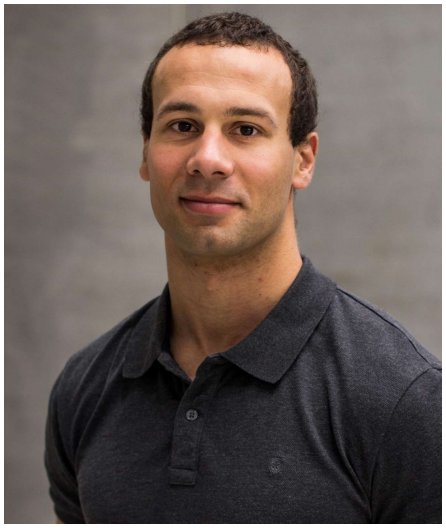
The 18th IBPSA International Conference and Exhibition
Building Simulation 2023
September 4-6, 2023
Shanghai, China

What Metrics Does the Building Energy Performance Community Use to Compare Dynamic Models?

Authors:
Hicham Johra
Markus Schaffer
Gaurav Chaudhary
Hussain Syed Kazmi
Jérôme Le Dréau
Steffen Petersen



Hicham Johra
Associate Professor
AALBORG UNIVERSITY
DENMARK

Motivations

- Comparing, validating and assessing the accuracy of dynamic models is crucial
- Dynamic simulation outputs are often in the form of time series
- Comparison methods and metrics for qualitative and quantitative analysis of these building model output time series are thus needed

Motivations

- Large reviews and discussions around comparison metrics and model key performance indicators exist in some research communities
- However, no such systematic review can be found for the field of indoor environment and energy in buildings
- Conclusions on the adequacy of a comparison metric can vary with the characteristics of the evaluated time series: sampling rate, amplitude, frequency spectrum, unit scale, value distribution, etc

Objectives

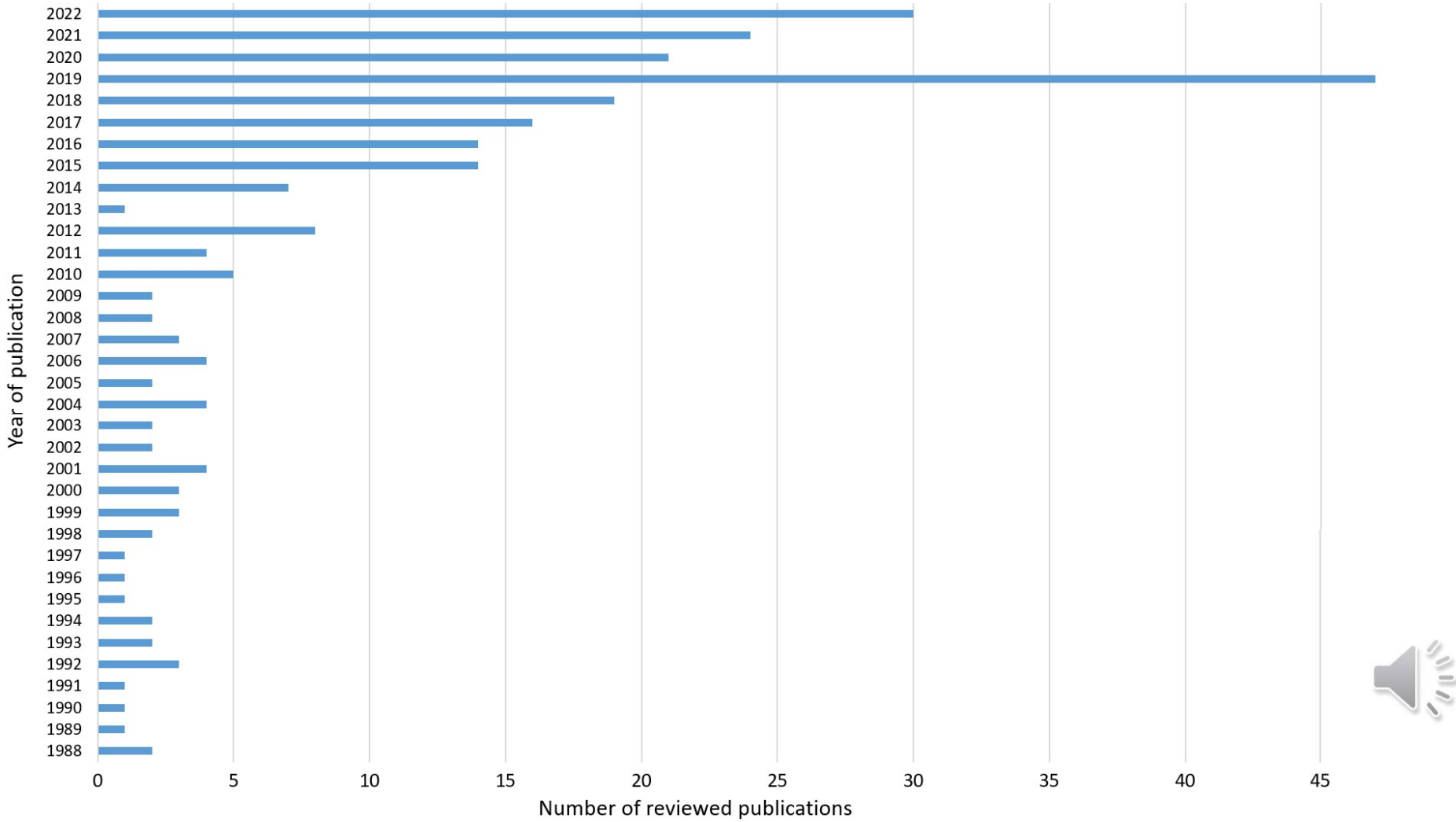
- Give an overview of what comparison methods and metrics are used in the building community to analyse simulation results (time series) of dynamic building models
- Discuss the issues of some common metrics
- Provide a unified definition and notation for the 48 metrics found in the review process

