ABSTRACT

Advanced building control technologies such as model predictive control rely on fast executing statistical models. These models likely use large quantities of training data to accurately predict a building's state. It is important to have evaluation techniques that are able to assess accuracy of both the aggregated and timestep predictions in order to have confidence in statistical model predictions. There is presently no standardized procedure for validation and verification of statistical models against test data. This paper presents two tools that have been developed for such high-resolution (e.g. 15 minute) statistical models evaluations: a Residual Analysis tool and an Absolute Percentage Error tool. These tools use a varying limit curve for timestep data and defines benchmarks for monthly/annual energy consumption and indoor air temperature accuracy.

INTRODUCTION

As Box and Draper (1987) have commented, “all models are wrong; the practical question is how wrong do they have to be to not be useful”. Model Predictive Control (MPC) strategies often use fast executing statistical models to operate. For example, Cole et al. (2013) used a linear regression statistical model to create a reduced order model from the EnergyPlus base model. Another example is Ma et al. (2012) who also used EnergyPlus based linear regression model to optimize the cooling energy use. May-Ostendorp et al. (2011) on the other hand employed a generalized linear models to optimize energy consumption in the building with both natural and mechanical ventilation.

These statistical models typically rely on large quantities of training data created from a building performance simulation engine; effectively a model of a model. One would expect “suitable” accuracy of both the aggregated and timestep predictions in order to have confidence in statistical model predictions.

ASHRAE Guideline 14 suggests two whole-building energy use statistical indexes for assessing model accuracy compared to real building data: Normalized Mean Bias Error (NMBE) and Coefficient of Variation of the Root Mean Square Error (CV(RMSE)) (ASHRAE, 2002). The NMBE measures how close the energy use or demand predicted by the simulation mode corresponds to actual building data, whereas CV(RMSE) quantifies the degree of dispersion of a set of predicted values around the mean of the observed values. Guideline 14 states limits of 5% and 15% for NMBE and CV(RMSE) on a monthly level, respectively, and 10% and 30% on an hourly level, respectively.

To reapply Guideline 14 for assessment of a statistical model-of-a-model would be lacking. This is because one would expect the allowances to be less (i.e. greater accuracy), and because it does not have timestep evaluation. The latter is critically important to MPC as it must make control decisions in real time.

At present, there is no standardized procedure for validation and verification of statistical models intended for MPC and created from simulated training data. This paper presents two tools intended to facilitate discussion towards a standardized procedure.

METHODOLOGY

Two tools for validation of the high-resolution statistical building response models have been developed: Residual Analysis (RA) tool and Absolute Percentage Error (APE) tool. To increase tools applicability, the RA and APE have been developed by using pivot tables and charts within the Excel platform. The tools estimate and graphically present discrepancy between high resolution predictions (e.g. 15 minute timesteps) of a detailed building energy model and a statistical model. They include two categories of interest for the verification of the statistical model, energy and temperature (with individual metrics proposed) and are separated into three sub-categories: timestep, monthly, and annual.
The RA tools is based on the residual calculations, whereas APE tool is based on the mean absolute percentage error (MAPE) and absolute percentage error (APE). The calculation of residual is presented in Equation 1 and the calculation for MAPE is presented in Equation 2, where $y$ is the known value and $y_i$ is the estimate. The calculation for APE is the same as MAPE, but without the summation.

$$\text{Residual} = \text{Statistical Model} - \text{Building Energy Model}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y}{y} \right| \times 100$$

The tools consist of three main sections: Data, which is common for both tools, Results and Acceptability Curves. The APE relies on the below described Acceptability Curves which contains defined limitations for the calculations.

**Data sheet** (see Figure 1) – contains the high resolution data including:

- Date and time
- Outdoor ambient air temperature
- Energy consumptions (electricity and thermal energy)
- Energy consumption residual
- Energy consumption absolute error
- Energy consumption absolute percentage error
- Setpoint temperatures (heating and cooling)
- Setpoint temperatures residual
- Absolute setpoint temperatures residual
- Change in internal zone temperature from the previous timestep, DeltaT
- Absolute DeltaT

An example of Data sheet is presented in Figure 1 (see end of paper). Orange columns with black text represent input data and white columns with orange text calculated data. To analyse data and develop Pivot charts, time and date are broken-down to Months (January to December), DayOfWeek (Monday to Sunday), and Hour&Minute. Along with timestep values, outdoor air temperature is also expressed by using three previously defined temperature bins: below 0°C, 0°C to 10°C, and above 10°C. These represent heating season, shoulder season, and cooling season for commercial buildings. For an easier interpretation energy consumptions are from Joules (J) transformed into kilowatt-hours (KWh), while residuals and absolute errors are calculated as well.

