Creating common exercises for modelling building and district energy systems: lessons learnt from the IBPSA Project 1 - DESTEST

Hicham Johra[1*], Tohid Jafarinejad[2], Arash Erfani Beyzaee[2], Hauke Hirsch[3], Clemens Felsmann[3], Dirk Saelens[2]
[1]Aalborg University, Department of the Built Environment, Aalborg, Denmark
[2]Katholieke Universiteit, Leuven, Belgium
[3]Technische Universität Dresden, Dresden, Germany
* *corresponding author: hj@build.aau.dk*

Abstract

The District Energy Simulation Test (DESTEST) is a series of common exercises about modelling building stocks and district heating networks aiming at testing, benchmarking and verifying different urban-scale energy system simulation tools. For each common exercise, participants are modelling a case with well-defined characteristics, grid topology and boundary conditions. The DESTEST allows participants to discuss common mistakes and pitfalls and define guidelines from the experience and feedback. These common exercises can also be used for training purposes. This article discusses the development process of these common modelling exercises and presents the main lessons learnt during the creation of the DESTEST.

Introduction

To tackle the current challenges in terms of sustainability, greenhouse gas emissions and energy supply reliability, most countries lean towards the development of smart grids coupling different energy networks (e.g., electricity grid, district heating (DH), district cooling, gas distribution) that can operate with a large share of renewable energy sources. Future grids must match the production, transport and distribution of decentralized mostly-intermittent energy sources with the demand of energy end-users. This requires optimising the use of centralized energy storage facilities, demand-side management and building-to-grid services (demand response and energy flexibility strategies). The planning, design, operation and forecasting of such complex systems rely heavily on dynamic multi-domain numerical models and urban-scale digital twins comprising thousands of buildings connected to energy distribution networks.

The trustworthiness of such simulations greatly depends on the accuracy of the employed numerical modelling tools and the skills, guidelines and good practices of the modellers/engineers. To ensure and improve the latter ones, common modelling exercises (CMEs) can be used to benchmark the different numerical modelling tools and serve as training tutorials and practice. The BESTEST (Judkoff and Neymark, 1995; ANSI/ASHRAE Standard 140-2017) is probably the most-known series of modelling cases intended to be used for the validation and benchmarking of building envelope and HVAC (heating, cooling & air conditioning) systems.

Inspired by the BESTEST, special series of CMEs were developed for testing certain specific aspects of the building physics, such as the CME series of the IEA EBC Annex 41 which focuses on the moisture transport and interaction in the building (Rode and Woloszyn, 2009). More recently, a CME was carried out for the comparison of DH network pipe models against measured data on a simple network configuration with controlled boundary conditions (Schweiger et al., 2018).

Under the umbrella of the IBPSA Project 1 (https://ibpsa.github.io/project1/), a series of CMEs has been developed for building energy systems and DH networks. They are denominated as the District Energy Simulation Test: DESTEST. The DESTEST aims at testing, benchmarking and verifying urban-scale energy system simulation tools. The DESTEST is also the occasion for participants to discuss common mistakes and pitfalls that are encountered when modelling such systems. The experience and feedback from these CMEs will be gathered into guidelines for good modelling practices and can serve as training for researchers, engineers and students working with dynamic simulations of urban-scale energy systems. For each CME, different participants are modelling and simulating a given case of buildings and/or energy grid with well-defined characteristics, topology, weather conditions, and boundary conditions. The participants can use any suitable commercial and non-commercial simulation tools (Saelens et al., 2019; Johra et al., 2021).

The process of creating CMEs for building energy systems and district energy networks is much more difficult and time-consuming than one could first think. However, the thought process, discussions and decision-making behind this iterative and collaborative work are not well documented. The aim of this article is thus to report and discuss some of the development processes for the DESTEST CMEs, such as how to structure and define a logical progression in the complexity of the different CMEs, the detail level and constraints in the CMEs' description, or how to compare the results of the different participants and provide feedback. The authors hope that this paper can help future CME initiatives, especially in the building and district energy system community.