Methodology

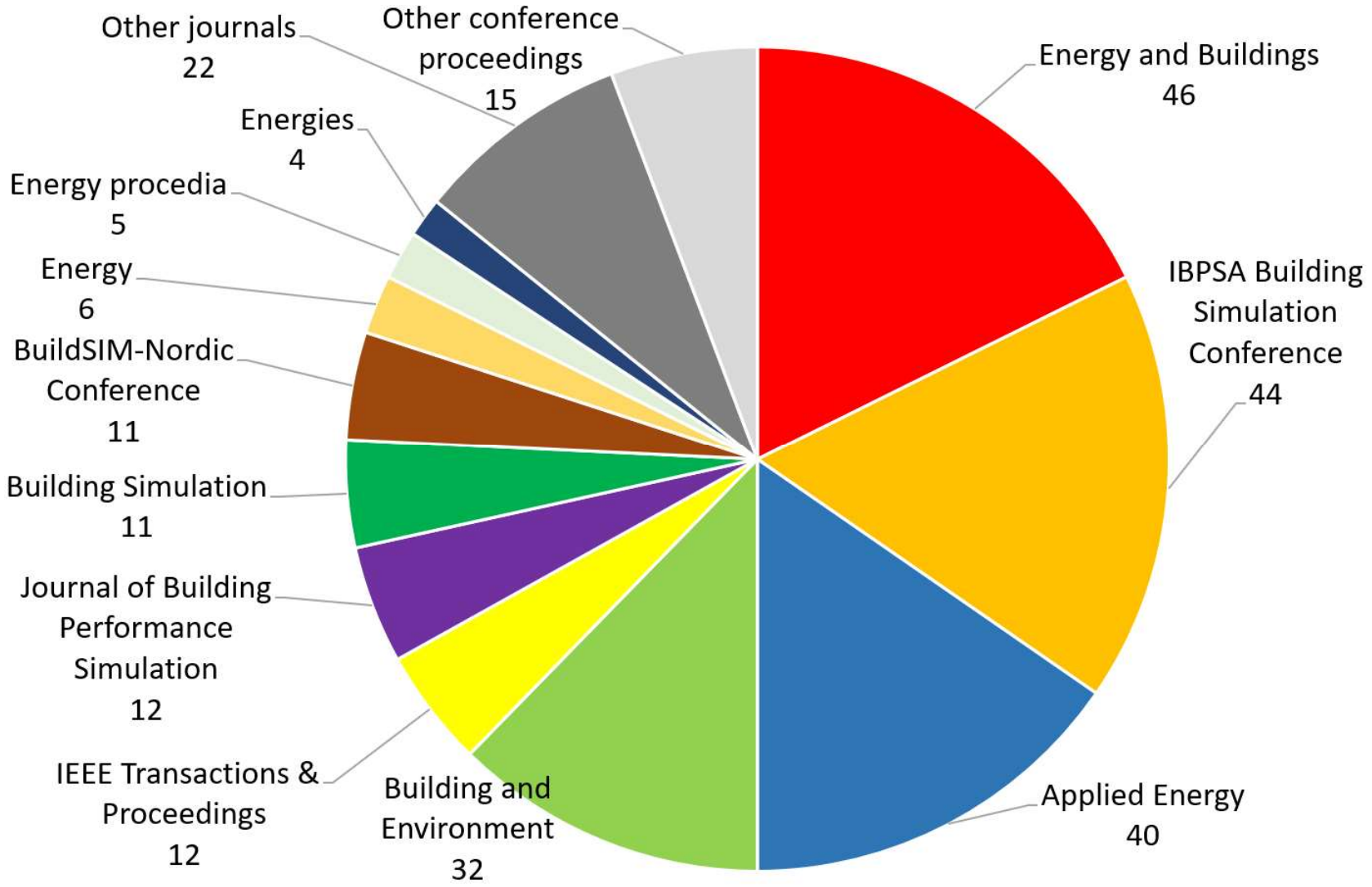
- Review of 259 papers about numerical modelling of building energy demand and indoor environment

- Systematic search for metrics and figures comparing the time series output of models for comparison or validation purposes

