**Contents**

1. Problem Overview……………………………………………………………………………………………………………………… 2
2. Support Vector Regression Explained ……………………………………………….………………………………………. 3-4
3. Kernel Trick………………………………………………………………………………………………………………………………… 5
4. Data Setup…………………………………………………………………………………………………………………………………. 6-8
5. Variable Selection using LASSO…………………………………………………………………………………………………. 9- 10
6. Performance Stats & Graphs………………………………………………………………………………………………………11-17
7. Conclusion…………………………………………………………………………………………………………………………………. 16

**Problem Overview**

Climate changes should be incorporated while designing buildings now and, in the future, to make them durable and resilient over a long period. Up to now, building envelopes are designed to withstand historical climate loads which are assumed to be static. But there is enough evidence that climate has been warming globally, thereby causing more frequent and extreme climate events. This can have significant impacts on building infrastructure, particularly the durability of building envelopes components. Therefore, there is a need to consider climate resiliency when designing new or retrofitting old building envelopes.

Performing field and laboratory tests to examine the limit of building materials under conditions like those that are expected in the future can be expensive, time consuming and requires specialized materials to test on and equipment’s to measure the performance. An alternative approach is to run the hygrothermal simulation models such as DELPHIN, WUFI, COMSOL, etc. which provide results from which to infer the durability and climate resiliency of building envelopes components. Although this mitigates the cost associated with a lab test, this approach is time-consuming since each simulation requires initial conditions setup and can takes days to run for 31 consecutive years. Moreover, the uncertainties associated with the future projected climate data results in several different sets of runs of simulation for the same year in a city. Hence computational intelligence and machine learning techniques are often looked upon by either accurately representing all the years by some combination of selected years called Moisture Reference Years or by reasonably well predicting the entire simulation results for all the years.

Considering the time constraint associated with the simulations especially considering in 2D and 3D ,where it can take even longer for the simulation to run, and success of the modern machine learning algorithms, this report presents an approach based on the Support Vector Regression (SVR) to predict the hygrothermal performance of the wall assembly, particularly focusing on the moisture performance or mould index of OSB or CLT over 31 years in three cities in Canada. The main idea is to construct and train the model on the first 5 consecutive years, extrapolate for the remaining 26 years, and validate with the actual simulation results from Delphin.

**Support Vector Regression Explained**

The support vector machine was first introduced by Vapnik and Cortes in 1995. [6] It was first used as a classification technique and later adopted for regression problems. Here we will focus on epsilon-Support Vector Regression. Epsilon-SVR works by minimizing the upper bound on the sum of training data called Structure Risk Minimization (SRM) as opposed to minimizing the sum of training error itself called Empirical Risk Minimization. This means SVR has a benefit over traditional machine learning approaches as it does not try to overfit on training data rather allow for some error threshold to capture the general trend in the data.

Given a set of data points {} which are generated from an unknown function. SVR estimates the relation between the explanatory variable x and response y by the following.

where

* <.,.> denotes the dot product
* is called the coefficient vector and b the bias term and are unknown and can be estimated by minimizing the regularized risk function.
* is a function that maps the input space x to a high dimensional space in order to overcome the nonlinear relationship between x and y.

In *ε*-SV regression (Vapnik 1995), our goal is to find a function *f* (*x*) that has at most *ε* deviation

from the obtained targets *yi* for all the training data, and at the same time is as flat as possible. In other words, we do not care about errors as long as they are less than *ε* but will not accept any deviation larger than this.

This objective of finding the flattest function yields the following:

subject to

The magnitude of act as a regularization term that controls the flatness of the solution. For example, if the same data can be modeled by a linear equation with the same accuracy as with a quadratic equation, then the flatter of the two model i.e. the linear model is more desirable.

The assumption in the above optimization problem is that that such a function *f* actually exists that approximates all pairs (*xi , yi* ) with *ε* precision, or in other words, that the convex optimization problem is *feasible*. Sometimes, however, this may not be the case, or we also may want to allow for some errors. This is where slack variables are introduced to cope with the otherwise infeasible constraint. The concept of slack variables is simple: for any value that falls outside of ϵ, we can denote its deviation from the margin as ξ. We know that these deviations have the potential to exist, but we would still like to minimize them as much as possible. Thus, we can add these deviations to the objective function.

Now the objective function becomes:

subject to

Where C is a constant that determines the trade off between minimizing training error and minimizing model complexity.

**Figure 1: Visualization of Support Vector Regression Objective Function**



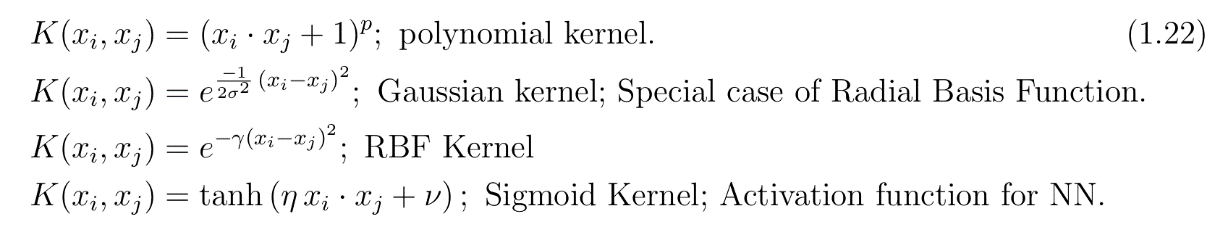
The right side of the image shows soft margin loss setting for SVR. Meaning any values between -ϵ and + ϵ are assigned a loss 0 and values outside the range is assigned a loss of .

**Kernel Trick**

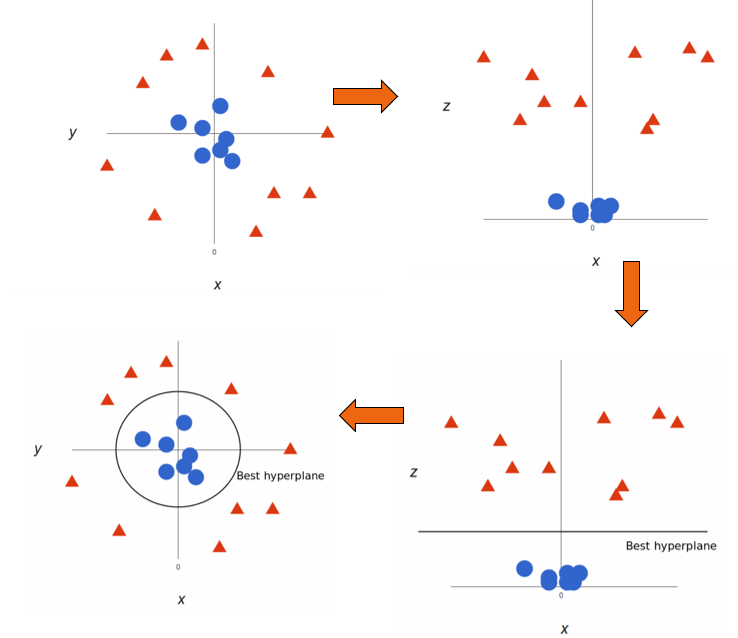
To make SVR adapt to nonlinear cases kernel are used. The idea is mapping the non-linear separable dataset into a higher dimensional space (feature space) where we can find a hyperplane that can separate the samples.

In the formulation of SVR, the complexity of a function’s representation by Support Vectors (SVs) is independent of the dimensionality of the input space *X*, and depends only on the number of SVs. Moreover, note that the complete algorithm can be described in terms of dot products between the data. Even when evaluating *f* (*x*) we need not compute *w* explicitly. These observations comes in handy for the formulation of a nonlinear extension. The difference to the linear case is that *w* is no longer given explicitly. Also note that in the nonlinear setting, the optimization problem corresponds to finding the *flattest* function in *feature* space, not in input space.

Common Kernel Function Choices:



**Figure 2: Visualization of Kernel Trick**



In this report the RBF kernel is used which is a simple function that can model systems of varying complexity. It is an extension of the linear kernel. The parameter gamma must be optimized when use this kernel. (See Grid Search)

**Data Setup**

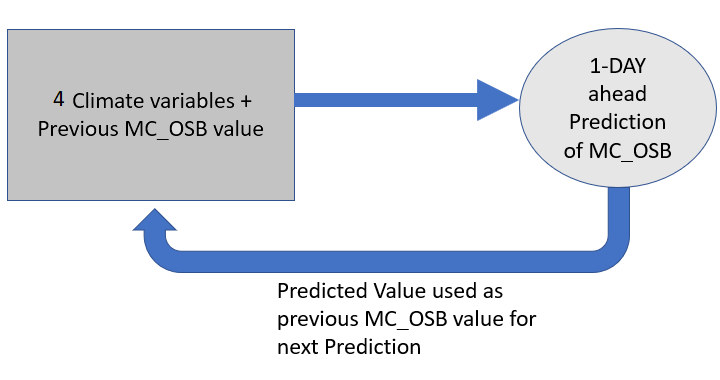
**Possible Explanatory Variables:**

Candidate Variables included Temperature, Relative Humidity, Wind Speed, Wind Direction, Direct Radiation, Diffuse Radiation, Wind Driven Rain (WDR). LASSO Regression was performed to do variable selection among the possible explanatory variables. Four variables including Temperature, Relative Humidity, Direct Radiation & Wind Driven Rain (WDR) were found to be the most important and were used to model the responses.

**Two Response Variables are modelled:**

1. Mould Index: To calculate mould index, Temperature & Relative Humidity in the wall assembly were predicted using Temperature, Relative Humidity, Direct Radiation & Wind Driven Rain (WDR). Hourly values were used.
2. Moisture Content: Moisture Content in the wall assembly were predicted using Temperature, Relative Humidity, Direct Radiation, Wind Driven Rain (WDR) & past values of Moisture Content. Daily values were used.

**Figure 3: Procedure for Predicting Moisture Content**



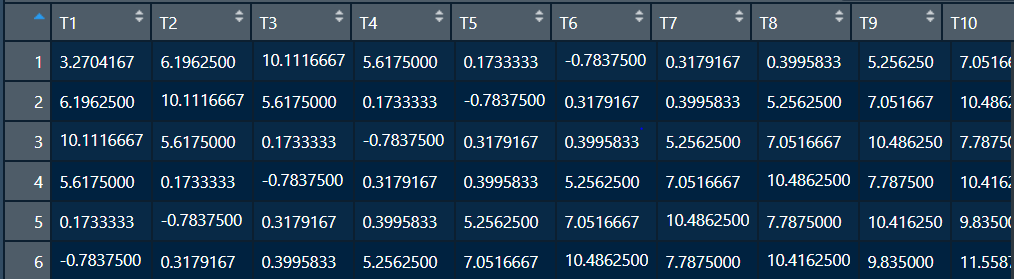
**Data Used:**

1. Mould Index: 3 claddings – Brick, Fibreboard, CLT with both Historical & Future for Ottawa, Calgary & Vancouver (North Orientation) resulted in total 18 cases i.e. 3 (Claddings)\*3(Cities)\*2(Historical & Future).
2. Moisture Content: Brick, Fibreboard, CLT with both Historical & Future for Ottawa, Calgary & Vancouver (North Orientation) resulted in total 18 cases i.e. 3 (Claddings)\*3(Cities)\*2(Historical & Future). Additional few cases for default Orientation were also included.

**Optimal Lag:** Support Vector Regression (SVR) is a tool used for supervised learning problem i.e. it tries to find relationship between explanatory variables and response variable. However, in the current context the effect of climate variables is not immediately reflected on the OSB since there a lag of certain time period (depending on the response variable) when the actual effects start to show. To take this behavior into account, a sliding window approach is used to make SVR suitable for this application. Instead of preparing the data in a row for only the current day, certain previous days of data is also included.

Example: To predict the first day of the 6th year, look at the past 10 values of each climate variable (Only Past 10 Days of Temperature Shown Below)

**Figure 4: Data Setup using Sliding Window**



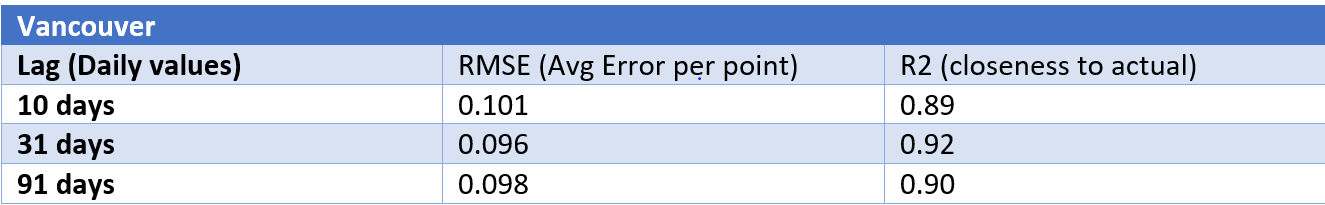
Different Lags of explanatory variables were used to model the response in few cases to assess the performance in terms of RMSEP & R2. In the case of

1. Mould Index: Past 48 hours (2 days) for Temperature & past 168 (7 days) for Relative Humidity were found to be optimal. Table below shows the performance of different lags on Temperature and Relative Humidity for one such case. Note for RH, performance gain can be made by using an even higher lag but requires more computational power.

|  |  |  |
| --- | --- | --- |
| Ottawa - Historical | Temperature Surface CLT out | |
| Lag (Hourly values) | RMSEP | R2 |
| 24 hrs | 0.36 | 0.96 |
| 48 hrs | 0.33 | 0.99 |

|  |  |  |
| --- | --- | --- |
| Ottawa - Historical | Relative Humidity Surface CLT out | |
| Lag (Hourly values) | RMSEP | R2 |
| 24 hrs | 7.03 | 0.80 |
| 72 hrs | 6.01 | 0.84 |
| 120 hrs | 5.33 | 0.89 |
| 168 hrs | 5.03 | 0.90 |

1. Moisture Content: Past 31 days were found to be optimal for forecasting Moisture Content. Table below shows the performance of different lags on Moisture Content for one such case.



