Evaluation of Thermal Inertia of Urban Buildings from Metered Heat Energy Use

Muhammad Hafeez Saeed

2020-08-09

Abstract

This paper highlights the potential of black box forecasting tools to model thermal inertia of the urban buildings by using their metered heating data. The aim of this research is to examine how accurately an arima model can represent the thermal inertia of the buildings. This research employs hockey shaped energy signature regression model, to predict the energy data based upon temperature values. Furthermore, an extensive sensitivity analysis has been carried to check the impacts of various attributes such as time window, forecast length, time series length etc. on the model's performance. Finally, the paper also presents a detailed error analysis and validates our claim about potential of blackbox models in representation of building's thermal characteristics.

Introduction

Metered heating data of a total of 97 buildings in Stockholm, Sweden was used to build our analysis. To study the behavior of individual buildings, one building's energy data was analyzed at a time. Using the energy signatures model presented by Pasichnyi, Wallin and Kordas (2019), each meter's energy data was estimated using a hockey shaped univariate regression model, with ambient temperature as exogenous variable. Finding the difference between modelled and actual energy data, residual data was obtained. The notion behind usage of residual data was to preclude the influence of temperature from energy consumption data and to have a microscopic attention on the internal behavior of the individual building that could potentially provide meaningful insights to consumption patterns of urban buildings.

Following are the sequential steps that would demonstrate the precedure opted for examining the balckbox model approach for Stockholm residential buildings

1. Selection of meter:- Each residental building meter, was expecetd to show many overlapping attributes with other buildings yet, the building's construction uniqueness and ability to retain the heat energy within, was to be examined by choosing one meter at a time.
2. Hockey Shaped Energy Signatures:- Each building's metered heating data was modelled using hockey shaped univariate regression model, with ambient temperature as exogenous variable. This approach gives relatively more precise regression fit as compared to naive linear or quadratic regression fits. The function generateResiduals(), returns the residuals calculated from the equaltion:

Residuals = ES - Actual Readings

3. Selection of time period:- Also known as analysis window, the time period refers to the bracket of time and dates of the year for which the building's energy data is analyzed. This is an important attribute which later on has also been used for carrying out sensitivity analysis.

```
start_date <- ymd_hms("2012-01-01 00:00:04")
#days_vector <- c(15, 31) # in days
days_vector <- 31 # in days
```

4. Select the length of forecast period:- For our arima model of the residual data, the foreacast length is a requirement that needs to be provided. To have a meaningful forecast length, we considered 12 hours and 24 hours options.

```
forecast_length = 12 # in hours
#forecast_length = c(12, 24) # in hours
```

5. Parallel processing option:- A total of ninty seven (97) meters with each each having around 8700 metered readings, the serial implementation of the code gets to a complexity of O(n^3), which means too much computational time. therefore, in order to speed up the calculationss, an option for parallel processing hs been provided. The code was run on a computer with an octa-core processor, therefore seven cores were used when the parallel option was selected to true.

```
parallelization = TRUE
```

6. Modifying the residual dataframe according to various selected paramters: To modify the the residuals data frame according to the selected analysis windows, forecast length, etc., an O(n^3) complexity loop was implemented.