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

Setting goals and ambitions for the common modelling exercises

Modellers participating in building and district energy systems CMEs are usually doing it voluntarily, on the side of another research project (often Ph.D. or Postdoc researchers) or as a side task within an engineering company. It is thus rare that they have a budgeted time for these CMEs. Consequently, it is difficult to attract many participants for a series of CMEs. This should be considered when setting the time frame and goals of such activity: how long time is given to complete each CME; how much time in between each CME round (pacing the starting point of each CME); how many participants are expected to deliver exploitable results.

Probably the most important point to address at the start of the creation of CMEs is to clearly define the goals of these CMEs. The presentation of these goals may serve as an introduction to the series of CMEs' descriptions and will ensure that creators of the CMEs are on the same page and understand the point of starting such activities. Similarly, this clear general introduction should serve as a motivation for the modelling participants to engage with reading the CMEs' descriptions and delivering results. At the beginning of each CME's description, a short introduction should state the specific aim and focus of the case. In the case of the DESTEST, the general goal is to offer a comparison and analysis method to test and improve numerical tools for building and district energy systems and help modellers to improve their modelling skills. Each CME then focuses on a specific aspect, such as pressure drops and temperature distribution in a DH network, or the influence of ground modelling details on heat losses in a DH network.

In terms of workload and timeframe, it is suggested to leave around one month for the modellers to complete a single CME in the series. The first CME of a series is usually the longest to complete since the following CMEs are/should be an expansion of the previous ones. The total time to complete and export results of a single CME should be around half a day and up to one day of full-time work (for experienced modellers). A higher workload would probably demotivate the participants to a series of CMEs. The total duration of a CME campaign (conducting, analysing and reporting the entire series of CMEs) should not extend over more than a year. It is, however, appreciable to have the possibility for late participants to join in after and add their results to that of the first cohort of modellers and include them in the analysis and reporting.

As aforementioned, gathering many dedicated participants for an entire series of CMEs is quite difficult. The ambitions and expectations should thus be set accordingly. A balance should be kept between the modification of a CME to include as many numerical tools as possible, and the necessity to draw the line that excludes several tools but preserve the main goal and focus of the CME.

It is suggested that a series of CMEs on a particular topic should not comprise more than 5-10 different CMEs/variations. For each CME, having at least 10 participants providing exploitable results from different numerical tools should be considered a great achievement in the field of building and district energy systems.

Structure of the series of common modelling exercises

To ease the completion of the CMEs series, consecutive CME should be built upon the previous ones. When modelling buildings and district energy systems, the generation of the building geometries and network topology/layout is commonly very time-consuming. The latter should thus be generated in the first CME and have minimum alteration in the following CMEs. The series of CMEs should optimize the reusability of the models created in the previous CMEs and only have a minimum set of parameters being varied each time. A logical progression with an incremental complexity and boundary condition variations thus allows for rapid completion of the CMEs, but also ensures a good testing power of the entire series of CMEs: if too many parameters are changed at once from one CME to the next one, it is difficult to assess which one of these modifications has a significant impact on the simulation results and could cause discrepancies in between the different numerical tools and modellers. The structure and progression overview of the CMEs series should be presented in a schematic and/or summary table at the beginning of the CMEs document (*Figure 1*). It is also suggested to include a list of all the participants to the different CMEs with information on the tool they use.

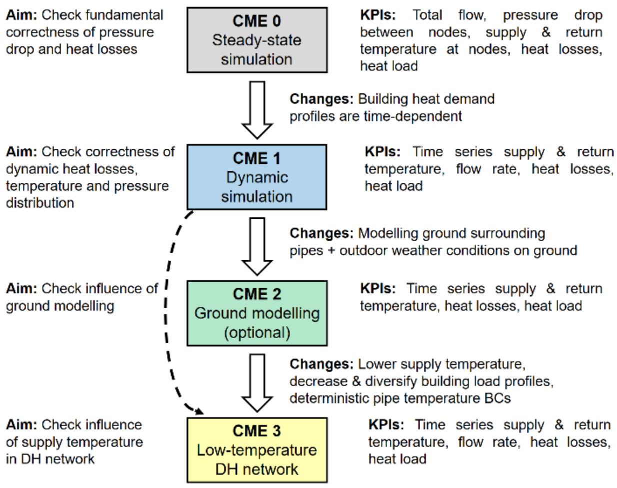

Figure 1: Overview of the series of common modelling exercises for district heating networks.

