Statistical Methodologies for Verification of Building Energy Performance Simulation

Amin Nouri[1], Jérôme Frisch[1], Christoph van Treeck[1]
[1]RWTH Aachen University, Institute of Energy Efficiency and Sustainable Building (E3D), Aachen, Germany

Abstract

Building energy performance simulation tools are being increasingly deployed by researchers and professionals to predict the thermal behavior of buildings. Validation methods are applied to ensure the accuracy of simulation results. Results and methods described in this paper are resulting from a research project funded by the German Federal Ministry for Economic Affairs and Energy (BMWi), which addresses the development of quality standards for building and systems energy performance simulations. The objective of this project is to develop a validation methodology, to define standards for simulation applications and to transfer them into planning practice. Another aspect of the research project is the development of a platform to provide a facility for defining individual test cases, to create individual simulation code, to perform a comparative validation, and to evaluate the accuracy of the simulation tools. The first part of this paper interprets the sources of error and uncertainty in building simulation and presents five statistical indices Mean Bias Error (MBE), Normalized Mean Bias Error (NMBE), Root Mean Squared Error (RMSE), Coefficient of Variance of Root Mean Squared Error (CVRMSE) and Coefficient of Determination (R^2), which are used in the verification and validation of the building energy performance simulation tools. The second part describes a systematic set of test cases based on the ANSI/ASHRAE Standard 140 and discusses the simulation approach in Modelica/Dymola. The third part presents the development and implementation of statistical indices and evaluates these indices' ability to deduce deviations in simulation results based on their interpretation. It is shown that the R^2 give a different interpretation of the discrepancies between predicted and reference values and it is recommended to apply this index only along with other statistical indices to correctly evaluate the model accuracy.

Key Innovations

- Definition of three test cases to assess statistical indices in building energy performance simulation.
- Investigation into the five statistical indices to assess their sensitivity to simulation results concerning heating and cooling loads, while the two parameters “infiltration rate” and “internal heat gains” are changed.
- The coefficient of determination (R^2) should be used only along with other statistical indices to correctly grasp the discrepancies between predicted and reference values.

Practical Implications

There are lots of uncertainty sources during the simulation, which affect the simulation results. Therefore, validation and verification of the results is inevitably. In this study, systematic test cases based on the ANSI/ASHRAE Standard 140 are defined. To evaluate and quantify the accuracy of the model, statistical indices are used.

Introduction

Over the past decades, Building Energy Performance Simulation (BEPS) tools have been taken into consideration for the scientific community as well as industrial society to predict the thermal behavior of buildings. In general, modeling and simulation techniques are employed to determine the thermal characteristics of buildings. However, studies (Jensen, 1995; Strachan, 2008; Mantesi et al., 2018; Haves et al., 2019) have revealed that there are substantial discrepancies between predictions of various BEPS tools. A sufficient degree of confidence in predictions is absolutely vital to go through the whole decision-making process. Validation and verification procedures are essential elements within the development of BEPS programs to assess reliability of the simulation results as well as to find and eliminate eventual bugs in the simulation software itself. Judkoff et al. (1982) developed an overall validation methodology which contains three different kinds of techniques: empirical validation, analytical verification, and software-to-software comparative tests. Numerous studies have been applying these techniques to ensure the accuracy of simulation results for years (Witte et al., 2001; van Treeck et al., 2009; Ryan and Sanquist, 2012; Mazzeo et al., 2019; Nouri et al., 2020). Furthermore, several standardized procedures exist to assess building performance simulation tools, such as ANSI/ASHRAE Standard 140 (2017) as well as the German VDI 6007 and VDI 6020 guidelines.

The BEPS is almost always accompanied by vast number of uncertainties such as uncertainties in the physical model and building geometry, discretization errors, round-off errors, programming errors, uncertainties in semantic interpretation and incorrect data transformations. These uncertainties may cause the apparent discrepancies between predicted and reference values. A challenging task in the field of verification and validation is the assessment and analysis of the substantial gap between simulation results. It is therefore essential to characterize the accuracy of the predictions. Statistical indices are used to quantify the accuracy of the models (also referred to as goodness-of-fit). There are various types of statistical indices which have been frequently used. Lauster et al. (2014) used the Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) to compare simulation results of the indoor air temperature to the Guideline VDI 6007. Ruiz and Bandera (2017) evaluated the most common validation measurements (uncertainty indices) in the calibration process. Mantesi et al. (2018) applied the absolute difference and the Normalized Root Mean Square Error (NRMSE) to analyze the annual heating and cooling energy consumption and hourly surface temperature of a wall for three different wall constructions within different BEPS tools. Markovic et al. (2019) assessed the discrepancies between the measured and simulated indoor air temperatures for each time-steps applying the indices Mean Bias Error (MBE), RMSE and Coefficient of Variance of Root Mean Squared Error (CVRMSE). Risch et al. (2021) evaluated calibrated models using R^2 and the coefficient of variation of the root mean squared error (CVRMSE).

