Amin Hosseini Nejad,

am2606ho-s@student.lu.se

I. INTRODUCTION:

In this project, we should construct a model to predict temperature of a city in 3 different ways; with and without using any external inputs as models A and B, and using Kalman filter to build a recursive predicting model as model C. I even tried to simplify my Kalman-filter-based model by excluding the parameters whose effects are not significant in the model performance as simplified model C.

The measured temperature of Sturup is considered as the observed data, y_t, and at the first try, its predicted data is considered as the input, u_t, to build model B to see if it can outperform the Naïve predictors; if not Svedala measured data can be set as the second input due to its closeness to Sturup and being more correlated with it compared to other cities whose data is provided. The predicted data of Sturup has lots of missing samples at the beginning of the dataset, so considering the fact that the time period the missing samples are part of is not of interest to build the models based on, which is the data related the beginning of 1993, we can simply remove them. The same strategy should be implemented on the measured data, y_t. Plus, the number of datapoint in y_t is more than u_t; hence, being the last datapoints in y_t not useful in our modelling, we can remove them to make both y_t and u_t of the same size.

The modelling data starts in week 49 in 1993, which is selected in a way that it locates in the rising part of the data in 1994, and to have space to have the validation data in the same rising trend as the modelling data to build good models based on. The modelling data starts at a time step such that the validation data starts at one of the measured data points, not at the interpolated ones which are less reliable, to validate the models on. Figure 1 plots the y_t data and its different parts and the close-up views of both validation and test starting points.

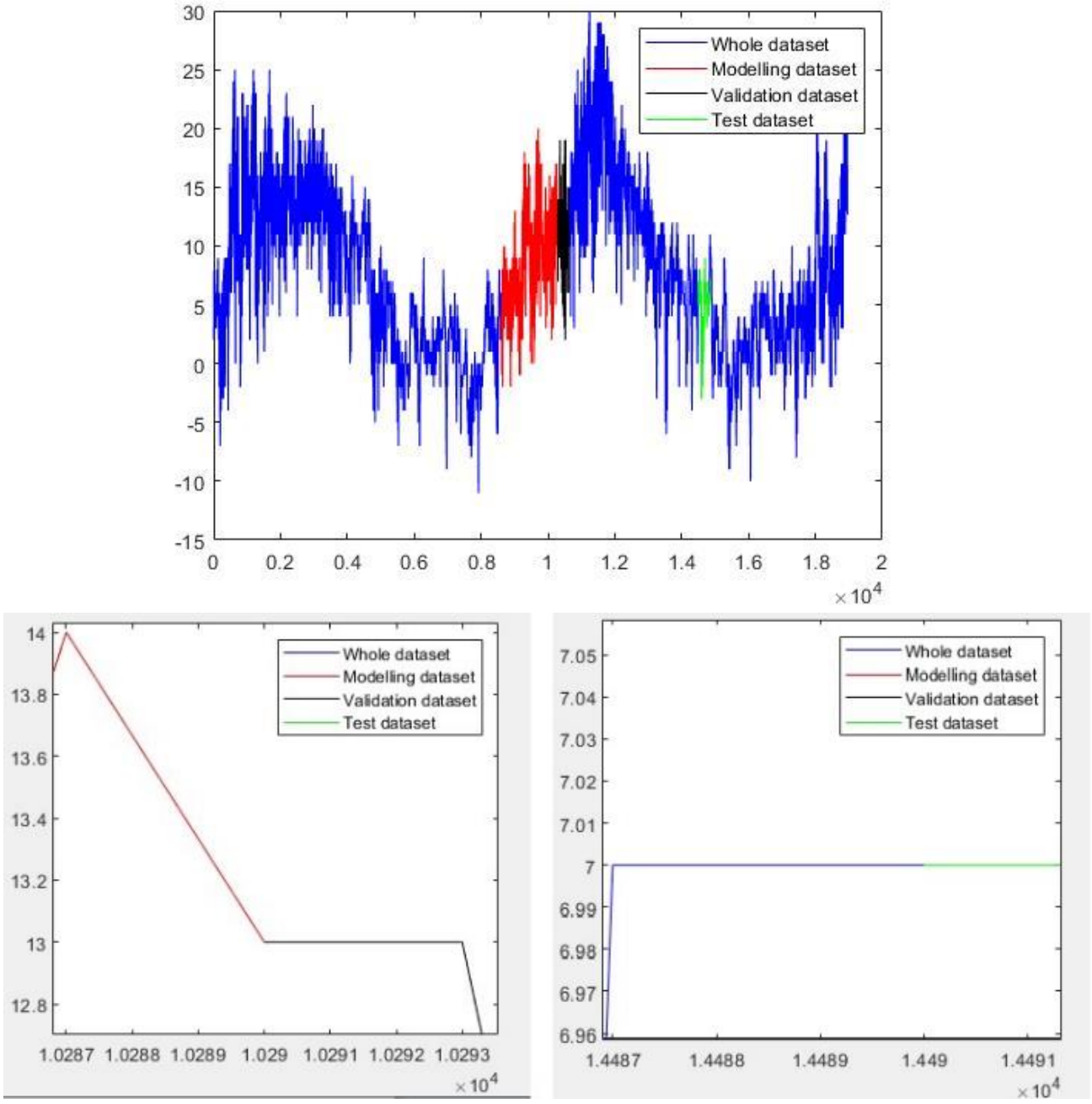

Figure 1. Measured temperature data in Sturup

To check if there are any outliers in y_t, both TACF and ACF plots are provided as Figure 2. As obvious, they are almost the same, so there are no outliers to deal with.

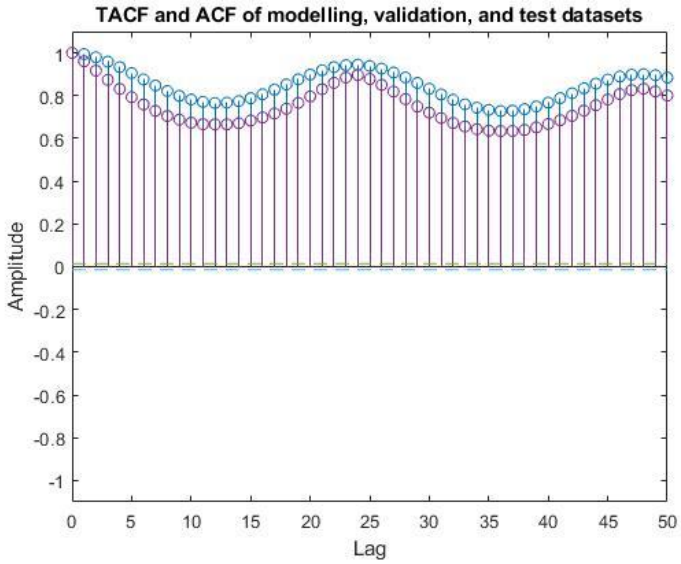

Figure 2. ACF and TACF plots for y_t

Using *normplot* function, Figure 3, we can see that y_t has a normal distribution. Using Box-Cox criterion, whose plot is in Figure 4, we can figure out that we need to log transform the data, but as there are both negative-valued datapoints, first we need to shift all the datapoints to be positive. The transformed data is plotted in Figure 5.

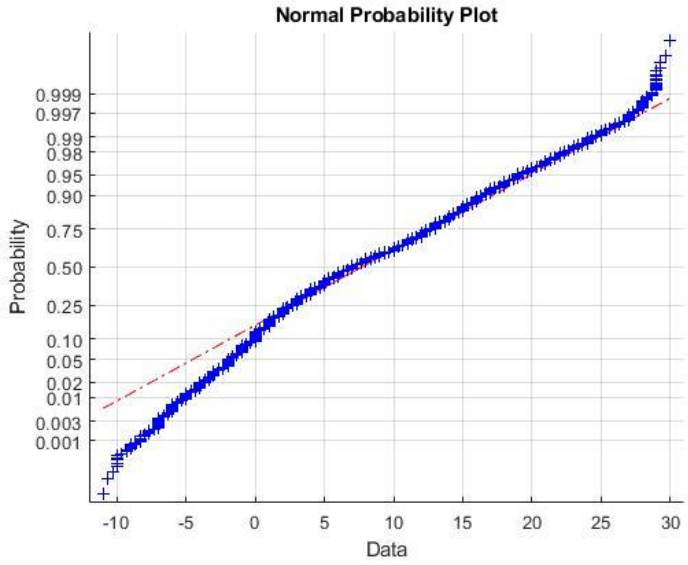

Figure 3. Normplot of y_t

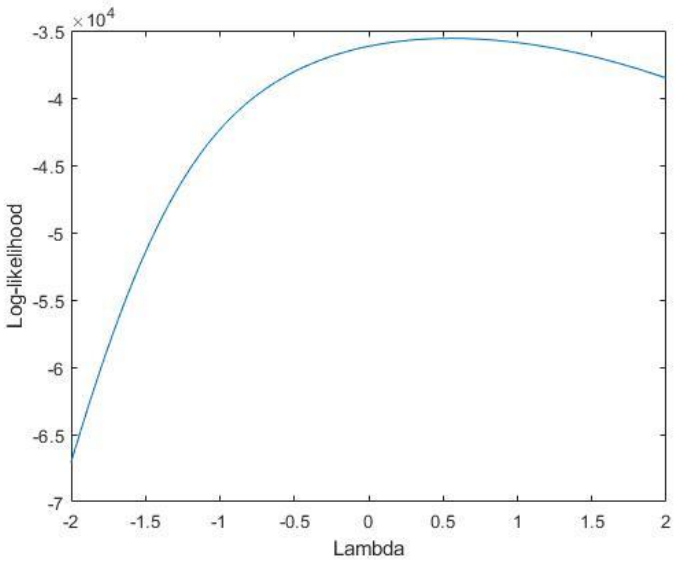

Figure 4. Box-Cox plot

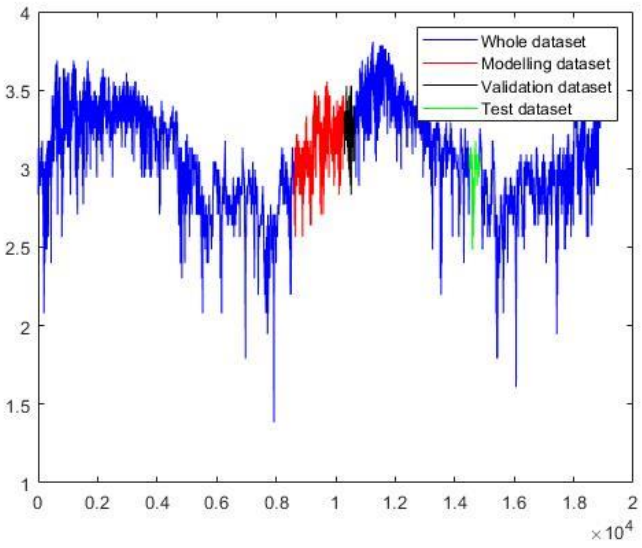

Figure 5. Transfomed y_t data

To evaluate the models, two Naïve predictors are designed for both the prediction horizons of interest. The 1-step Naïve predictor is designed in a way that it reflects the same temperature as the previous time-step. The 9-step Naïve predictor is constructed with the logic of reflecting the temperature of the time-step on previous day at the same time as the 9-step horizon reflects. The relations of either the Naïve predictors are stated as below, (1) for the 1-step and (2) for the 9-step predictors, respectively. They are shown in Figure 6.