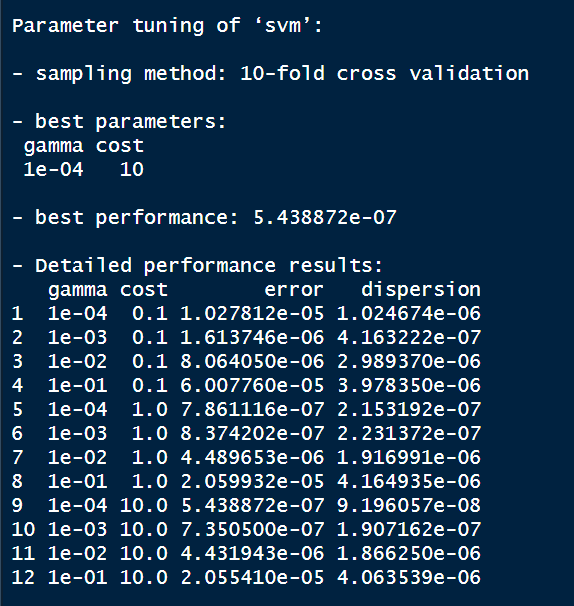
**Grid Search**

To optimize the parameter associated with SVR such as epsilon, gamma & Cost, grid search was performed with epsilon fixed at 0.1. Different combination of gamma & Cost was varied with 10-fold cross validation performed on the training set and the error calculated. Parameters combination which yielded the lowest error were then used for forecasting on the test set.

Set parameter search ranges as follows:

* Cost - from 0.1 to 10 in multiples of 10.
* Gamma - between one of the following values: 10-4 to 10-1 in multiples of 10.

**Figure 5: Finding optimal parameters using Grid Search**



**Packages**

The choice of Package is different depending on the response because Mould Index requires hourly value which is computationally expensive to train and forecast compared to average daily values for Moisture Content.

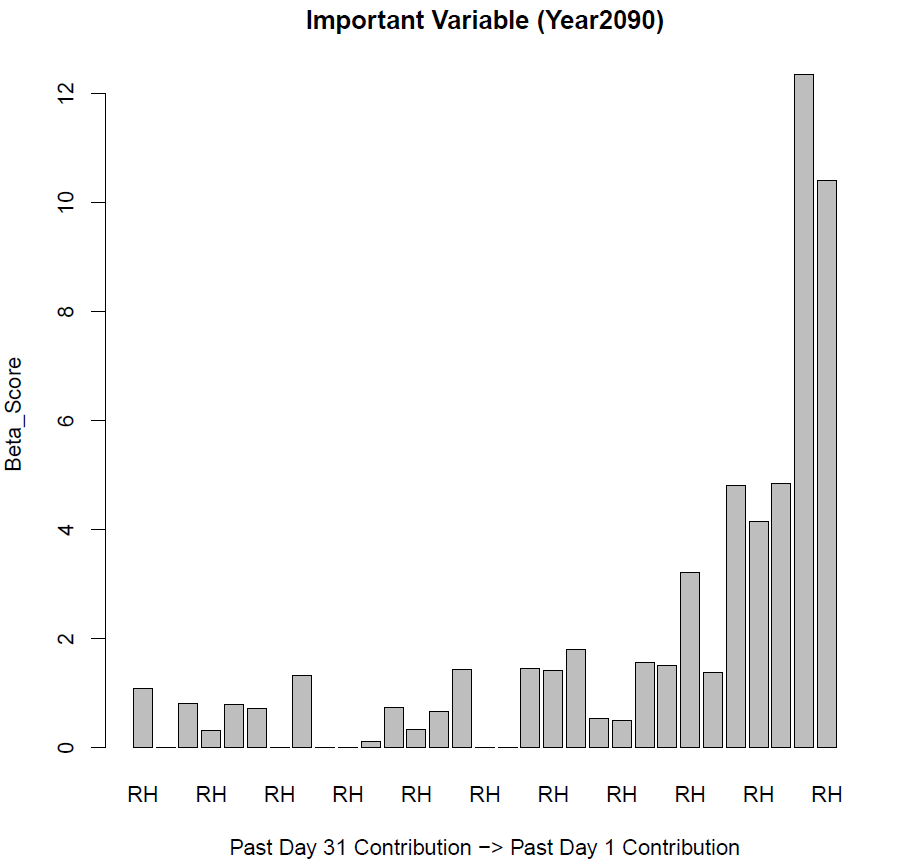
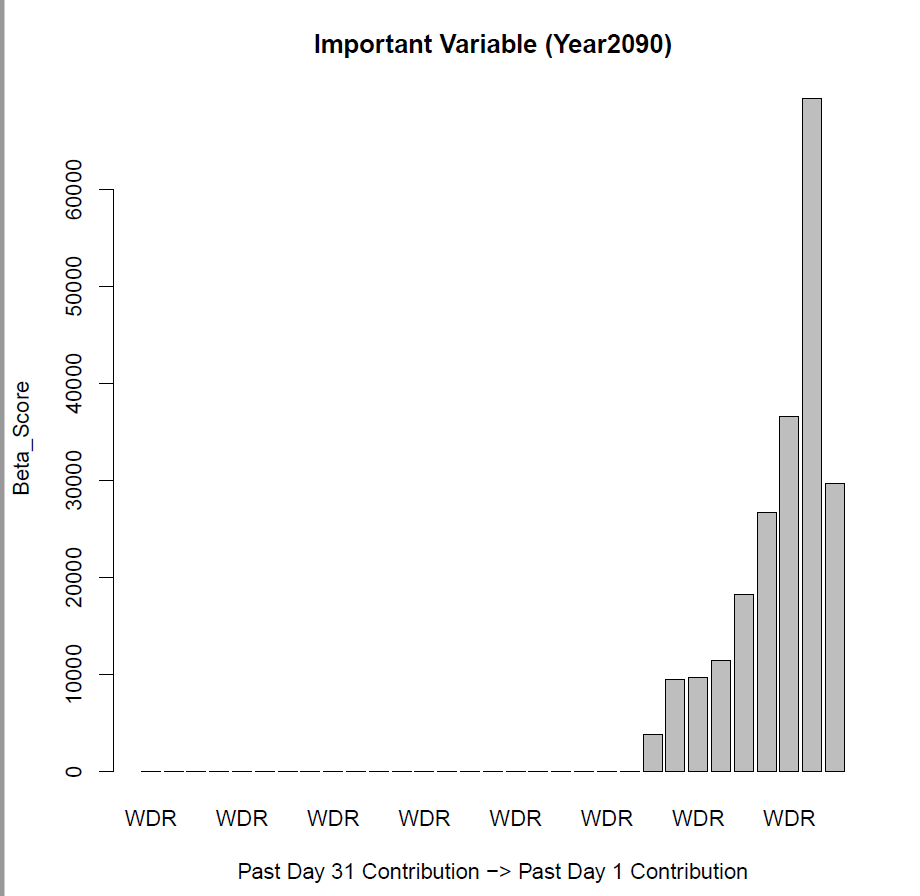
1. Mould Index: Package **ThunderSVM** in R was used for forecasting Temperature & Relative Humidity. This package exploits NVIDIA GPUs and multi-core CPUs to achieve high efficiency.
2. Moisture Content: Package **e1071** in R was used to forecast Moisture Content. Custom Code was written to incorporate the previous response in predicting the current response.

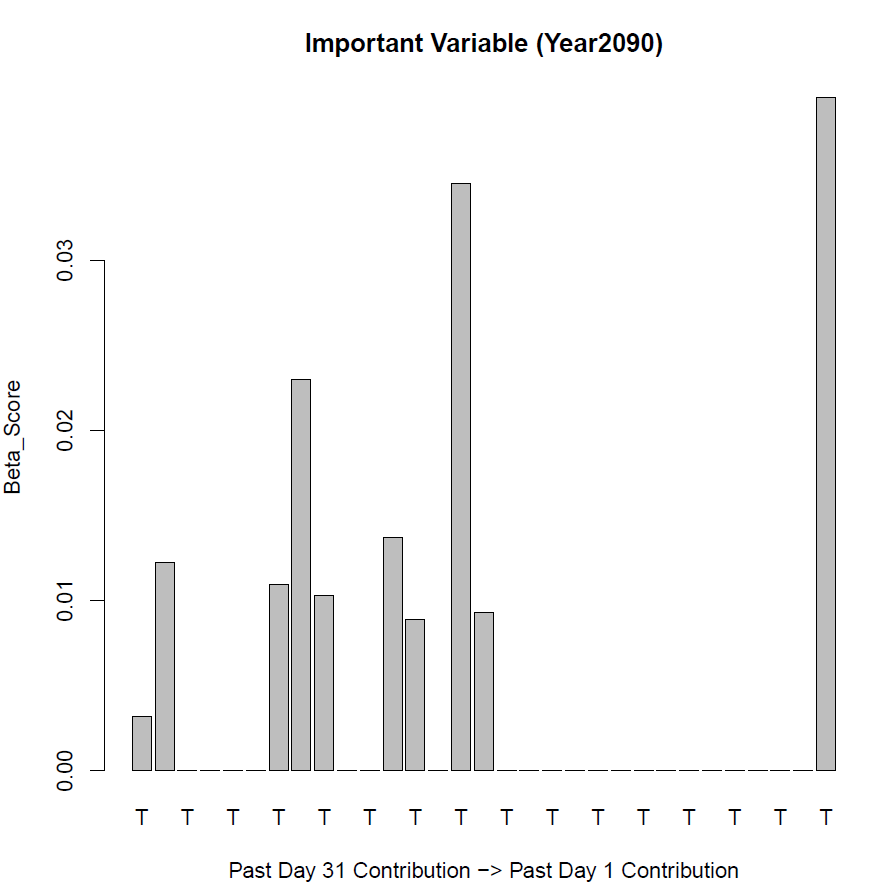
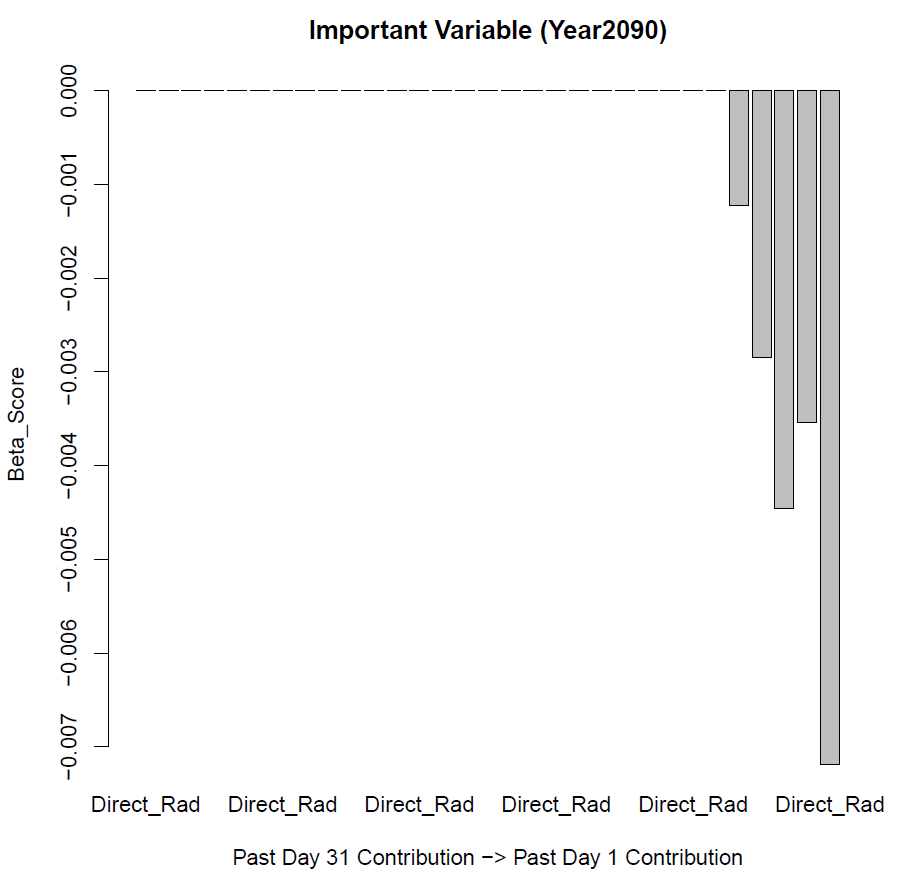
**Variable Selection Using LASSO**