However, each statistical index may have different properties and emphasize different aspects of deviations in simulation results. Reddy and Claridge (2000) conceptually illustrated that there are considerable differences between the R^2 and CVRMSE, i.e., the value of R^2 may indicate a good model, while the value of CVRMSE may suggest a poor model. Hence, it is not recommended to rely only on one statistical index to assess the model accuracy. The first part of this paper discusses the sources of uncertainty in BEPS and represents five statistical indices in the context of the validation of simulation results. The second part describes three test cases based on the ANSI/ASHRAE Standard 140 provided by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) which will apply the indices from part one. The corresponding test cases are modelled and simulated in Modelica/Dymola using the AixLib library (Müller et al., 2016). The simulation settings and the version of the library used for the simulation are described in the following sections. The third and last part evaluates the statistical indices and indicates the emphasis of each index on the simulation behavior before concluding the paper with an outlook.

Uncertainties in BEPS

During modeling and simulation, considerable uncertainties may arise form a wide variety of sources. De Wit and Augenbroe (2002) classified the sources of uncertainty in the context of building energy performance simulation into four major categories:

(i) Modelling uncertainties, which are directly related to the errors arising from the modeling assumptions and simplifications of the complexity of the physical processes.
(ii) Specification uncertainties, which emerge from the incorrect or missing information on the properties of the building, including building geometry and the properties of the materials and components.
(iii) Numerical uncertainties, which are associated with the errors arising from the discretization and simulation of the model.
(iv) Scenario uncertainties, which introduce errors in external conditions of the case study (i.e. building), including outdoor climatic conditions, internal heat gains in the building (from people, lighting and equipment), behavior of occupants and solar shadings control.

In this paper, the uncertainties of the two parameters “infiltration rate” and “internal heat gains” are taken into account that may belong to modelling and scenario uncertainties, respectively. To quantify and ascertain the discrepancies between predicted and reference values, statistical indices may then be applied which are described in the following section.

Methods

The accuracy of BEPS predictions can be evaluated by statistical indices. This paper introduces five statistical indices including Mean Bias Error (MBE), Normalized Mean Bias Error (NMBE), Root Mean Squared Error (RMSE), Coefficient of Variance of Root Mean Squared Error (CVRMSE) and Coefficient of Determination (R^2) which are commonly used in literature. As discussed later, each of these metrics grasps different behaviors and has different and specific implications on discrepancies in the BEPS results. In the following sections, the indices used in this study are described in detail.

Mean Bias Error (MBE)

The MBE index indicates the average of the model errors, where the term “error” refers to the difference between the reference value and the predicted value. The MBE is calculated by the equation (1):

$$MBE = \frac{\sum_{k=1}^{n}(r-p)}{n} \tag{1}$$

where r is the reference value, p is the predicted value, and n is the total number of values (i.e., number of data points).

This index provides information about the average underprediction or overprediction, i.e., it can represent both positive and negative values. A value of MBE close to zero indicates that the average of the predicted and reference values is very close together, which is obviously desirable. The values of the MBE have the same units as the datasets. The main drawback of MBE is that positive and negative errors can cancel each other out. The cancellation of the errors can decrease the overall MBE value.

Normalized Mean Bias Error (NMBE)

The NMBE is described by dividing the MBE by the mean of the reference values and is calculated according to equation (2):

$$NMBE = \frac{MBE}{\bar{r}} * 100 = \frac{\sum_{k=1}^{n}(r-p)}{n*\bar{r}} * 100 \qquad (2)$$

where $\bar{r}$ is the average of the reference values. This index is a non-dimensional index. However, the issue of compensation of positive and negative deviations remains. According to the ASHRAE Guideline 14 (2002) and Measurement and Verification of Federal Energy Projects (FEMP, 2015), the model can be assumed to be calibrated if the NMBE lies within ±10% when using hourly data or within ±5% with monthly data.