$$E\{y_{t+1\,|\,t}\} = y_t \qquad (1)$$

$$E\{y_{t+9\,|\,t}\} = y_{t-24+9} \qquad (2)$$

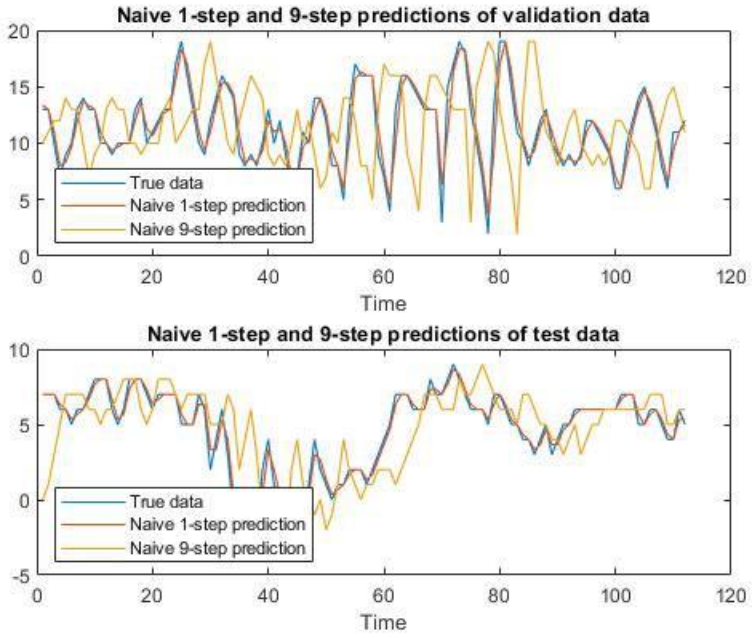

Figure 6. 1-step and 9-step Naïve predictions for both the validation and test data

II. MODEL A:

As it can be seen, there is a linear trend in the modelling data, that should be detrended as ∇y_t. Furthermore, plotting the ACF and PACF graphs of the detrended data, Figure 7, we can see the seasonal trend of 24 hours as its periodicity to be detrended.

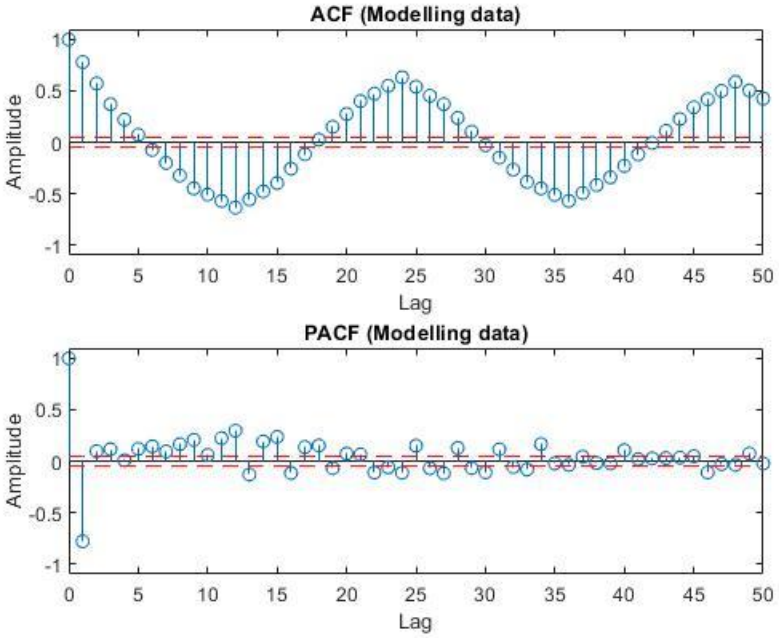

Figure 7. ACF and PACF plots of y_t detrended

Modelling the modelling data gives us an ARMA model, stated as:

$$A(z) = 1 - 0.7992\ (\pm 0.0192)\ z^{-1} - 0.2565\ (\pm 0.04446)\ z^{-3} + 0.2327\ (\pm 0.03505)\ z^{-4} + 0.02962\ (\pm 0.033)\ z^{-9} + 0.01182\ (\pm 0.02705)\ z^{-10} - 0.15\ (\pm 0.02744)\ z^{-18} + 0.09723(\pm 0.02548)\ z^{-19}$$

$$C(z) = 1 - 0.8268\ (\pm 0.03085)\ z^{-3} - 0.867(\pm 0.01272)\ z^{-24} + 0.7111\ (\pm 0.03033)\ z^{-27}$$

There are some insignificant parameters in the model, such as a_9 and a_{10} , which can be removed without making the modelling residual non-WN, so the final model is as:

$$A(z) = 1 - 0.8056(\pm 0.01917)\ z^{-1} - 0.2869(\pm 0.03801)\ z^{-3} + 0.2693(\pm 0.0281)\ z^{-4} - 0.1575(\pm 0.026)\ z^{-18} + 0.1025(\pm 0.02536)\ z^{-19}$$

$$C(z) = 1 - 0.8504(\pm 0.02161)\ z^{-3} - 0.8648(\pm 0.01287)\ z^{-24} + 0.7302(\pm 0.02328)\ z^{-27}$$

The Moni test proves that the modelling residual is WN (as $12.45 < 31.41$). Figure 8 demonstrates the ACF and PACF of the modelling residual of model A.

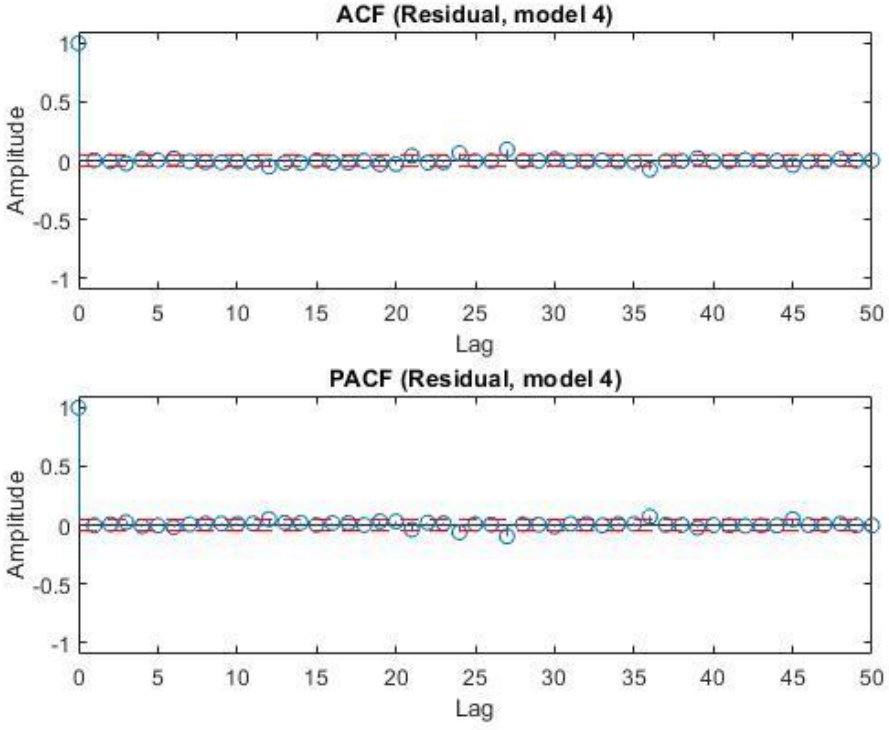

Figure 8. ACF and PACF of the modelling residual of model A

To evaluate different predictor models on validation and test data, variances of their prediction errors for the true measurements, and not for the interpolated ones, should be scaled by variances of true measurements of the validation and test, which are 11.9 °C and 8.2 °C, respectively. It is worth noting that as during filtering process in prediction, some initial samples of the predicted data corresponding to the test data are ruined, those samples should be completely ignored in any working with the test data. The number of the ruined initial samples is 54, which is set as being a multiple of 3 so that the first initial test sample is again a true observation and not an interpolated one.

Predicting with 1-step and 9-steps horizons and comparing the predictors outputs with the validation data shows that the model A for both prediction horizons outperform their corresponding Naïve predictors notably. Figure 9 illustrates both the predictions for both validation and test data compared to the true data. The model seems to do well on validation data, but not on test data.

The scaled variance of the 1-step and 9-step prediction errors on true observations of the validation data are 0.096 and 0.53, respectively; variance of the 1-step prediction error is almost the same as its corresponding Naïve predictors, which is 0.095, but the 9-step one is smaller than its corresponding Naïve one, which is 2.58.

The scaled variance of the 1-step and 9-step prediction errors on the test data are 0.18 and 1.53, respectively, which both do not outperform their corresponding Naïve predictors, which is 0.02, but the 9-step one is bigger than its corresponding Naïve one, showing that model A is not good enough on the test data.