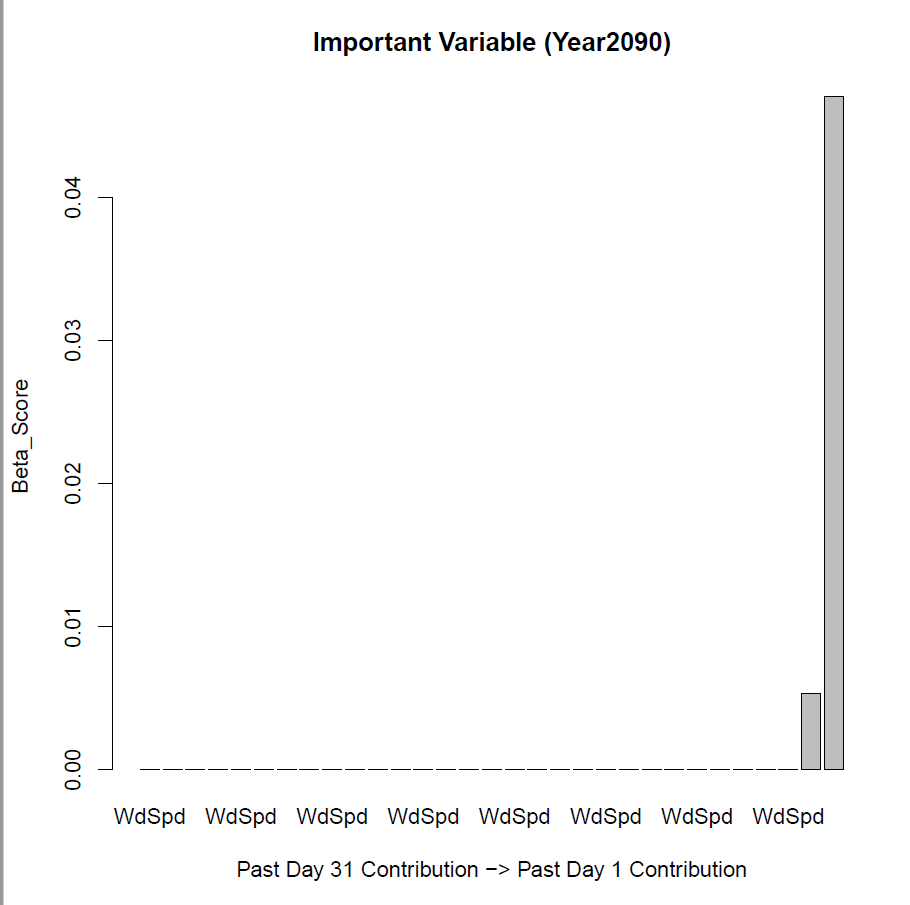
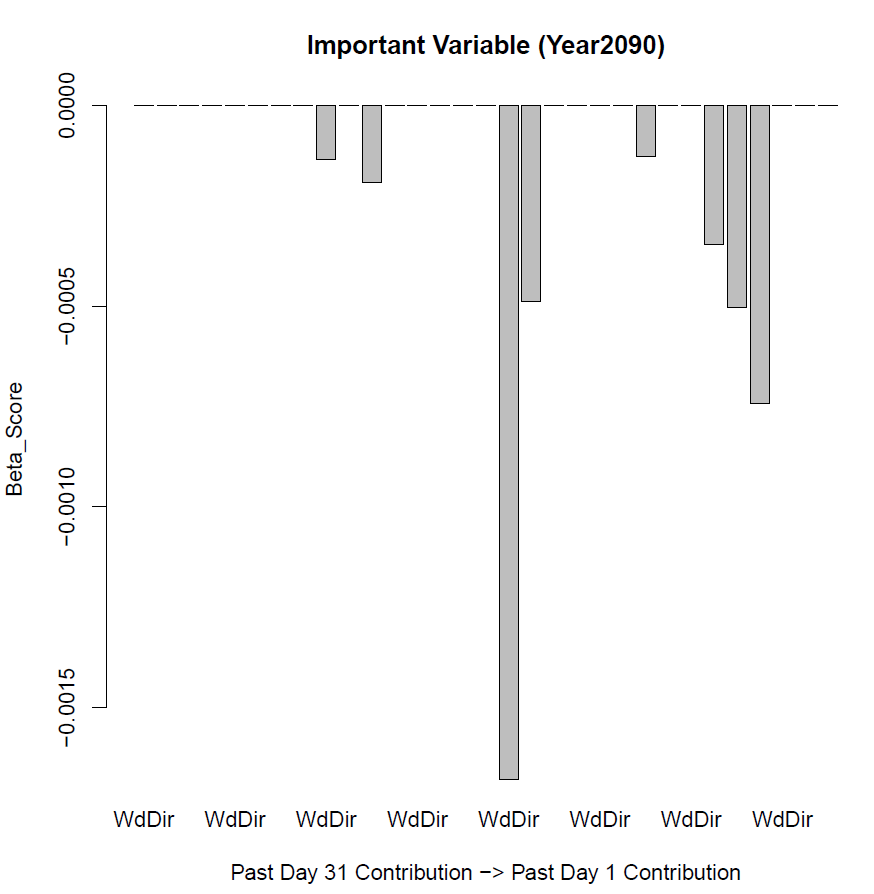
Lasso Regression was performed to determine which variables including Temperature, Relative Humidity, Wind Speed, Wind Direction, Direct Radiation, Diffuse Radiation, Wind Driven Rain (WDR) explain the significant variation in the Relative Humidity out OSB. Data was organised using the past 31 days of each explanatory variable to predict the response at current time step. (Note Average Daily values of RH and explanatory variables are used here, hence the lag is higher).

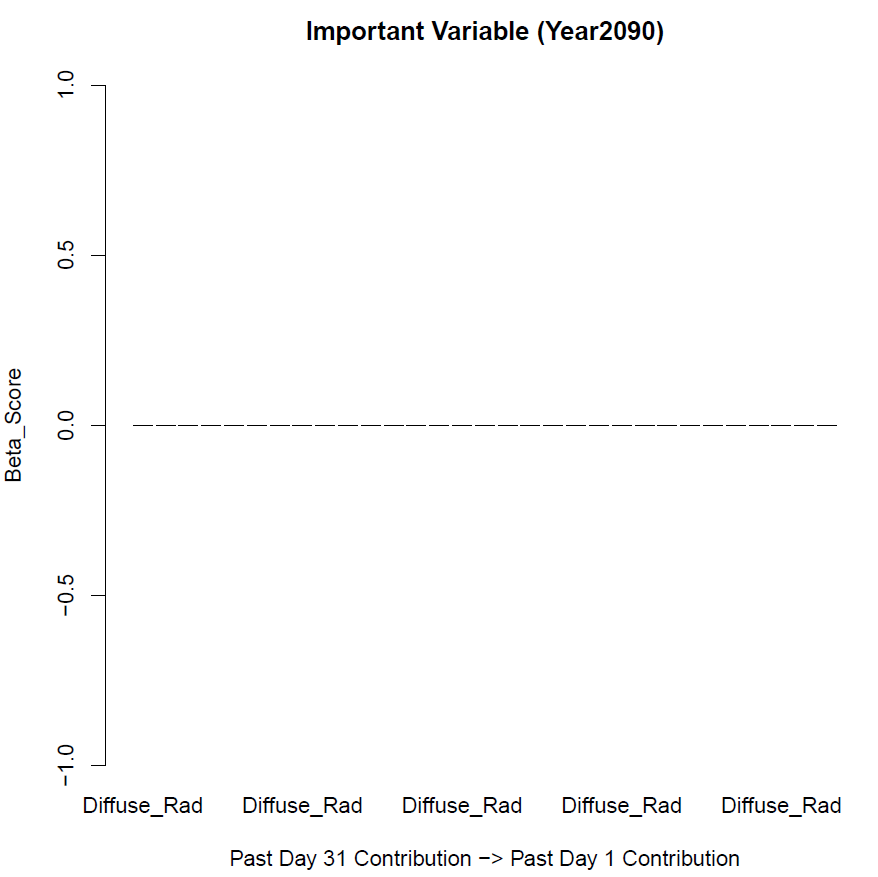
The case shown here is for Ottawa City for the Future Year 2090. Here on the Y-AXIS the beta score is shown. Lasso recognizes the importance of a variable contribution but assigning it a beta score. Important Variable are given a higher Beta Score. Wind Driven Rain, Relative Humidity, Temperature and Direct Radiation show the most contribution and are thus selected.

Note to select optimal lamba in lasso regression, 5 fold cross validation were performed using a grid of possible lamba values. Then the lamba associated with the 1 SE rule was selected. The below graphs show the regression coefficent associated with this lambda.





**Performance Statistics & Graphs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Brick- RH\_out\_OSB | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 8.29 | 0.03 | 7.89 | 0.05 |
| Calgary | 2.73 | 0.57 | 6.67 | 0.3 |
| Vancouver | 5.21 | 0.55 | 3.4 | 0.57 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Brick- T\_out\_OSB | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.68 | 0.99 | 0.70 | 0.99 |
| Calgary | 0.56 | 0.99 | 0.61 | 0.99 |
| Vancouver | 0.52 | 0.99 | 0.53 | 0.99 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fibreboard- RH\_out\_OSB | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 6.16 | 0.31 | 5.56 | 0.63 |
| Calgary | 5.58 | 0.42 | 5.44 | 0.58 |
| Vancouver | 3.24 | 0.55 | 2.94 | 0.63 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fibreboard- T\_out\_OSB | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.85 | 0.99 | 0.88 | 0.99 |
| Calgary | 0.73 | 0.99 | 0.76 | 0.99 |
| Vancouver | 0.63 | 0.99 | 0.65 | 0.99 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CLT-Surf-RH | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 5.33 | 0.89 | 5.71 | 0.86 |
| Calgary | 6.20 | 0.88 | 6.35 | 0.88 |
| Vancouver | 8.70 | 0.36 | 9.54 | 0.26 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CLT- Surf-T | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.35 | 0.99 | 0.37 | 0.99 |
| Calgary | 0.37 | 0.99 | 0.41 | 0.99 |
| Vancouver | 0.63 | 0.99 | 0.40 | 0.99 |

|  |  |  |  |
| --- | --- | --- | --- |
| Response | Lag | Training Time on 5 years | Prediction Time on 26 years |
| Temperature | Past 48 Hours | 5 min | 3 min |
| Relative Humidity | Past 120 Hours | 12 min | 5 min |

**Moisture Content**

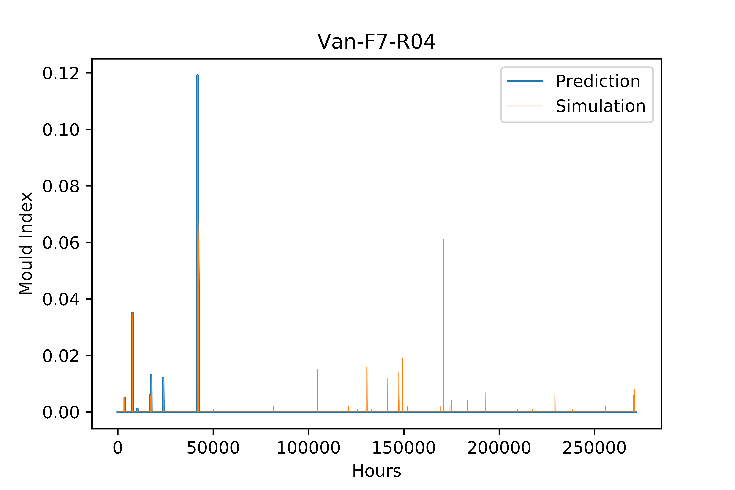
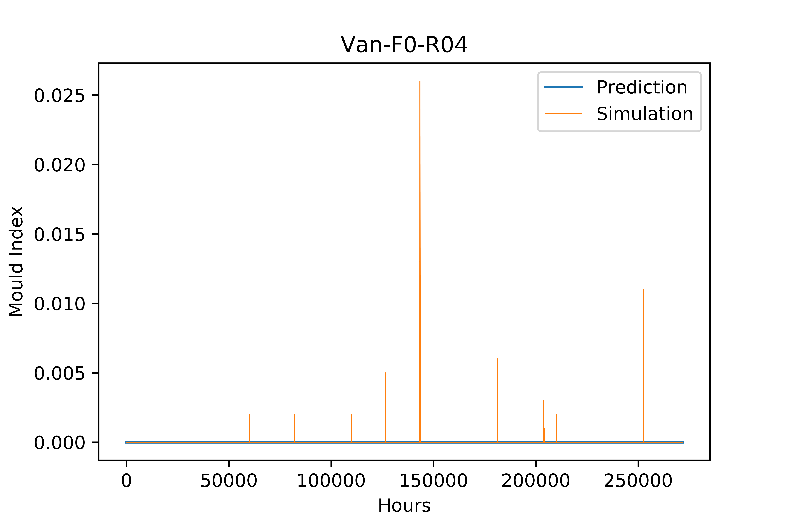
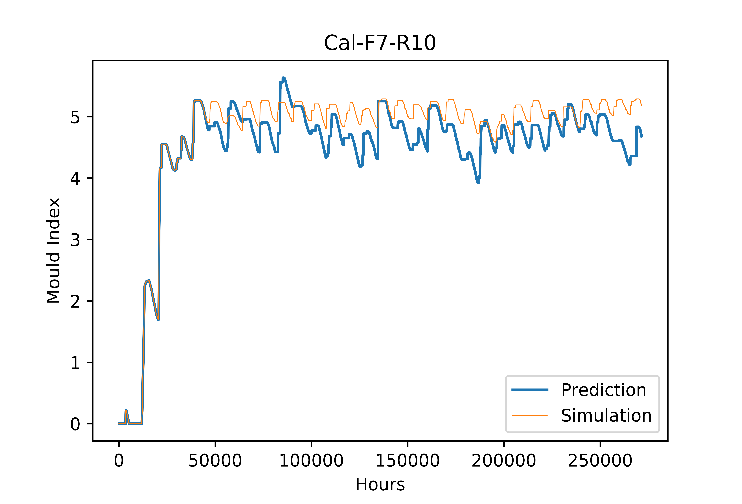
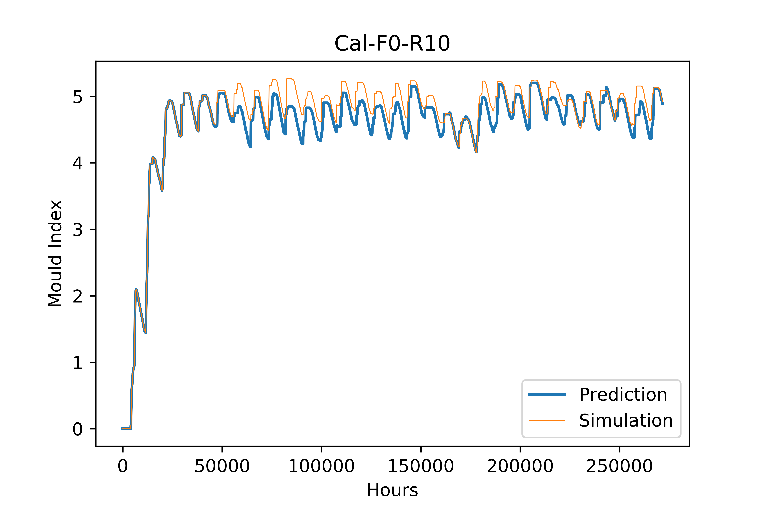
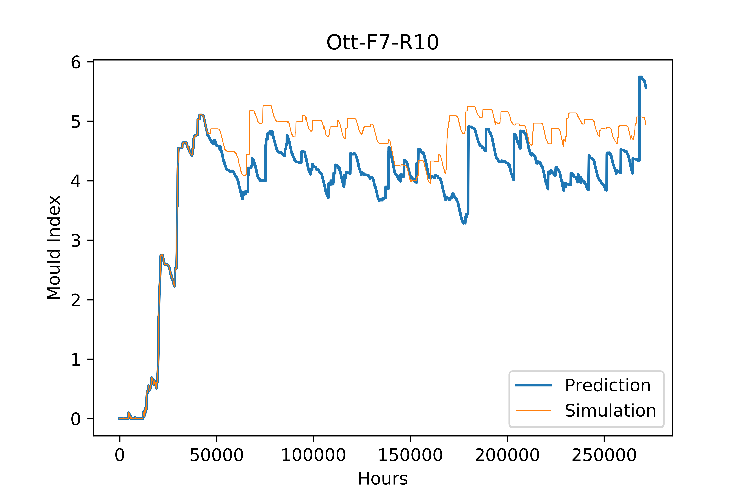
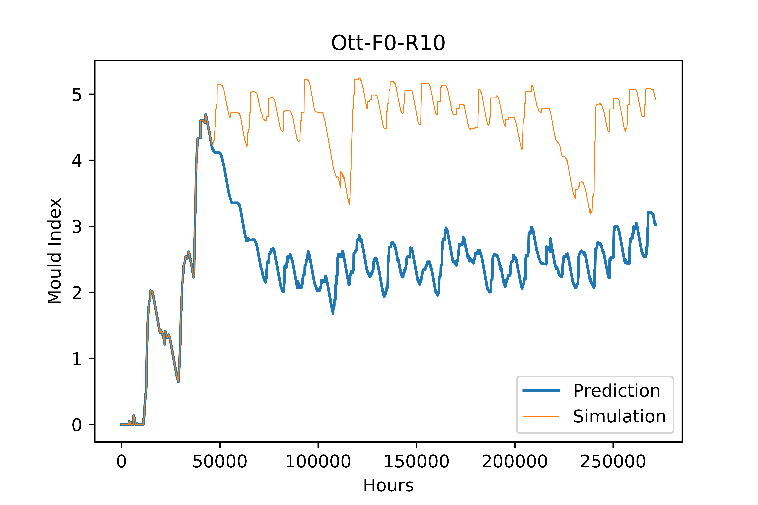
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Brick- MC\_OSB (kg) | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.27 | 0.01 | 0.21 | 0.16 |
| Calgary | 0.17 | 0.26 | 0.33 | 0.07 |
| Vancouver | 0.03 | 0.51 | 0.04 | 0.45 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fibreboard- MC\_OSB OSB (kg) | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.15 | 0.23 | 0.08 | 0.65 |
| Calgary | 0.03 | 0.89 | 0.03 | 0.87 |
| Vancouver | 0.01 | 0.74 | 0.02 | 0.80 |