Root Mean Squared Error (RMSE)

The RMSE indicates the square root of the average of the squared model errors. Therefore, this index always returns positive values and is calculated according to equation (3):

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}(r-p)^2}{n}} \qquad (3)$$

A lower value of the RMSE means that the prediction values fit the reference values better. This index has the same unit as the datasets. The significant advantage of using the RMSE is that there is no cancellation of positive and negative errors. However, RMSE is highly sensitive to outliers, i.e., even few large errors may result in a substantial increase in the value of RMSE.

Coefficient of Variance of Root Mean Squared Error (CVRMSE)

The CVRMSE is a normalized form of RMSE and is represented by dividing the RMSE by the mean of the reference values $\bar{r}$ and is calculated according to equation (4):

$$CVRMSE = \frac{RMSE}{\bar{r}} * 100 = \frac{\sqrt{\frac{\sum_{k=1}^{n}(r-p)^2}{n}}}{\bar{r}} * 100 \qquad (4)$$

This index indicates a non-dimensional result and there is no compensation of errors. Therefore, the CVRMSE together with the NMBE are recommended by ASHRAE Guideline 14 (2002) and FEMP (2015) to evaluate the accuracy of the model. These guidelines suggest a value of CVRMSE within ±30% for hourly data or ±15% for monthly data. Here, a larger range is allowed compared to the NMBE. The reason can be that the errors in each data point are squared and there is no cancellation of positive and negative values. Therefore, the CVRMSE can increase rapidly in some data points.

Coefficient of determination (R^2)

The coefficient of determination R^2 describes how well the prediction values (i.e., the regression line) fit the reference values and is defined according to equation (5):

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(r-p)^2}{\sum_{k=1}^{n}(r-\bar{r})^2} \qquad (5)$$

The value of R^2 is ranging between 0 and 1, where a value of 0 presents no correlation between the predicted and reference values, whereas a value of 1 indicates that the predicted values perfectly match the reference ones, that is, the closer R^2 is to 1, the greater predictive accuracy may be ensured. When the value of R^2 is close to 1, it indicates that the data points may lie around the y = x line (i.e. perfect fit line). It is nevertheless noteworthy that a perfect fit can lead to the value of R^2=1, even if there is a bias (systematic error) in the predicted values.

Test cases, modeling, and simulation

In order to properly assess the previously described statistical indices in the building energy performance simulation context, three test cases are defined which are based on the ANSI/ASHRAE Standard 140 (2017). Test case 1 (TC01) represents exactly Case 600 of the ANSI/ASHRAE Standard 140. In this study, TC01 is considered as a reference model and establishes the base case for two further test cases.

TC01 consists of a rectangular building with one single zone and two windows on the south surface without any shading. The geometry of the building model is shown in Figure 1. Each window has an area of 6 m^2. All walls are modeled as exterior walls with dimensions of 2.7 m height, 8 m width and a 6 m length. The building has a lightweight construction, which means that TC01 utilizes lightweight walls, floor, and roof. Table 1 describes the material properties for the TC01. The insulation of the floor has been opted extremely thick to thermally decouple the floor from the ground and to diminish the impact of heat losses to the ground on the heating and cooling loads. The infiltration rate is set to 0.41 air changes per hour (ach). The internal heat gains are set to 200 W (100% sensible), and the radiative and convective portion of the internal gains are taken as 60% and 40%, respectively. In this test case, the thermal zone is assumed heated and cooled by an ideal HVAC system (i.e., 100% effective efficiency). The heating set-point temperature is 20 °C and the cooling set-point temperature is set to 27 °C.

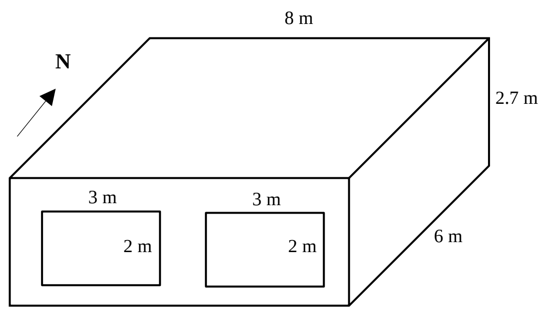

Figure 1: Building geometry of test cases based on the ANSI/ASHRAE Standard 140.

In this paper, two further test cases (test case 2 and test case 3) are introduced which are based on TC01 except for small changes in input parameters. It is important to note that these changes are consciously made to evaluate the uncertainties and to analyze the sensitivity of the indices in test case 2 (TC02) and test case 3 (TC03). Based on the presented literature review, it was identified that two parameters "infiltration rate" and "internal heat gains" may have a significant impact on the simulation results concerning heating and cooling loads (de Wit and Augenbroe, 2002; Hopfe and Hensen, 2011). Hence, in this study, these two parameters were selected.