Black Box Models

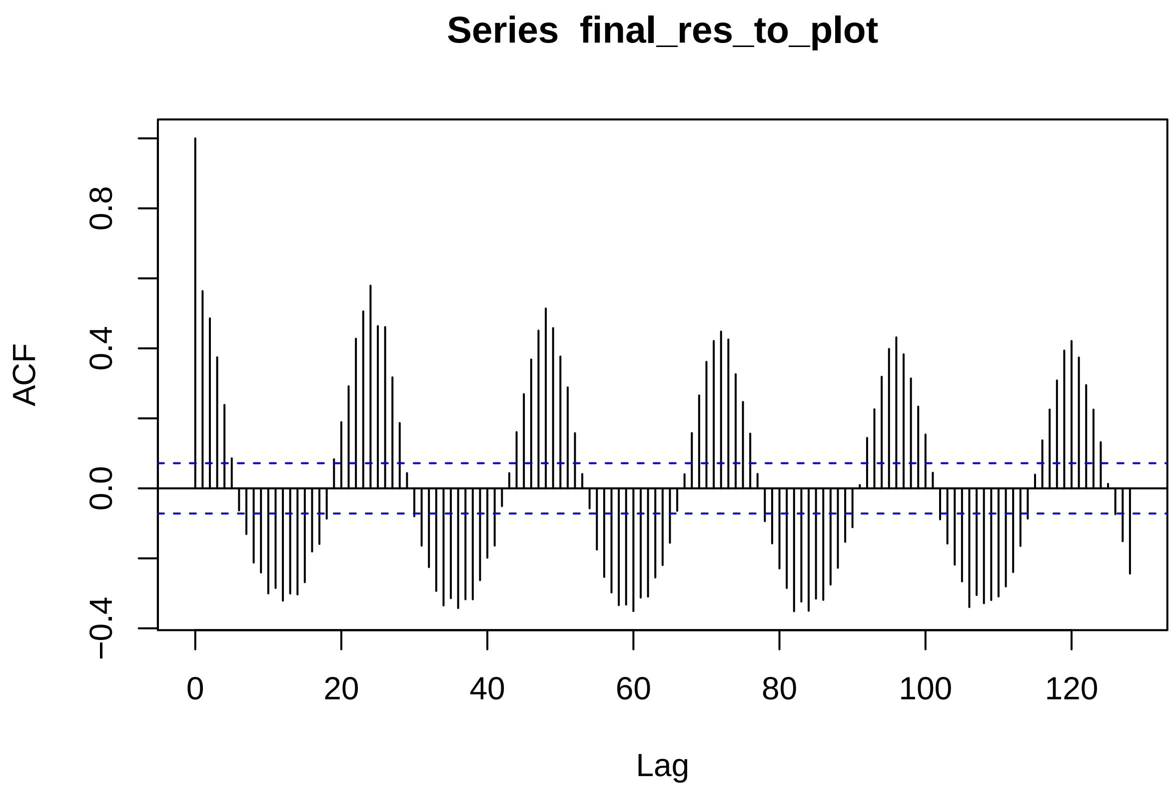
There are different notions attached to suitability of blackbox models in comparison to white box models. The most important of all is balance between intrepretability and accuracy. The blackbox models generally outperform whitebox models in many applications which are complex and involve highly non-linear behavior, eg: a deep learning neural network used to classify the healthy, COVID-19 patients would certainly be having some hidden layers that are tough to be explained yet are successful in classifying healthy individuals and patients. However, as discussed, blackbox models lack intrepretability which means in order to reassess or re-train models there is almost from the scratch training or analysis that needs to be done. Therefore, principally in applicatio where it is possible to explain the underlying phenomenon it is scientifically preferred to use whitebox models as they inherently entail better explainability and logic.

In this research on urban buidings, we already mentioned that we are anlayzing the behavior of residuals derived from the modelled and actual metered heat energy data and want to understand the knitty gritty patterns in it to better understand the thermal characteristics of the buildings, which could help in smart planning and efficient operations of these urban infrastructures.

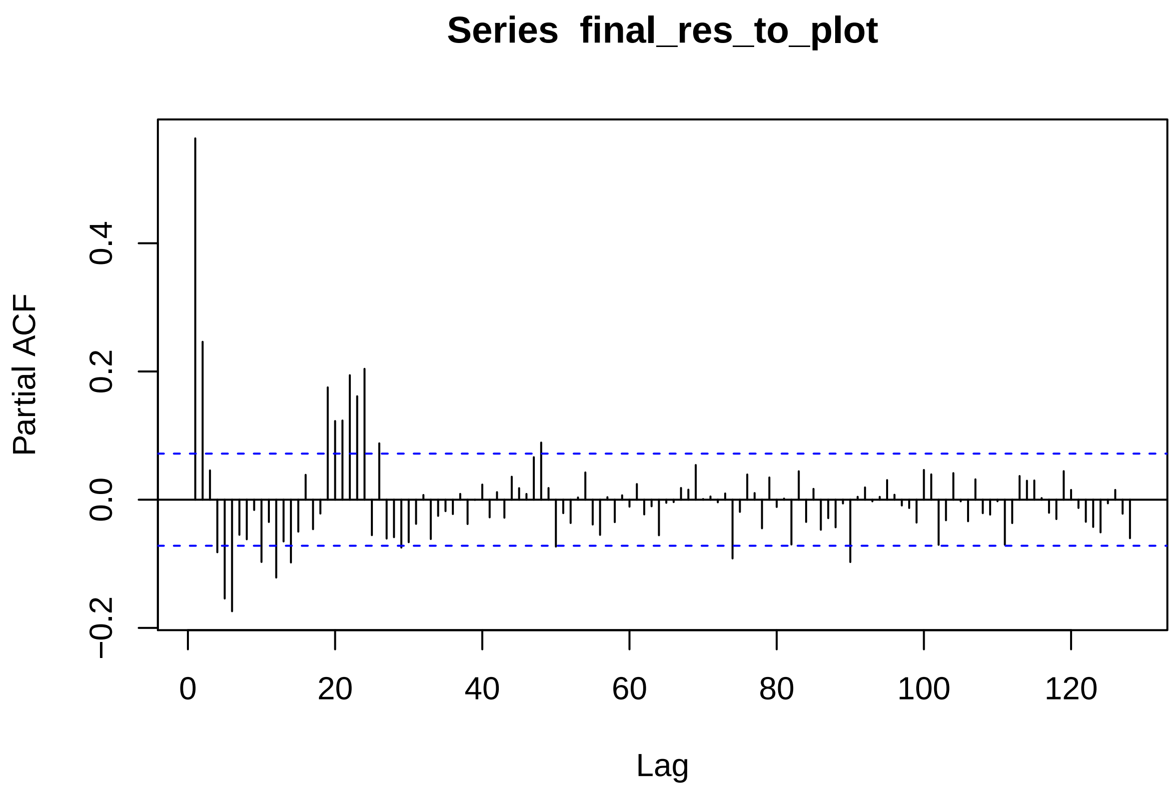
There are different kinds of black box methods: a. Linear Models including Autoregressive Models, Moving Average Models, Arima Models, Seasonal Arima Models, etc. b. Non-Linear Blackbox Models. Eg: Deep Neural Networks, Non-linear Ordinary Differential Equations based models, etc.