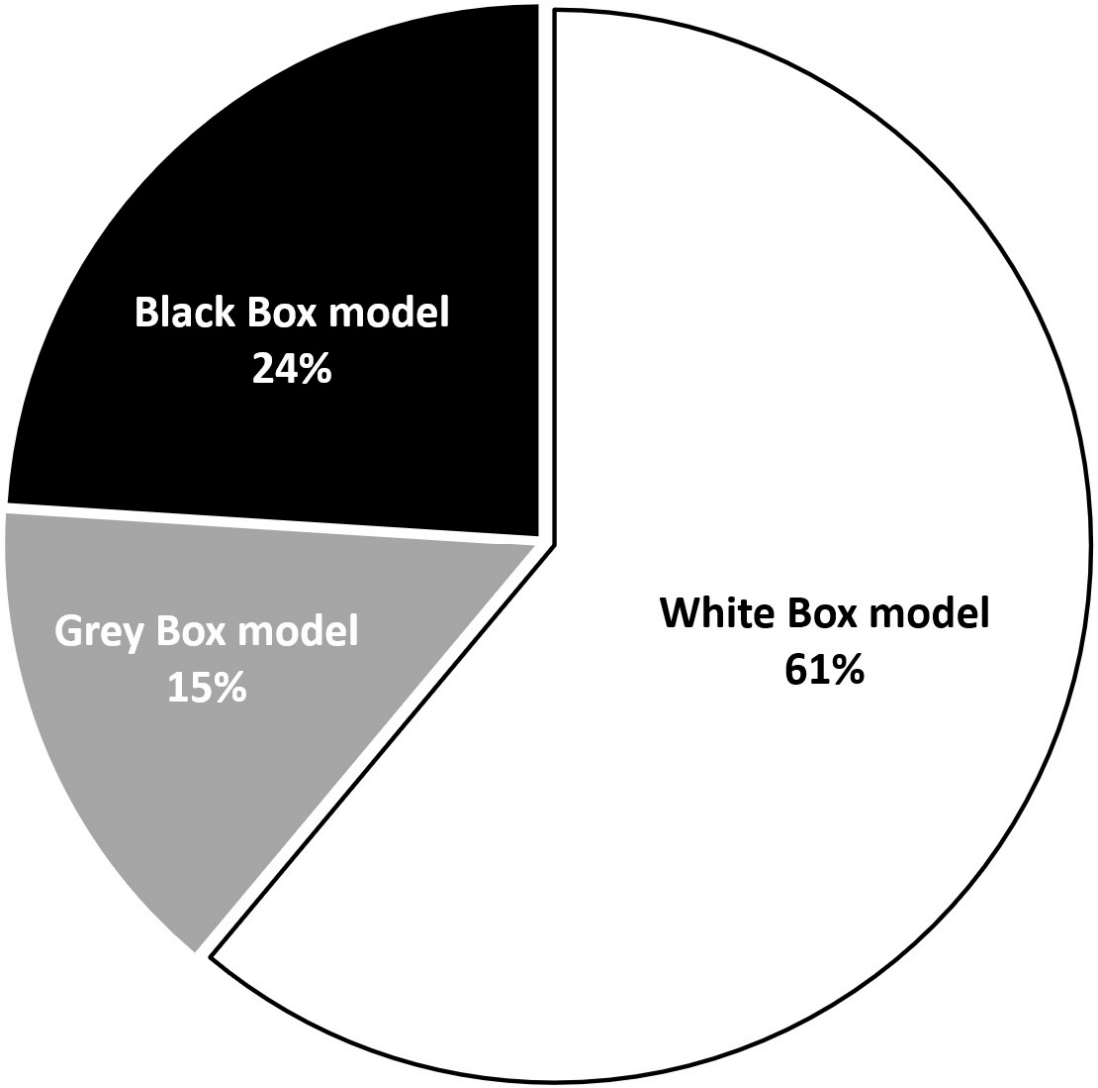
Methodology



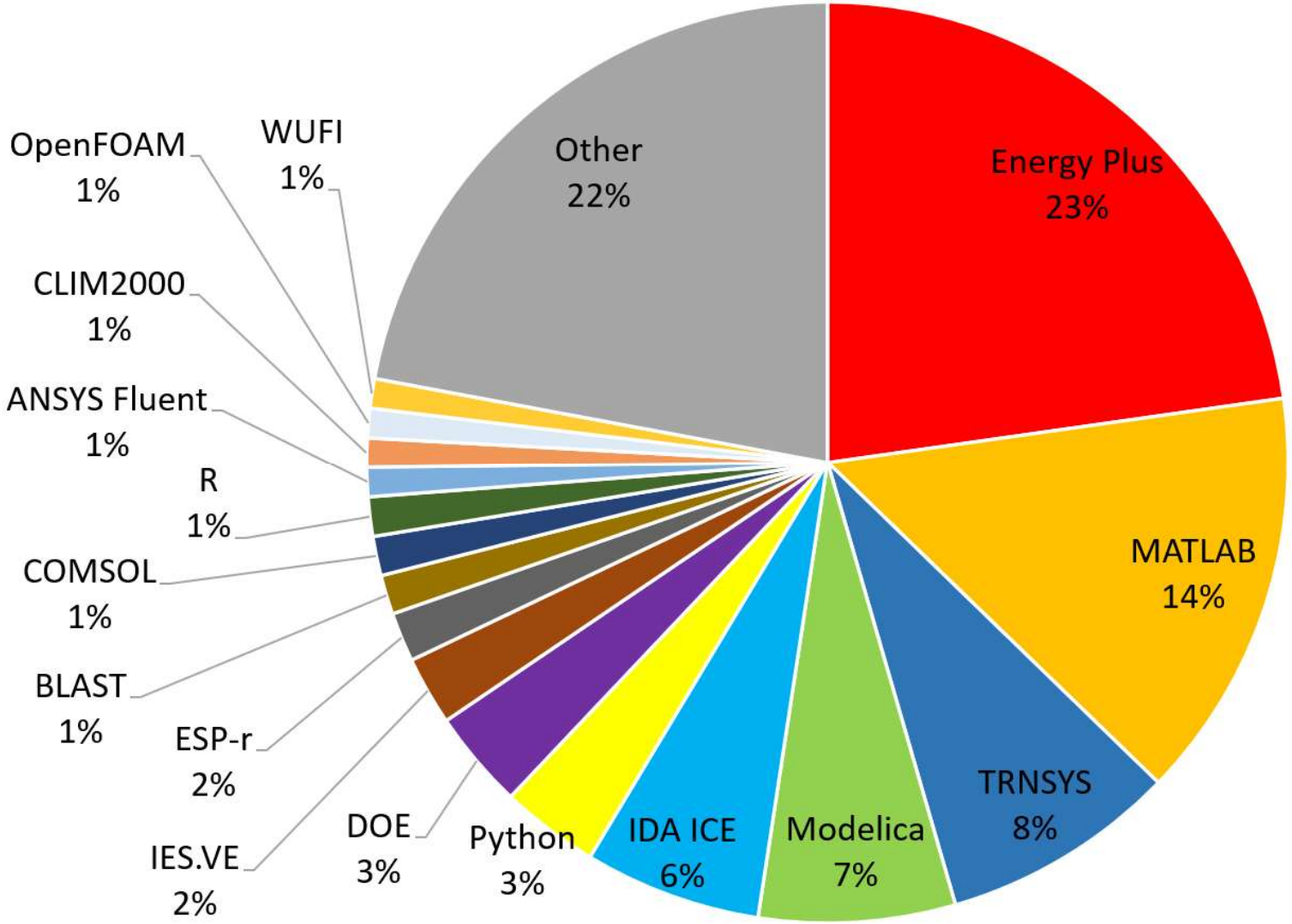
Methodology



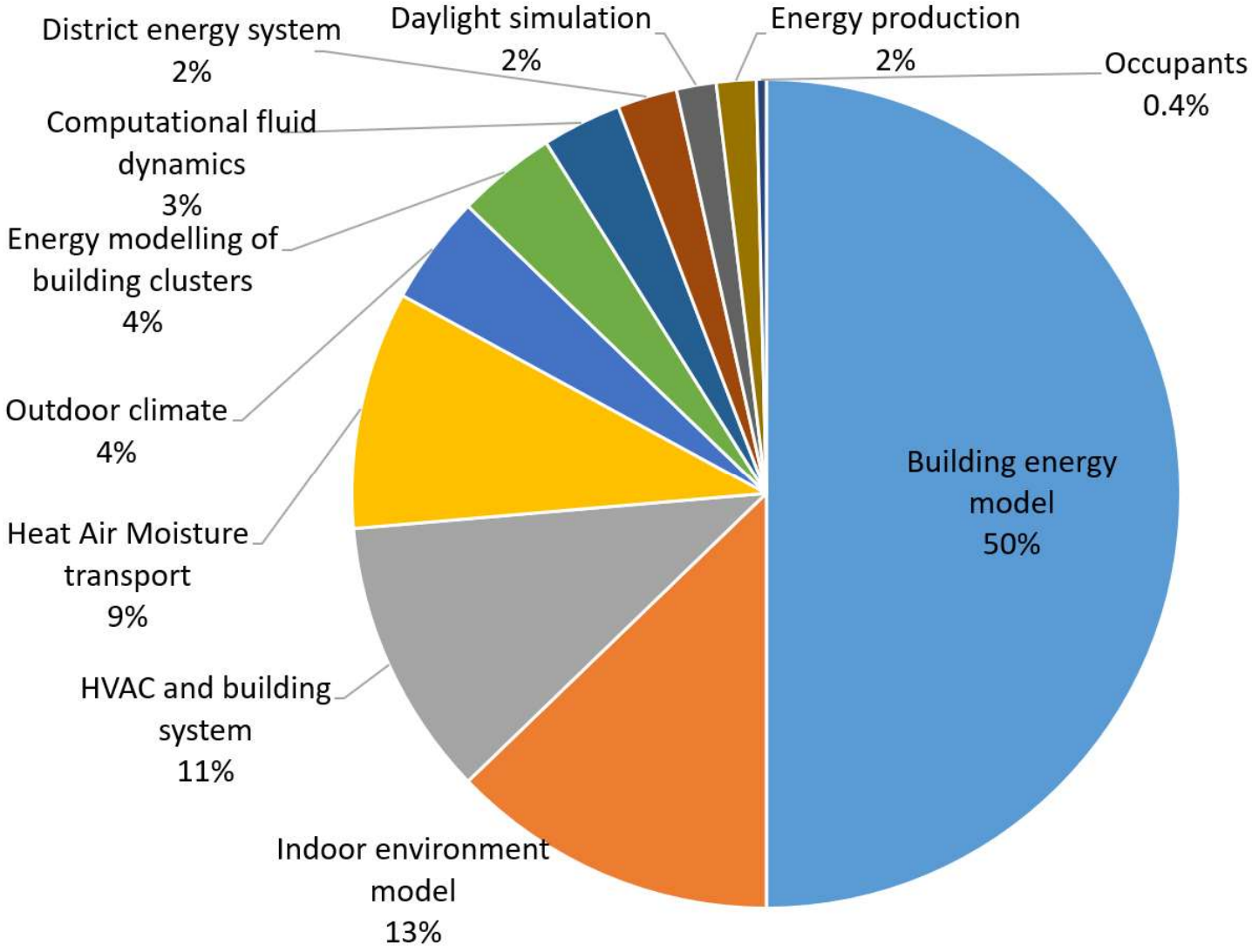
Trends in building modelling



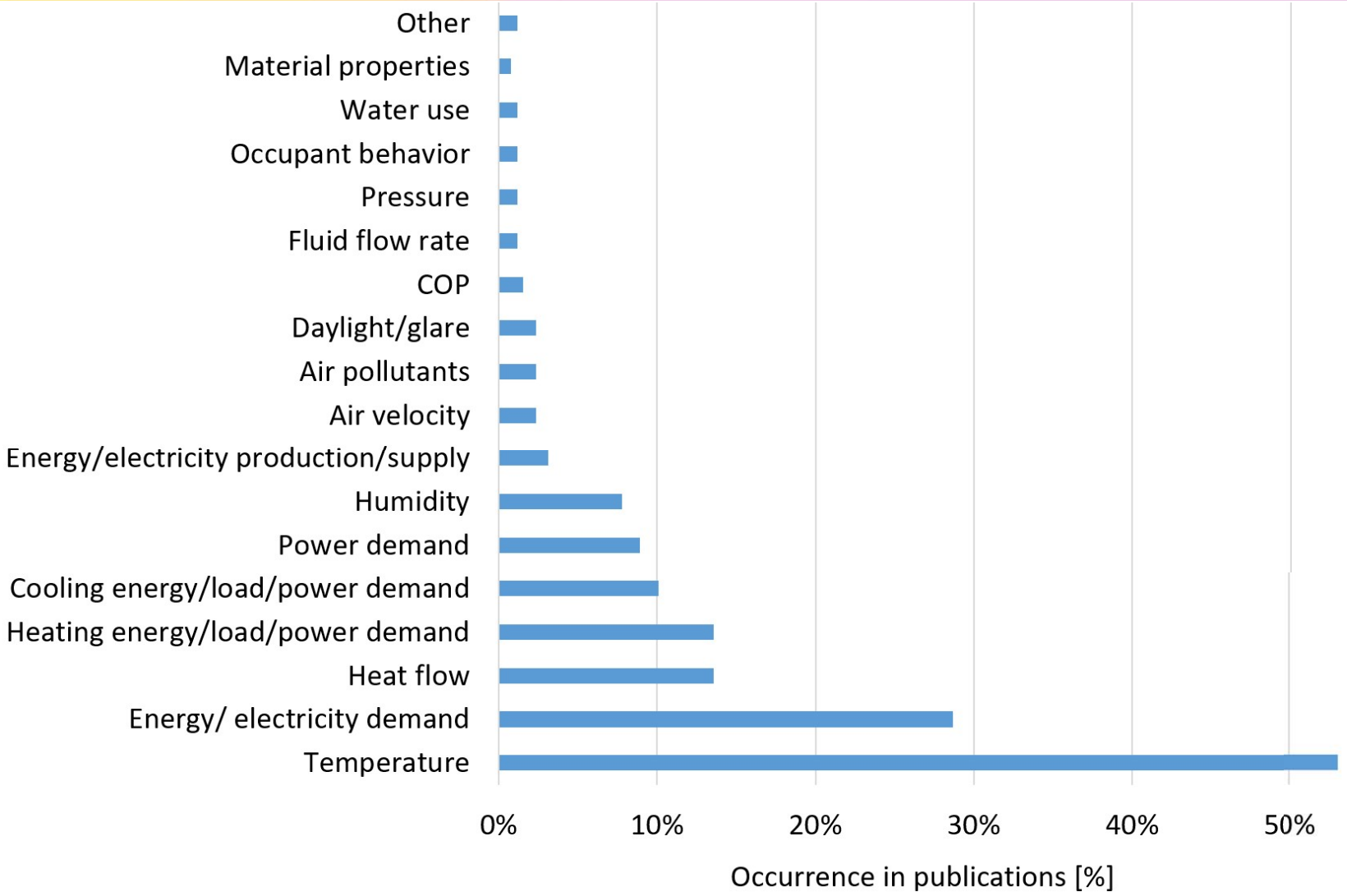
Trends in building modelling



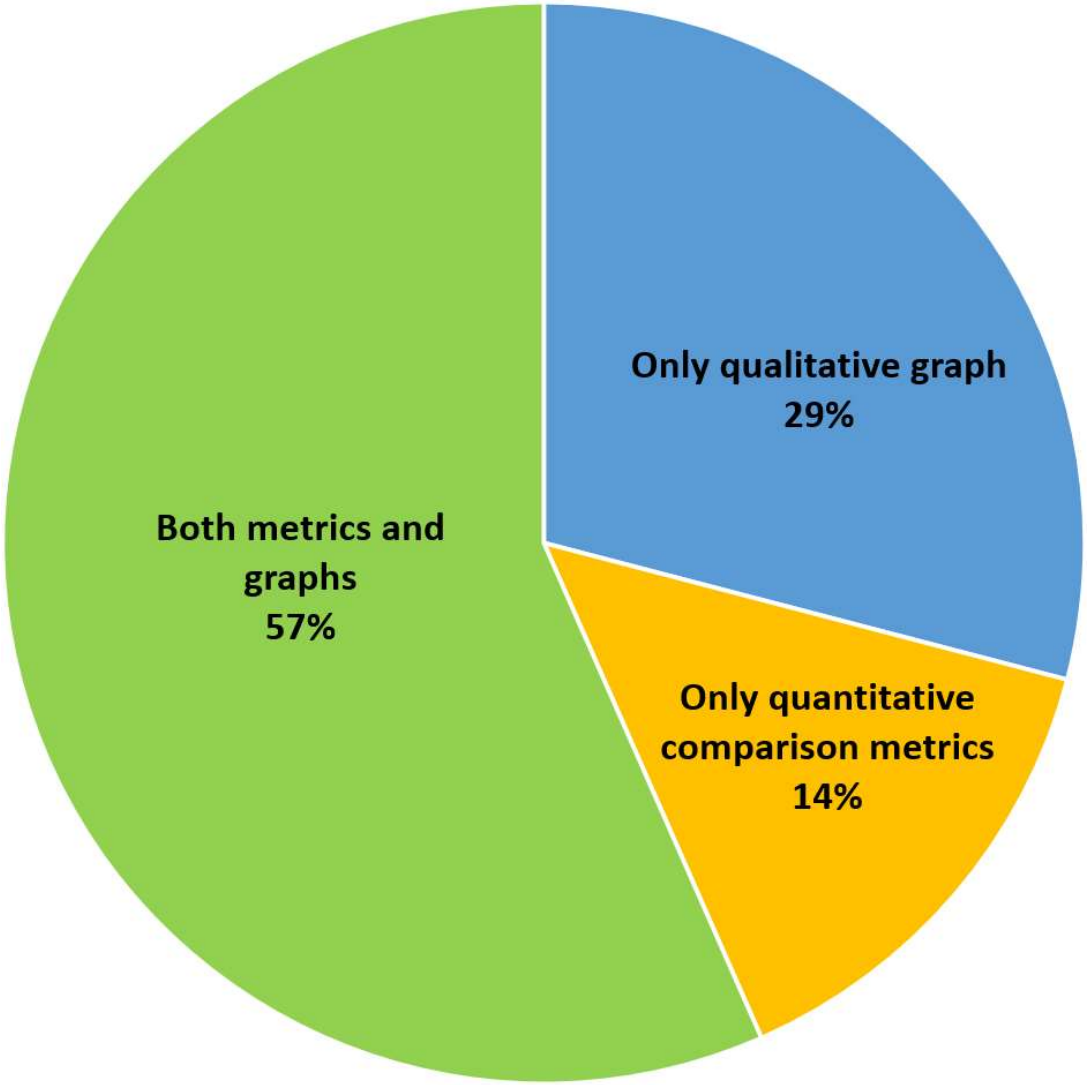
Trends in building modelling



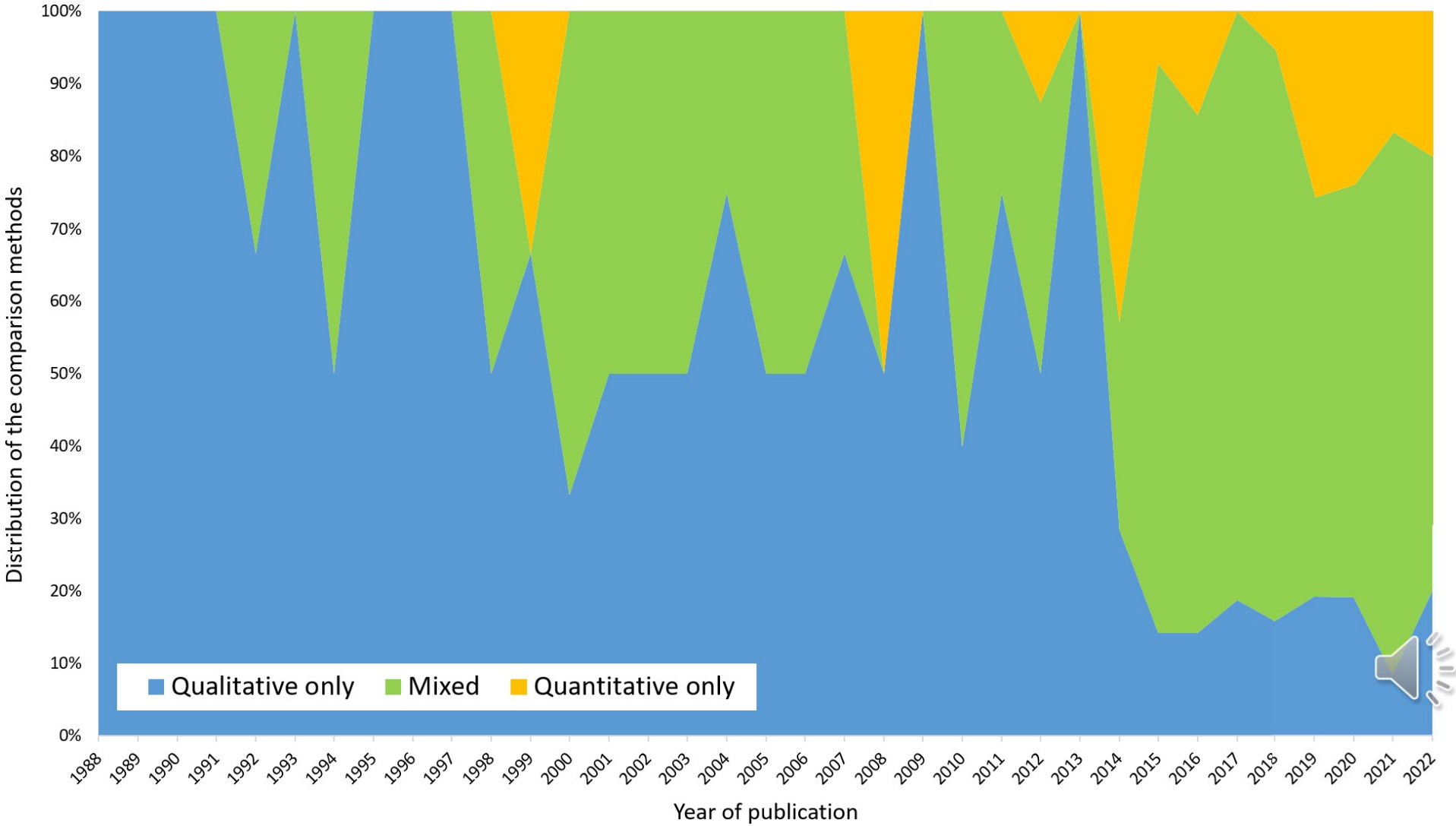
Trends in building modelling



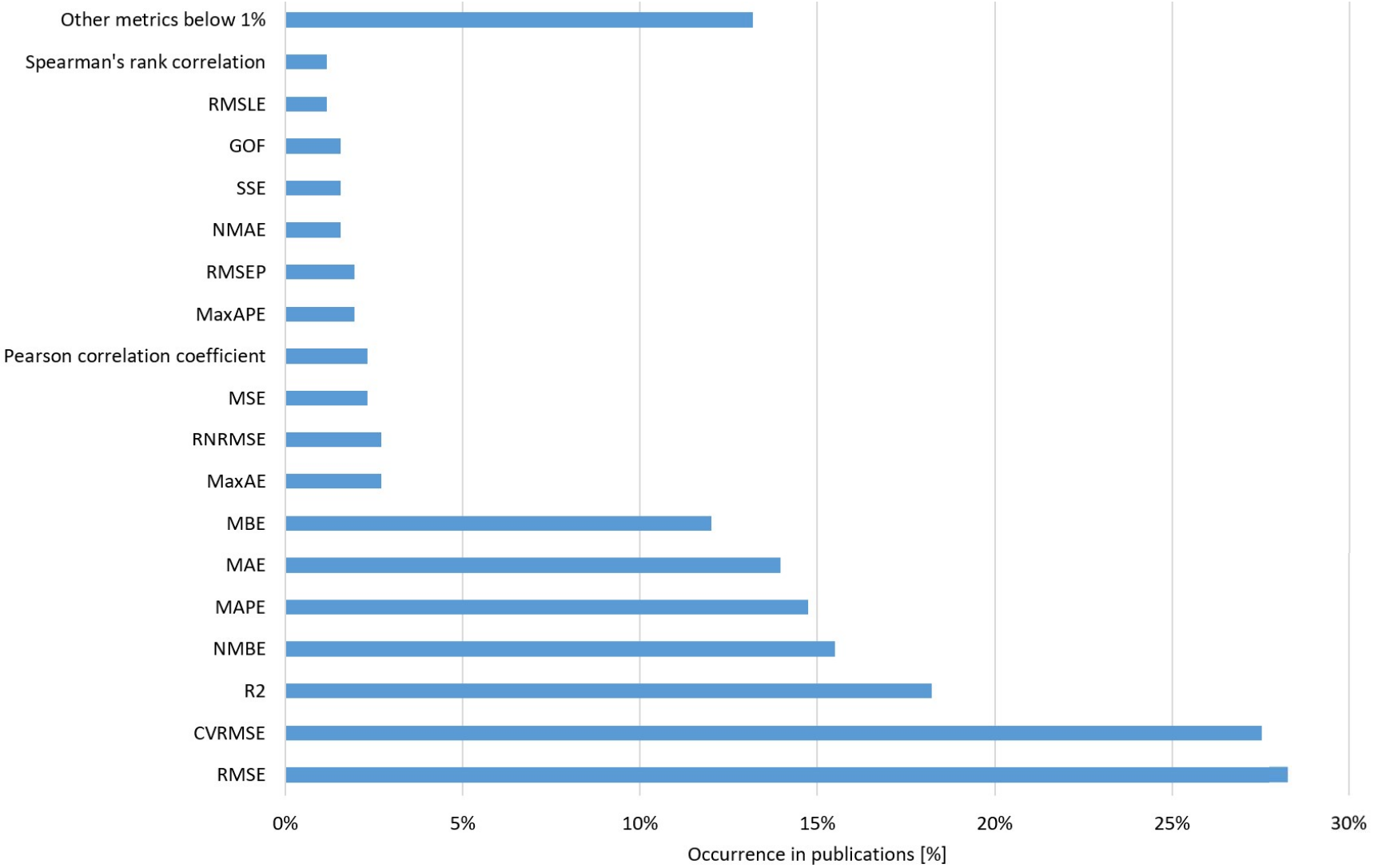
Comparison of time series



Comparison of time series



Comparison of time series



Discussions

- Large influence of recent guidelines (e.g., ASHRAE 14) on the adoption of MBE, NMBE, RMSE and CVRMSE
- These guidelines also suggest thresholds for model validity