In TC02, the parameter "infiltration rate" is changed to 0 ach, while in TC03, there are no "internal heat gains" (i.e., the value of the internal gains is set to 0 W). All other input parameters of TC02 and TC03 remain unchanged compared to TC01. Therefore, the variations in the output of the simulations in TC02 and TC03 are due to the changes in two parameters "infiltration rate" and "internal heat gains", respectively.

Table 1: Material properties for the test cases.

Components (outside to inside)	**k (W/mK)**	**C_p (J/kgK)**	**Thickness (m)**	**Density (kg/m³)**
Exterior Wall				
Wood Siding	0.14	900	0.009	530
Fiberglass quilt	0.04	840	0.066	12
Plasterboard	0.16	840	0.012	950
Floor				
Insulation	0.04	0.0001	1.003	0.0001
Timber flooring	0.14	1200	0.025	650
Roof				
Roof decking	0.14	900	0.019	530
Fiberglass quilt	0.04	840	0.1118	12
Plasterboard	0.16	840	0.010	950

All test cases are modelled in Modelica/Dymola using AixLib library version 0.7.3 and are simulated for an entire year. The simulation results of the heating and cooling loads are analyzed by the previously mentioned statistical indices for the month of January. In the following section, the simulation results and the assessment of statistical indices are described.

Results and Discussion

The hourly simulation results of the ambient temperature, heating and cooling loads for the month of January are depicted in Figure 2, Figure 3 and Figure 4, respectively. In the following sections, a detailed comparison between TC01 (reference model) and the two other cases (TC02 and TC03) is made. In addition, the ability of the described statistical indices to represent deviations in simulation results is discussed.

Comparison of TC01 (reference model) and TC02 (no infiltration rate)

From Figure 3, it can be qualitatively seen that the heating load of TC02 represents the deviation from the reference model. Thus, the statistical indices are applied to investigate these discrepancies.

Table 2 presents the results of the statistical indices for the heating load, while the scatterplot of reference versus predicted values of TF02 for the heating load is depicted in Figure 5. The vertical axis represents the heating load of the predicted model and the horizontal axis indicates the values of the reference model. The grey line presents the perfect fit (y = x line).

The value of the MBE = 291.19 W indicates the average of the difference between reference and predicted values. The positive value of MBE indicates that the average of the predicted values (TC02) during the time period (i.e. in January) is lower than the average of the reference values. The value of the RMSE is 387.20 W and is higher than MBE. Here, the reason is that the difference between reference and predicted value in each data point is squared, in accordance with equation (3). The indices NMBE and CVRMSE show relatively high values and they are outside of the range of the values recommended by ASHRAE Guideline 14 and FEMP. Nevertheless, the R^2 evaluates as 0.87 which can be considered as an acceptable value ($R^2 > 0.75$), as shown in Figure 5. In this case it can be inferred that the R^2 presents a different interpretation of the predicted values and may not describe the deviations from reference model realistically. The reason is that the R^2 can quantify only the dispersion of the values and the underprediction or overprediction of the model cannot be considered. Therefore, the value of R^2 can be close to 1, although the accuracy of the model is not good.

Figure 5 shows that there are significant deviations for values below 1500 W. The reason is that in this range, the infiltration rate is the main driving factor. Here, the simulation results of TC02 (predicted values) are lower than the reference values, since the infiltration rate is set to zero and lower heating loads are needed. In the range above 1500 W, the outdoor temperature plays a more important role.

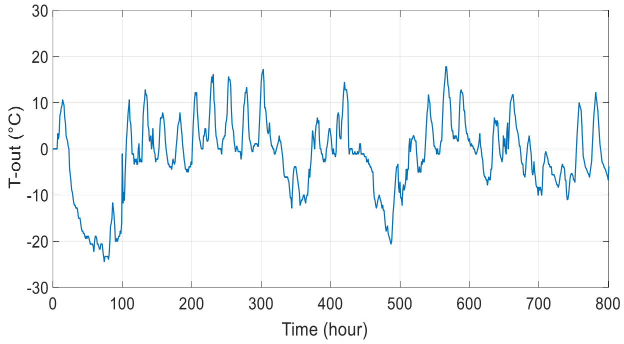

Figure 2: Hourly ambient temperature for the month of January.