For this research we analyzed Seasonal Arima Models for modelling metered heating data.

The decision of using seasonal version in R was made solely based on Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots.



Series final_res_to_plot
ACF
0.8
0.4
0.0
−0.4
0
20
40
60
80
100
120
Lag



The ACF plot presented in the above figure 1 shows a clear repetition in the form of lagged cofficients apprearing after 24 lags suggesting a daily seasonality behavior in the energy consumption time series data of an individual meter. Since the ACF plot is not converging so the possibility of fitting an MA model is ruled out. However, to check whether AR model is applicable, the PACF plot was referred (also presented above in the figure) , it was evident that there were only two significant error cofficients before the first zero crossing. This suggested that AR model with two lagged cofficients incorporated is appropriate. Therefore, a daily (24 hrs) seasonality pattern and a two lagged AR model was chosen and implemented using Sarima function in Rstudio:

1. p = 2, d = 1, q = 0 (The regular/conventional arima paramteres fed to the sarima function)
2. P = 1, D = 0, Q = 0, m = 24, (The seasonal parameters fed to sarima function)

In regular parameters, note that a first order differencing (d = 1) has been applied to convert the series into a stationary series i.e. constant mean and variance.

In the seasonal paramteres, P = 1 means that only first seasonally offset observations are considered in model, e.g. (t -mx1)th or (t-12)th observation. In case of P=2, model would have used last two seasonally offset observations (t - mx1)th, (t - mx2)th and so on.

Quality analysis

To check the test accuracy of the Seasonal Arima Series(Sarima) model, we have considered two types of error metrics :-

