2007). Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications. *HVAC&R Research 13*(6), 887–905.

Morawska, L., J. W. Tang, W. Bahnfleth, P. M. Bluyssen, A. Boerstra, G. Buonanno, J. Cao, S. Dancer, A. Floto, F. Franchimon, C. Haworth, J. Hogeling, C. Isaxon, J. L. Jimenez, J. Kurnitski, Y. Li, M. Loomans, G. Marks, L. C. Marr, L. Mazzarella, A. K. Melikov, S. Miller, D. K. Milton, W. Nazaroff, P. V. Nielsen, C. Noakes, J. Peccia, X. Querol, C. Sekhar, O. Seppänen, S. ichi Tanabe, R. Tellier, K. W. Tham, P. Wargocki, A. Wierzbicka, and M. Yao (2020). How can airborne transmission of COVID-19 indoors be minimised? *Environment International 142*, 105832.

Morozova, N., R. Capdevila, F. Trias, and A. Oliva (2018). Towards Real-Time CFD Simulation of Indoor Environment. In *Proceedings of 10th International Conference on Computational Fluid Dynamics*. Barcelona (Spain), 9-13 July 2018.

Morozova, N., R. Capdevila, F. X. Trias, and A. Oliva (2019). On the feasibility of CFD for transient airflow simulations in buildings. In *Proceedings of Building Simulation 2019: 16th Conference of IBPSA*. Rome (Italy), 2-4 September 2019.

Morozova, N., F. X. Trias, R. Capdevila, and A. Oliva (2021). Investigating the capabilities of CFD-based data-driven models for indoor environmental design and control. In *Proceedings of 14th WCCM & ECCOMAS Congress 2020*. Virtual congress, 11-15 January 2021.

Morozova, N., F. X. Trias, R. Capdevila, C. D. Pérez-Segarra, and A. Oliva (2020). On the feasibility of affordable high-fidelity CFD simulations for indoor environment design and control. *Building and Environment 184*, 107144.

Phan, L. and C.-X. Lin (2017). Reduced order modeling of a data center model with multi-Parameters. *Energy and Buildings 136*, 86–99.

Ruano, A. E., S. Pesteh, S. Silva, H. Duarte, G. Mestre, P. M. Ferreira, H. R. Khosravani, and R. Horta (2016). The imbpc hvac system: A complete mbpc solution for existing hvac systems. *Energy and Buildings 120*, 145 – 158.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*(56), 1929–1958.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological) 36*(2), 111–133.

Trias, F. X., D. Folch, A. Gorobets, and A. Oliva (2015). Building proper invariants for eddy-viscosity subgrid-scale models. *Physics of Fluids 27*(6).

Trias, F. X. and O. Lehmkuhl (2011). A self-adaptive strategy for the time integration of Navier-Stokes equations. *Numerical Heat Transfer, Part B: Fundamentals 60*(2), 116–134.

Verstappen, R. W. C. P. and A. E. P. Veldman (2003). Symmetry-preserving discretization of turbulent flow. *Journal of Computational Physics 187*(1), 343–368.

Warey, A., S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan (2020). Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results. *International Journal of Heat and Mass Transfer 148*, 119083.