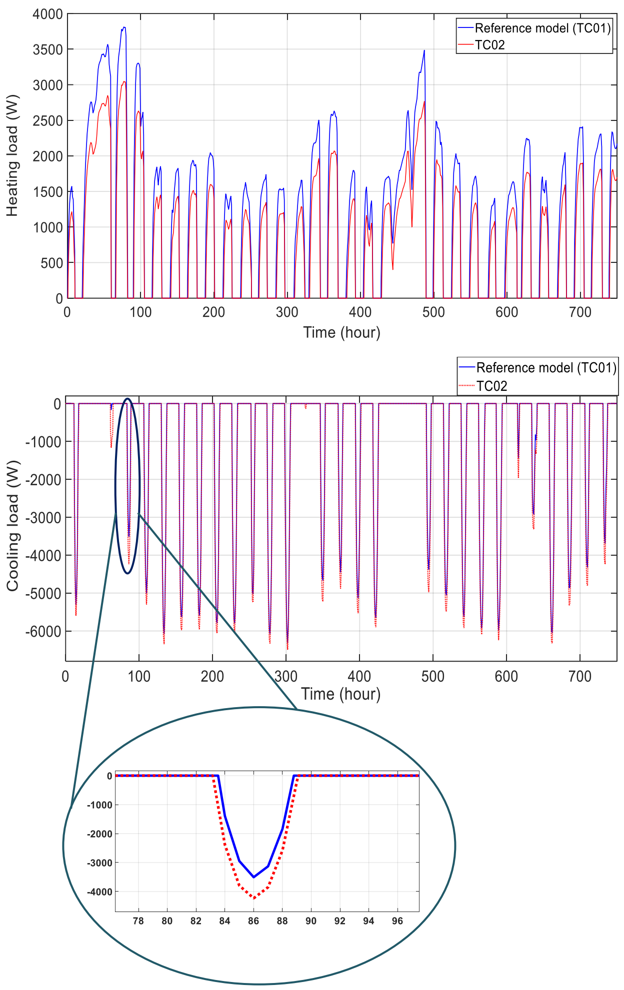

Figure 3: Hourly heating and cooling load of TC01 (reference model) & TC02 for the month of January.

Figure 4 illustrates the discrepancies between predicted and reference values for the cooling load. The results of the implemented statistical indices for the cooling load are shown in table 3. The values of NMBE and CVRMSE are -14.43% and -34.23%, respectively. Here, these indices are outside the range of the values recommended by ASHRAE Guideline 14 and FEMP. However, the value of the R^2 is close to one (R^2 = 0.97). As with the heating load, the R^2 interprets the predicted values in a different manner i.e. R^2 can take the dispersion between the predicted and reference values into account and cannot deal with the bias in the predicted values. Therefore, in this case, die indices NMBE and CVRMSE are more sensitive to represent the deviations from the reference model. Figure 6 shows the scatterplot of reference versus predicted values of TF02 for the cooling load. It should be noted that in this study, the cooling load is considered as a value with a negative sign. Therefore, the average of the reference values is negative. For this reason, the NMBE is negative, in accordance with equation (2).

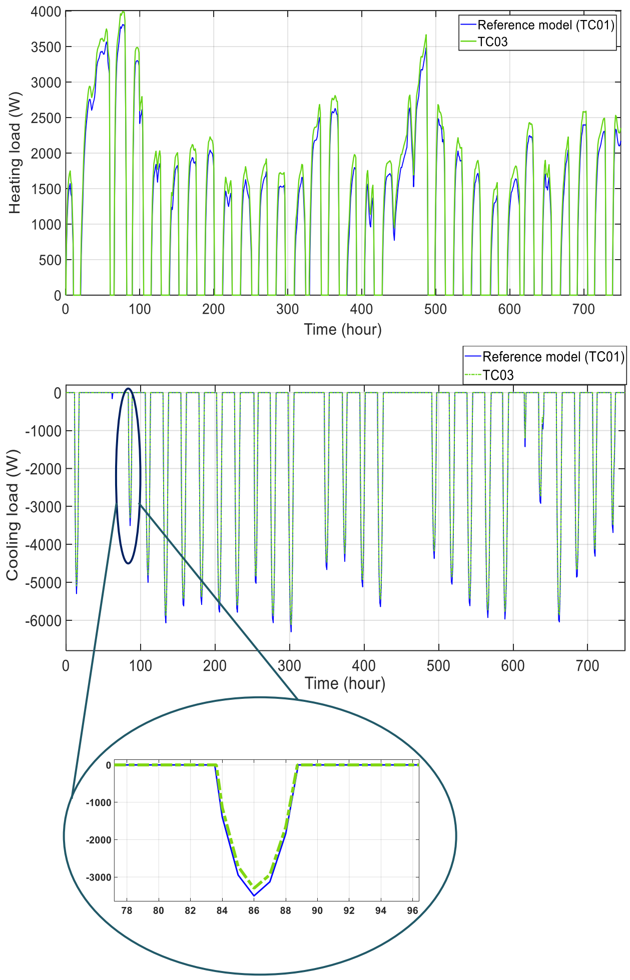

Figure 4: Hourly heating and cooling load of TC01 (reference model) & TC03 for the month of January.