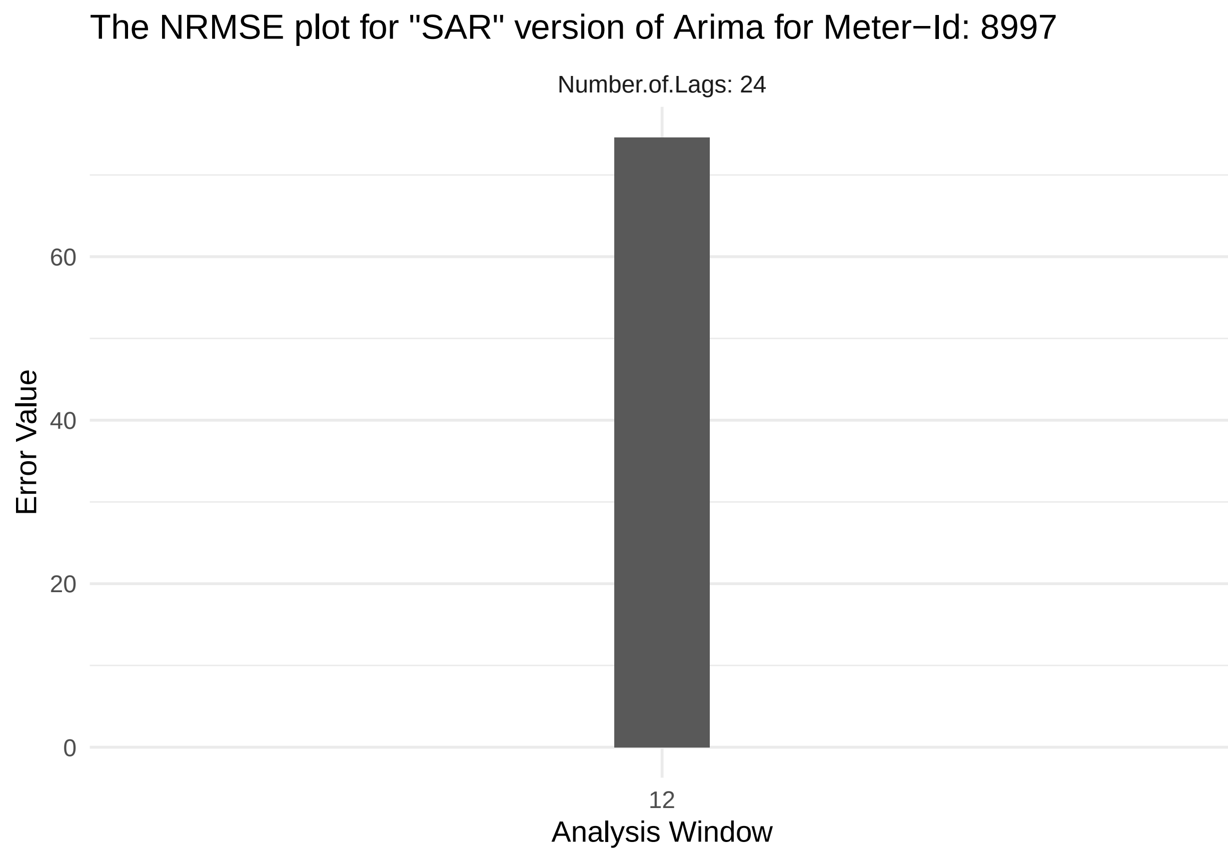
1. NRMSE

Root Mean Squared Error is a formal way to measure the error of a model in predicting quantitative data. The root mean square error (rmse) is a measure of how much individual observations on average deviate from the value predicted by the model. To standardise all estimates of rmse, we used the normalised rmse (nrmse) which relates the rmse to the observed range of the variable and gives a sliding scale measure of deviation between 1 (min)-100 (max).

2. Mean Error

The mean error is an informal term that usually refers to the average of all the errors in a set. An "error" in this context is an uncertainty in a measurement, or the difference between the measured value and true/correct value. The more formal term for error is measurement error, also called observational error.

The plots below show the error results for different analysis windows for different meters.



Next step is to evaluate the NRMSE difference between AR and Auto type of Arima model for each meter_id for a given forecast length and analysis window.

The final dataframe with all the required error metrics and the sarima coefficients, called "experiments", is shown in the below presented table. A clear improvement in the NRMSE and MAE values can be noted for SAR model in comparison the AutoArima model. The table can be extended according to the desired number of meters, forecast lengths, number of lags and analysis windows.

experiment_id

meter_id

Arima.Type

Analysis.Window

Forecast.Length

Number.of.Lags

Metric

Value

Coefficients

NRMSE_Diff

1

8997

SAR

744

12

24

NRMSE

74.6000000

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

1

8997

SAR

744

12

24

NRMSE

74.6000000

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

1

8997

SAR

744

12

24

NRMSE

74.6000000

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

1

8997

SAR

744

12

24

NRMSE

74.6000000

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

2

8997

Auto

744

12

24

NRMSE

116.1000000

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

2

8997

Auto

744

12

24

NRMSE

116.1000000

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

2

8997

Auto

744

12

24

NRMSE

116.1000000

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

2

8997

Auto

744

12

24

NRMSE

116.1000000

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

3

8997

SAR

744

12

24

Mean_Error

0.8484027

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

3

8997

SAR

744

12

24

Mean_Error

0.8484027

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

3

8997

SAR

744

12

24

Mean_Error

0.8484027

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

3

8997

SAR

744

12

24

Mean_Error

0.8484027

c(ar1 = -0.5476, ar2 = -0.2688, sar1 = 0.3324)

-41.5

4

8997

Auto

744

12

24

Mean_Error

2.4824972

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

4

8997

Auto

744

12

24

Mean_Error

2.4824972

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

4

8997

Auto

744

12

24

Mean_Error

2.4824972

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5

4

8997

Auto

744

12

24

Mean_Error

2.4824972

c(ar1 = 0.814687884341854, ma1 = -0.686604394708911, ma2 = -0.129016821646532)

-41.5