**Results** – consist of 35 sheets with various statistical analysis including:

**RA tool:**

- Residual analysis in relation to energy defined energy bins
- Distribution analysis
- Monthly residuals
- Monthly absolute error analysis
- Seasonal (winter and summer) hourly residual analysis

**APE tool:**

- Monthly energy consumptions in relation to acceptable MAPE
- Timestep energy consumptions in relation to acceptable APE
- Monthly energy consumptions in relation to acceptable MAPE
- 95th, 99th and MAPE energy consumption in relation to acceptable APE
- Annual energy consumption APE as a function of time of day
- Timestep temperature in relation to allowable error
- Temperature absolute error as a function of time of day

**Acceptability curves** – define allowable APE, MAPE and annual limits

When considering energy we would expect that during operation about the average building power level a reasonable level of prediction accuracy for monthly and annual totals, with an increase in timestep level errors. The error will consist of both a bias and a scale. It is also important to recognize that percentage errors are inflated at lower power levels for the same absolute error due to bias, thus it is important to have a metric able to account for difference in power levels.

The proposed metric accounts for all of these features by employing an exponential decay function as the limit for timestep absolute percentage error (APE) that passes through a maximum acceptable percentage error. These are as a function of power that is presented as the ratio of annual or seasonal average power (i.e. a percent function). We have selected 10% error limit at the average annual or seasonal power level, and a 2% error limit at high power levels (e.g. 7 times the annual average). A static value of 5% is applied to aggregate values such as annual and monthly, in similar reason.
to ASHRAE Guideline 14. The curves developed are shown in Figure 2 (see end of paper), with the MAPE line representative for larger aggregate values (monthly, annual, etc.) and the APE line for timestep data points and smaller aggregate periods while also being presented in Table 1.

These maximum error limits were selected as we are comparing the performance of a black box statistical model against a calibrated building energy model. Thus, ASHRAE Guideline 14 (ASHRAE, 2002) is intended for comparing an energy model to measured building data. We would expect that a black box statistical model should perform with little error compared to the building energy model. This is important, because using the statistical model with MPC means that errors presented below are in excess to those intrinsic to the building energy model versus the real building performance.

Table 1 Allowable error levels for energy

<table>
<thead>
<tr>
<th></th>
<th>Timestep</th>
<th>Monthly</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allowable % error at average building power</td>
<td>10%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

While energy can be considered as a function of the power level for acceptable error, a different approach is needed for temperature. This is due to buildings operating at different steady state levels based on occupancy and/or seasons. To overcome this challenge, the proposed metric is a function of the change in internal zone temperature from the previous timestep, DeltaT. This allows for small errors about a nominal operating point, and a larger error during transition periods, where a step change in control setpoint has occurred.

The proposed curve structure shown in Figure 3 (see end of paper) is based on a logarithmic curve. The timestep error limit begins at 0.5 °C for no DeltaT to account for the resolution of the predictive building control options, measurement accuracy, and levels of temperature change noticeable by occupants. It then increases to approximately 1.5 °C for larger DeltaT’s. Only timestep level analysis has a curve, as the change in temperature metric does not apply over a period such as a month or year. The monthly and annual metrics are stricter than the timestep metric, and are outlined in Table 2, equal to 0.25 °C bias.

Table 2 Allowable error for temperature

<table>
<thead>
<tr>
<th>Period</th>
<th>Timestep</th>
<th>Monthly</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allowable Absolute Error for 0 °C Delta Temperature</td>
<td>0.5 °C</td>
<td>0.25 °C</td>
<td>0.25 °C</td>
</tr>
</tbody>
</table>

RESULTS

The following section presents application of developed tools on an example building, that of the Mona Campbell academic building at Dalhousie University in Halifax, Nova Scotia.