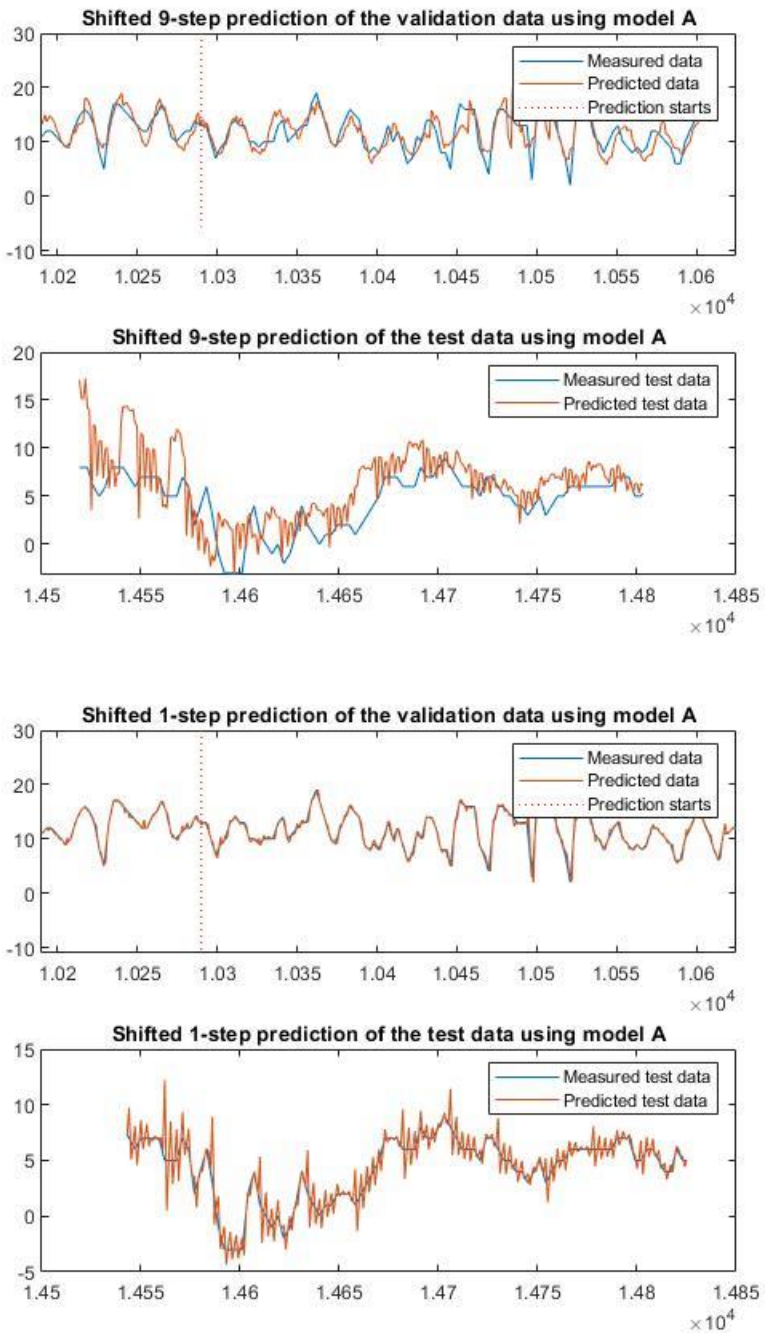

Figure 9. 1-step and 9-step predictions using model A on both the validation and test data

Figure 10 plots the prediction error using model A for both the validation and test true measurements for both prediction horizons. As it is obvious, the variance for validation data is less than it is for test data. It is worth noting that the plots show only 112 true measurements, not 336 as the number of hours in a 2-week period.

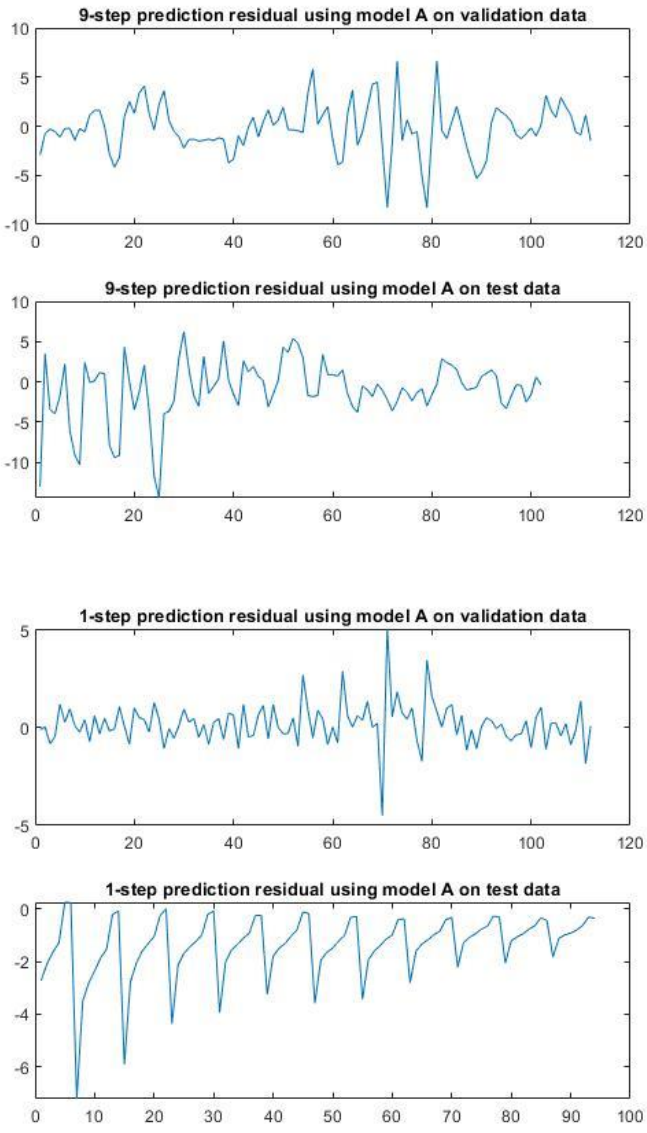

Figure 10. prediction residual using model A for both the validation and test true measurements

The ACF plot of the prediction residual using model A for 1-step and 9-step predictions on validation data is shown in Figure 11. Ideally, the prediction residual should look like an MA(k-1) model. For 1-step prediction, it looks like a sparse MA(16), which has only 3 fairly significant components at lags1,6 1nd 16; while for 9-step one, it looks like an MA(15), but the 15^{th} component is not highly significant, so it can be seen as an MA(1) model.

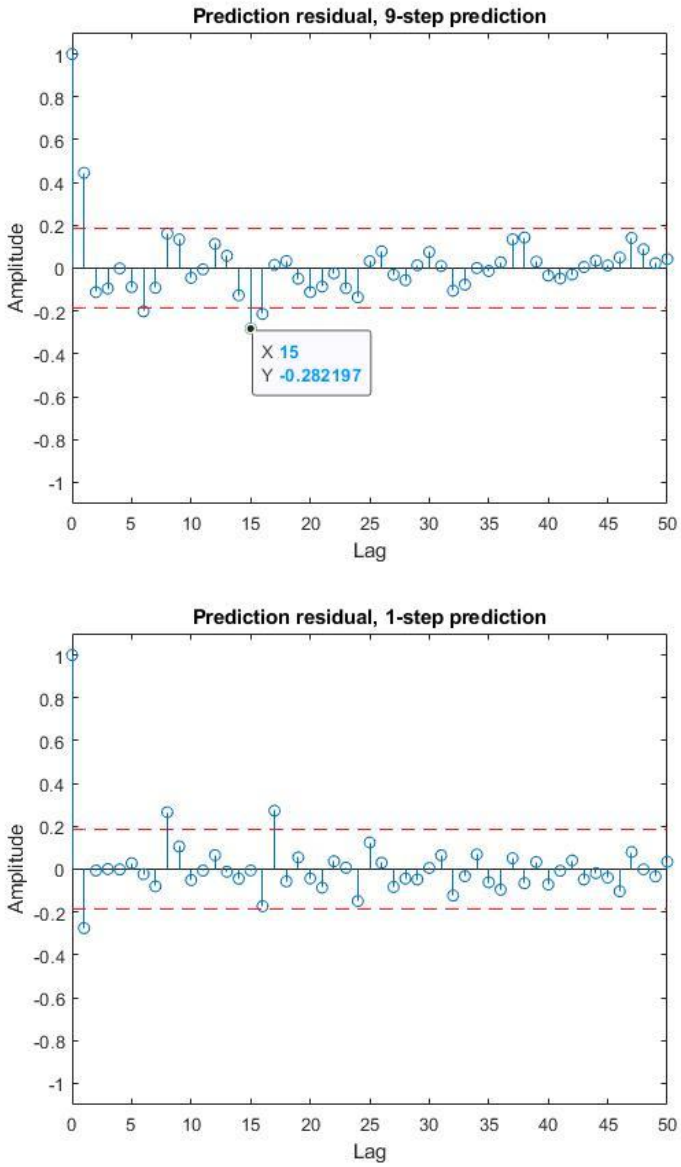

Figure 11. ACF plot of the prediction residual for model A; the top one for 9-step and the below one for 1-step predictions

III. MODEL B:

As stated before, a BJ model is designed as model B to see if any better prediction could be obtained by adding an external input, as u_t , whose correlation with the observed data, y_t, is big. As Sturup temperature is considered as y_t, its prediction data provided by SMHI is considered as u_t. Figure 12 shows the input data. The same data investigation and data processing as y_t should be done on u_t. By comparing its TACF and PACF plots, which are almost the same, we can figure out that there are not outliers in this input. By checking its *normplot*, we can see that it has a normal distribution, and by checking the Box-Cox criterion, we can see that it could be useful to log-transform the input data.

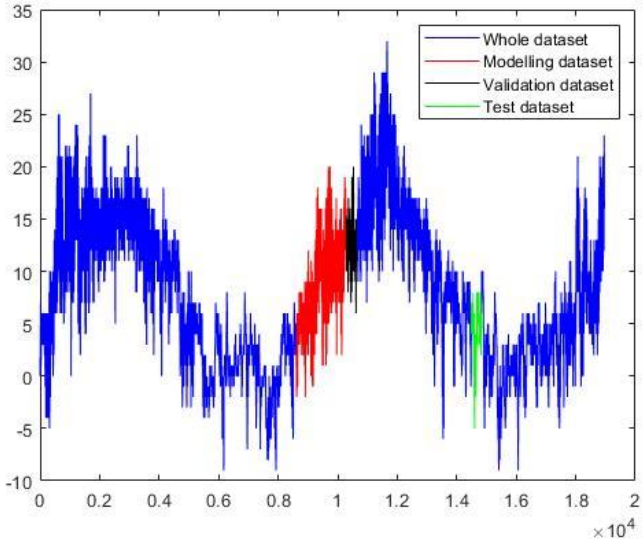

Figure 12. the input data

First the BJ model on the modelling data should be designed and then, the model B predictor is designed to evaluate on the validation and test data. To construct the BJ model, first the input should be modelled. We can see some linear and 24-hour seasonal trends in the input that should be detrended. The model designed for the input is an ARMA model, Monti-test (as 21.60 < 31.41), stated as:

$$A_3(z) = 1 - 0.6593\ (\pm\ 0.02374)\ z^{-1} - 0.3453(\pm\ 0.06985)\ z^{-3} + 0.333(\pm\ 0.05231)\ z^{-4} - 0.1566(\pm\ 0.03863)\ z^{-6} + 0.1718(\pm\ 0.02829)\ z^{-7} + 0.07532(\pm\ 0.02331)\ z^{-14} - 0.1439(\pm\ 0.02494)\ z^{-18} + 0.07265(\pm 0.02537)\ z^{-19}$$

$$C_3(z) = 1 - 0.7065(\pm\ 0.06486)\ z^{-3} - 0.7556(\pm\ 0.01751)\ z^{-24} + 0.5083(\pm 0.05277)\ z^{-27}$$

Figure 13 depicts the ACF and PACF plots of modelling residual of the input signal. Figure 14 shows the cross-correlation between u_t and y_t. It suggests a delay of d=8 and s=9 (s+d = 17). There does not seem to be much of a decay, so r=0.