The DESTEST states that it is expected from the modellers to strictly follow the order of the CMEs and complete them one after another. Each CME has a specific aim/focus and purpose which is ideally outlined in a clear diagram (*Figure 1*). It is important to have identified the sources of discrepancies in a given CME before moving on to the next one. A series of CMEs would typically start with a basic case (e.g., steady-state boundary conditions) and gradually integrate dynamics of the different elements, or explicit modelling of elements that were previously considered as boundary conditions (e.g., explicitly modelling part of the ground around the DH network pipes instead of applying temperature boundary conditions at the surface of the pipes).

In some cases, a well-established testing procedure already exists for part or entirety of the considered system. It is thus important to mention that procedure in the CME description and encourage the modellers to validate their numerical tool with that procedure before conducting the series of CMEs. In the case of building CMEs, for instance, it should be recommended to validate the employed building modelling tool with the well-known BESTEST procedure (ANSI/ASHRAE Standard 140-2017) before starting the series of CMEs.

The overall structure of a CMEs series may be rigid and linear with each CME executed one after the other. In that case, all the redundancy of information from one CME's description to the next one should be strictly limited to avoid confusion of the readers and force the participants to execute the CMEs in the right order. Typically, the first CME's description will comprise all the detailed information to create the geometry of a building or the layout of a thermal network. This information will not be repeated in the next CMEs as the base model is reused and only slightly modified. Only the variations in the characteristics of the study case are described in the next CMEs. It is recommended not to place crucial information in appendices that are shared among different CMEs so that the modellers do not have to browse back and forth in the document and get confused.

In certain cases, one would want to include an optional CME, e.g., when a certain detail or phenomenon is worth being investigated but considered too specific to be inclusive enough among the participants. Branching a CME or making it optional should be clearly indicated in the summary schematic of the CME series (e.g., CME 2 in *Figure 1*). If the modifications of an optional CME are included in the following CMEs, all the necessary substitutes should be provided to the participants. For instance, in the DESTEST DH network CME series, the CME 2 focuses on the modelling of the ground around the pipes. This ground model is supposed to also be included in the following CME 3. However, for the participants using a numerical tool that does not allow for ground modelling, the CME 2 can be skipped and the description of CME 3 contains a link to pre-calculated time series which can serve boundary conditions to the pipe network and substitute an actual dynamic ground model. The modeller can thus carry out the CME 3 without any need of conducting or reading the optional CME's description.

The structure of each CME should always be the same from one to another. Structure consistency throughout the whole description of the CMEs is very important to not confuse the participants who will read the document over a long period of time. The structure adopted for each CME of the DESTEST can be used for inspiration:

- *Introduction and goals*: Specify the focus of that particular CME. It also summarizes the main differences between the current CME and the previous one (if any).
- *General description*: Description of the study case.
- *General assumptions and simplifications*.
- *Specific description*: Several sections describe the different elements of the modelled case in detail, e.g., for a DH network: *DH network layout*, *Pipe properties*, *Substation configuration*, *Heat-carrier fluid properties*, etc. Short specific modelling guidelines and "tips & tricks" should be included to help the participants and reduce the occurrence of mistakes.
- *How to compare and report results*: Include a list of KPIs and simulated variables to extract and report for the comparison of the simulations, together with instructions for formatting the result file and analysing it with a dedicated Python code.
- *Results*: A series of figures for the result comparison and analysis of the current cohort of participants. It includes the list of all participants with the indication of what modelling tools have been used and optional comments or descriptions of the specificities of their models.
- *Discussions: model differences, important points of interest and common modelling mistakes*: A collection of guidelines and discussions around the important points of interest and common modelling mistakes/pitfalls (e.g., incorrect time zone, wrong location coordinate, erroneous geometries, improper decimal delimiter) that have been identified during the execution of the CME by the participants. This is particularly important for the analysis of the results and for the new participants who would like to carry out the CME. These guidelines can also provide good modelling practices for new trainees.
- *Appendices*: Additional discussions, results or guidelines being of interest to certain modellers.