A statistical model is compared with tester data against the Energy Plus model that was used to create the training data upon which it is based. The statistical model was created using the Random Forest package for R developed by Breiman and Cutler (Liaw and Wiener, 2002) in collaboration with industry research partner Green Power Labs Inc. The Random Forest model was chosen as it provided the best accuracy when compared with linear regression and neural network models for the same data.

While developed tools produce plots for electricity, thermal energy and temperatures due to the space limitations only results for electricity consumption will be presented.

Residual Analysis (RA) tool

Annual analysis

An example of annual residual plots for energy consumption (i.e. electricity) are presented in Figure 4 and Figure 5 (see end of paper). Figure 4 illustrates the distribution of electricity residuals throughout a single year of rule based control. As show, there is no skew in the residuals, with the majority lying between -1 and +1 kWh. There is also no distinct difference in the different temperature split models as they are apportioned the same.
Figure 5 (see end of paper) compares the electricity residual and number of occurrences as a function of power level. This analysis provides a comprehensive insight into the accuracy of the statistical model as it provides answers to the four important questions:

1) At what power levels the statistical model is the most inaccurate? (see power level)

2) Does it under-predict or over-predict? (see bias)

3) At which outdoor temperature statistical model is inaccurate? (see colors)

4) What are the typical power levels?

As is presented, the majority of the time building uses between 40-70 kWh (see counts). The charts also shows that the statistical model under-predicts medium energy usage from 50-60 kWh and the highest from 80-90kWh. On the other hand, the model over-predicts the lowest energy consumption below 50 kWh and the energy use from 70-80 kWh.

As is presented, the majority of the time building uses between 40-70 kWh (see counts). The charts also shows that the statistical model under-predicts medium energy usage from 50-60 kWh and the highest from 80-90kWh. On the other hand, the model over-predicts the lowest energy consumption below 50 kWh and the energy use from 70-80 kWh.

Figure 6 and Figure 7 show a monthly residual analysis and absolute error analysis, respectively. The monthly residual analysis enables us to determine which energy carrier is the least accurately predicted (electricity or thermal), during which months and whether model under-predicts or over predicts. This analysis adds to the previous residual analysis as it identifies the months were the biggest gaps occur. The absolute error analysis provides information about the magnitude of these discrepancies, as they may be both over or under-estimated.

Figure 6 illustrates a peak deviation of 0.1% for electricity and 1.2% for thermal energy. The electricity shows no apparent seasonal trend, while the thermal energy over predicts during the winter and under predicts during the summer. Figure 7 (see end of paper) demonstrates the absolute error analysis, with a peak mean absolute percentage monthly error of 0.7% for electricity and 2.4% for thermal energy. The distribution on mean absolute error by month is flat, with a minimum of 0.4% for electricity and 1.6% for thermal energy.

**Hourly analysis**

Figure 8 and Figure 9 present an example of timestep plots for total electricity. Hourly analysis allow us to identify specific days and hours when the biggest discrepancies occur as well as to determine magnitude of these gaps. This provides information and knowledge about the accuracy of the statistical model and times of the day where it should be improved. The pivot charts are setup with filters to easily navigate from one day to another.

The 2 weeks analyzed are February 4-8 for the winter season, and July 15-19 for the summer season. These weeks correspond to the extremes found in the weather file (e.g. lowest and highest temperatures for winter and summer, respectively). As shown, the statistical model tracks the Energy Plus results, with a peak winter error of 2 kWh and summer error of 3 kWh, for the sample weeks. No distinct patterns occur in either the winter or summer electricity residual profiles.
Figure 9 Thermal energy timestep profiles for July 15-19

**Absolute Percentage Error (APE) tool**

*Annual and monthly analysis*

Sample monthly and annual analysis plots for electricity are provided in Figure 10. Each individual month is labelled to help identify potential areas of further investigation. All MAPE values should reside below the acceptable MAPE level line, with lower values indicating better performance. It is apparent in the figures that the annual data point “Ann” (yellow dot) lies at 100% of average value, which is expected. Summertime electricity months are at higher power level due to space cooling. All values are well within the limits.

