# Extrapolating the seasonal efficiency of a boiler using clustering: A case study

**P BALASUNDARAM*[1,3], B DELINCHANT[2] , S PLOIX [1] ,C MURESAN[3]**

**[1] Grenoble INP – GSCOP [2] Grenoble INP – G2ELab [3] ENGIE Crigen Labs**
***prasaant.balasundaram@external.engie.com**
**benoit.delinchant@G2Elab.grenoble-inp.fr**
**stephane.ploix@grenoble-inp.fr**
**cristian.muresan@external.engie.com**

*RESUME.*

*La simulation hybride en temps réel (RTHS) offre une procédure intuitive pour la validation et l'évaluation des produits de technologie du bâtiment, testés dans des conditions dynamiques. Le seul inconvénient de ces RTHS est qu'ils sont sujets à des contraintes de temps, lorsque des produits comme les chaudières, les pompes à chaleur, etc. sont testés pour leur performance annuelle / saisonnière. Dans une telle situation, on ne peut pas tester le produit pendant une longue période, on peut plutôt réduire le cycle de simulation sur la base de quelques jours qui représentent toute l'année. Le choix des jours représentatifs doit se faire sur la base d'algorithmes de clustering. Cette performance obtenue dans un cycle court est extrapolée à une année entière. Le résultat extrapolé est ensuite vérifié par simulation annuelle et leur degré de proximité est représentatif de l'exactitude du cycle de simulation réduit. L'ensemble de la simulation se fait sur la DYMOLA (plateforme commerciale Modelica), qui permet à un utilisateur de modéliser les composants du système au plus près de la représentation du monde réel. Une étude approfondie de la littérature est effectuée, suivie d'un exemple d'étude de cas avec un bâtiment simple à zone unique qui a des besoins de chauffage d'espace pendant la saison de chauffage.*

*Mots clés: simulation hybride en temps réel, cycle de simulation réduit, extrapolation, clustering, simulation de construction.*

*ABSTRACT. Real Time Hybrid Simulation (RTHS) offers an intuitive procedure for the validation and evaluation of building technology products, tested under dynamic conditions. The only downside of such RTHS is that they are prone to time constraints, when products like boilers , heat pumps, CHP units etc. are tested for their annual/seasonal performance. In such situation, one cannot test the product for a long time , rather one can reduce the simulation to a few days whose performance should be representative of a whole year. The choice of representative days has been done based on a clustering algorithm. The few days performance obtained is extrapolated to a whole year. The extrapolated result is then verified with annual simulation and their degree of closeness is representative of correctness of the reduced simulation cycle. The entire simulation is done on the DYMOLA (commercial Modelica platform), which allows a user to model precisely the components of a home/office building system. A thorough literature study is done followed by a sample case study with a simple single zone building model that has space heating requirements during the heating season.*

*Key Words: Real time hybrid simulation, reduced simulation cycle, extrapolation, clustering, building simulation.*

## 1. INTRODUCTION

Heating Ventilation and Air Conditioning (HVAC) systems and products are often characterized by dynamic operating conditions, the complexity- the cost and the time required for such type of experimentation are major obstacles. Right now a simple steady state method to test the products is widely followed. Despite the simplicity of steady-state characterization, the behavior during a steady-state test may be quite far from the real performance of the component when it is installed in a complex system as explained in (M ALBARIC). Hence - a dynamic whole system testing or (RTHS) was developed to test small scale and several residential systems (ANDREASBELDERBOS, 2017). Figure 1 shows the different blocks of the RTHS, which can be divided into three parts (S FURBO, 2004). The simulation level models the virtual components and these are simulated in real time. The emulation level acts as a coupling between the real and the virtual components. It includes devices and components for the measurement of the hardware behavior and to emulate the simulated behavior of the virtual components. At the hardware level real components of a complex system are installed in an emulator. There also exists a communication interface for the exchange of data between the emulator (Data Acquisition Unit) and the simulator.