For large CME projects covering very different simulation aspects or domains, it might be beneficial to structure different series of CMEs into an organized collection of CMEs series. For instance, the DESTEST is a collection of CMEs series (see *Figure 2*). Each series is specific to one aspect of the energy communities/district:

- Series of CMEs for building clusters without simulated grids.
- Series of CMEs for grids alone (DH network).
- Series of CMEs for dynamic simulation of the building cluster connected to the grid (either by co-simulation or with tools that can model both the buildings and the energy networks at the same time, e.g., Modelica or TRNSYS).

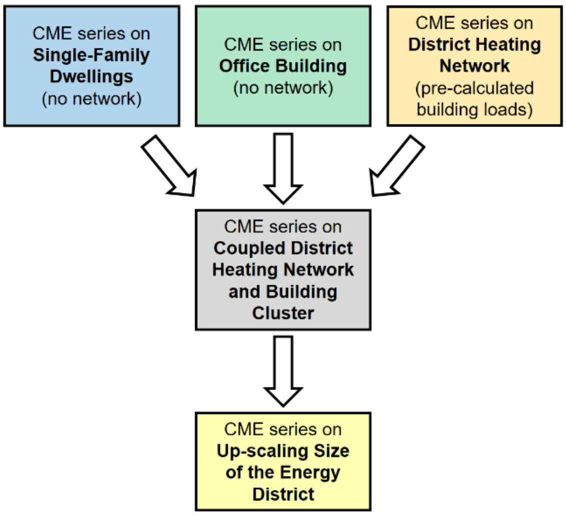

Figure 2: Structure of a collection of CMEs series for testing numerical tools for energy district modelling.

In the first phase, the CMEs series focus on comparing single-domain models (simulation of the buildings without explicit modelling of the network; simulation of the network without explicit modelling of the connected buildings). In the second phase, the CMEs series then investigates multi-domain simulations in which dynamic models of building clusters are coupled with a dynamic model of the DH network. The scalability of such coupled models can thus be tested, e.g., by assessing how the computation time is increasing with the size of the building cluster and network.

Development of the common modelling exercise cases

For each CME, different participants are modelling and simulating the same case of buildings and/or energy grids with well-defined characteristics, geometries, grid configuration, and boundary conditions. The choice or creation of a study case for a given CME or series of CMEs is quite complicated and should, ideally, be a collaborative and iterative process among several potential future modelling participants. The complexity and parameter variations (with regards to the previous CME, if any) should serve the specific purpose of that CME (studying a particular aspect of the simulation) and fit within the logical progression of the series of CMEs. During this design, one should always keep the CME's objectives in mind to justify or limit the complexity of the exercise so that it remains motivating for the participants. The final decision on the CME's case should be a balance between the different expectations, visions, expertise levels and personal interests of the design team. If the study case is too narrow-focussed, it might be restricted to a small number of numerical tools and will not attract enough dedicated and motivated modelling participants.

Ideally, CME's case should not just be an academic (unrealistic) situation, but reflect real-world challenges. For that matter, it is important to reach out to field professionals and experts in the design team. This will also help to make decisions more rapidly for fixing the characteristics of the test cases. However, the geometry of the buildings and the layout of energy networks should be kept as simple as possible to avoid time-consuming model generation, mistakes and frustration.