Table 2: Results of statistical indices for heating load.

Indices	Comparison of TC01 (reference model) and TC02	Comparison of TC01 (reference model) and TC03
MBE [W]	291.19	-122.60
NMBE [%]	26.17	-11.02
RMSE [W]	387.20	155.20
CVRMSE [%]	34.80	13.95
R^2	0.87	0.98

Table 3: Results of statistical indices for cooling load.

Indices	Comparison of TC01 (reference model) and TC02	Comparison of TC01 (reference model) and TC03
MBE [W]	120.64	-43.99
NMBE [%]	-14.43	5.26
RMSE [W]	286.20	175.28
CVRMSE [%]	-34.23	-20.97
R^2	0.97	0.99

MBE and NMBE indicate that the average of the predicted values (TC03) in January is higher than the average of the reference values, according to equations (1) and (2). The values of the RMSE and CVRMSE are 155.20 W and 13.95 %, respectively. Due to calculating the square root of the average of the squared errors (equations (3) and (4)) the values of the RMSE and CVRMSE are positive. The NMBE is outside of the range of the criteria given by ASHRAE Guideline 14 and FEMP, however CVRMSE = 13.95% lies within the range of the defined criteria (±30%). The scatterplot of reference versus predicted values of TF03 for the heating load is shown in Figure 7. The R^2 evaluates as 0.98. In this case, the R^2 is more reliable than in the previous example.

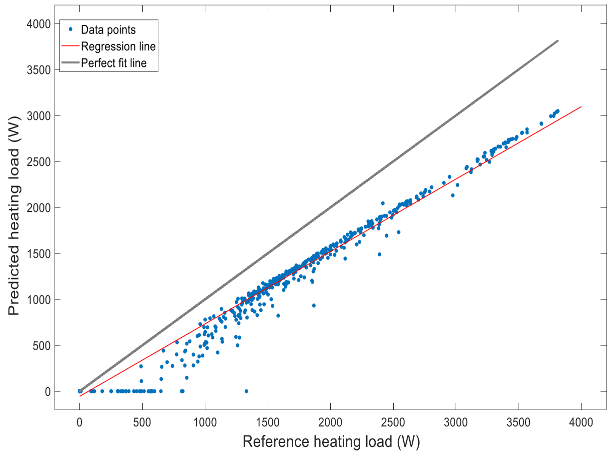

Figure 5: Scatterplot of reference (TC01) vs. predicted (TC02) values for heating load (R^2=0.87).

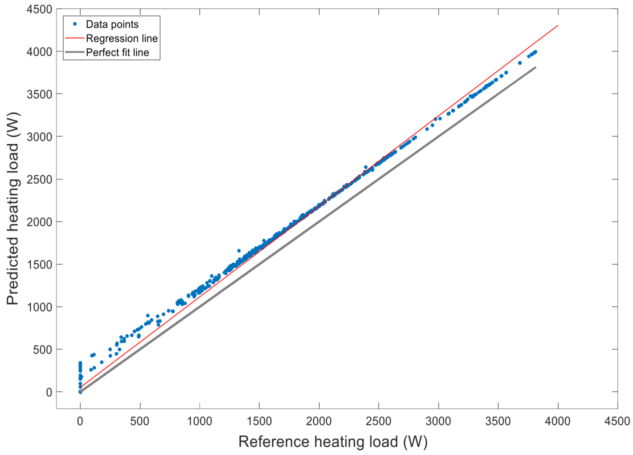

Figure 7: Scatterplot of reference (TC01) vs. predicted (TC03) values for heating load (R^2=0.98).