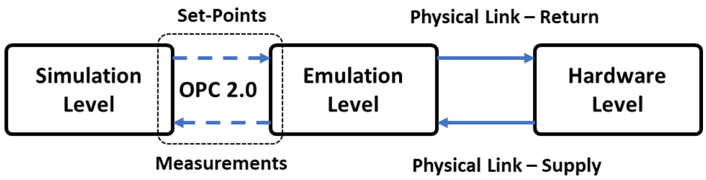


*Figure 1: Scheme of RTHS*

Say if one needs to test a product whose annual or seasonal efficiency is to be obtained then it becomes infeasible to test the product for a whole year. In such cases it is always better to reduce the simulation time(simulate for a few days) and then extrapolate the quantity of interest to one whole year. Section 2 gives an overview of some methods to reduce the simulation time. Section 3 gives an idea on how a reduction in simulation is done for a case study in a virtual environment. Section 4 discusses how further improvements could be made and also the future scope.

## 2. SEQUENCE REDUCTION METHODS

The sequence reduction methods can be split into three large categories namely heuristic methods, iterative methods and grouping algorithms (HASAN SAYEGH, 2019) . In short, heuristic method is a manual method , with the selection criteria based on the annual data. The idea is to select the number of periods with different load or meteorological conditions in order to capture a variety of different events. In (DIEGO MENEGON, 2017), the author selected specific days to contain the hours with meteorological and load events to characterize a typical system behavior. The heuristic method is used to reach immediate goals but not an optimal one. Iterative method searches for a best optimal solution based on repetitive action and compares the quality of the result in each iteration. The Short Cycle System Performance (SCSPT) is a method developed by the French Commission of Alternative Energies and Atomic Energy (CEA) , where a twelve day reduced sequence is used to represent the whole year. The sequence was able to reproduce the annual performance with a good degree of accuracy applicable for different models (M ALBARIC). Clustering algorithms employ an advanced technique of grouping days with similar attributes into clusters. The most commonly used method is the k – means clustering as explained in (THOMASSCHÜTZ, 2018). The days close to the cluster centers are chosen for the short sequence. One such application is demonstrated in (MY HALLER, 2013). The method of time reduction process which directly depends on the way the extrapolation is done. The literature suggests that there are two different ways, one consists in multiplying the result obtained in the reduced cycle by a proportion (commonly adopted in heuristic and iterative methods) while the other method is to multiply the result obtained in the reduced cycle by weight of the group represented by a day (usually adopted for the clustering methods) (HASAN SAYEGH, 2019). The clustering algorithm is chosen at this point since it provides an easy way to represent different representative periods given the complexity and problem size. Also , in the literature there are several different methods of sequence reduction that lead to the evaluation of the performance index of a particular product under test. The most prominent among them are Concise Cycle Test(CCT), Combi test , Short Cycle and System Performance Test (SCSPT) (MY HALLER, 2013), Prescribed Load-Performance Extrapolation(PLPE) (DIEGO MENEGON, 2017) . In our case we follow the same path as the above methods and only there will be a difference in the following parts:

- Number of test days
- Level of detail of the building model
- Data exchange rate

The method we follow will allow us to model larger buildings with a high level of detail in each zone/part of the building (in this sample use case we just use one single zone) with dynamic boundary conditions. The data exchange rate could be one second or less. Moreover, some other clustering algorithms (k -medoids , k -medians , etc.,.) could be adopted in the future where there is significantly lesser number of days. These are the significant improvements that are foreseen in the near future while the actual Hardware in Loop simulation is performed, leading us to the HIL evaluation of the component under test.

