A Comparative Study of Computational Algorithms Used in the Automatic Generation of Reduced-Order Models from CFD Simulations

Thibault Marzullo – National University of Ireland Galway – t.marzullo1@nuigalway.ie
Sajjad Yousefian – National University of Ireland Galway – s.yousefian2@nuigalway.ie
Marcus M. Keane – National University of Ireland Galway – marcus.keane@nuigalway.ie
Marco Geron – Queen's College of Belfast – m.geron@qub.ac.uk
Rory F.D. Monaghan – National University of Ireland Galway – rory.monaghan@nuigalway.ie

Abstract

In indoor thermal environment modelling applications where dynamic local effects of fluid flows are critical, classic zonal models are not always suitable. On the other hand, CFD simulations can give accurate solutions at very high computational cost. Reduced-order models (ROMs), extracted from CFD simulations, can preserve CFD model accuracy while being characteristically of low computational cost. The authors propose a method, known as CFD-ROM, capable of rapidly and automatically generating, from CFD simulations, zones, mass, and heat flows, and boundary conditions (BCs) for ROMs. This paper presents a comparative study of automatic zone generation algorithms as a necessary initial step to developing the CFD-ROM method. Zone generation algorithms compared in this paper are: (1) Mean Values Segmentation; (2) Classic Watershed; and (3) Coarse Grid Interpolation. The methods were compared on the bases of their accuracy against the original validated CFD simulation results, and their time to zone generation. The Mean Values Segmentation method yields promising results, providing a mean error below 0.2 K for 15-zone models generated in under 28 seconds. The next immediate steps for the development of CFD-ROM are (i) construction of a ROM solver, and (ii) testing its ability to predict thermal conditions when CFD BCs and ROM BCs differ.

1. Introduction

The buildings sector and people's activities in buildings are responsible for 31 % of the global final energy demand, one-third of energy-related CO_2 emissions, two-thirds of halocarbon emissions, and 25–33 % of black carbon emissions (GEA, 2012). Sustainability, as defined by the ASHRAE Handbook of Fundamentals (ASHRAE, 2013), is "providing for the needs of the present without detracting from the ability to fulfil the needs of the future". A key approach to improve the sustainability of energy use is the reduction of waste through careful planning, optimization and management of demand. In the built environment, the use of models of building thermal conditions during both the design and operation phases is recognised as a powerful method to increase energy efficiency and reduce energy demand (Harish and Kumar, 2016).

These computer models can be grouped into three categories: (1) physics-based models, (2) experimental models, and (3) mixed models (Harish and Kumar, 2016). Physics-based (theoretical) models describe in detail the studied system and its subsystems, and define an output based on mathematical equations constrained by physical laws. They are typically used during the preliminary design and energy audit phases. Experimental models on the other hand are tailored to a particular system. Through experimentation, the system's response to various inputs is evaluated and the model is developed accordingly. Finally, mixed models are physics-based models for which parameters are estimated using statistical and/or experimental analysis.

Among physics-based models, Computational Fluid Dynamics (CFD) is a powerful and increasingly widespread tool for simulating fluid domains, which yields models of high fidelity (van Hooff and Blocken, 2013; Tamura and Van, 2014; Antonioni et al., 2012). However its large computational expense frequently renders it too time-consuming for design tasks. Designers and operators therefore often choose lower fidelity but faster methods when

available (Li and Wen, 2014). Unfortunately lower fidelity methods lose information that could be otherwise beneficial to the overall system performance. In order to overcome some of these issues we propose a method called CFD-ROM (Computational Fluid Dynamics – Reduced Order Model) which will automatically extract ROMs from CFD simulations and solve them rapidly for a wide range of conditions, including those for which no CFD solutions are available. The method aims to develop ROMs that will retain a high level of accuracy, similar to the CFD simulations, but at a significantly lower computational cost (Mullen et al., 2015; Tan and Glicksman, 2005; Mora et al., 2003; Lucia et al., 2001). A flowchart for the CFD-ROM method is shown in Fig. 1. The method uses as its input, results obtained from a CFD simulation (Step 1). It then clusters computational cells together to create zones (Step 2). Zones' physical properties and inter-zone interactions are calculated (Step 3). Afterwards, a multi-zone ROM is generated and passed to the ROM solver (Step 4). It is solved (Step 5) and the results are remapped to the original CFD domain (Step 6). Operations performed by the CFD-ROM method can be categorised as online or offline. Online operations are tasks that need to be repeated to obtain a new ROM solution, while offline operations are those tasks executed only once for a given indoor environment. Zone generation (Step 2) represents a major portion of offline method computational expense and is executed only once to generate a set of ROMs. ROM solving, on the other hand, is an online operation, as yielding new results involves computing ROMs for new boundary conditions.