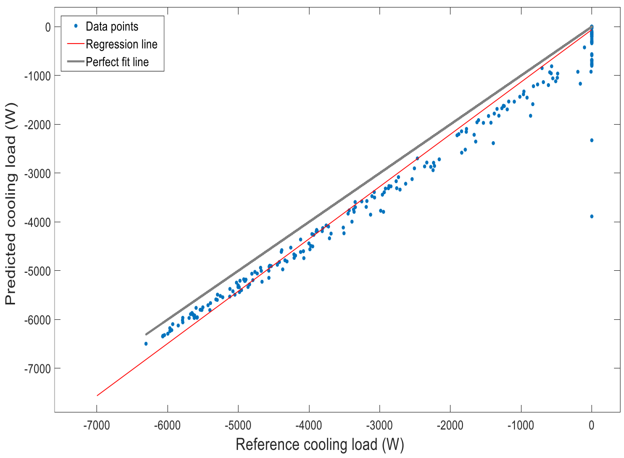

Figure 6: Scatterplot of reference (TC01) vs. predicted (TC02) values for cooling load (R^2=0.97).

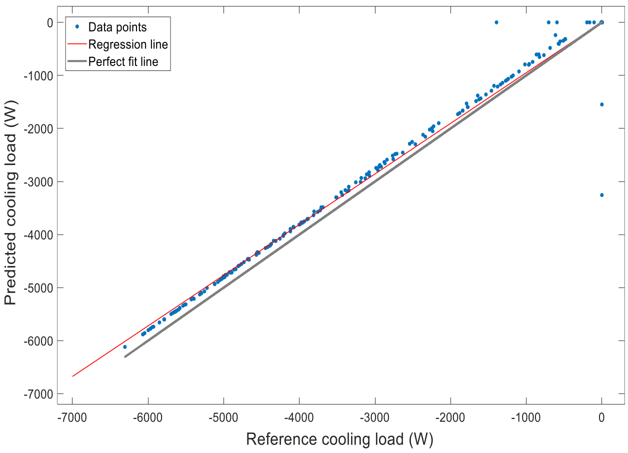

Figure 8: Scatterplot of reference (TC01) vs. predicted (TC03) values for cooling load (R^2=0.99).

Comparison of TC01 (reference model) and TC03 (no internal heat gains)

In TC03, the input parameter "internal heat gains" is changed to 0 W. Taking the heating load into account, the MBE of the comparison of TC01 and TC03 is -122.60 W and the NMBE is -11.02 %. The negative values of the

With respect to the cooling load, the MBE was calculated as -43.99 W and the RMSE as 175.28 W, where it shows a considerable difference between MBE and RMSE. The reason can be due to a cancellation of positive and negative errors in the calculation of the MBE, in accordance with equations (1). The values of NMBE and CVRMSE are 5.26% and -20.97%, respectively. Here, both NMBE and CVRMSE are within range of the criteria recommended by ASHRAE Guideline 14 and FEMP. Figure 8 presents the scatterplot of reference versus predicted values of TF03 for the cooling load. The value of the R^2 results as 0.99, which is very close to one. In this case, the index R^2 together with NMBE and CVRMSE can be considered to interpret the predicted values correctly.

Conclusion and Outlook

Building simulation tools are used to predict the energy performance of buildings. However apparent discrepancies can occur between predicted and reference values. The purpose of this research study is to assess these discrepancies using five statistical indices MBE, NMBE, RMSE, CVRMSE and R^2. Three test cases based on the ANSI/ASHRAE Standard 140 were defined and simulation approaches in Modelica/Dymola were discussed. A detailed comparison between reference model and the two other cases was made to determine the impact of changes induced by modification of two input parameters “infiltration rate” and “internal heat gains” on simulation results and furthermore to evaluate the statistical indices.

The comparison between reference model and TC02, where the infiltration rate was set to zero, indicated that the indices MBE, NMBE, RMSE, CVRMSE were reliable to describe the deviations from reference values for both heating and cooling load. However, the coefficient of determination R^2 gave a different interpretation of the discrepancies and could not realistically represent the deviations from the reference model in both heating and cooling loads.

Taking the input parameter “internal heat gains” into account, the comparison between reference model and TC03 (internal heat gains were set to zero) presented that the index R^2 was more sensitive to indicate the deviations from reference values.

In conclusion, it is recommended that in both cases (TC02 and TC03) the R^2 should be used only along with other statistical indices to correctly grasp the discrepancies between predicted and reference values.

The research project “SimQuality” deals with the development of a validation methodology for the building energy performance simulation and carries out multiple comparative analyses using several simulation tools, which is beyond the scope of this paper. The interested reader is referred to Nouri et al., 2020 as well as SimQuality-website, available at https://simquality.de/, for a comprehensive overview of the whole project.