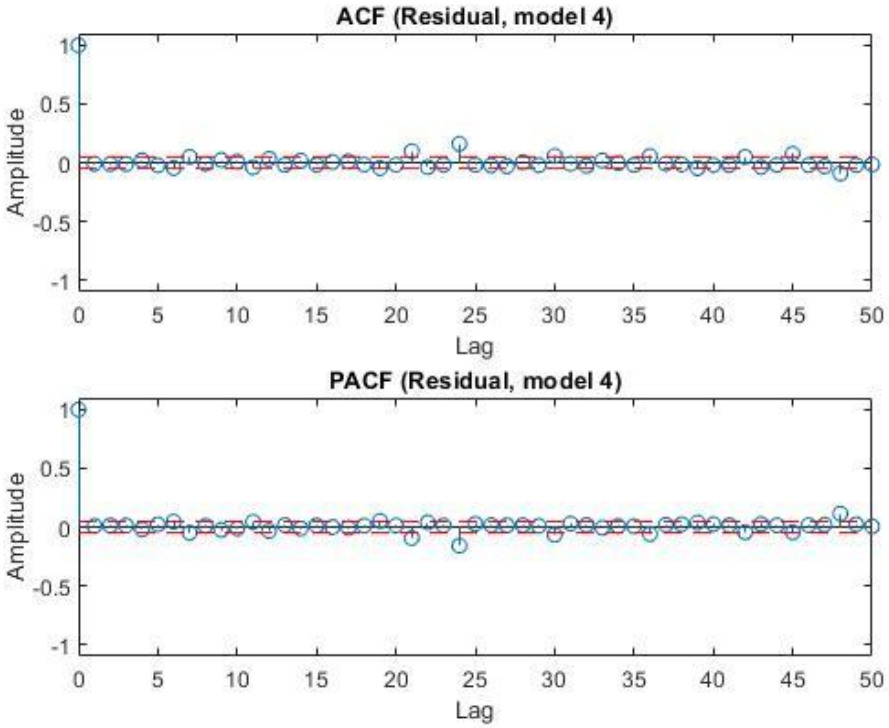

Figure 13.the ACF and PACF for the input

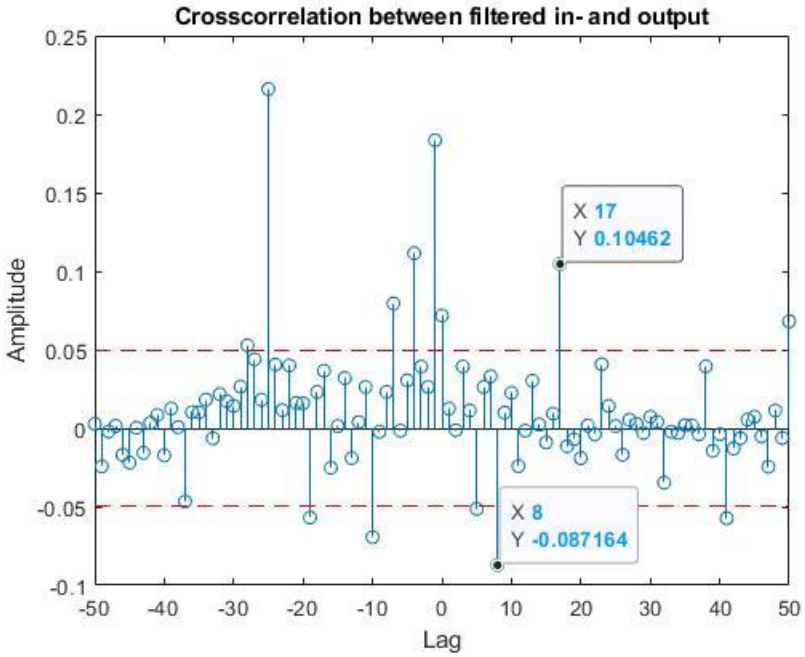

Figure 14. cross-correlation between u_t and y_t.

The BJ model found is stated as below, Monti-test(as $21.95 < 31.41$):

$$A_2(z) = 1$$

$$B(z) = -0.1117\ (\pm\ 0.02226)\ z^{-8} + 0.06431(\pm\ 0.01933)\ z^{-17}$$

$$C(z) = 1 - 0.8041(\pm\ 0.01837)\ z^{-3}$$

$$A_1(z) = 1 - 0.785(\pm\ 0.01936)\ z^{-1} - 0.06676(\pm\ 0.02348)\ z^{-3} + 0.07718(\pm\ 0.01735)\ z^{-6} - 0.06091(\pm\ 0.01705)\ z^{-18} + 0.1027(\pm 0.01703)\ z^{-21}$$

Figure 15 illustrates the ACF and PACF of the BJ modelling residual on the observed data. There seems to be some significant components at lags 24 and 48, but that might be due to the data being real. Cross-correlation between the modelling residual and the input is shown in Figure 16. It is obvious that they are not quite uncorrelated, so the found model on some simulated data is investigated to check the reliability of the model.

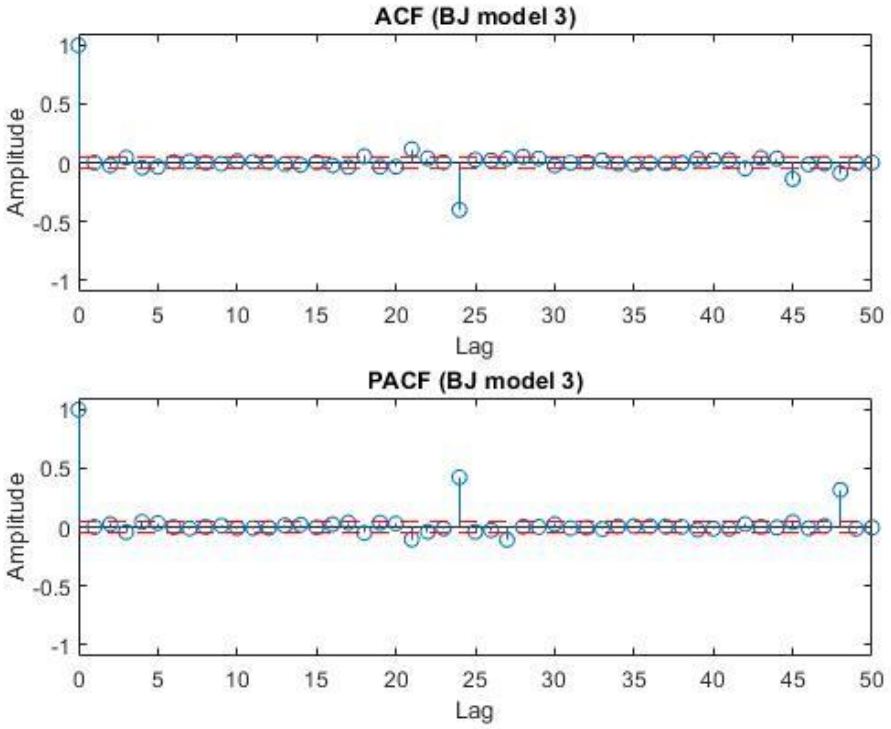

Figure 15. ACF and PACF of the BJ modelling residual on real data

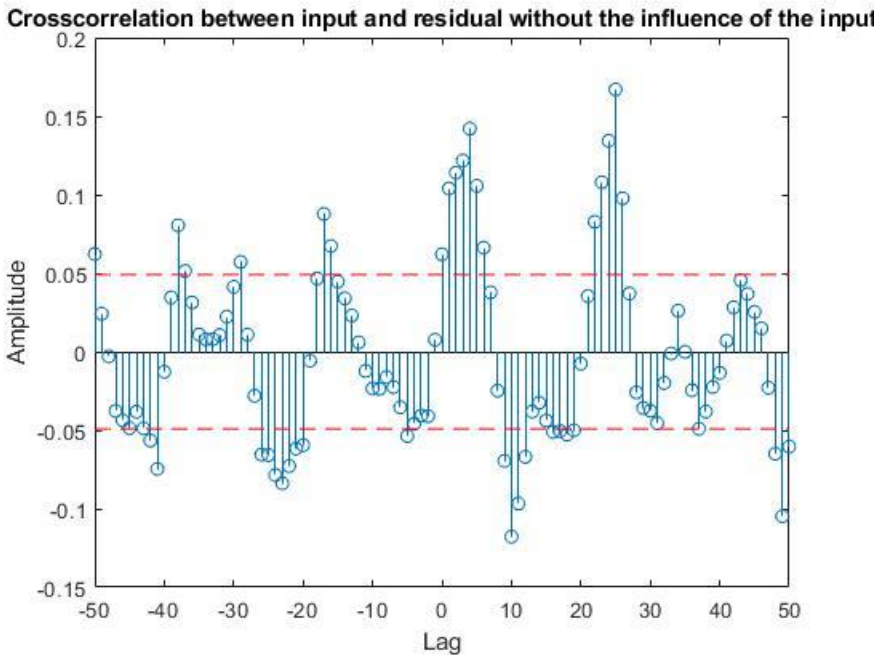

Figure 16. Cross-correlation between the modelling residual and the input

The ACF and PACF of the BJ modelling residual on simulated data is shown in Figure 17, which proves that the modelling residual is quite a WN, Monti-test ($8.63 < 31.41$).

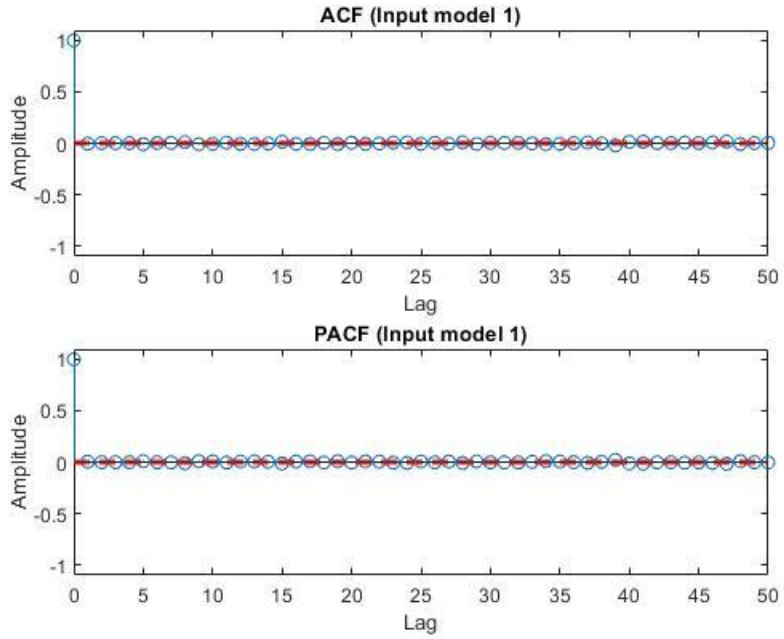

Figure 17. ACF and PACF of the BJ modelling residual on simulated data

Figure 18 depicts the cross-correlation between the modelling residual and the input as per simulated data. Some correlations can be still seen, while noting the amplitude of the peaks being small compared to the confidence interval, we can make sure that the model is good enough.