*Table 1: Overview of state of the art Real Time Hybrid Simulations*

| Acronym | CCT | Combitest | SCSPT | PLPE |
|---|---|---|---|---|
| Number of Days | 6 | 6 | 12 | 6 to 24 |
| Data Exchange Rate | 1/32h | 0.3 minutes | 1 minute | 5 seconds |

# 3. EXTRAPOLATION OF BOILER EFFICIENCY

One can exemplify the approach by an use case, that consists of a single zone building with 40m2 of floor area and 40 percentage of window area on all four walls. All the walls have a single layer of insulation outside followed by a concrete layer. The choice of weather data is Trappes city near Paris , in France .The peak power requirement during winter is around 13kW and the boiler used for space heating has a nominal power of 23kW. The heating season is for 153 days and this detail is obtained from the annual simulation (because only the space heating of the building is taken into account and not the domestic hot water consumption). The sample use case in modelica is shown in figure below.

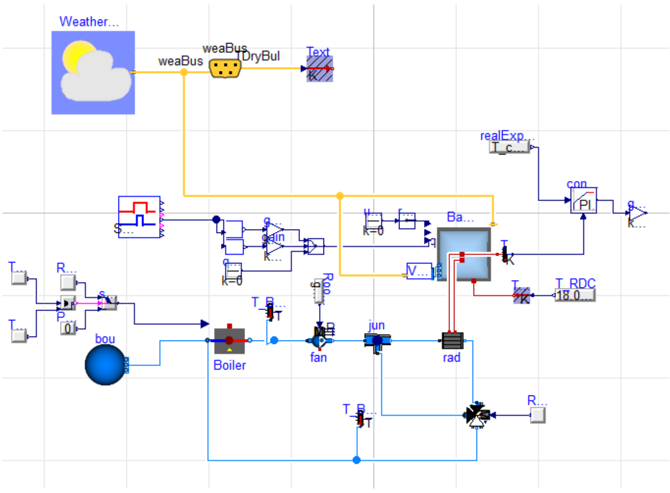


*Figure 2: Modelica model of the sample use case*

The next idea is to find out the most important features that have a direct effect on the power consumed for space heating in order to perform k-means clustering. The features can be selected based on the combination of expert knowledge and also based on correlation. From the boiler point of view the power is directly affected by the load which in turn depends directly on the Outside Air Temperature. Also the type of control also affects the power (inside vs outside air temperature control), in this case an internal air temperature control is used for the single zone building (the internal temperature is affected by both the outside air temperature and also on the Global Horizontal radiation as shown in the correlation heat map ). The simulation is then run for the heating period and a correlation analysis is performed between the various factors vs the power consumed as explained in (Yan Chen, 2019) which explains why correlation analysis actually a good indicator in identifying the parameters that affect the power consumed .The heat map in Figure 3 is constructed by collecting the simulation result for the one whole year , the power required is fixed as the target variable and all the parameters affecting it are ranked using cross correlation, as shown in (Myers.J) the top two or three parameters are chosen from the top 10 parameters affecting the target variable and then a heatmap representation is formed. From Figure 3 it can be seen that the external temperature and the global horizontal irradiation(GHI) have the most significant effect on the thermal power consumed. It can also be seen that the GHI also has a direct influence on the ambient temperature because of the 40 percentage of window area on all four walls of the building. To select the representative days let us consider the three features namely T-external , GHI and the internal loads(the internal loads are chosen because the internal loads also give us the idea on the occupancy which in fact triggers the boiler for heating). The time resolution for weather data is 1 hour. Partitional clustering algorithms are greedy and thus, the k-means algorithm only converges to local solutions. We repeat the process with 10,000 initializations and pick the clustering with minimal SSD (Sum of squared distance between al point to the cluster centroid within one cluster or WCSS

Within Cluster Sum of Squares) across all initializations (this allows for reproducibility of results). The same process is then repeated for different number of clusters (n), which gives us an initial idea on what are the number of clusters to choose, plus the elbow method Figure 5. The choice of number of clusters is also made based on how long in real time the simulation can be done , if one intends to perform a hardware in loop simulation , 7 days seemed a decent trade off considering all the previously mentioned factors (a combination of least error and the number of days for real time simulation is used to decide the number of clusters).