Future work will focus on the following four key aspects:
(1) considering other sources of uncertainty such as the ones related to the construction of the building (e.g. thermal conductivity), building geometry, solar radiation, and weather conditions,

(2) developing further test cases, and

(3) applying and evaluating other statistical indices to building energy performance simulation cases.

(4) evaluating the impact of validated BEPS results on the thermal comfort.

Acknowledgement

This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi), under the research project “SimQuality” (promotional reference 03ET1570B).

References

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

ASHRAE Guideline 14 (2002). *Measurement of Energy and Demand Savings*. American Society of Heating, Refrigeration and Air-Conditioning Engineers, Atlanta.

FEMP (2015). M&V guidelines: *measurement and verification for performance-based contracts* - version 4.0. U.S. Department of Energy Federal Energy Management Program. Washington, D.C., USA.

De Wit, S. and Augenbroe, G. (2002). Analysis of uncertainty in building design evaluations and its implications, *Energy and Buildings 34*, 951–958

Haves, P., Ravache, B., Fergadiotti, A. and Kohler, C. (2019). Accuracy of HVAC load predictions: Validation of EnergyPlus and DOE-2 using an instrumented test facility. *Proceedings of the 16th IBPSA Conference*, 2-4 September, Rome.

Hopfe, C.J. and Hensen, J.L.M. (2011). Uncertainty analysis in building performance simulation for design support. *Energy and Buildings 43*, 2798–2805.

Jensen, S.Ø. (1995). Validation of building energy simulation programs: a methodology. *Energy and Buildings 22*, 133-44.

Judkoff, R., Wortman, D. and Burch, J. (1982). Empirical validation of building energy-analysis simulation programs: a status report. U.S. Department of Energy, Washington, D.C. USA.

Mazzeo, D., Romagnoni, P., Matera, N., Oliveti, G., Cornaro, C. and De Santoli, L. (2019). Accuracy of the most popular building performance simulation tools: experimental comparison for a conventional and a PCM-based Test Box. *Proceedings of the 16th IBPSA Conference*, 2-4 September, Rome.

Mantesi, E., Hopfe, C.J., Cook, M.J., Glass, J. and Strachan, P. (2018). The modelling gap: Quantifying the discrepancy in the representation of thermal mass in building simulation. *Building and Environment 131*, 74-98.

Markovic, R., Grintal, E., Nouri, A., Frisch, J. and van Treeck, C. (2019). Right on Time – Exploring Suitable

Time Discretization for Occupant Behavior Co-Simulation. Proceedings of the 16th IBPSA Conference, 2-4 September, Rome.

Müller, D., Lauster, M., Constantin, A., Fuchs, M. and Remmen, P. (2016). AixLib - An Open-Source Modelica Library Within the IEA-EBC Annex 60 Framework. *Proceedings of BauSIM 2016*, September 14-16. Dresden

Nouri, A., Nicolai, A., Krämer, B., Hirth, S., Agudelo, J., Seifert, C., Malhotra, A., Madjidi, M., Frisch, J. and van Treeck, C. (2020). Entwicklung von Qualitätsstandards für die energetische Gebäude- und Anlagensimulation als Planungswerkzeug. *Proceedings of BauSIM 2020*, 23-25 September. Graz.

Reddy, T.A. and Claridge, D.E. (2000). Uncertainty of “measured” energy savings from statistical baseline models. HVAC&R Research 6 (1), 3-20.

Risch, S., Remmen, P. and Müller, D. (2021). Influence of data acquisition on the Bayesian calibration of urban building energy models. *Energy and Buildings 230*.

Ruiz, G.R. and Bandera, C.F. (2017). Validation of calibrated energy models: common errors. *Energies 10*, 1587.

Ryan, E.M. and Sanquist, T.F. (2012). Validation of building energy modeling tools under idealized and realistic conditions. *Energy and Buildings 47*, 375–382.

Strachan, P.A. (2008). Simulation support for performance assessment of building components. *Building and Environment 43* (2), 228-236.

van Treeck, C., Frisch, J., Egger, M. and Rank, E. (2009). Model-adaptive analysis of indoor thermal comfort. *Proceedings of the 11th IBPSA Conference*, 27-30 July, Glasgow.

Witte, M.J., Henninger, R.H., Glazer, J. and Crawley, D.B. (2001). Testing and validation of a new building energy simulation program. *Proceedings of the Seventh IBPSA Conference*, 13-15 August, Rio de Janerio, Brazil.