The purpose of the study presented in this paper is to perform an evaluation of different zone generation algorithms to inform the development of the CFD-ROM method. This is important as the zoning algorithm has an influence on the quality of the information passed to the ROM solver and preliminary studies showed that this step can represent a large portion of the offline total time to solution (Mullen et al., 2015).

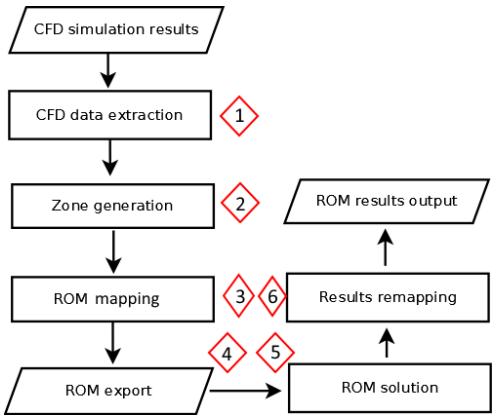

Fig. 1 – CFD-ROM method flowchart

2. Description of Zoning Algorithms

As previously stated, a crucial step in the development of the CFD-ROM method is the selection of a zone generation algorithm to develop a multi-zone ROM. It is important to note that these algorithms do not output a complete ROM; their scope is limited to the optimal clustering of cells into zones of similar physical properties. Three methods have been considered in the present study: (1) Mean Values Segmentation (MVS), (2) Classic Watershed (CW), and (3) Coarse Grid Interpolation (CGI). All three algorithms, which are explained in the following sections, have been coded in Python v2.7 (Python Software Foundation.). MVS has been developed from the results obtained during preliminary CFD-ROM development (Mullen et al., 2015) as it provided satisfactory proof-of-concept results. CW is directly adapted from the eponymous image processing method (Soille and Vincent, 1990). Finally, CGI interpolates the CFD solution to a very coarse mesh.

The use of higher computational power such as Graphic Processing Unit (GPU) processors and parallel computing are not within the scope of this study, but their use will be investigated in the future once a zoning method has been selected.

2.1 Mean Values Segmentation

This method successively splits the domain depending on the zone criteria. Zone criteria are the physical properties that are used to identify uniform regions of the simulation domain. Examples of zone

criteria include temperature, air velocity and carbon dioxide concentration. For the purpose of the present comparative study, temperature is the only zone criterion used. Zone generation is achieved through three main steps described by Mullen et al. (2015): (1) zone-type generation, (2) zone creation, and (3) zone number reduction. A zone-type is an interval containing all the cells sharing similar values of the zone criteria independently from their spatial position.

Fig. 2.a describes the MVS method. The algorithm extracts the values of the zone criteria from the entire domain (Step 1), computes the average value of the zone criteria (Step 2), then divides the domain into two zone-types (Step 3). The first zone-type containing cells with zone criteria values greater than the average value, and the second containing the remaining cells. The algorithm successively splits the zone-types until the required number of zone-types has been achieved (Step 4).