The number of parameters to be varied from one CME case to the next one is a trade-off between the testing power of the CMEs series and the total number of CMEs in a series. On the one hand, many parameters changing at the same time will make it difficult to interpret the results and identify what is causing errors or deviations. On the other hand, if too few parameters are changed, the CME might appear repetitive and the series might comprise too many of them. This may lead to the participants losing interest or getting overwhelmed by too many CMEs to complete. To optimize the testing power of the CME, adequate key performance indicators (KPIs) should be computed on a limited number of simulated variables and variable time series.

The description and instructions to complete a CME should be as clear and detailed as possible to allow participants to use any suitable commercial and non-commercial simulation tools without extra efforts to convert or calculate necessary parameters. One should keep in mind that modellers might have different habits and tools might have different modelling paradigms: e.g., in the case of the geometry generation of a building, some modellers first set the internal dimensions of a room and extrudes the walls outwards while other modellers first set the external dimensions of a room and extrude the wall inwards. The description should thus include additional information (more than what a specific modeller might need) with detailed schematics and tables that present all the necessary information, characteristics and dimensions from different perspectives (see *Figure 3* and *Figure 4*). Certain additional metrics, e.g., the window-to-wall ratio for buildings or design flow rate for DH network pipes, are very informative and should also be included in the case description.

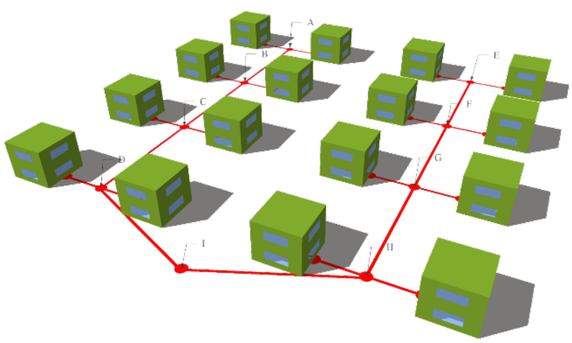

Figure 3: Overview of the DESTEST energy district study case.

If possible, links to download the IFC (Industry Foundation Classes) and/or the GIS (Geographic Information System) files of the study cases should be included in the CME description. All necessary weather and boundary conditions files should be included in different standard formats, such as tabulated text files or .csv files.

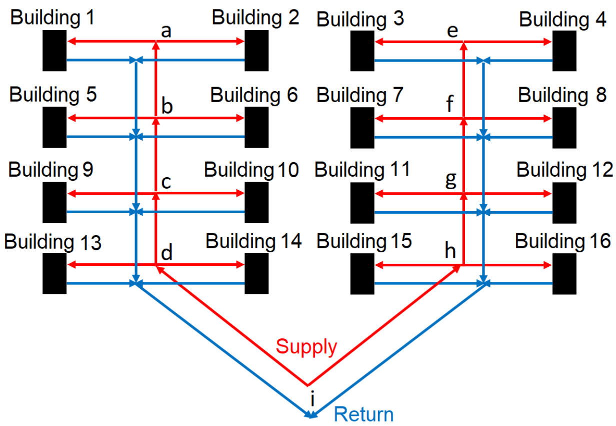

Figure 4: Detailed topology of the DESTEST district heating network study case.

To avoid ambiguity, it is important to state that all the information might not be useful for all modellers. When an important parameter is not specified, it should be clearly indicated in the description that the modellers are free to make the necessary assumption according to their modelling practices or tools' guidelines. However, suggesting general approaches and directions for that particular CME would limit the dispersion of the results. One should not forget to also state what aspects or physics phenomena are not included in the CME, e.g., moisture transport in the indoor environment and building envelop.

If a CME is marked as optional but is the basis for the rest of the series, participants should be given alternatives to skip this case and move on to the next CME. For instance, if an optional CME concerns the modelling of a sub-system that is present in the rest of the series, an appropriate replacement boundary conditions time series should be provided to emulate this sub-system in the subsequent exercises for the participants who are not able or willing to take that particular optional CME.

Finally, it can be interesting to encourage participants to submit suggestions of new cases that can be integrated into the CMEs series as a follow-up or an optional/branched study case.