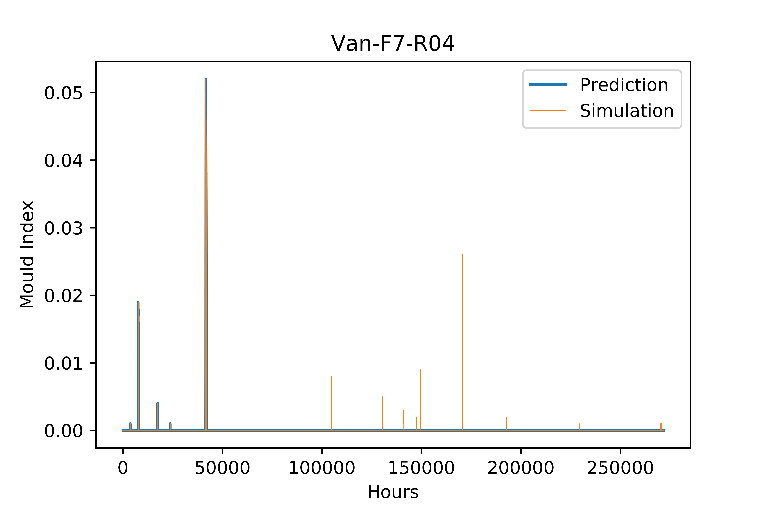
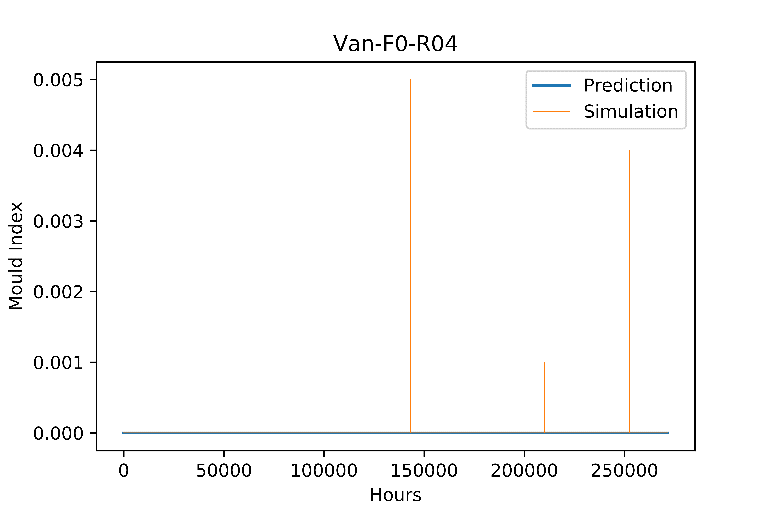
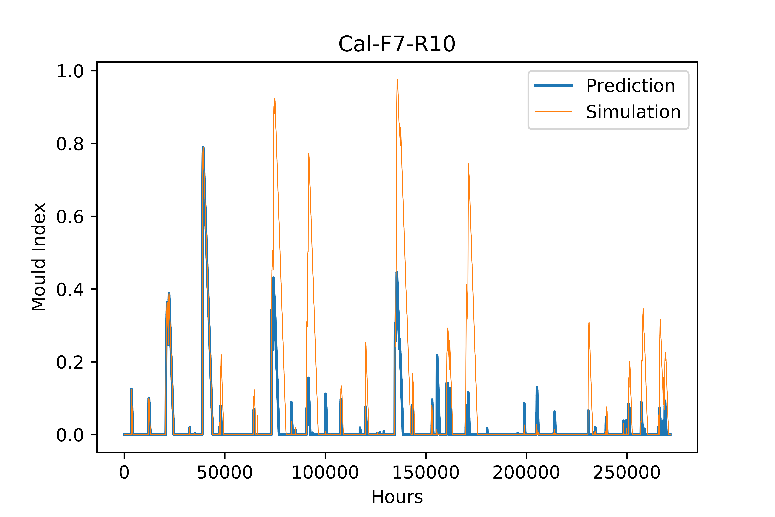
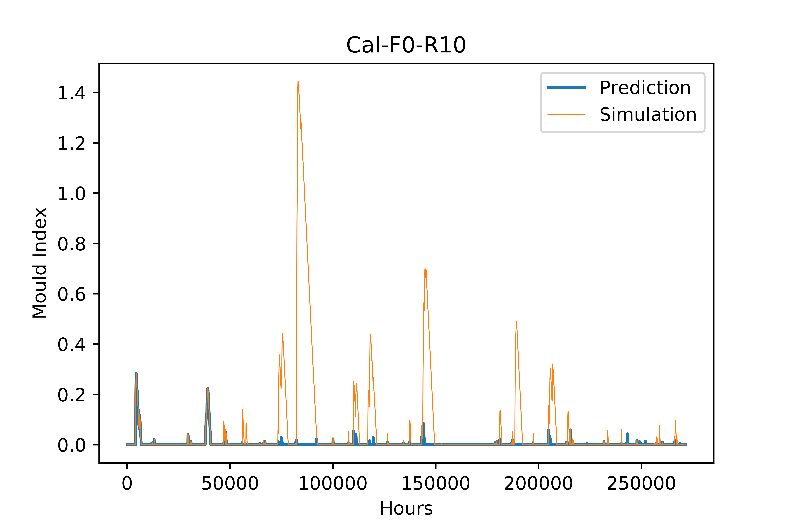
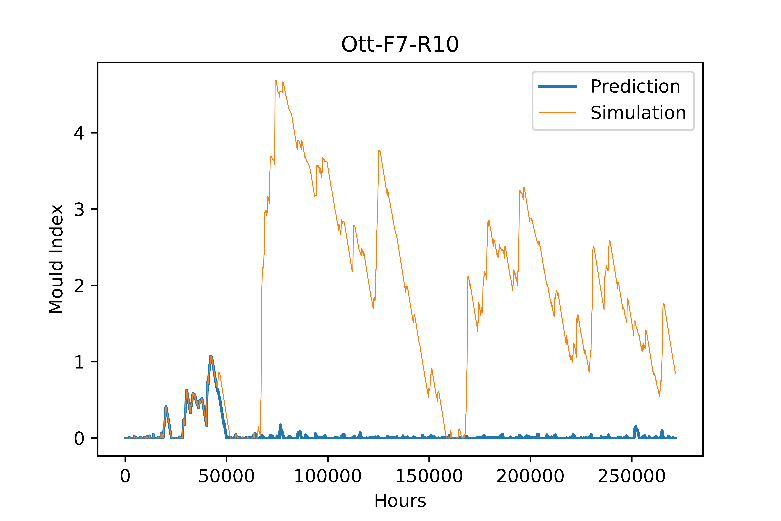
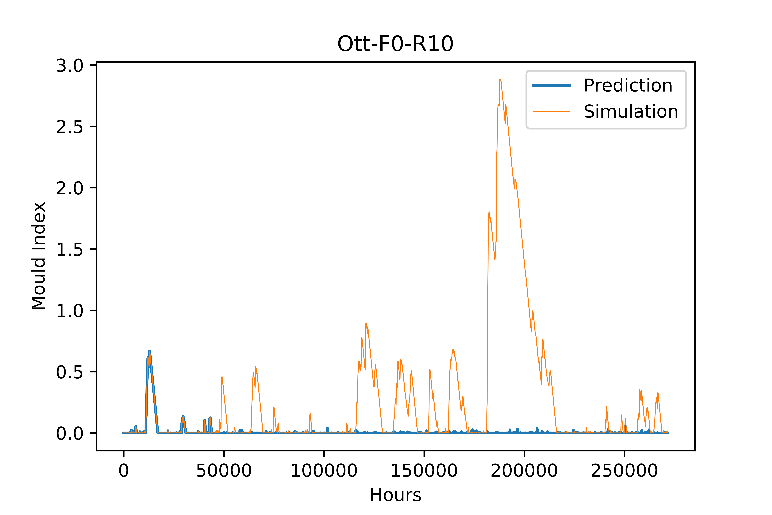
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CLT (kg/kg) | | | | |
| City | **Historical** | | **Future** | |
|  | RMSEP | R2 | RMSEP | R2 |
| Ottawa | 0.02 | 0.83 | 0.03 | 0.70 |