- It clearly improves the simulation performance reporting and comparison between different studies
- However, these common metrics can present some issues

Discussions

MBE and NMBE:

- Global bias of the model
- Prone to cancellation or compensation effects
- Can lead to low MBE or NMBE despite large local discrepancies
- Squared difference-based or absolute value-based metrics do not have this problem

Discussions

Normalised metrics:

- Normalised metrics should be preferred: e.g., CVRMSE
- Necessary for model comparison on datasets of different sizes or unit scale
- Some confusion in the naming and definition of normalised metrics

Discussions

Normalised metrics:

- Normalisation by the total mean average of the entire dataset: bias towards periods of high magnitudes
- Can be problematic for time series with strong seasonality like heating/cooling demand

Discussions

Metric robustness:

- Some metrics are very sensitive to values close to the 0 of the time series unit scale
- Some metrics (e.g., MAPE) are mathematically undefined when the quantity is 0
- Very problematic for time series of building energy demand

Discussions

Metric robustness:

- Some metrics are very sensitive to outliers (e.g., CVRMSE)
- Can cause over-penalization: problematic for global model validation
- RMSE of log error (RMSLE) is less sensitive to large outliers but is asymmetrical: penalises more under-estimations than over-estimations

Discussions

Metric definition and implementation:

- *n-1* or *n* samples in the computation of comparison metrics?
- Various definitions and implementations of R^2, coefficient of determination

Conclusions

- Many different comparison metrics with various definitions, implementations and flaws causing mistakes or misunderstandings
- Quantitative comparison metrics must be reported in studies
- Qualitative comparison with time series graphical visualization should also be included with different time scales
- Normalized metrics are preferred over absolute metrics for quantitative comparison

Conclusions