Collecting, analysing and reporting results

A CME can be based on the comparison of simulation results with empirical results from the measurement of a well-documented real study case. However, such empirical data with sufficient quality, details and documentation is usually very rare in the building and energy community. An alternative to an empirical reference is to use a specific model as a ground truth reference. This might be possible with the analytical solutions of specific cases or when a certain modelling approach is considered significantly more accurate than any of the tested models in the cohort of participants. Unfortunately, this is hardly the case for the vast majority of the practical and realistic configurations of building and district energy systems. One solution to that limitation is to use the mean average (or median or another aggregation method) of the results from all CME participants as a reference. This "average" reference should not be considered a ground truth but it can be used to identify model outliers and discrepancies within the cohort, and point out in certain directions to reduce the result dispersion (Johra et al., 2021).

The comparison of the different simulation results should be performed with a limited number of KPIs, preferably less than 10 to 20 KPIs or output results to report. These KPIs should be chosen to support the analysis and comparison focus of the CME. A very large number of KPIs would lessen the analysis comprehensibility of the CME. The direct comparison of simulated dynamic variables is a distance assessment between two time series: the tested model output and the reference time series. In that case, it is recommended to use common point-to-point normalized comparison metrics for time series, such as the NMBE (Normalized Mean Bias Error) and CVRMSE (Coefficient of Variation of Root Mean Square Error). However, some other time series comparison metrics could be of interest, such as the CVRMSE of the daily amplitude of the data, which does not over-penalize small time shifts and informs about daily dynamics (Johra et al., 2021).

Although time series distance metrics are useful to give a synthetic assessment of the models' output fitting, it is important to visualize the data on a plot. This enables graphical qualitative comparison for a specific period of time. The latter should be chosen so that it maximizes the observability of certain phenomena of interest or expected mistakes and discrepancies (see *Figure 5*).

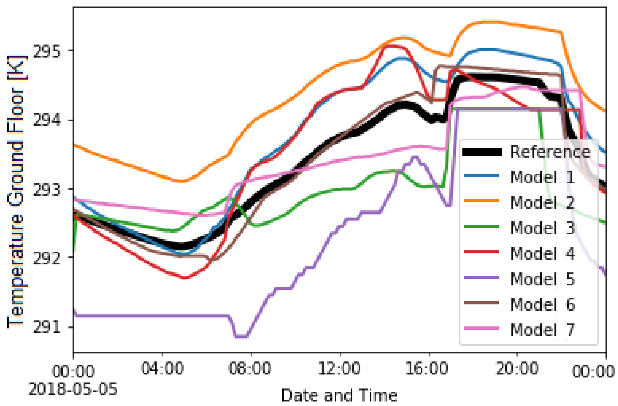

Figure 5: Plot of all models output (simulated variable of interest) on a specific day.

If the goal of the CME is to verify the accuracy of models and/or comprises a large number of (accuracy) KPIs, one could create a meta-KPI (weighted average of all KPIs, or rank-based grade on each KPI) to give a synthetic overview of the performance of the cohort of models.

The actual collection, compilation and comparison of all the results from the different models can be very

https://doi.org/10.1051/e3sconf/202236210005

cumbersome if conducted with spreadsheet software like *Excel*, especially for large datasets. It is recommended to rather develop a simple open-source program or script (with a well-known programming language like *MATLAB* or, preferably, a licence-free one like *Python* or *R*) that will automate the data analysis of each participant's results and compare them with the results of the entire cohort. If the results from the participants are regularly uploaded on an open online repository (e.g., *Github*), the analysis and comparison script can be distributed among the participants so that each of them can directly compare new results with that of the cohort of modellers.

Some participants appreciate having an open-source code for results analysis, but many others might be intimidated or uncomfortable with programming languages, especially if it requires the installation of a specific IDE (integrated development environment) or libraries. It is important to distribute a tested and compiled executable version of the results analysis tool that can be rapidly executed on all common computer operating systems.