**Mould Index**

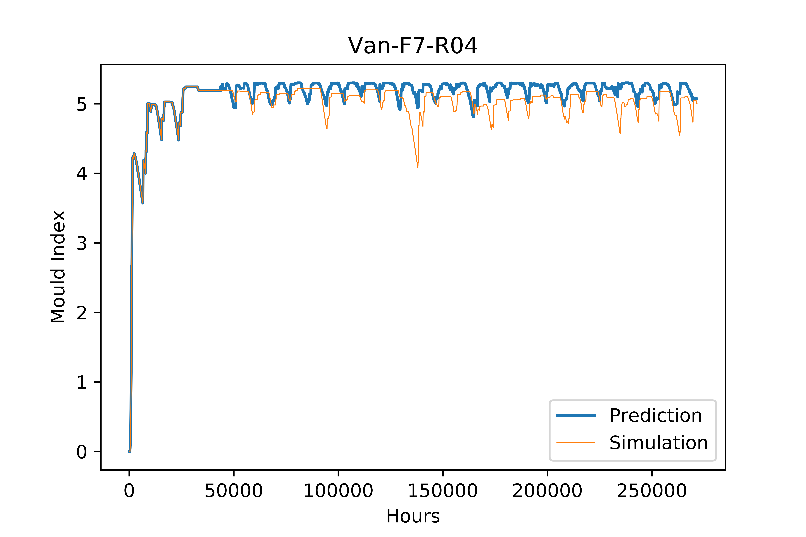
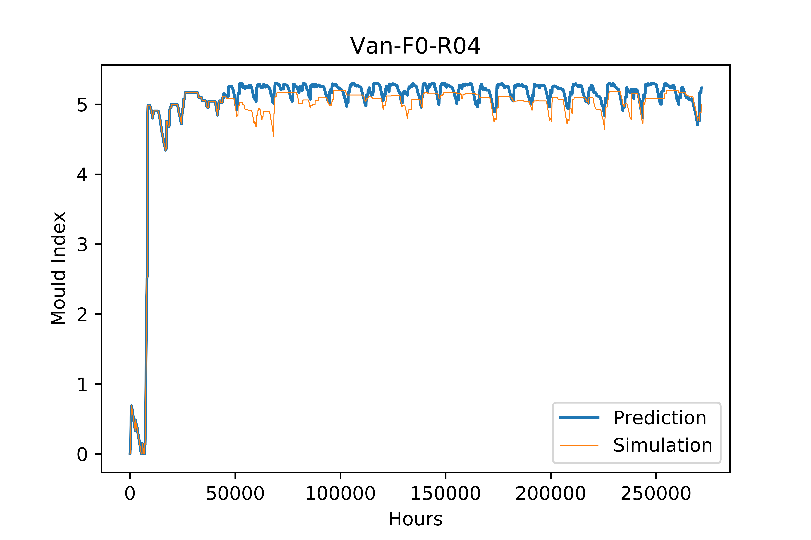
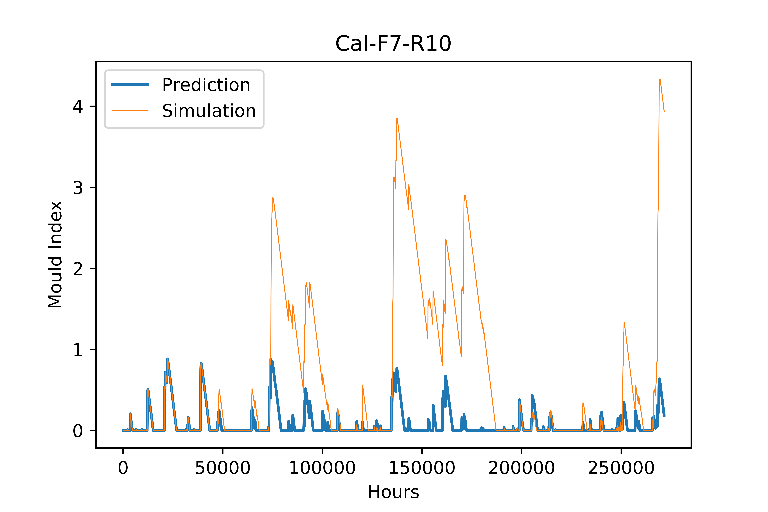
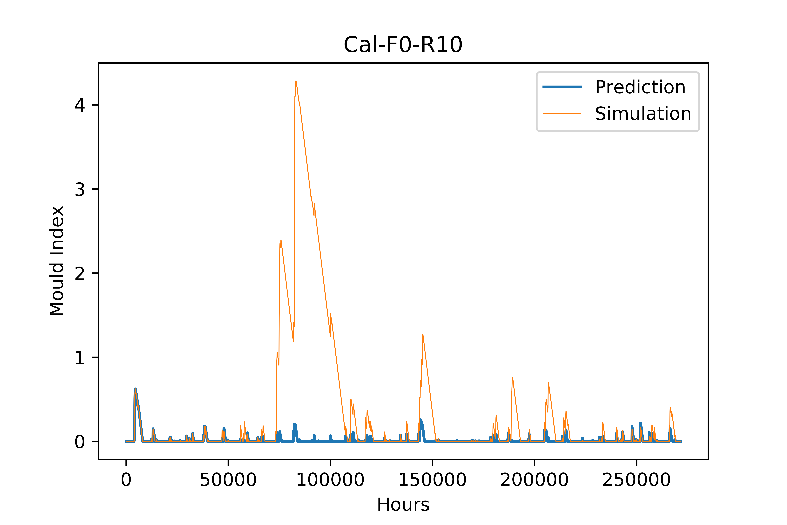
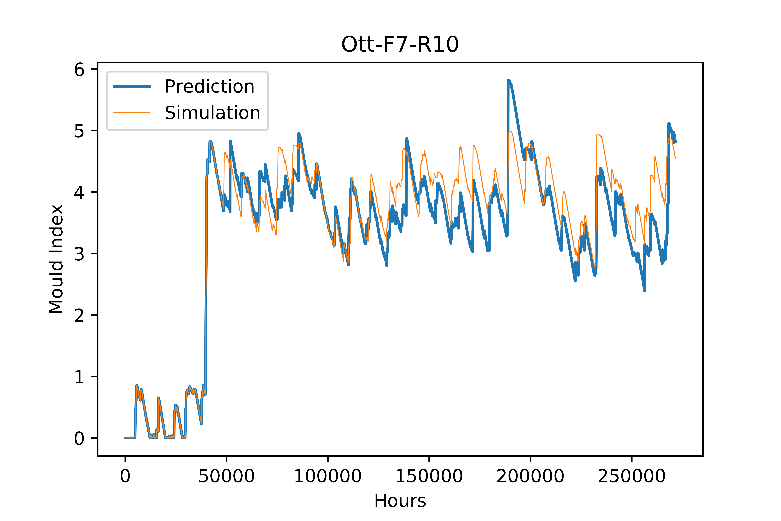
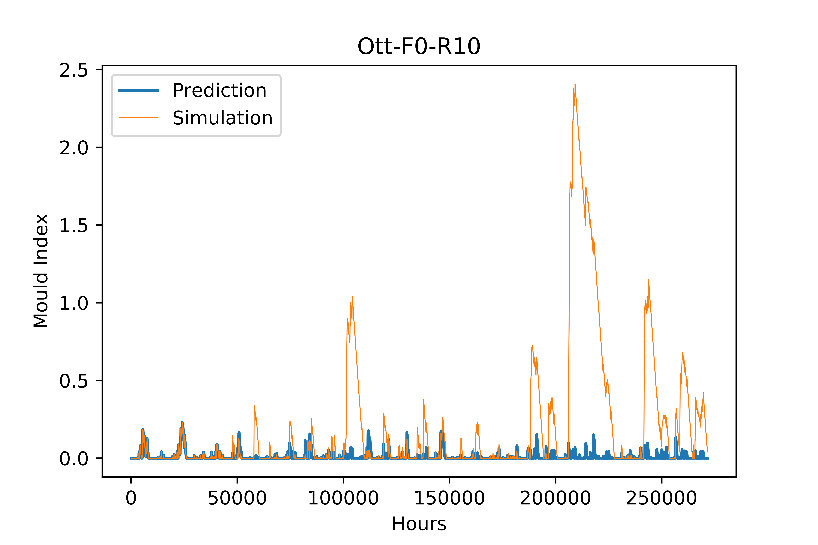
Brick

****

Fibreboard

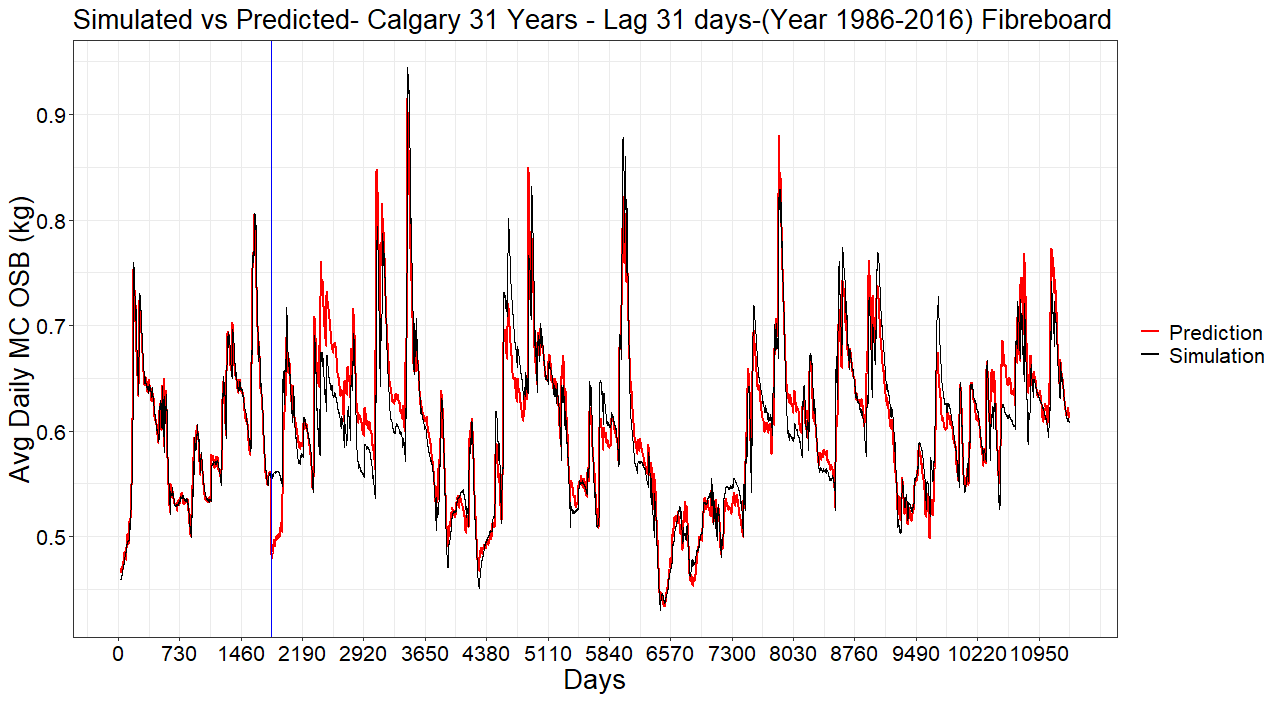
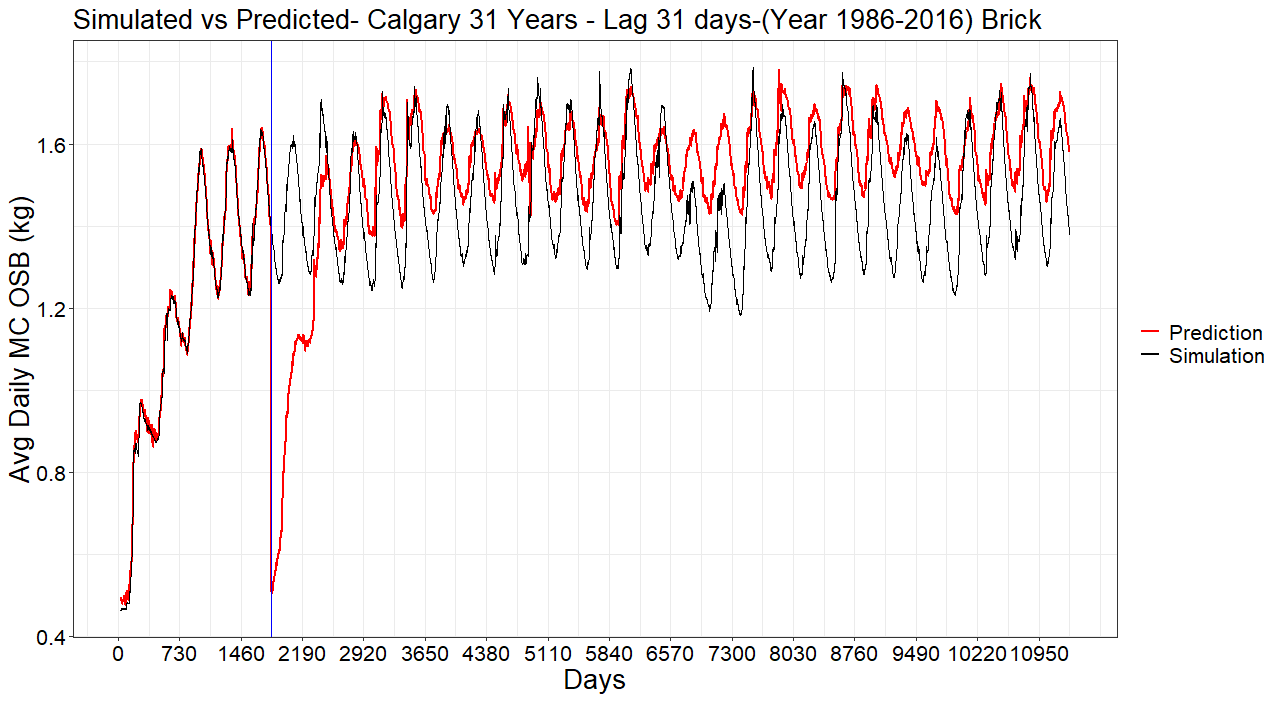