- The equation of the metrics should always be provided along with the evaluation period, and information on the data treatment for zero-values
- For error evaluation, CVRMSE, RMSE, MAPE and MAE are commonly used
- For bias evaluation, NMBE and MBE are commonly used
- Elastic distance metrics (e.g., Dynamic Time Warping or Frechet distance) should be considered for further analysis of time series with possible time-shifting

Conclusions

A unified and coherent definition and notation for the 48 reviewed metrics:

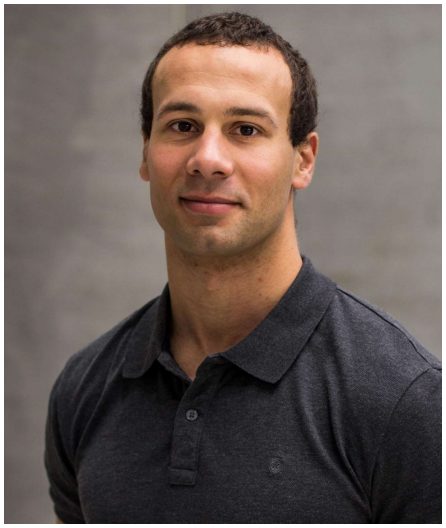
Johra, H., Schaffer, M., Chaudhary, G., Syed Kazmi, H., Le Dréau, J., & Petersen, S. (2023). *Coherent description of 48 metrics to compare, validate and assess accuracy of building energy models and indoor environment simulations*. DCE Technical Reports No. 314. https://doi.org/10.54337/aau533917780

Future work

- Continue testing and analysing comparison metrics for energy in building and indoor environment modelling applications
- Study common mistakes and pitfalls in the use of these metrics
- Make some recommendations for the comparison of building models with those metrics
- Looking into probability-based metrics for forecasting models

Thank you for your attention!

Please contact me if you have any questions



Hicham Johra
Associate Professor

AALBORG UNIVERSITY
DENMARK

Email: hj@build.aau.dk
LinkedIn ➔ ➔ ➔ ➔ ➔