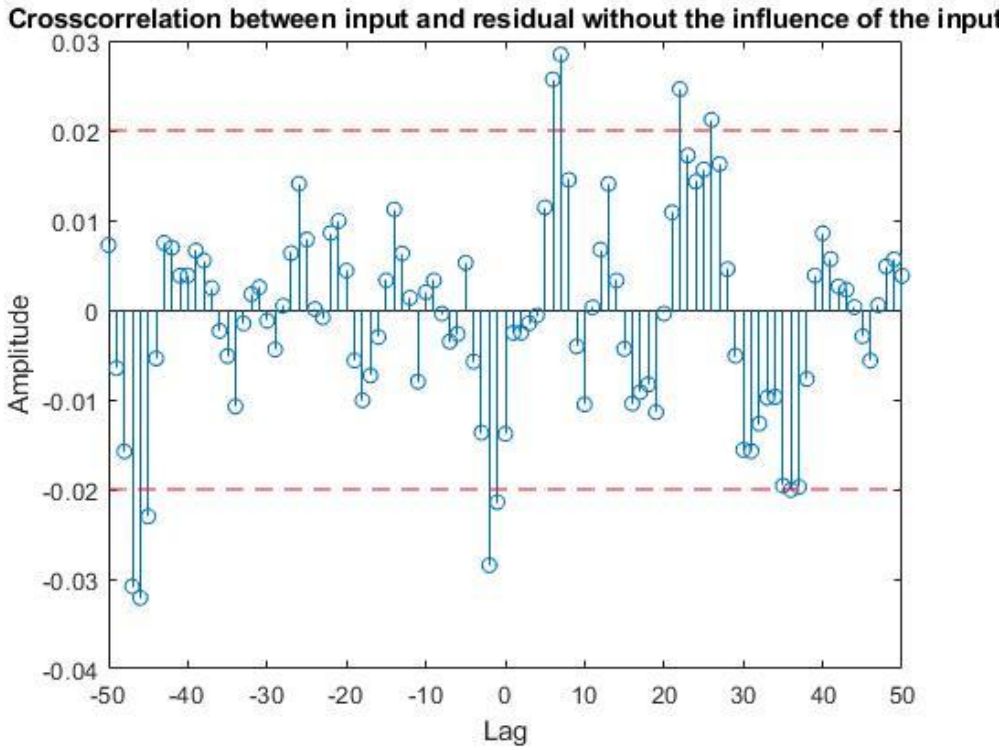

Figure 18. cross-correlation between the modelling residual and the input as per simulated data

To design the BJ predictor, we need design the prediction of the input signal first, which is the same as the input signal for 9-step prediction, for the prediction horizon here is bigger than the lag in the input, as being 7; while the 1-step prediction of the validation and test parts of the input signal is simply done and shown in Figure 19. The reason why we need to have a prediction model for the test part of the input signal is that it needs to enter calculations of the predictions for the test observation data. We should implement the same scenario for prediction of this part of the input.

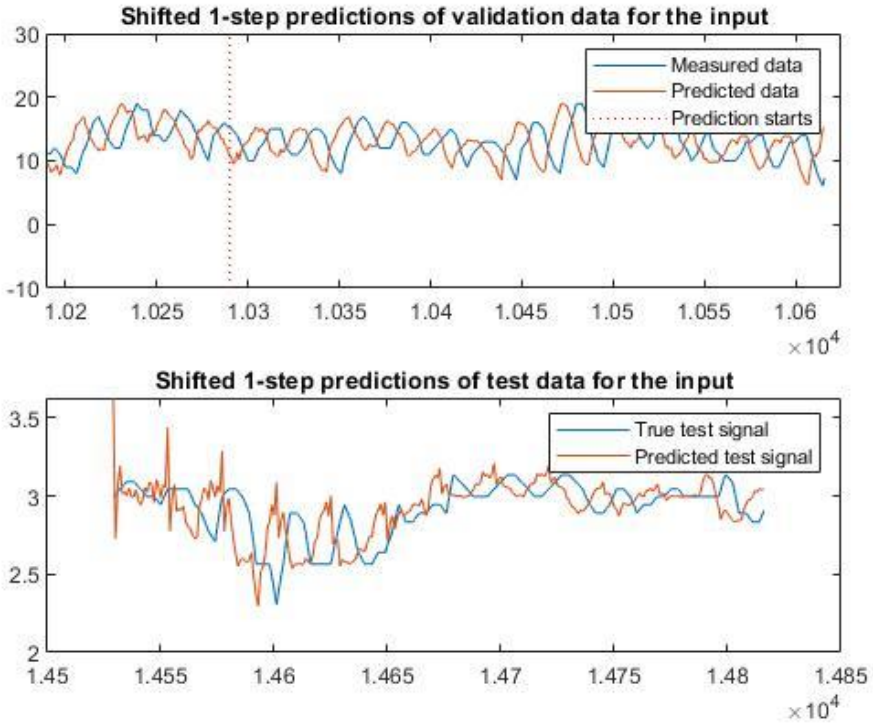

Figure 19. 1-step prediction of validation and test parts of the input signal

Prediction of both the validation and test data for both the prediction horizons using model B is plotted in Figure 20. For 9-step prediction, there seems to need more delay than the group delay calculated based on this model. However, the model can predict both the data fairly well. For 1-step prediction, on the other hand, the model works just perfect.

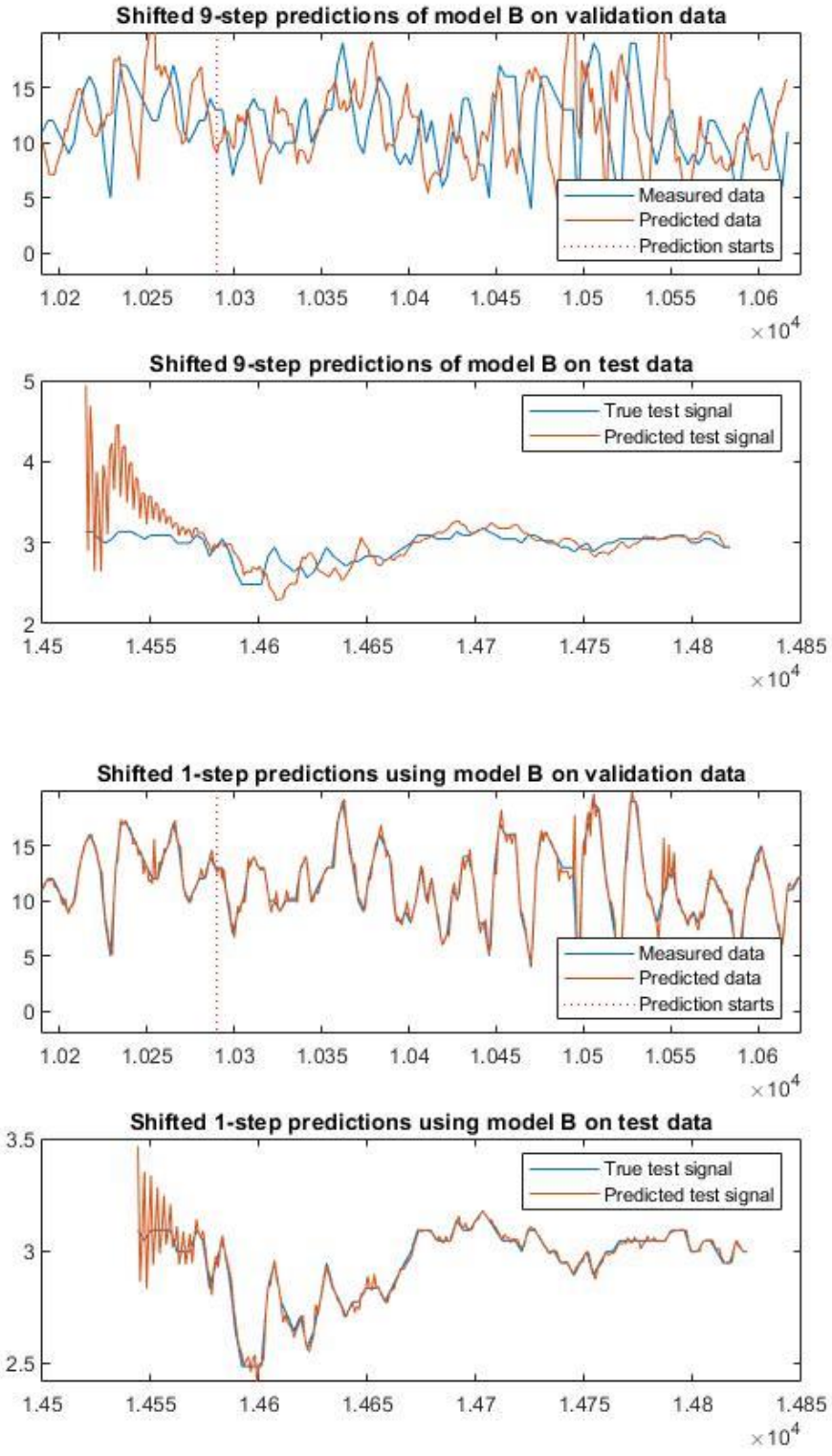

Figure 20. prediction of both the validation and test data for both the prediction horizons using models

The scaled variance of the 1-step and 9-step prediction errors on true observations of the validation data are 0.14 and 0.71, respectively. The 1-step prediction model does not outperform its corresponding Naïve predictor, which is 0.095, but the 9-step one is smaller than its corresponding Naïve predictor, which is 2.58.

However, the scaled variance of the 1-step and 9-step prediction errors on the test data are 0.18 and 1.53, respectively, which are quite bigger than them for their corresponding Naïve predictors, which is 0.02 and 0.89, respectively, showing that model B is not good enough for either 1-step prediction or 9-step prediction on the test data.

The ACF plot of the prediction residual of the validation data for both the prediction horizons are yielded in Figure 21. For the 9-step one, the residual looks like a sparse MA(15), while for 1-step prediction, it looks like an MA(16), but the 16th component is not highly significant, and the residual can be approximated as WN. It should be noted that the prediction residual of BJ models are not MA models when $x_{t+k} \neq E\{x_{t+k} \mid \theta\}$, for it is formulated as:

$$\varepsilon_{t+k|\,t}(\theta) = F(z)\, e_{t+k} + \hat{\tilde{F}}\,(z)\,[\, x_{t+k} - E\{x_{t+k} \mid \theta\}]$$

Hence, for 9-step prediction, the prediction residual should ideally look like an MA(k-1) model, but not for 1-step prediction. Since we deal with real data, the MA(15)-ish prediction residual, whose 15th and 14th components are not highly significant, and thus can be approximated as an MA(1) model, is acceptable.