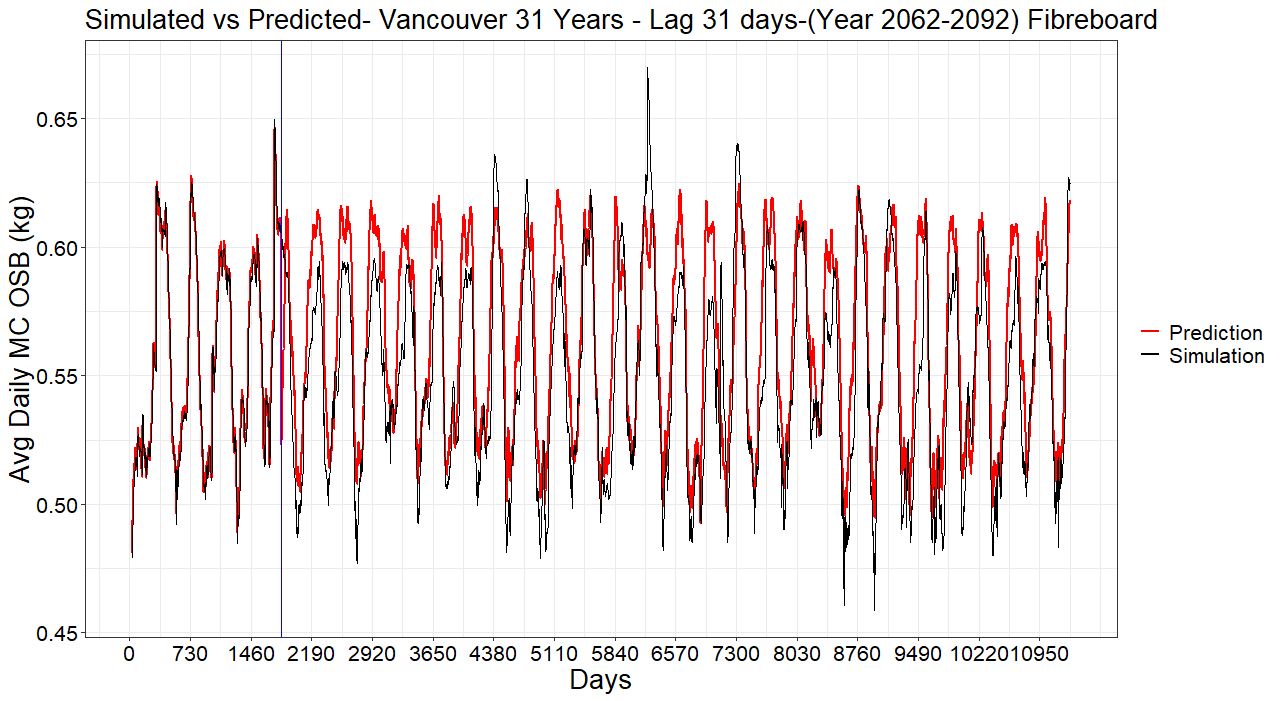
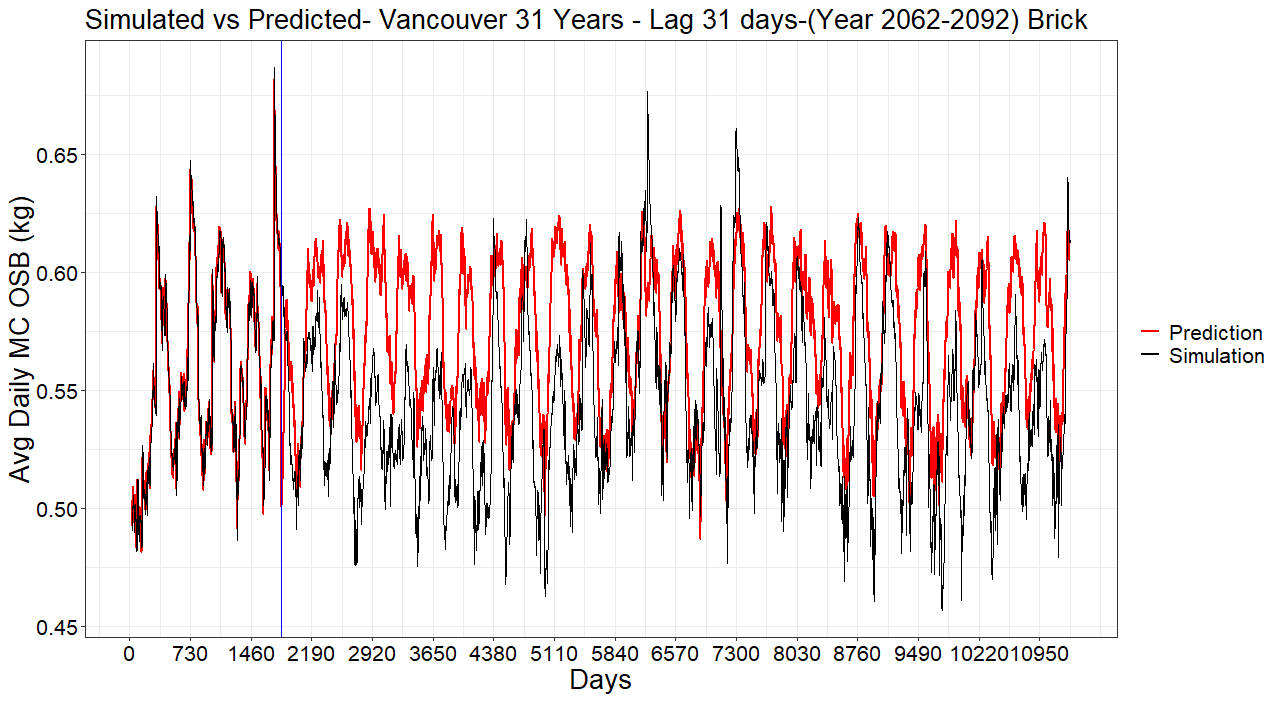
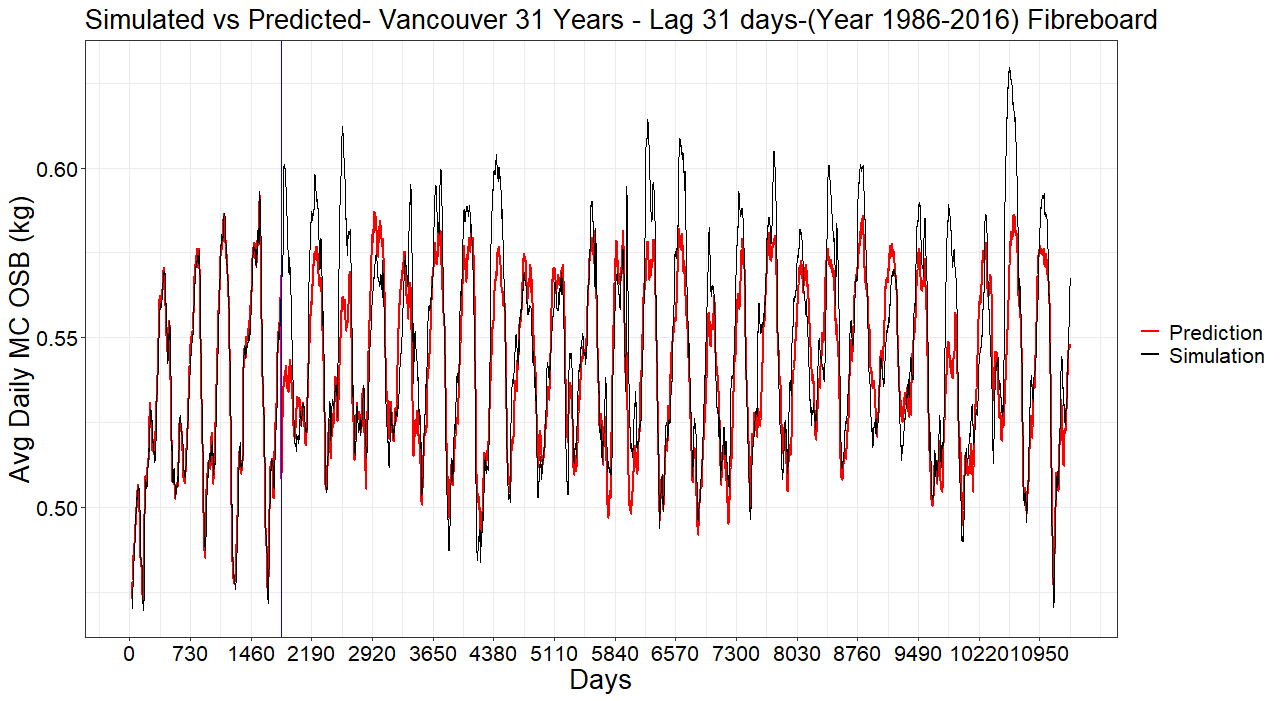
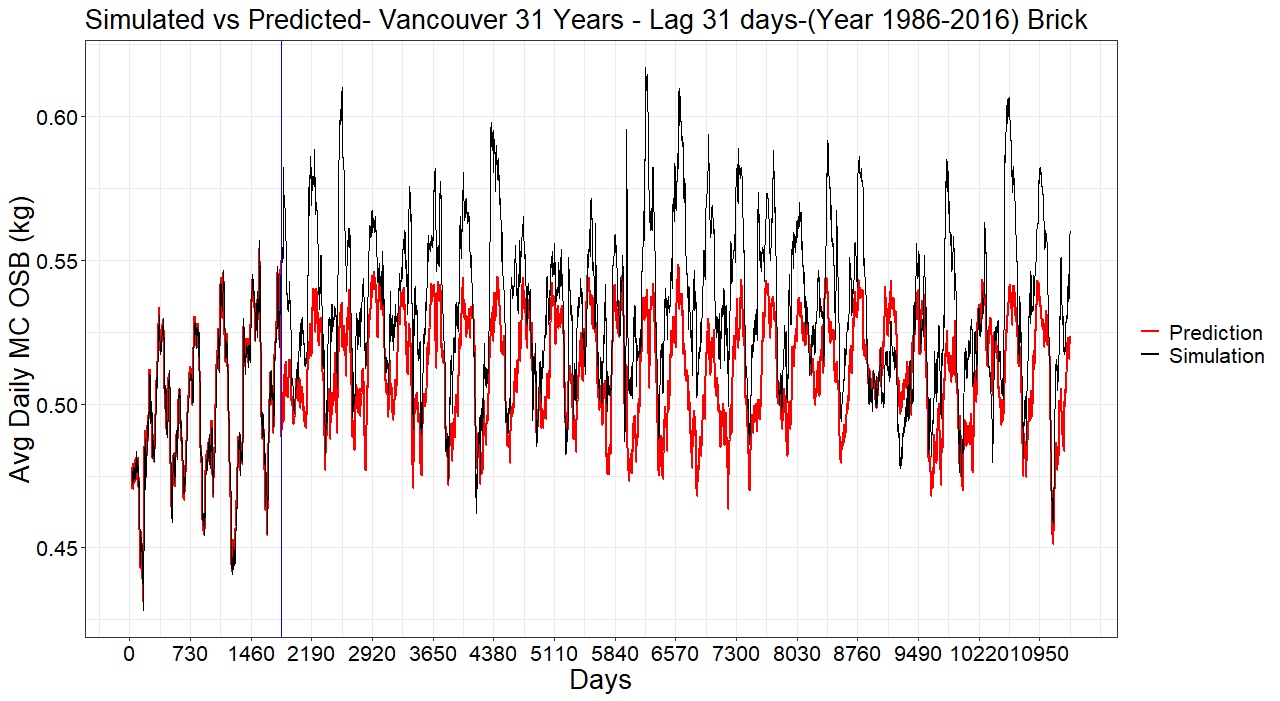
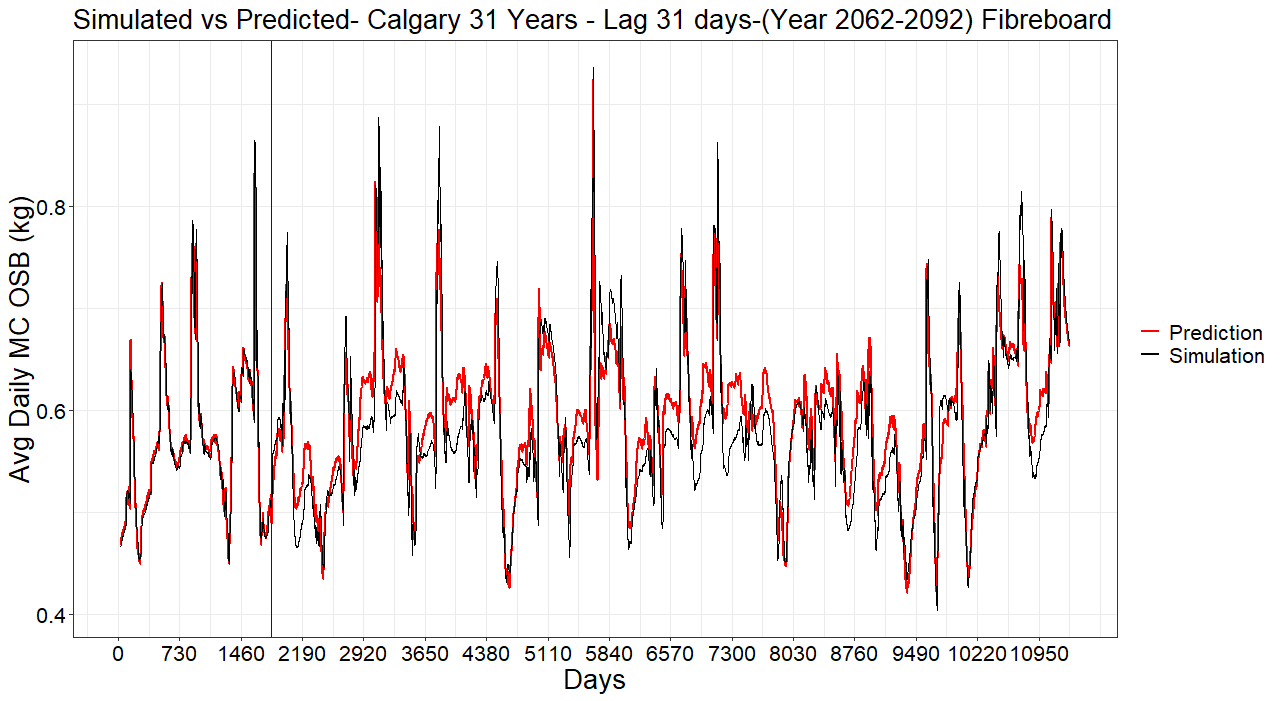
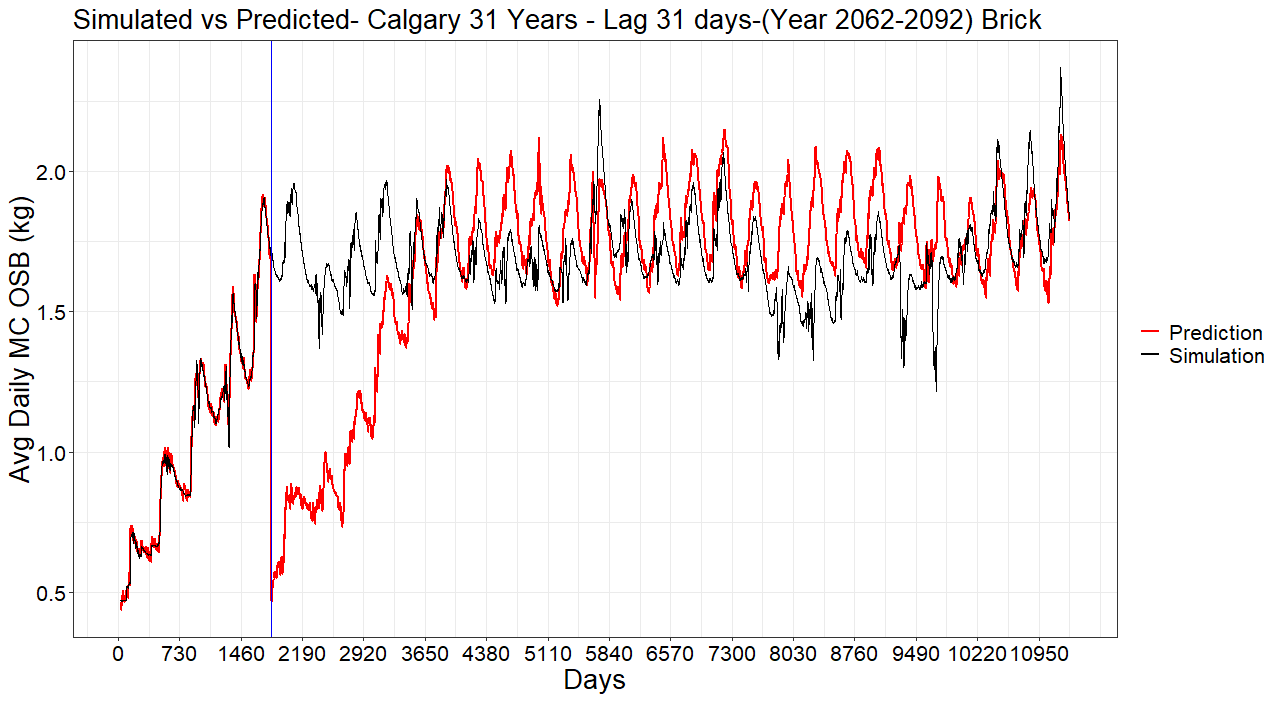
CLT