After generating the required number of zone-types the algorithm will cluster adjacent cells together within a zone-type. During this step a new zone is generated at each time a cell that has not been assigned a zone yet, is found. The algorithm then searches all adjacent cells within the same zone-type and assigns them the same zone. The process continues until no adjacent cells are found in the zone-type any more, at which point the process is repeated for a new zone until all cells within the zone-type have been assigned a zone. At this point

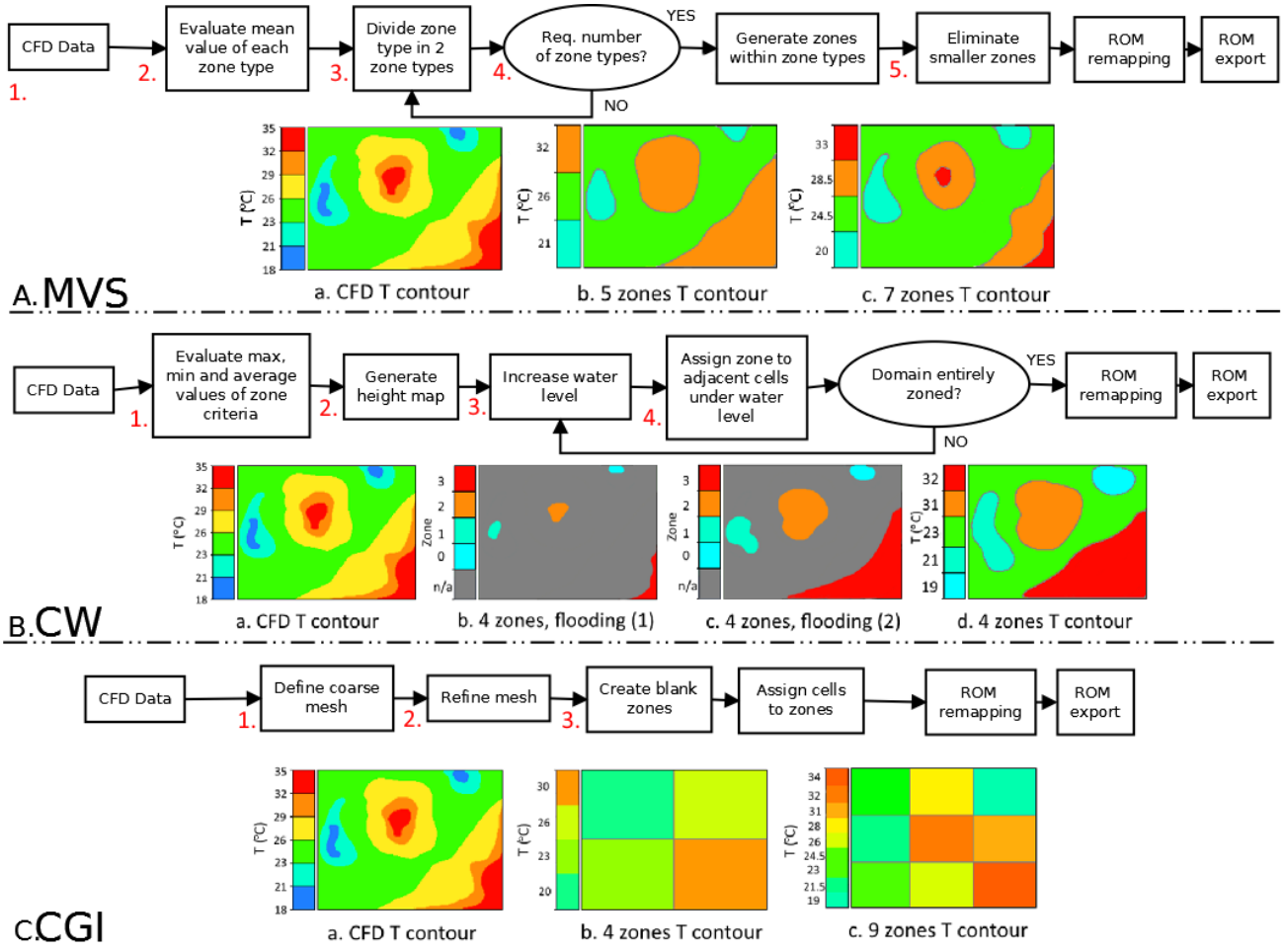

Fig. 2 – Flowcharts of the Mean Values Segmentation (A), Classic Watershed (B) and Coarse Grid Interpolation (C) algorithms

a large number of zones has been created. The algorithm scans the domain one last time to incorporate the smallest zones into larger ones (Step 5) until the desired number of zones is reached.