However, the most efficient way to currently distribute a simple data analysis procedure is to integrate it into a web application that can be opened and executed in any web browser. A simple but effective graphical user interface (GUI) for a web application can easily be created with *Python* (e.g., with the *Dash Plotly* library) or with *R* (*Shiny* package). This web application can also be used to upload and store the different contributions from the participants. This thus avoids manual collection and compilation work which can be quite demanding if the CME gathers many participants. In any case, the results comparison tool should help to benchmark their simulations against that of the others by means of comprehensive qualitative comparison plots (see *Figure 5*), quantitative comparison plots (see *Figure 6*) and summary tables (see *Figure 7*).

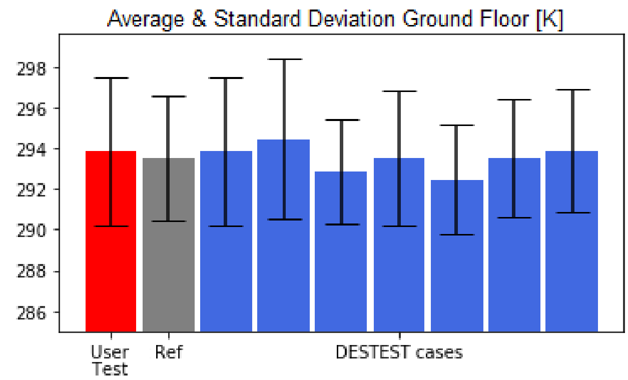

Figure 6: Bar plot of all models output mean average and standard deviation.

Each CME should include guidelines for the interpretation of results comparison so that the participants can detect potential mistakes (especially when their results are outliers within the cohort of modellers) and track the simulation results of their respective modelling tools from one CME to the next one.

Heating Power [W]	Reference	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Minimum	-0.4	0	0	0	0	0	-3.3	0
Maximum	13729.2	16573	16573	13404.7	15880.2	16972.2	16471	18223.9
Mean Average	2186.1	2080.4	1804.1	2190.9	2321.2	2822.8	2572.9	1510.7
Standard Deviation	2975.4	3102.4	3054.8	2885.7	3250.9	3506.6	3430.8	2316.7

Ground Floor Temperature [K]	Reference	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Minimum	289.1	289.1	289.1	289.1	289.1	289.1	289.3	289.1
Maximum	302.8	304.7	305.4	299.9	304	301.7	302.9	302.1
Mean Average	293.5	293.8	294.4	292.8	293.5	292.4	293.5	293.8
Standard Deviation	3	3.6	3.9	2.5	3.2	2.6	2.9	3

First Floor Temperature [K]	Reference	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Minimum	289.1	289.1	289.1	289.1	289.1	289.1	289.1	289.1
Maximum	303.8	304.5	305.1	302.2	305.6	303.9	305.2	301.3
Mean Average	293.5	293.7	294.2	293	293.7	293.3	293.6	293.1
Standard Deviation	3.1	3.4	3.7	2.8	3.4	3	3.1	2.9

Figure 7: Summary table of all models output minimum, maximum, mean average and standard deviation.

The description of the CME should also include simple instructions about how to format the results output file of the different participants. The name of the results file should make it possible to identify the modelling tool used for the CME, the institution of the participant, and some key features or characteristics of the model: e.g., "*Modelica_Buildings_plugflow_AAU_CME_0.csv*".

Finally, all the figures and tables generated by the results analysis tool comparing all the simulations of the modelling cohort should be included (and updated) in the *Results* section of the CME description.

Dissemination and execution of the common modelling exercises

Once the description of the series of CMEs has been refined, reviewed and revised, it must be distributed among the scientific community to attract as many modellers as possible. The use of well-established professional networks such as *LinkedIn* and *ResearchGate* can be very effective. The CME activity can also be shared among large international projects on specific topics, such as the IEA EBC Annex 81, 82, 83 and 84 for building and district energy systems.

Furthermore, these large international projects, such as the IEA EBC Annex projects, usually have activities and subtasks that could accommodate and benefit from an existing CME. It is thus more interesting to adapt or branch an existing CME rather than start one from scratch. This can create a large synergy and considerably expand the community of participants. The CME can also be suggested for a student modelling competition at an international conference within the field of interest.