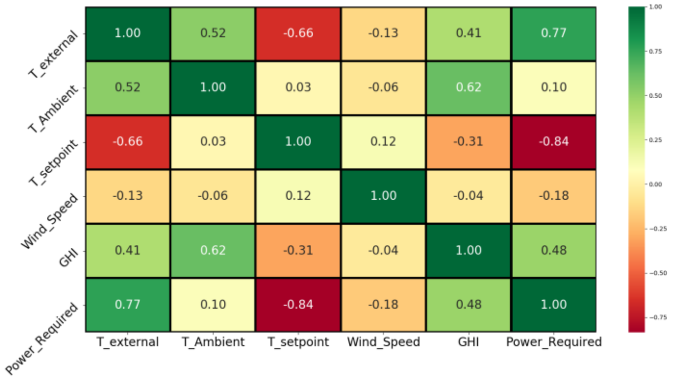


*Figure 3:Heatmap correlation*

Before performing the clustering algorithm it is important to prepare the data properly. First the data is to be normalized since it would be more useful to represent datasets of different scales. The internal loads, the external temperature for each hour and the GHI for each hour are collected into a vector of size 72 (24 X 3) for each day. Then a k-means clustering algorithm is run for the set of 153 vectors (since there are 153 days for the heating season).The days closer to the centroids are chosen as the representative days .The weight for each cluster were almost equal hence the scaling factor for extrapolation was chosen to be 153/7. The days that are closer to the cluster centroids are chosen and used in the modelica weather data reader. Then the simulation is run for one week. The results obtained for one week is the compared to the result obtained for one year after extrapolating the results obtained from one week simulation. The results are shown in Table 1, the comparison between the annual simulation and the weekly simulation are compared. Moreover , two other random set of days are considered to check how clustering significantly improves , in general the degree of the closeness between the extrapolated results and the annual result is one way to measure the validity of the approach.

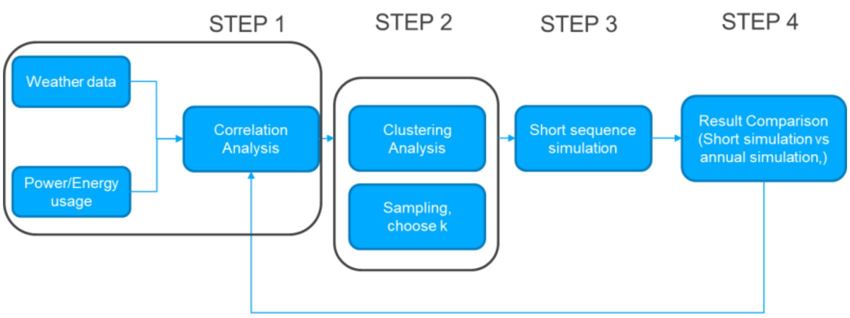


*Figure 4:Clustering procedure for choosing representative days*

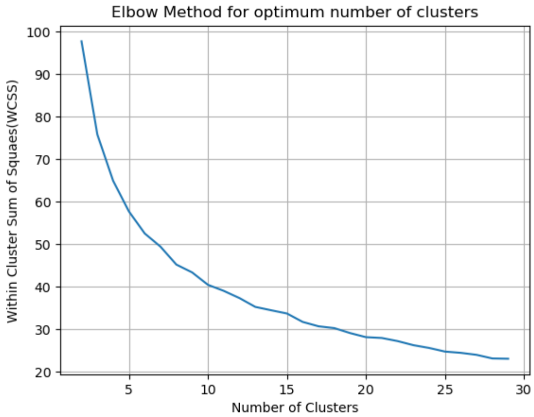


*Figure 5: Elbow method for choosing optimum number of clusters*

Another important thing to be noted is that by extrapolating using a factor of 153/7 is leading to overestimation of the indices considered (Input/output energy and Efficiency of the boiler). The annual simulation ran for a total time of 2 minutes and thirty seconds , while the one week simulation ran for few milliseconds. From the results it can also be seen that the extrapolation based on clustering could yield better results with less deviation (~ 0.5%) compared to the random choice of days which deviate more from the original annual simulation. The possibility that the results of the random sets of days producing the results closer to the extrapolated result could be due to the fact of running the three short sequences simultaneously one after the other which could mean the internal states of the building were adjusted to the simulation.