2.2 Classic Watershed

The Classic Watershed (CW) is a method used in image processing for image segmentation (Soille and Vincent, 1990; Tsukahara et al., 2008). It has been adapted by the authors to ROM generation because its principle is very similar to zone creation. In fact, the CW algorithm is mainly used for gray-scale image segmentation but its fundamental basis can be used for any n-dimensional dataset provided a height map which is defined below can be extracted.

The CW algorithm comprises 4 steps highlighted in Fig. 2.b and described in detail by Soille and Vincent (1990): (1) a height map is generated from the dataset, assigning lower heights to data points that are more relevant to the study; (2) working only with the height map, a water level is defined. The water level will rise (3) at each new iteration, and the data points belonging to the same basin (i.e. adjacent data points under the water level) are clustered together (4). Basins will grow until either they come in contact with another basin or a maximal height is reached. At this point, watersheds are raised.

The CW has many variations depending on the application, such as aerial photography (Wei and Xin, 2010) or medical imaging (Tsukahara et al., 2008).

In this study the height map is generated from temperature values and is normalized so that a lower height (i.e. greater importance) is assigned to local temperature maxima and minima. This ensures that the algorithm prioritizes sections of the domain where temperature differences are important.

Once the height map has been generated and it has been iteratively flooded, a set of clusters of cells belonging to the same basin is obtained. The clusters are remapped to the original domain to obtain a multi-zone model.

2.3 Coarse Grid Interpolation

In the CGI method the CFD results are interpolated to a coarse mesh. For this study the mesh was coarsened from 1.5 million cells to 2–30 cells (see the Results section). In this method, the only zone criteria are the X, Y, Z coordinates of each cell. The algorithm developed for this study is straightforward and is shown in Fig. 2c. The algorithm (1) scans the domain and defines a coarse mesh based on the desired number of zones, (2) refines the coarse mesh to ensure that boundary cells cannot be part of more than one zone, (3) creates empty zones, and (4) assigns cells to a zone depending solely on their coordinates. Similarly to the MVS and CW methods, the mean values of the zone criteria of each zone are computed and assigned to the zone.

3. Results

The data used in this comparative study are taken from the previously validated CFD model of a north-facing office in the Environmental Research Institute (ERI) building at University College Cork (UCC) (Mullen et al., 2015). CFD models were developed using the Phoenics modelling software to generate a database of test cases. For all simulations turbulence is modelled using the steady-state Reynolds Average Navier-Stokes (RANS) approach coupled with the Re-Normalisation Group (RNG) k-ε turbulence model. Air is modelled as an incompressible ideal gas. Phoenics utilises an immersed body technique and consequently the domain is discretized using a Cartesian structured grid with 1,572,165 cells (115x147x93). Constant temperature boundary conditions have been utilised for the ceiling, and the floor, and east and west walls. All other objects are considered adiabatic. All CFD simulations used have been validated with experimental data and previously published (Mullen et al., 2015). All the multi-zone models used to obtain the results are automatically extracted with the only input from the user being the required number of zones. As described in section 1, zone generation is one of the CFD-ROM method's key steps in creating ROMs. It is not possible to predict CFD-ROM's output based

solely on the results of the zone generation algorithm, so the bases of comparison for this study are (1) the accuracy of the multi-zone models, and (2) the time it takes to generate them.

3.1 Accuracy

Quantifying error of multi-zone models provides an indication on the quality of zone generation algorithms. The present study uses a weighted Mean Absolute Error, or WMAE (Equation 1) in Kelvin, to account for differences in cell volumes. The unitary error is defined by the AE or Absolute Error in Kelvin, as shown in Equation 2.