Figure 10 Mona Campbell Building - Monthly/Annual electricity verification

In addition to the monthly and annual error analysis, an aggregate timestep analysis can be conducted. In such analysis, values for a particular time (e.g. 06:00) are taken from the 261 occurrences of that time per year (it would be 365, but weekends are not presently utilized). This population of data at each time is then presented statistically by MAPE (average at that time) and by APE’s according to the 95th and 99th percentile of the distribution. We expect that MAPE and 95th percentile APE for a given time’s population should always reside within the maximum error limit, which to be generous here is taken as the timestep APE at average power (e.g. 10%). It is expected that the 99th percentile APE may exceed the limit, as these are rare occurrences.

Figure 11 illustrates the MAPE and APE distributions of electricity. The APE tool produces similar chart for thermal energy that is not presented here. The electricity is shown to have largest APE at 6h, 8h, and 22h, which are control transition periods; all of which remaining within the acceptable limits. The maximum value as shown occurs only once throughout the year and is provided to understand the peak possible error.

Figure 11 Annual time of day electricity analysis

**Hourly analysis**

As the statistical model is used as an integral component for predictive building control that acts on the timestep level, it is important that statistical model provides accurate results for individual predictions to accompany good aggregate performance. To better evaluate timestep performance, scatter plots and statistical spread of APE values for electricity are plotted in Figure 12 and Figure 13.

In Figure 12, the scatter plot shows the distribution of all APEs. It can be seen that the vast majority lie below the limit, and that there is decreasing APE as a function of power level. Because every point is shown, the maximum APE can be identified. In this example,
electricity is 10.6% APE occurring at essentially the annual average power value.

Figure 12 Sample electricity timestep validation

![Electricity timestep validation](image)

Figure 13 Sample electricity timestep validation - Percentile

![Electricity timestep validation - Percentile](image)

CONCLUSION

This paper presents two tools that have been developed for validation of the high-resolution (e.g. 15min) statistical building response models: RA tool and APE tool.

The RA tool identifies power levels, months, days and hours when the statistical model was the most inaccurate. In addition it determine magnitude of these discrepancies and whether they were positive or negative. The APE tool sets acceptable limits and provides additional information that are required to improve and the accuracy of the statistical model. Together they provide a powerful and fast tool that can quickly identify areas for further investigation and improvement.

The limitation of this tool is that requires readjustment of the scales once a new data of an order of magnitude are inserted.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance and support provided by Green Power Labs Inc. throughout this research. In addition, the authors appreciate major funding support provided by the Atlantic Canadian Opportunities Agency supporting innovative economic growth in Canada’s Atlantic Provinces. Additional funding was provided by MITACS.

REFERENCES


<table>
<thead>
<tr>
<th>Date/TimeExcel</th>
<th>Month</th>
<th>DayOfMonth</th>
<th>DayOfWeek</th>
<th>Hour&amp;Minute</th>
<th>Min &amp; Hour</th>
<th>Outdoor Temperature</th>
<th>Temperture Model</th>
<th>Electricity</th>
<th>Electricity</th>
<th>Electricity</th>
<th>Electricity</th>
<th>Natural Gas</th>
<th>Natural Gas</th>
<th>Natural Gas</th>
<th>Natural Gas</th>
<th>Heating Setpoint</th>
<th>Cooling Setpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-01-02 0:00</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>00:00</td>
<td>00:00</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 0:15</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>00:15</td>
<td>00:15</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 0:30</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>00:30</td>
<td>00:30</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 0:45</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>00:45</td>
<td>00:45</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 1:00</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>01:00</td>
<td>01:00</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 1:15</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>01:15</td>
<td>01:15</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
<tr>
<td>2006-01-02 1:30</td>
<td>Mon</td>
<td>1</td>
<td>0</td>
<td>01:30</td>
<td>01:30</td>
<td>32</td>
<td>Natural Gas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>26.7</td>
</tr>
</tbody>
</table>

Figure 1 Data sheet
Figure 2 Energy allowable error

Figure 3 Temperature DeltaT acceptable error
Figure 5 Electricity residual and occurrence as a function of power level

Figure 6 Monthly residual energy annul
Figure 7 Monthly absolute error analysis

Bars are Sum of Monthly Absolute Error
Bracketed values are Percent of Monthly Consumption