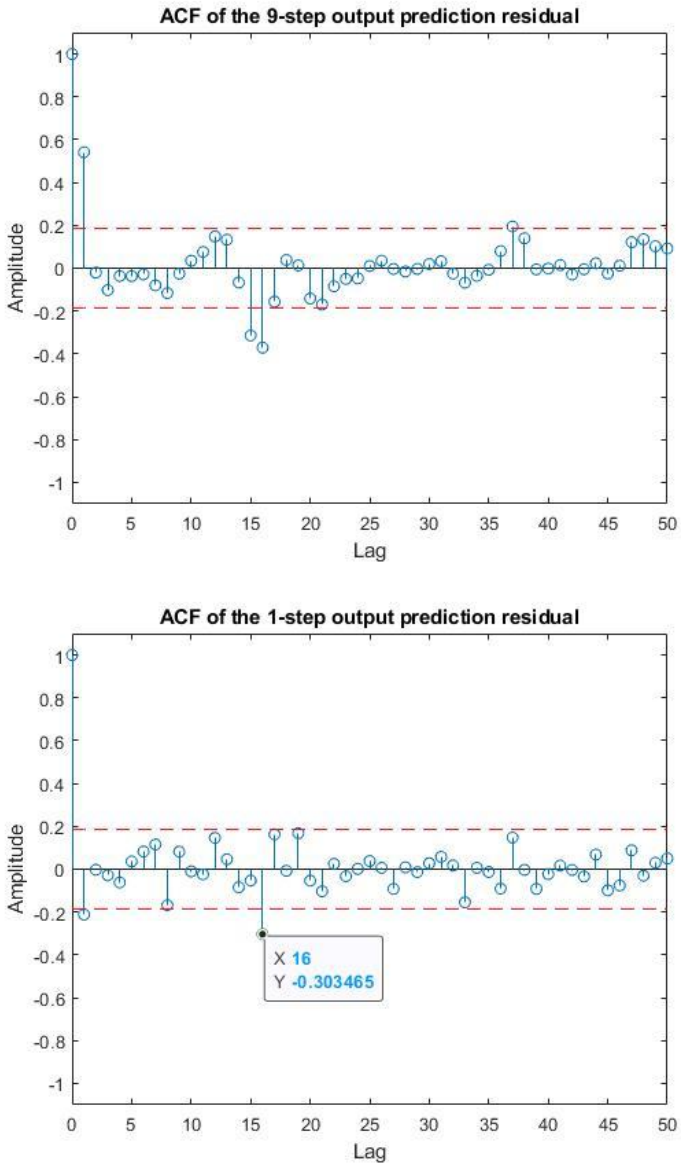

Figure 21. ACF plot of the prediction residual of the validation data using model B for both the prediction horizons

IV. MODEL C:

To design model C, which should be a recursive model, I implemented Kalman filter to update the model parameters for the whole measured and prediction datasets, y_t and u_t, not just their modelling sections, for Kalman filter is able to model data dynamically. Then, the evaluate the model on the validation and test data, the prediction performance and variance of the prediction residual of their corresponding parts of the data are considered.

As an idea, we can set the initial expected values of the model parameters and their variances by them in the non-recursive model designed as model B. This can help the filter make the model parameters converge to their true values faster and without spoiling so many initial samples.

The non-recursive model is a BJ one, which is formulated as:

$$y(t) = \frac{B(z)}{A_2(z)}\, u(t) + \frac{C(z)}{A_1(z)}\, e(t)$$

While to build the state-space form of the model to create the matrix C, we need to convert the BJ model to an ARMAx one, which is simply done by multiplying both $A_1(z)$ and $A_2(z)$ with either side of the above equation, resulting in:

$$A_1(z)A_2(z)\, y(t) = B(z)A_1(z)\, u(t) + C(z)A_2(z)\, e(t)$$

Being $A_2(z) = 1$, it is as:

$$A_1(z)y(t) = B(z)A_1(z)\, u(t) + C(z)\, e(t)$$

Thus, the initial expected values of the parameters and state-space matrix C are defined as:

$$\mathrm{x(t)} = [a_1, a_6, a_7, c_3, b_8, b_8 * a_1, b_8 * a_6, b_8 * a_7, b_{17} * a_1, b_{17} * a_6, b_{17} * a_7\ ]$$

$$C = [-y(t-1), -y(t-6), -y(t-7), \hat{e}(t-3), u(t-8), u(t-9), u(t-14), u(t-15),$$
$$u(t-17), u(t-18), u(t-23), u(t-24)]$$

Initial values of variances of the parameters are defined in a similar way to $\mathrm{x(t)}$. As we have $-y(t-1),\ -y(t-6),\ and - y(t-7)$ in the C matrix, if the prediction horizon is bigger than 1, 6, and 7 respectively, they should be approximated by their corresponding predicted values, $\hat{y}$. The Matlab code to do so is as below:

```
% Form the k-step prediction by first constructing the future C vector
    % and the one-step prediction.

    yk = zeros(1,k+1);  % to store predicted values in every iteration

    Ck = [ -y(t), -y(t-5), -y(t-6), h_et(t-2),u(t-7), u(t-8),...
        u(t-13), u(t-14), u(t-16), u(t-17), u(t-22), u(t-23) ];
    yk(1) = Ck*xt(:,t);

    for k0=2:k
        if k0 <= 6 %while k0 is less than 6, we still have observations for y(t+k0-6)
            Ck = [ -yk(k0-1), -y(t+k0-6), -y(t+k0-7), h_et(t+k0-3),u(t+k0-8), u(t+k0-
9),...
            u(t+k0-14), u(t+k0-15), u(t+k0-17), u(t+k0-18), u(t+k0-23), u(t+k0-24) ];
            yk(k0) = Ck*A^k*xt(:,t);

        % while k0 is 7, we do not have observations for y(t+k0-6),
        % but still have observations for y(t+k0-7)
        elseif k0 == 7
            Ck = [ -yk(k0-1), -yk(k0-6), -y(t+k0-7), h_et(t+k0-3),u(t+k0-8), u(t+k0-
9),...
            u(t+k0-14), u(t+k0-15), u(t+k0-17), u(t+k0-18), u(t+k0-23), u(t+k0-24) ];
            yk(k0) = Ck*A^k*xt(:,t);

        % when k0 is bigger than 7, we do not have observations for y(t+k0-6) and
y(t+k0-7),
        % so we need to use the predicted value of 6 and 7 time-steps
        % before, respectively, which were already saved in yk array
        else
            Ck = [ -yk(k0-1), -yk(k0-6), -yk(k0-7), h_et(t+k0-3),u(t+k0-8), u(t+k0-
9),...
            u(t+k0-14), u(t+k0-15), u(t+k0-17), u(t+k0-18), u(t+k0-23), u(t+k0-24) ];
            yk(k0) = Ck*A^k*xt(:,t);
        end
    end
    if k> 1
        yhatk_C_simp(t+k) = yk(k0;% Note that this should be stored at t+k.
    else
        yhatk_C_simp(t+k) = yk(1); % Note that this should be stored at t+k.
    end
```

Estimates of the model parameters are plotted in Figure 22. There seems not to be high changes with the parameters over time, showing that static prediction model work well on the whole data.

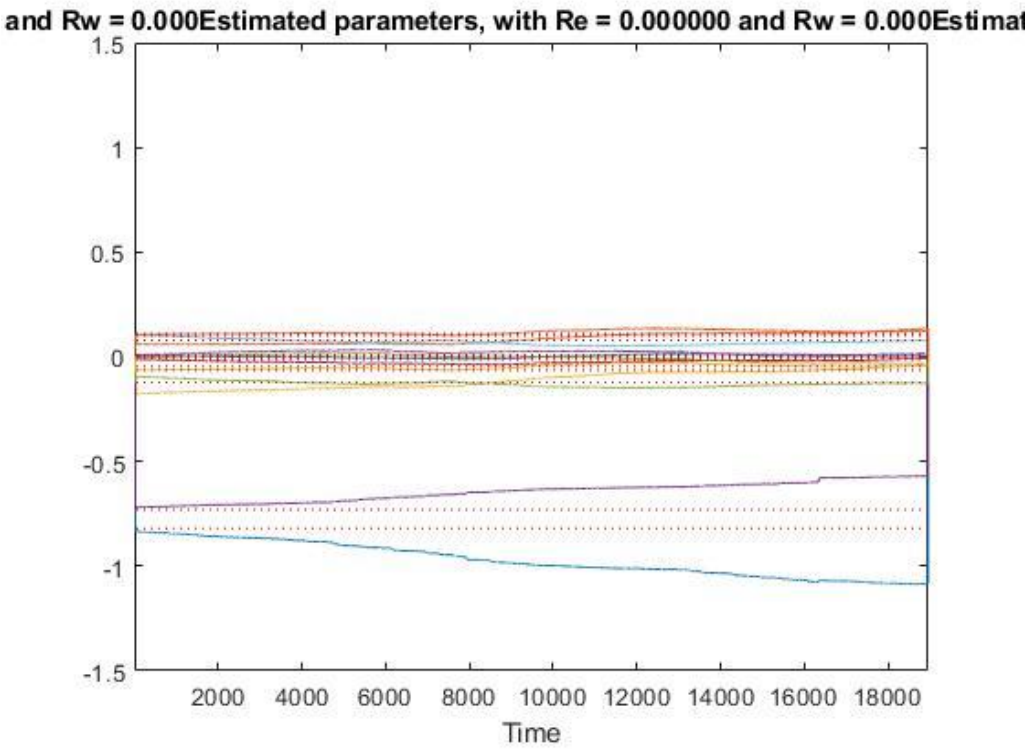

Figure 22. estimates of the model parameters

The resulting 1-step and 9-step predictions and prediction residuals using model C for both validation and test data are depicted in Figure 23a and Figure 23b, respectively. It is obvious that the model works quite well on both validation and test data for 1-step prediction, but not for 9-step prediction.