$$WMAE = \sum_{i=o}^{n} \frac{V_i \times AE_i}{n \times V_{domain}} \quad (1)$$

$$AE_i = |T_{CFDi} - T_i| \quad (2)$$

Where n is the number of cells in the domain, V_i is the volume of cell i, AE_i is the absolute error of cell i, V_{domain} is the total volume of the domain, T_{CFDi} is the original CFD temperature of cell i, and T_i is the temperature assigned to the cell after the zoning process is completed. The motivation for weighting the original MAE is to account for varying cell volumes, as the computational grid is finer close to boundaries, where the largest temperature differences also occur. The following results have been obtained for zone numbers ranging from 2 to 27 for CGI, from 3 to 27 for CW, and from 2 to 29 for MVS. Fig. 3 shows a plot of the WMAE versus number of zones N for the three algorithms described above. The results indicate that for this range of numbers of zones the CGI method produces multi-zone outputs in which accuracy is highly dependent on the geometry of the coarse grid. In fact, the more accurate results (for 3, 8, 12, 18, and 27 zones, as shown in Fig. 3) correspond to grids which were divided more finely in the vertical direction in order to better capture thermal stratification.

The mean error of the outputs produced with the CGI algorithm is still higher than the mean error obtained with the MVS algorithm. The mean error of the latter rapidly decreases as more zones are generated before gradually stabilizing. The results presented by Mullen et al. (2015) show a similar trend for the solved ROM.

The results obtained for the CW algorithm indicate that the mechanics of the algorithm are not optimal. The watersheds are placed whenever one basin comes into contact with another, which excludes the possibility of including an intermediate zone between two others to characterize a smooth transition from one zone to another: a zone of high temperature can come in direct contact with a zone of cold temperature without a transition zone in between. Future work includes the investigation of techniques that could address this issue.

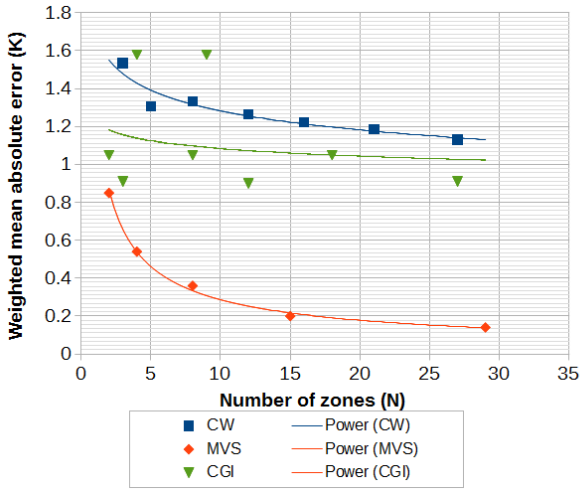

Fig. 3 – Weighted mean absolute error versus number of zones

3.2 Time to Generation

The time to generate multi-zone models is measured with Python's time() function found in the time library. Python 2.7 time() documentation states that whether the code is run on a UNIX system or not, "[time] is the function to use for benchmarking Python or timing algorithms" and that "the resolution is typically better than one microsecond" (Python Software Foundation). The timer is set to start after the initialization step and stops after the algorithm outputs the results in order to capture only the time for the zoning process. Each algorithm was run 15 times and an average time was calculated.

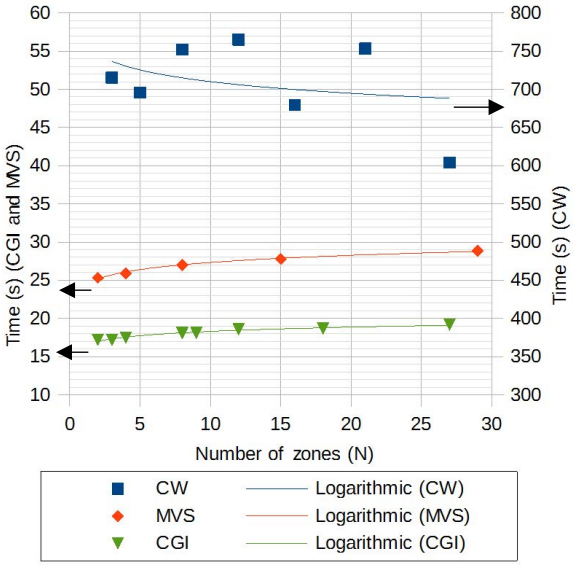

Fig. 4 – Time to generation versus number of zones

It is important to note that zone generation is considered an offline operation, meaning that each ROM has to be generated only once. This is in contrast to ROM solution, an online operation, which is performed multiple times. Therefore the time for zone generation cannot provide an indication of the overall ROM generation and solution performance.