A good way to motivate academics and researchers to participate in a CME and to provide in-depth analysis and discussions on the results is to plan a peer-reviewed scientific article to be the main output of a series of CMEs.

The presentation of the CME to a group of participants should be done, if possible, during an in-person meeting rather than via an online meeting. However, in-person meetings can be difficult to organize for an unfunded CME activity gathering modellers from all around the world. Once again, international conferences and IEA project meetings can be a good occasion to hold a special workshop or seminar dedicated to that CME and thus create a much larger engagement in that activity.

Regular online meetings to discuss the questions, comments, minor corrections and results of the CMEs series are very important to keep the motivation of the participants high. These meetings also serve as deadlines stimulating participants to produce results. Social interactions between the participants should also be considered one of the key elements of any joint project.

Once the first participants have completed a CME, it is important to include their results and analysis in the *Results* section to stimulate other modellers. Updating the list of participants for each CME with the name of numerical tools used and a list of other possible (but not limited to) tools that could be used for the CME reinforces the motivation among current and new participants as they feel being part of an active community.

Conclusion

The elaboration and execution of common modelling exercises are much more difficult than one could think in the first place. This paper presents a series of suggestions and recommendations to create CMEs for building and district energy systems. They are based on the lessons learnt during the design of the DESTEST, a series of CMEs for the comparison, benchmarking and thorough verification of district energy simulation tools developed in the framework of IBPSA Project 1.

The design of CMEs should be an iterative and collaborative process to capture many points of view from experts and include all possible participants with suitable numerical modelling tools.

Clear goals must be set and stated for each CME and the entire series of CMEs. The choices of the study cases should be made according to these goals and the consecutive CMEs must show a logical progression. The KPIs used to analyze the results should emphasize the modelling aspects that the CME focuses on.

The CME description document should follow a clear structure. It should also include all the necessary information from different perspectives so that modellers can carry out the CME with various modelling tools and paradigms. In addition, a good results analysis tool should be supplied to the participants.

Finally, the social aspects and the integration of the CMEs into a large international project or conference can greatly improve the dynamics of the modelling community and expand the visibility of such activities.

Acknowledgements

This work emerged from the IBPSA Project 1 (https://ibpsa.github.io/project1/), an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA).

References

American Society of Heating, Refrigerating and Air-Conditioning Engineers/American National Standards Institute (2017). Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs (ANSI/ASHRAE Standard 140-2017).

Johra, H., Mans, M., Filonenko, K., De Jaeger, I., Saelens, D., Tvedebrink, T. (2021). Evaluating different metrics for inter-model comparison of urban-scale building energy simulation time series. *Proceedings from the 17th IBPSA Conference BS2021*. Bruges (Belgium), 1-3 September 2021.

Judkoff, R., Neymark, J. (1995). Building energy simulation test (BESTEST) and diagnostic method. NREL/TP-472-6231. *National Renewable Energy Laboratory* (USA).

Rode, C., Woloszyn, M. (2009). Common exercises in whole building HAM modelling. *Proceedings from the 12th IBPSA Conference BS2009*. Glasgow, (Scotland), 27-30 July 2009.

Saelens, D., De Jaeger, I., Bünning, F., Mans, M., Vandermeulen, A., van der Heijde, B., Garreau, E., Maccarini, A., Rønneseth, Ø., Sartori, I., Helsen, L. (2019). Towards a DESTEST: a District Energy Simulation Test Developed in IBPSA Project 1. *Proceedings from the 16th IBPSA Conference BS2019*. Rome (Italy), 2-4 September 2019.

Schweiger, G., Heimrath, R., Falay, B., O'Donovan, K., Nageler, P., Pertschy, R., Engel, G., Streicher, W., Leusbrock, I. (2018). District energy systems: Modelling paradigms and general-purpose tools. *Energy* 164, 1326-1340.