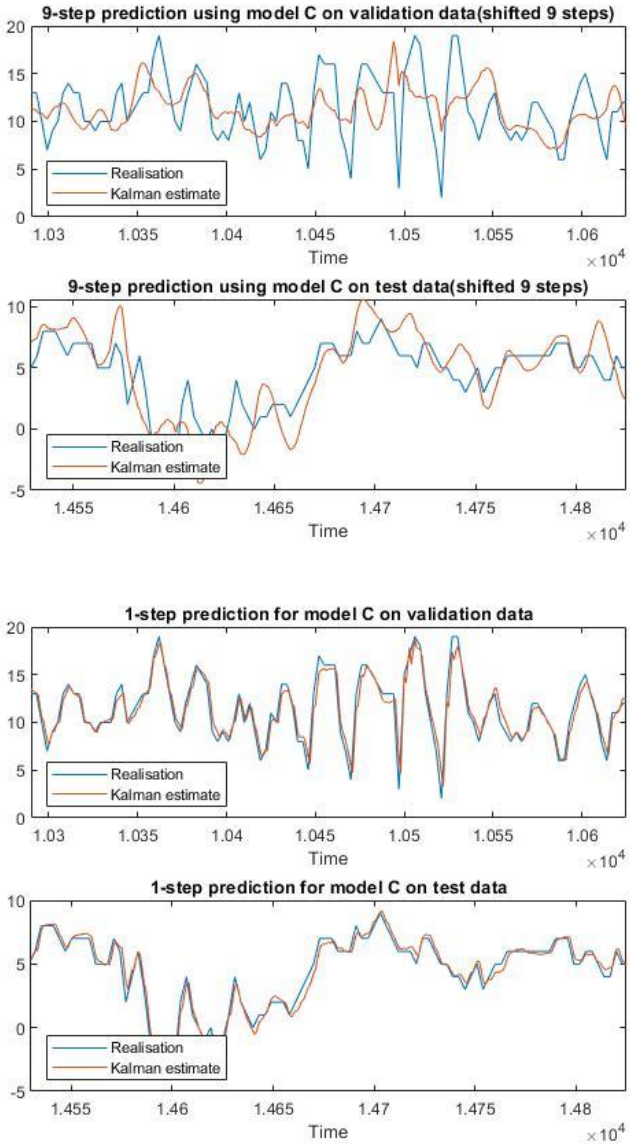

Figure 23a. resulting 1-step and 9-step predictions using model C for both validation and test data

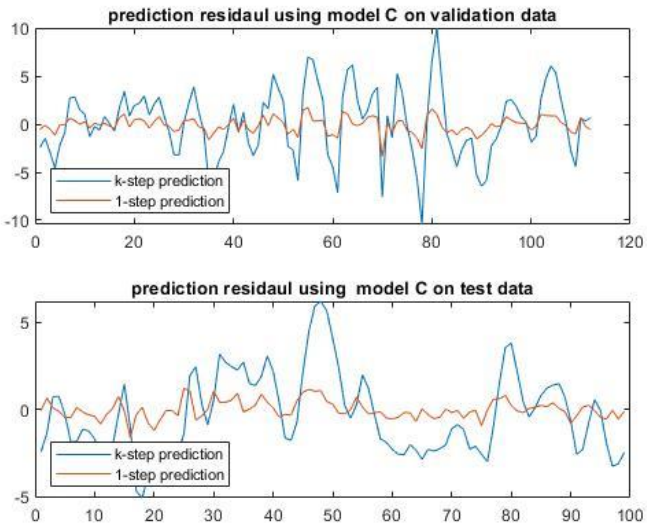

Figure 23b. resulting 1-step and 9-step predictions and prediction residual using model C for both validation and test data

The ACF and PACF plots of prediction residual using model C for both prediction horizons on the true observations of the validation data are illustrated in Figure 24. Having the true observations every 3 hours, the ACF shows the 24-hour periodicity, and the PACF shows an AR(19) with significant parameters at lags 1,2,3, 18 and 19, which resembles the AR part of model A. Both the ACF and PACF are normal, showing that the results are quite reliable.

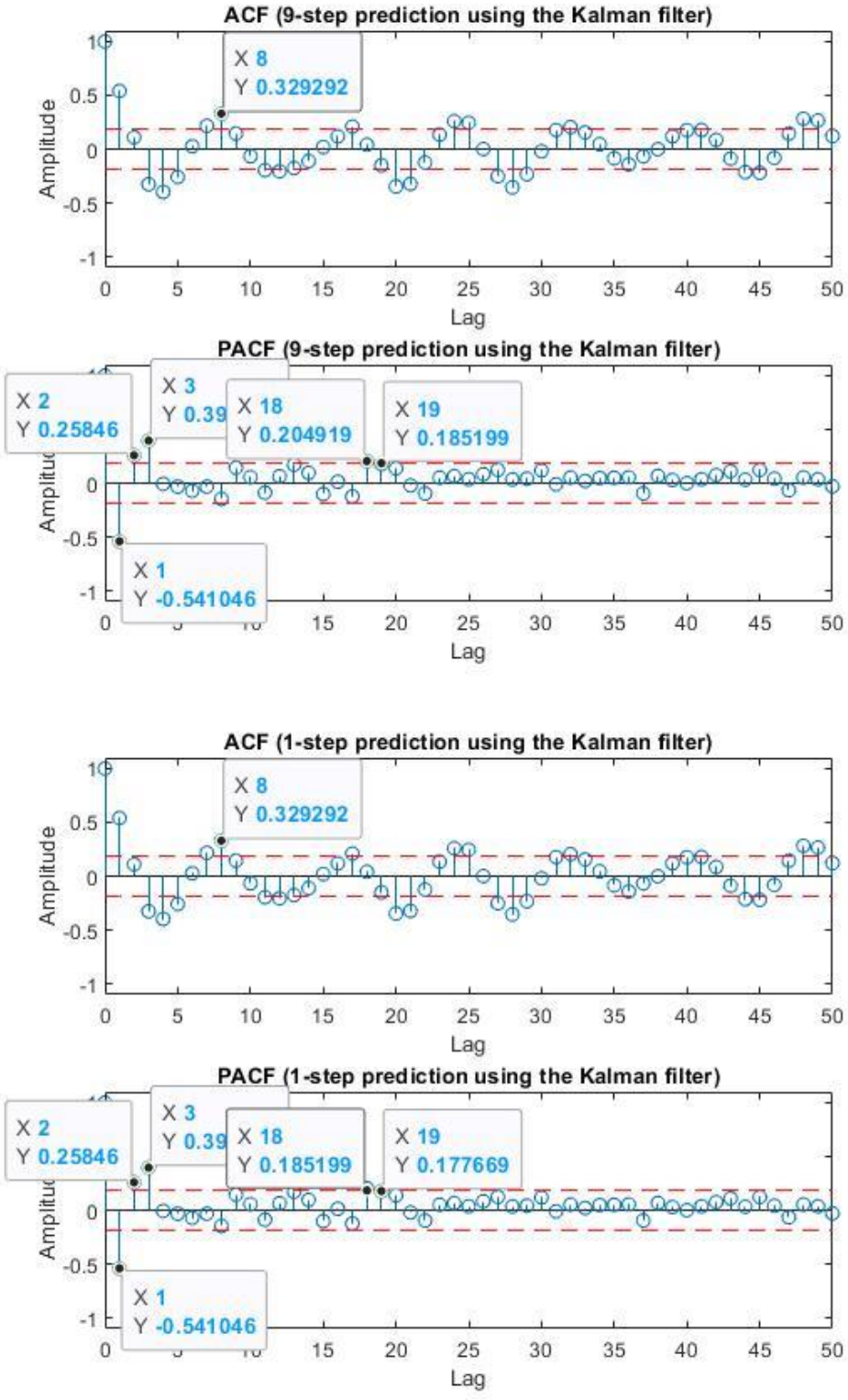

Figure 24. ACF and PACF plots of prediction residual using model C for both prediction horizons on the true observations of the validation data

The scaled variance of the 1-step and 9-step prediction errors on true observations of the validation data are 0. 057 and 1.06, respectively. They both are less than their corresponding Naïve ones. Plus, the scaled variance of the 1-step and 9-step prediction errors on the test data are 0.035 and 0.72, respectively, which are almost the same with them for their corresponding Naïve predictors, which are 0.02 and 0.89, respectively, showing that model C is good enough for either 1-step prediction or 9-step prediction on the validation and test data.

V. SIMPLIFIED MODEL C

It was tried to simplify the Kalman filter without reducing its performance too much, resulting in the simplified model C in what the parameters a7 and b17 having low values are excluded from the model. Estimates of the model parameters are given in Figure 25. They are almost constant over time.

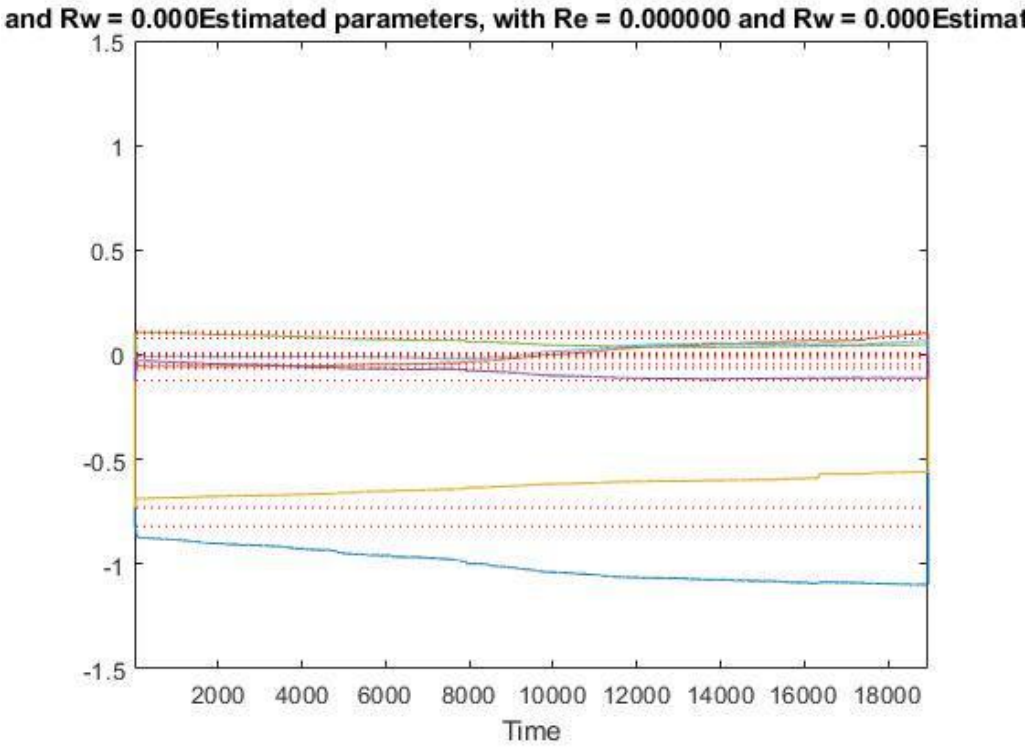

Figure 25. estimates of the model parameters

The resulting predictions and prediction residuals for both the validation and test data are shown in Figure 26a and 26b. As expected, the model does better with 1-step prediction compared to the 9-step one, while there is no significant difference between the validation and test data in terms of prediction performance.