The time to completion for the CGI algorithm is relatively constant, varying only from 17.2 s to 19.2 s across the range of 2–27 zones as shown in Fig. 4. This result was expected, as the CGI algorithm does not divide the domain according to cell conditions but only according to cell coordinates. MVS on the other hand demands more time to generate multi-zone models as the number of zones increases.

This study found times to generation ranging from 25.3 s to 28.9 s for 2–29 zones. Finally, CW in this configuration proved to be very time consuming (right-hand y-axis). The iterative mechanism of the algorithm involves much longer times to completion, as shown in Fig. 4. The decrease in time to completion with increasing zone numbers for CW is explained by the involvement of a height map in zone generation and the creation of watersheds. As watersheds are created during zone generation, some zones will stop growing in size and will then be ignored at the next iteration. The watersheds are automatically placed at different locations depending on the desired number of zones, therefore the size of each zone is not inversely proportional to the number of zones. An improvement in the method could be the creation of watersheds on small slopes (i.e. areas with low temperature differences) rather than points of contact between two zones.

4. Discussion

This study has shown encouraging results for the MVS algorithm, as it yields multi-zone models with a lower error than the CGI algorithm with comparable times to generation. The CW algorithm could benefit from a modification in the mechanics of watershed creation when considering the WMAE, but its high times to generation may prove problematic if its accuracy is not greatly improved. It should however be noted that in the case of a ROM being generated only once, but being solved with different BCs a large number of times, the time for zone generation might be of secondary importance compared to that for ROM solution. Estimates of ROM solution times will be needed to clarify this. Likewise the characterization of error based on the generated zone temperatures rather than the solved ROM temperatures represents only a first step in comparing these algorithms. Additionally it will be important to characterize the ability of ROMs generated by the different algorithms to predict thermal conditions in cases where the ROM BCs differ from the CFD BCs. This is an important capability for the CFD-ROM method as it would enable the replacement of large numbers of CFD simulations with a smaller number of CFD simulations complemented by very large numbers of ROMs spanning the conditions between them.

Nevertheless this comparison has proven extremely valuable as a first step towards the development of a truly systematic CFD-ROM method that can reduce the need for computationally expensive CFD simulations in determining thermal conditions in realistic indoor environments.

The principal limitation of this study is the restriction to zone generation. In fact, the final goal of the CFD-ROM method is the ability to generate

multiple ROMs from a reduced set of CFD simulations. The algorithms described in this study could present different results when solved (1) for the same BCs and (2) for different BCs.

5. Conclusion

The main objective of this study was to compare zone generation algorithms to be used in a novel CFD-ROM method for thermal modelling of indoor environments. At this stage, MVS was found to be the most suitable method. However, the multi-zone models generated by these algorithms could yield different results when solved for sets of BCs different from the original CFD BCs.

Future plans for the CFD-ROM methods include, in order: (1) the development of a ROM solver in the Modelica language, (2) the validation of the CFD-ROM method for the extraction of a single ROM from a CFD simulation, (3) the extension of the method to multiple zone criteria, and (4) the evolution of the method to allow ROM generation for off-design BCs (multi-ROM approach).

This study is the first step towards the ultimate goal which is a rigorous and systematic design of the CFD-ROM method.

Acknowledgements

The authors wish to acknowledge the following funding. T. Marzullo is supported by a Scholarship from the College of Engineering and Informatics at NUI Galway and by HIT2GAP (EU/H2020 Grant Agreement No.: 680708). S. Yousefian is supported by a Scholarship from the College of Engineering and Informatics at NUI Galway and by the Science Foundation Ireland-backed Research Centre for Marine and Renewable Energy (MaREI) with industrial support from the Gas Networks Ireland Gas Innovation Group.