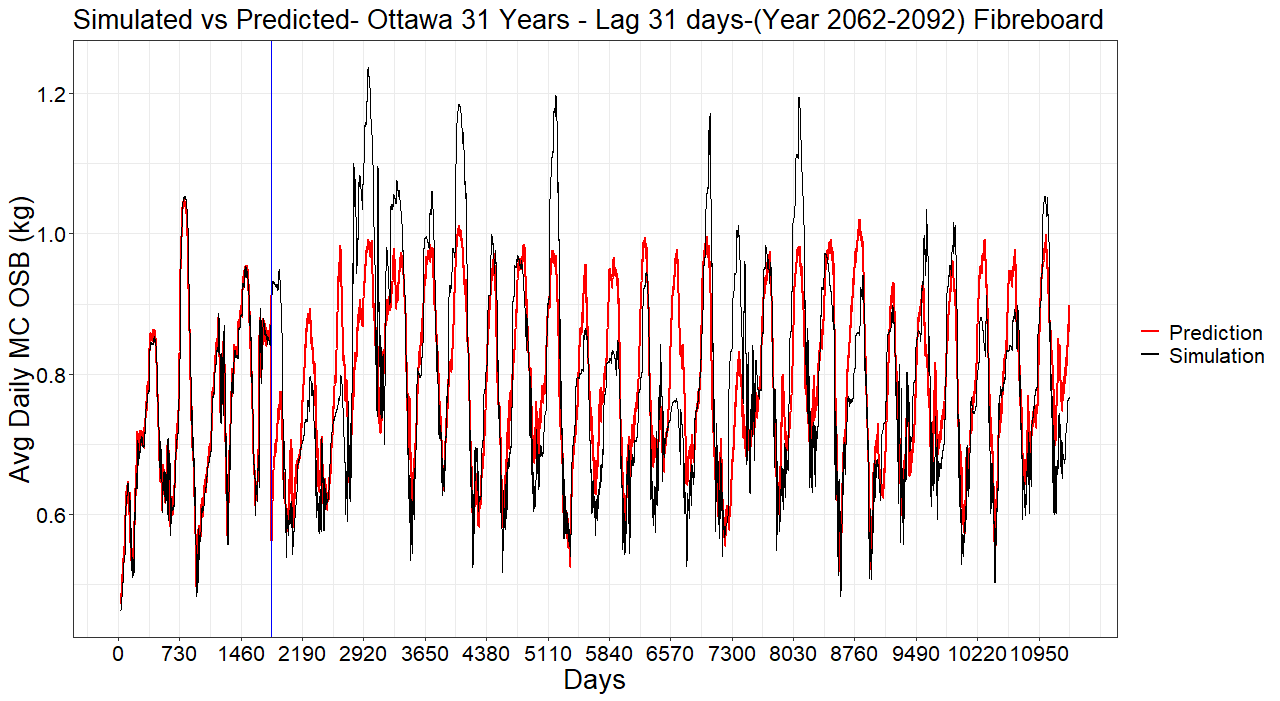
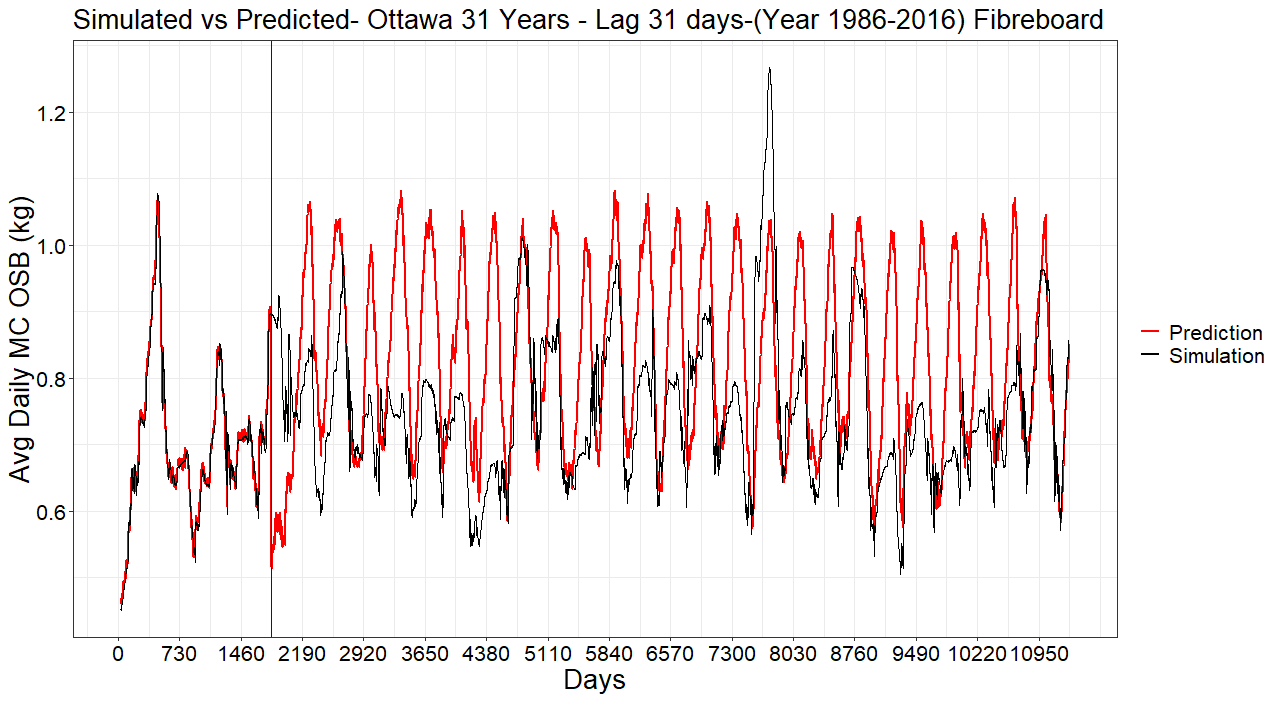
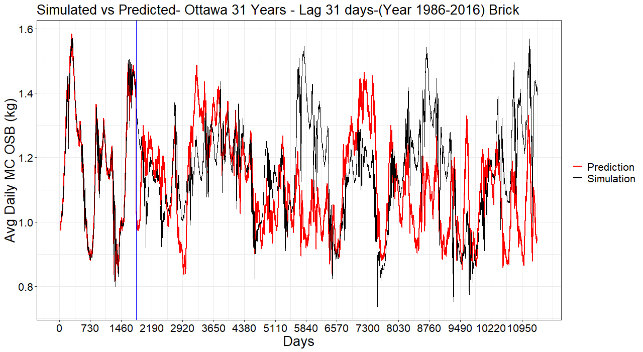
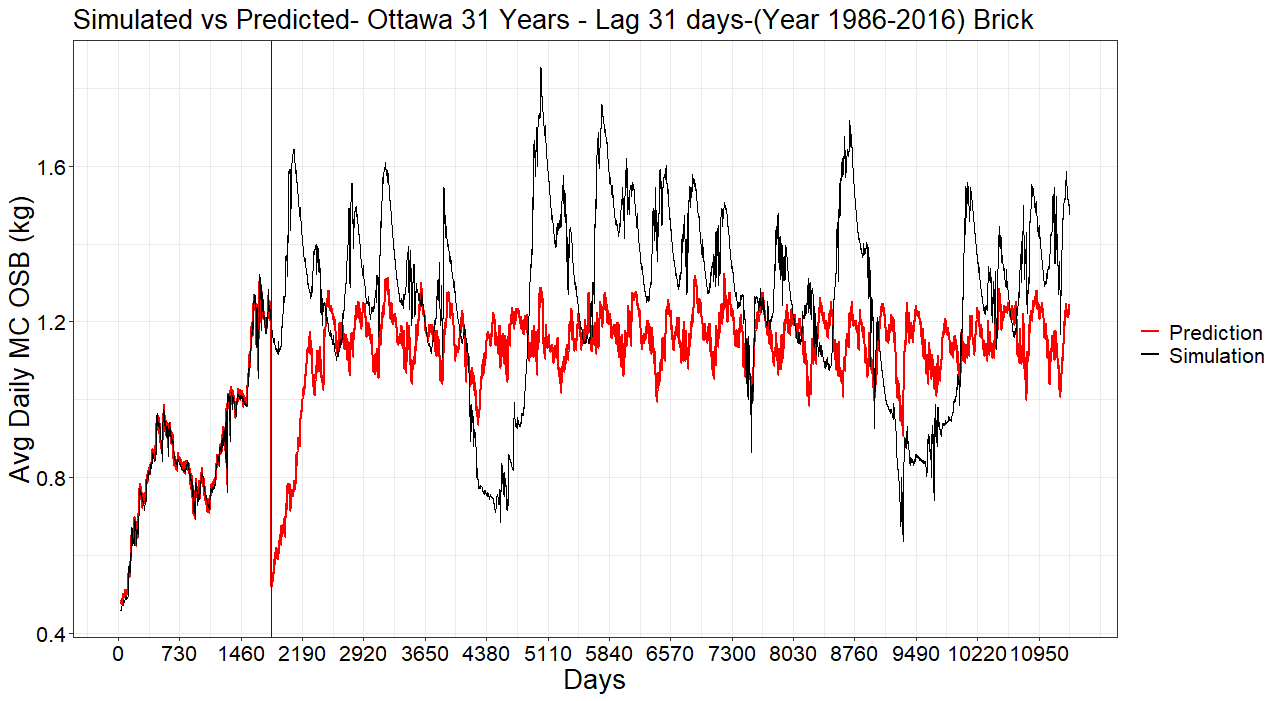