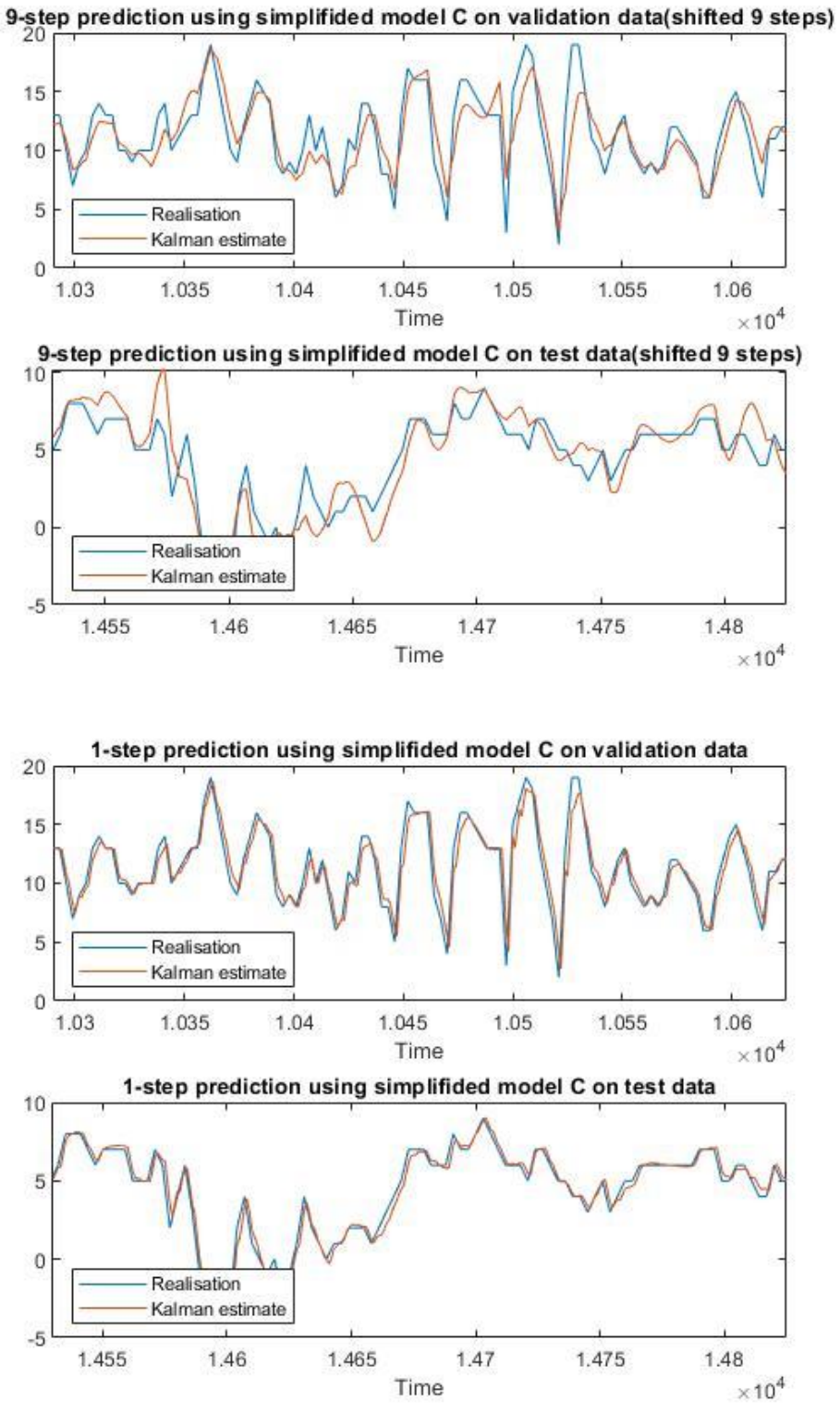

Figure 26a. resulting predictions for both the validation and test data

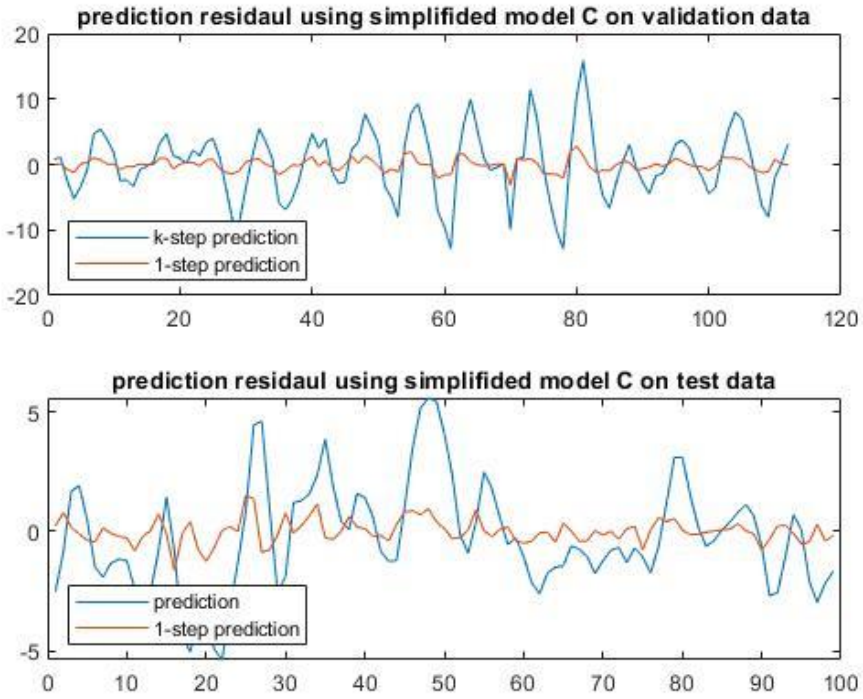

Figure 26b. resulting prediction residuals for both the validation and test data

The ACF and PACF plots of prediction residual using simplified model C for both prediction horizons on the true observations of the validation data are illustrated in Figure 27. Having the true observations every 3 hours, the ACF shows the 24-hour periodicity, and the PACF shows an AR(19) with significant parameters at lags 1,2,3, and 18, which resembles the AR part of model A. ACF has not normal distribution, but PACF is, showing that the results for PACF is quite reliable.

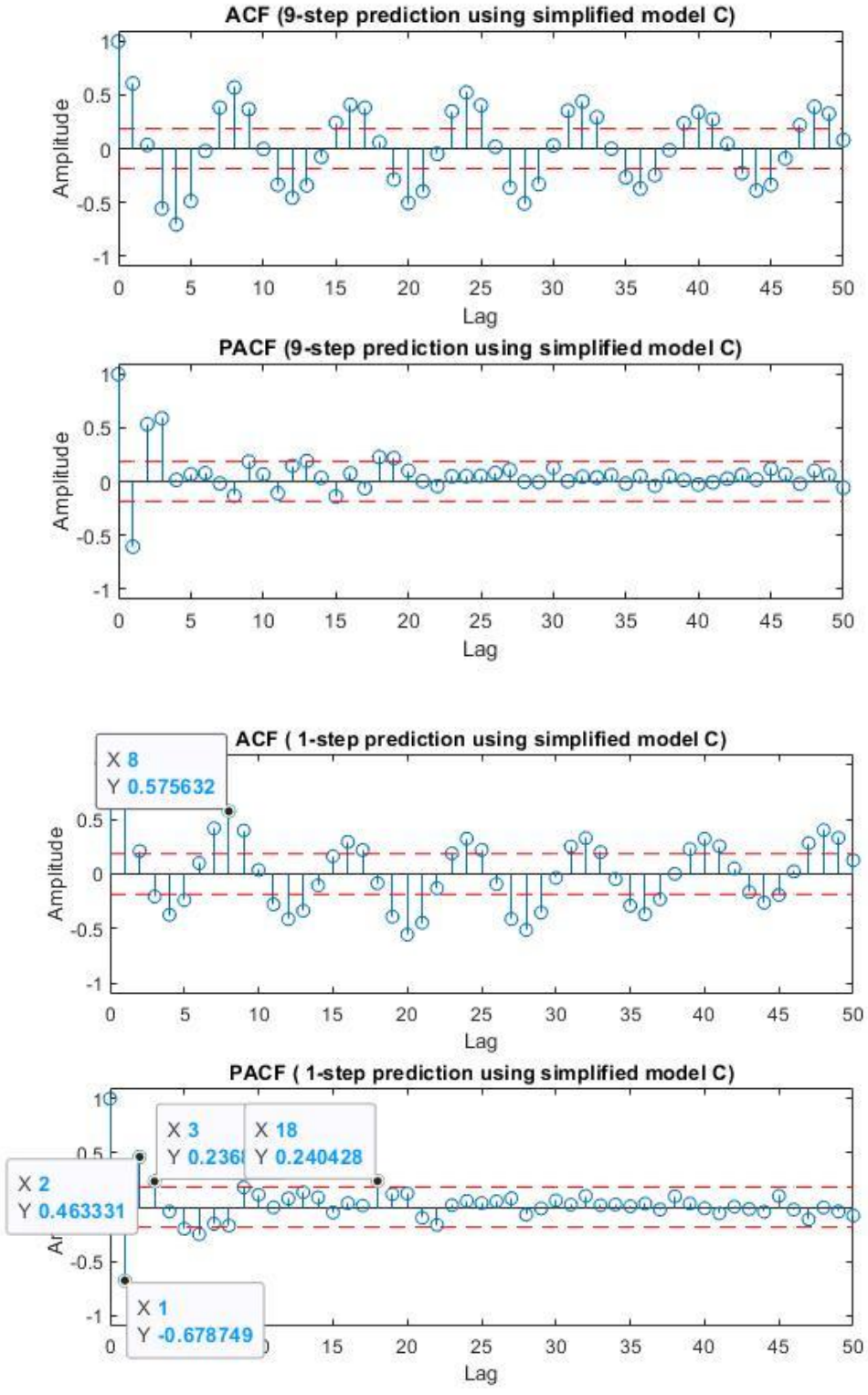

Figure 27. ACF and PACF plots of prediction residual using simplified model C for both prediction horizons on the true observations of the validation data

The scaled variance of the 1-step and 9-step prediction errors on true observations of the validation data are 0.076 and 2.23, respectively. They both are less than their corresponding Naïve ones, but not model C. On the other hand, the scaled variance of the 1-step and 9-step prediction residuals on the test data are 0.032 and 0.62, respectively, which are almost the same with them for their corresponding Naïve predictors, which are 0.02 and 0.62, respectively, but a bit better than model C, showing that simplified model C is good enough for either 1-step prediction or 9-step prediction on both the validation and test data.

VI. CONCLUSION:

In this project, different predictive models with and without external inputs as models A and B, a recursive model as model C and a simplified recursive model using Kalman filter as simplified model C were designed to beat their corresponding Naïve predictors. Both the observed and input data are temperature in Sturup.

Variances of prediction residuals using different models for true observations in the validation data, Table 1, and in the test data, Table 2, are presented. As can be seen, model A does very well on the validation data, but on the test data. Model B is not a good one for either of the data, but model C outperforms the Naïve predictors for both the validation and test data as per both the prediction horizons. Simplified model C is the best to me, for it has far less parameters than model C, but still outperforms the Naïve predictors. Less number of model parameters is more desirable, for the risk of overfitting will be less, and manufacturing cost and limitations of smaller models are less than bigger ones.

Model	$\sigma^2_{t+1\mid t}$	$\sigma^2_{t+9\mid t}$
Naïve models	0.095	2.58
Model A	0.096	0.53
Model B	0.14	0.71
Model C	0.057	1.06
Model C (simplified)	0.076	2.23

Table 1. variances of prediction residuals using different models for true observations in validation data

Model	$\sigma^2_{t+1\mid t}$	$\sigma^2_{t+9\mid t}$
Naïve models	0.02	0.89
Model A	0.18	1.53
Model B	0.30	1.83
Model C	0.035	0.72
Model C (simplified)	0.032	0.62

Table 2. variances of prediction residuals using different models for true observations in the test data