Nomenclature

Symbols

AE	Absolute Error (K)
n	Number of computational cells
T	Temperature (K)
V	Volume (m^3)

Subscripts

CFDi	Computational cell i of the CFD solution
domain	Domain
i	Computational cell i of the multi-zone model

Acronyms

BC	Boundary Condition
CFD	Computational Fluid Dynamics
CFD-ROM	Computational Fluid Dynamics – Reduced Order Model
CGI	Coarse Grid Interpolation
CW	Classic Watershed
GPU	Graphic Processing Unit
MAE	Mean Absolute Error
MVS	Mean Values Segmentation
ROM	Reduced Order Model
WMAE	Weighted Mean Absolute Error

References

Antonioni, G., S. Burkhart, J. Burman, A. Dejoan, A. Fusco, R. Gaasbeek, T. Gjesdal, et al. 2012. "Comparison of CFD and Operational Dispersion Models in an Urban-like Environment". *Atmospheric Environment* 47: 365–372. doi:10.1016/j.atmosenv.2011.10.053.

ASHRAE. 2013. *ASHRAE Handbook Fundamentals*. Atlanta, U.S.A.: American Society of Heating, Refrigerating, and Air Conditioning Engineering, Inc.

GEA. 2012. *Global Energy Assessment - Toward a Sustainable Future*. Cambridge, UK and New York, U.S.A.: Cambridge University Press; and Laxenburg, Austria: International Institute for Applied Systems Analysis.

Harish, V.S.K.V., A. Kumar. 2016. "A Review on Modeling and Simulation of Building Energy Systems". *Renewable and Sustainable Energy Reviews* 56: 1272–1292. doi:10.1016/j.rser.2015.12.040.

Li, X., J. Wen. 2014. "Review of Building Energy Modeling for Control and Operation". *Renewable and Sustainable Energy Reviews* 37. Elsevier: 517–537. doi:10.1016/j.rser.2014.05.056.

Lucia, D.J., P.S. Beran, W.A. Silva. 2001. "Reduced-Order Modeling - New Approaches for Computational Physics". *Progress in Aerospace Sciences* 40(1-2): 51–117. doi:10.2514/6.2001-853.

Mora, L, A.J. Gadgil, E. Wurtz. 2003. "Comparing Zonal and CFD Model Predictions of Isothermal Indoor Airflows to Experimental Data". *Indoor Air* 13(2): 77–85. doi:10.1034/j.1600-0668.2003.00160.x.

Mullen, D.T., M.M. Keane, M. Geron, R.F.D. Monaghan. 2015. "Automatic Extraction of Reduced-Order Models from CFD Simulations for Building Energy Modelling". *Energy and Buildings* 99: 313–326. doi:10.1016/j.enbuild.2015.04.015.

Python Software Foundation. *Python Language Reference, Version 2.7*.

Soille, P., L. Vincent. 1990. "Determining Watersheds in Digital Pictures via Flooding Simulations". *Visual Communications and Image Processing* 1360: 240–250. doi: 10.1117/12.24211.

Tamura, Y., P. Pham Van. 2014. "Development of CFD and Applications: Monologue by a Non-CFD-Expert". *Journal of Wind Engineering and Industrial Aerodynamics* 144: 3–13. doi:10.1016/j.jweia.2015.05.003.

Tan, G., L.R. Glicksman. 2005. "Application of Integrating Multi-Zone Model with CFD Simulation to Natural Ventilation Prediction". *Energy and Buildings* 37(10): 1049–1057. doi: 10.1016/j.enbuild.2004.12.009.

Tsukahara, M., S. Mitrović, V. Gajdosik, G. Margaritondo, L. Pournin, M. Ramaioli, D. Sage, Y. Hwu, M. Unser, T.M. Liebling. 2008. "Coupled Tomography and Distinct-Element-Method Approach to Exploring the Granular Media Microstructure in a Jamming Hourglass". *Physical Review E* 77(6).

van Hooff, T., B. Blocken. 2013. "CFD Evaluation of Natural Ventilation of Indoor Environments by the Concentration Decay Method: CO2 Gas Dispersion from a Semi-Enclosed Stadium". *Building and Environment* 61: 1–17. doi:10.1016/j.buildenv.2012.11.021.

Wei, W., Y. Xin. 2010. "Rapid, Man-Made Object Morphological Segmentation for Aerial Images Using a Multi-Scaled, Geometric Image Analysis". *Image and Vision Computing* 28(4): 626–633. doi:10.1016/j.imavis.2009.10.002.