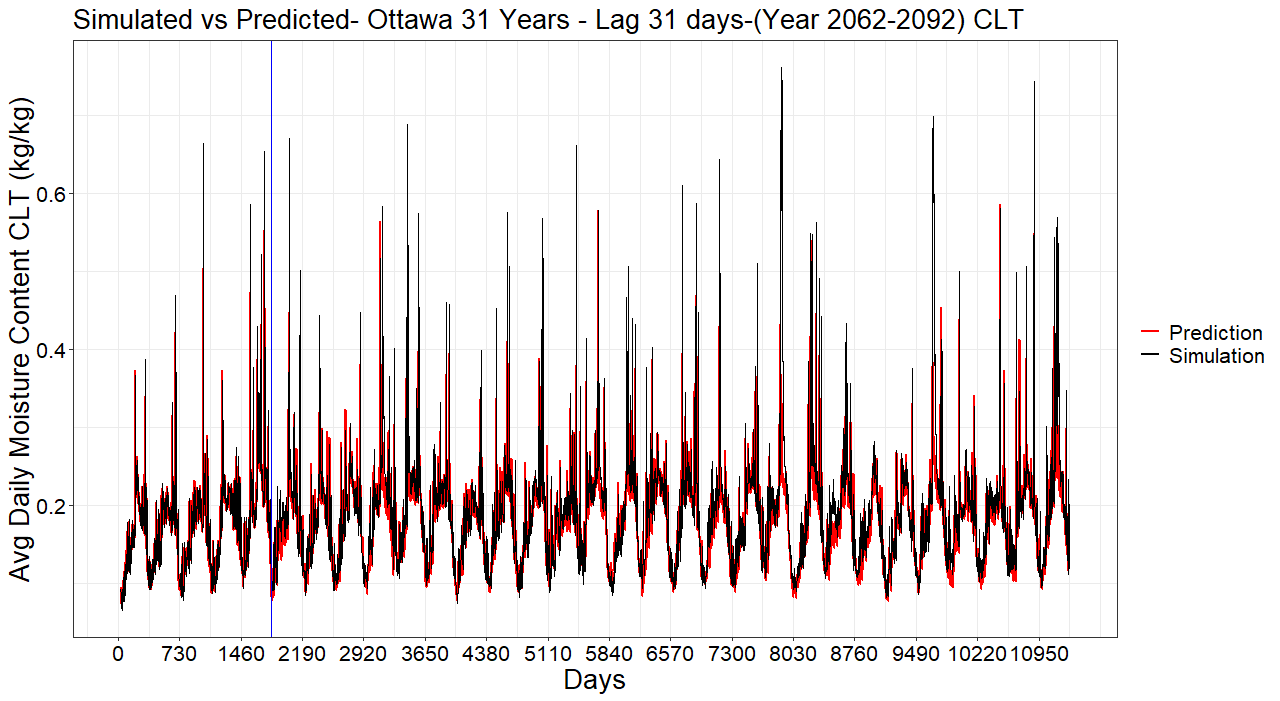
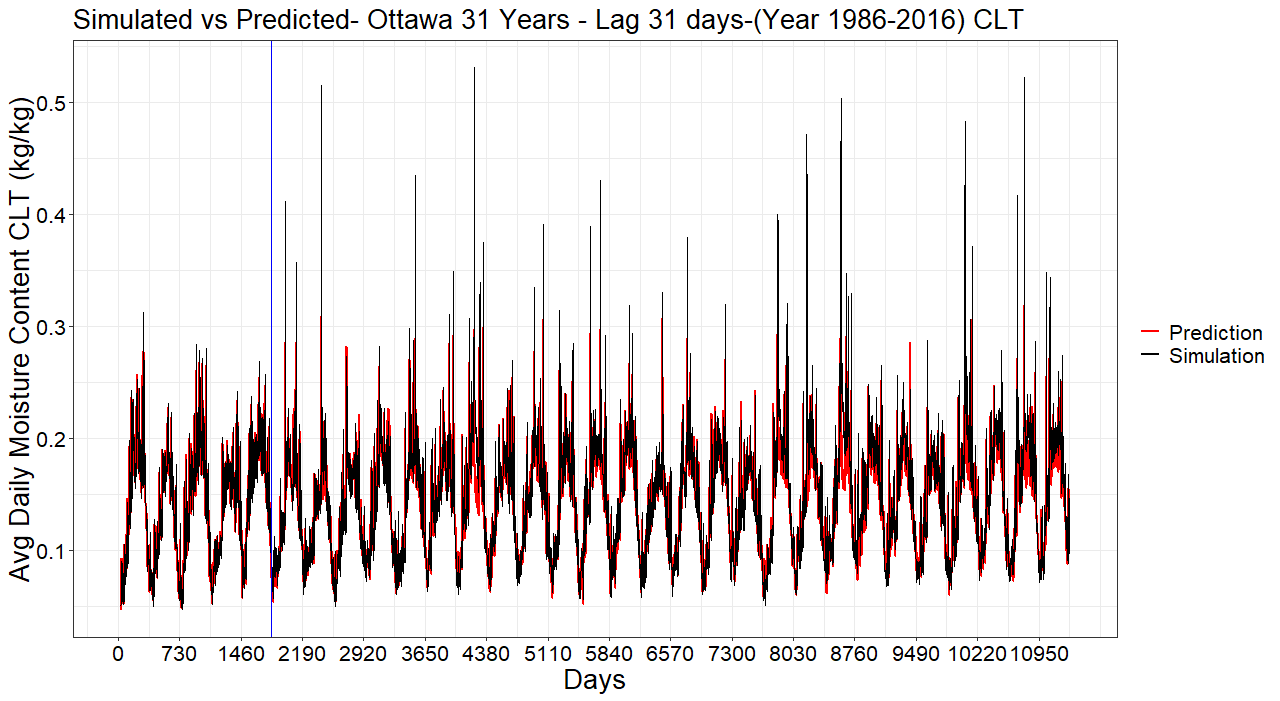
**Moisture Content**

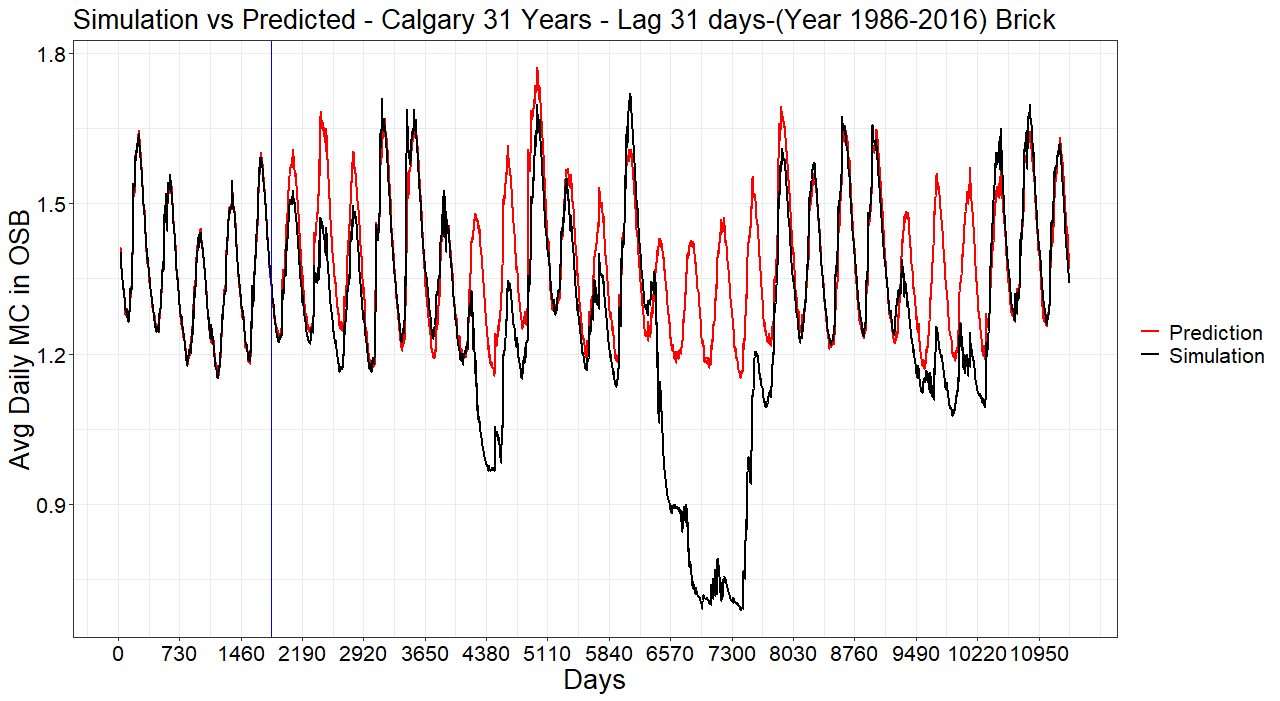
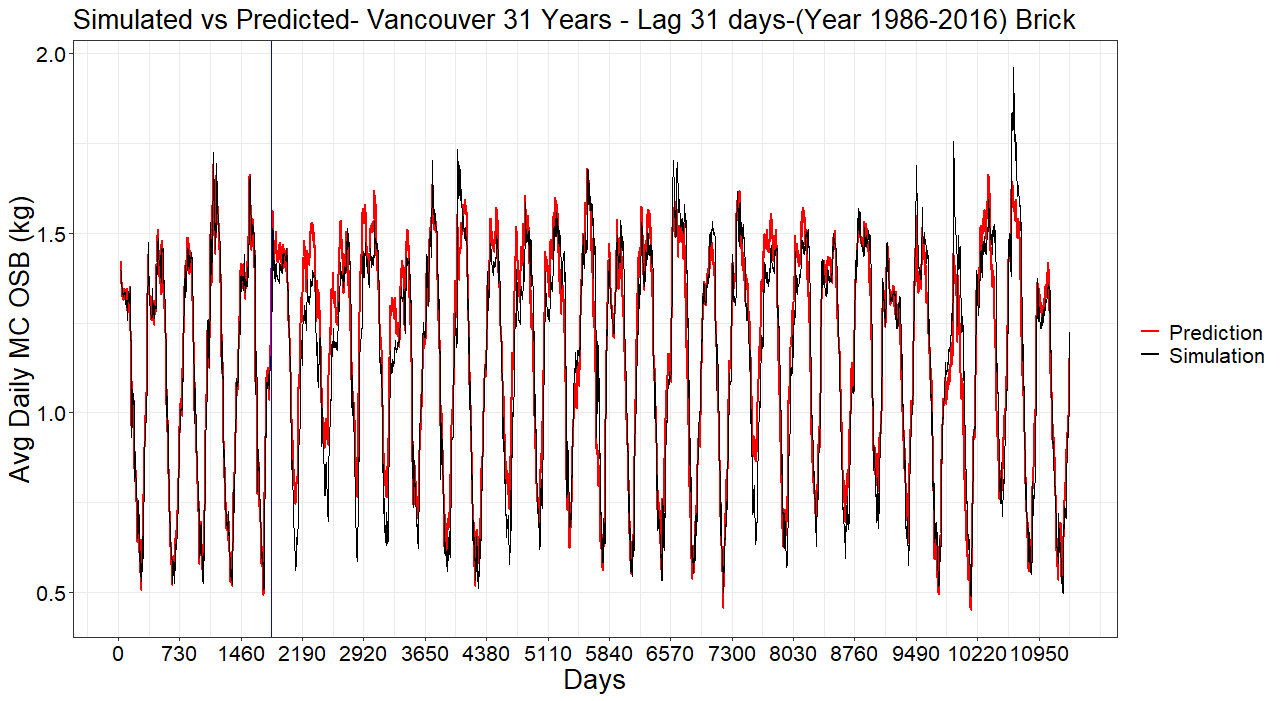
North Orientation

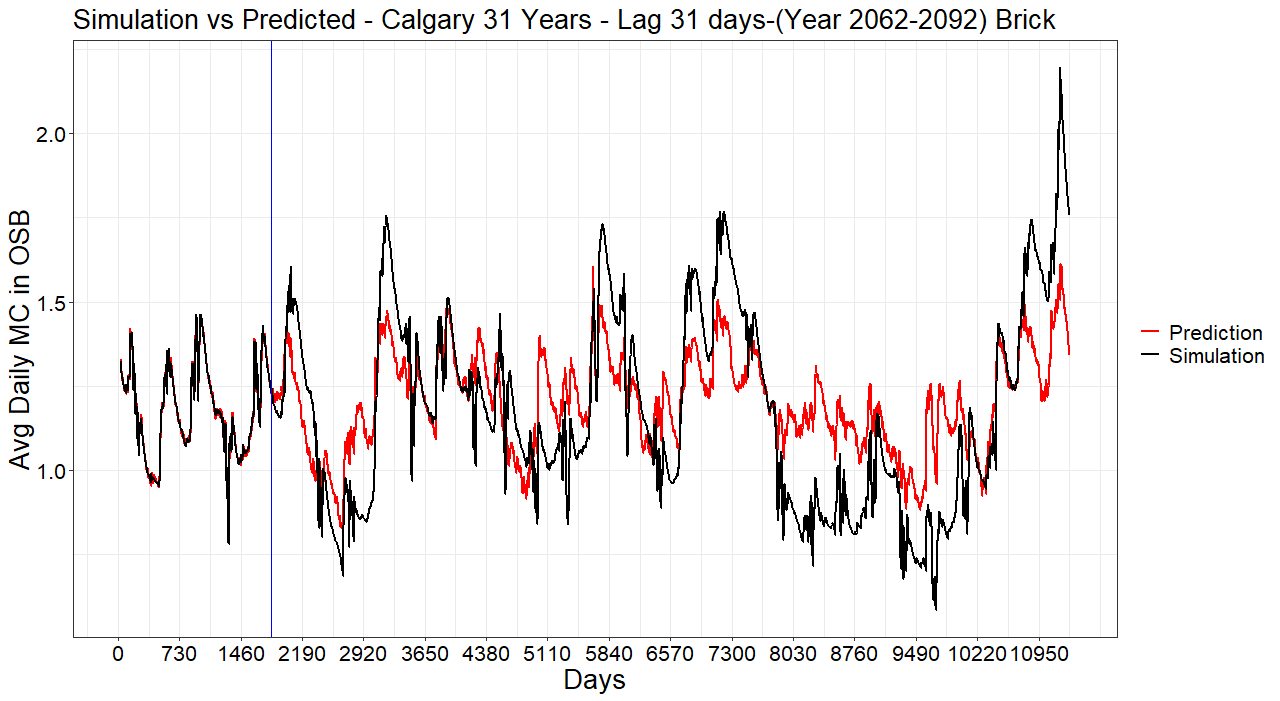






Default Orientation





**Conclusion**

The objective of this study was to explore the potential of a machine learning algorithm, the Support Vector Machine Regression (SVR), to forecast long-term hygrothermal time series (e.g. temperature, relative humidity, moisture content) of building envelopes. Simulations were performed using a 31-year long series of climate data in several cities across Canada. Then, the first 5 years of the series were used in each case to train the model which was then used to forecast the performance indicator for the remaining years of the series. A sliding window approach was used to incorporate the dependence of the response on the past climatic condition which allowed SVR to capture time implicitly.

Results for Moisture Content & Mould Index showed SVR can be effectively used to forecast hygrothermal responses on a long series of climate data given all the variation in the respective response is present and accounted for during the training phase i.e. the first 5 years. Cities like Vancouver where there is no drastic change in climate throughout the 31-year series are predicted well for Moisture Content while Calgary, where there is a significant dry period that occurs later in the 31-year series, have huge discrepancies. For Mould Index, the temperature was predicted well for all cases irrespective of cladding and cities. This is attributed to the stationary nature of the series and short lag time i.e. temperature reacts quickly to change in outdoor climate conditions and thus requires less information from the past compared to relative humidity. Actual Time Series of Relative Humidity also exhibit a conditioning phase in the first few years before it reaches an equilibrium point. This influenced the predictions since the model was trained on the conditional phase where the magnitude of relative humidity is lower compared to later in the 31- year series when it reaches equilibrium. This caused the model to underpredict the Mould Index.