*Table 2: Comparison of results*

| | **Annual** | **Random 1** | **Random 2** | **Extrapolated** |
|---|---|---|---|---|
| **Energy Supplied (in kWh)** | 11471 | 13001 | 13463 | 12456 |
| **Energy Utilized (in kWh)** | 9857 | 11339 | 11665 | 10756 |
| $\eta_{boiler}$ | 0.859 | 0.872 | 0.866 | 0.863 |
| **Energy Supplied Deviation %** | - | 12% | 15% | 8% |
| **Energy Utilized Deviation %** | - | 13% | 18% | 8% |
| $\eta_{boiler}$ **Deviation %** | - | 1.47% | 0.820% | 0.49% |

## 4. CONCLUSION AND FUTURE WORKS

From the above process it is evident that the clustering method chooses the most representative days. Moreover a lot of knowledge base and data is required, also such short sequences are always constructed keeping the entire system in mind but not necessarily from a product point of view. From expert knowledge the outside temperature and the control determines the operational efficiency of the boiler . Therefore, the next step will be to construct a short sequence based on the limited expert knowledge and also on the signal features of the annual weather data which would include the extreme days and the most representative days as well, which would result in shorter testing time and a new method of generating a short sequence from the product point of view. In the above result it would be interesting to know if the clustering approach would produce the same set of results with the same deviation for different parameters in the same building (window size and the surface area). The discontinuity between the representative days chosen could present a significant problem if the application involves a domestic hot water application. Such discontinuities could lead to an energy shifting in the buildings, and also the energy shifting in the thermal storage. Rather than taking more features for clustering like external temperature and GHI it could also be interesting if the equivalent temperature could be considered as a single feature. This could also lead to utilizing different methods in defining the sequence of days for testing like in *Table **1*** . It would also be useful if there could be a possibility of reducing the number of representative days. Finally this approach could be experimentally verified using a Hardware in Loop testing. In addition as a future scope the uncertainty posed by the components in the test bench on this extrapolated performance could be analyzed. Finally based on the knowledge obtained from the above case study the next step is to choose representative where significant knowledge on the building is not available.

## 5. BIBLIOGRAPHIE


AndreasBelderbos, E. D. (2017). Accounting for flexibility in power system planning with renewables. *International Journal of Electrical Power & Energy Systems*, 33-41.

Diego Menegon, A. (2017). Development of a new dynamic test procedure for the laboratory characterization of a whole heating and cooling system. *Applied Energy*, 976-990.

Hasan Sayegh, A. L. (2019). Determination of a Short Simulation Sequence for the Multi-Criteria Optimization of Buildings : A case study. *IBPSA.* Rome.

M Albaric, J. N. (n.d.). Thermal performance evaluation of solar combisystems using a global appraoch,. *Eurosun Conference* . 2008.

MY Haller, R. H. (2013). Dynamic whole system testing of combined renewable heating systems – The current state of the art. *Energy and Buildings*, 667-677.

Myers.J, W. L. (n.d.). Research Design and Statistical Analysis . *Research Design and Statistical Analysis.*

S Furbo, C. B. (2004). *Combitest. A new Test Method for Thermal Stores used in Solar Combisystems PhD Thesis.* DTU.

ThomasSchütz, M. H. (2018). Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. *Renewable Energy*, 570-582.

Yan Chen, S. B. (2019). Assesing Building Energy Perfromance via Selection of Represntative Simulation Days . *IBPSA* , (pp. 4825-4832). Rome .