Time Series Final Report

Rujia Li, Xinxu Yao, Group 49

Introduction and data preparation

In the final project, we have a dataset of the history data of the Air temperature and some related features, which are Reflected shortwave radiation, Net radiation and Earth heat flux. Our main task is to predict the Air temperature and build a predictor.

Firstly, we select our training data, validation data and test data. For training data, this is where we build our model, so we need this data to be stationary, which means the mean value should be a constant. So we choose data points from 30025 to 31369, which is a 8-week long period in summer as the training dataset. Then we choose the data point right after this from 31370 to 31874 as the validation dataset. Then we choose the 31875 to 32043 as the first test dataset, then we randomly choose 34350 to 34518 which is in winter as the second test dataset, finally we choose the 50001 to the end as the third test dataset which is long enough and contains more difference like trending. And it is reasonable to believe that if the predictor does well in the whole third test dataset, it works fine in most part of this sequence.

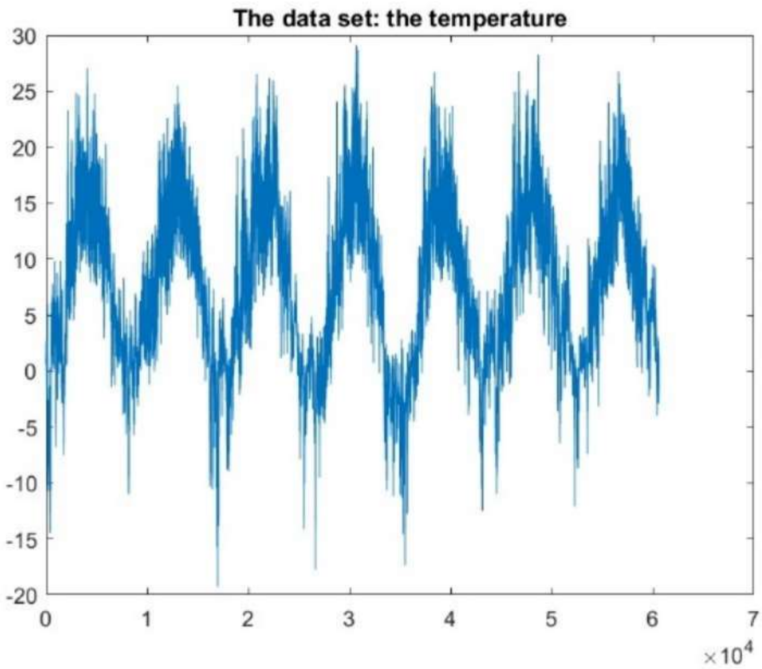


Figure 1: Air temperature for the whole dataset

Part A: ARMA model

1. Transformation

We plot the training data to see if there is a period or trending. In this figure, the data is periodic and relatively stable, so this can be a good training dataset.

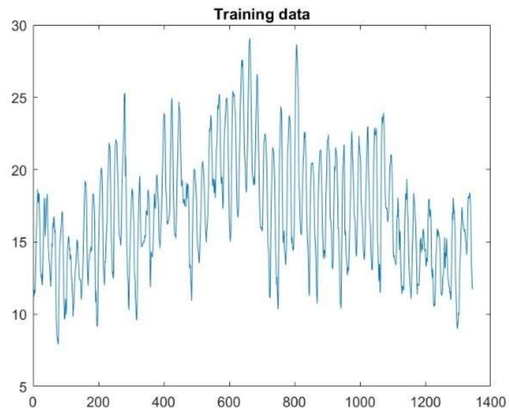


Figure 2: original training data

Then we use periodogram to see the period, the first peak is at 0.083, so the period is 24 which is reasonable.

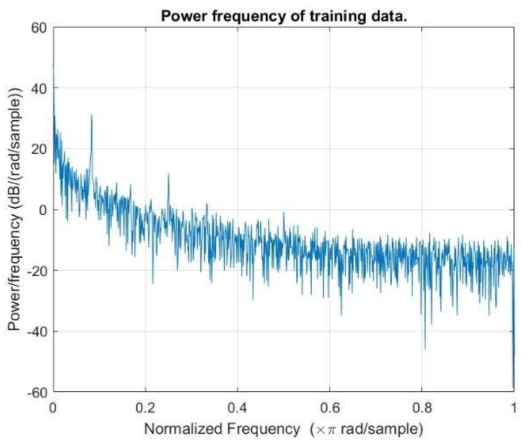


Figure 3: Power frequency of training data

So we do 24th order difference. And Fig.4 is the differentiated training data, it seems stationary and aperiodic. So we can move on to the modeling part.

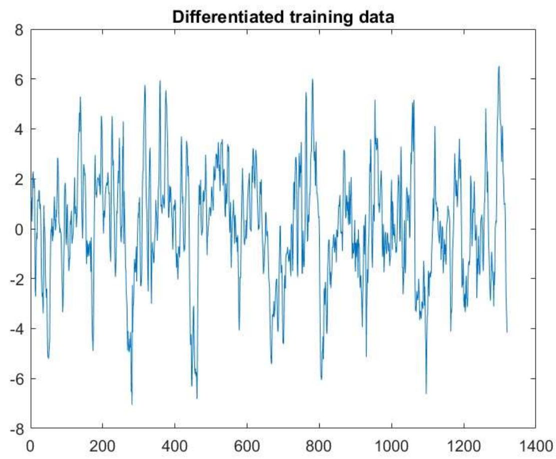


Figure 4: Differentiated training data

2. Modeling and Residual Analysis

We first plot the ACF and PACF of the differentiated training data. In the Fig.5, start with the PACF, we can see 1 and 2 are significantly not zero, so we choose AR(2) model first.

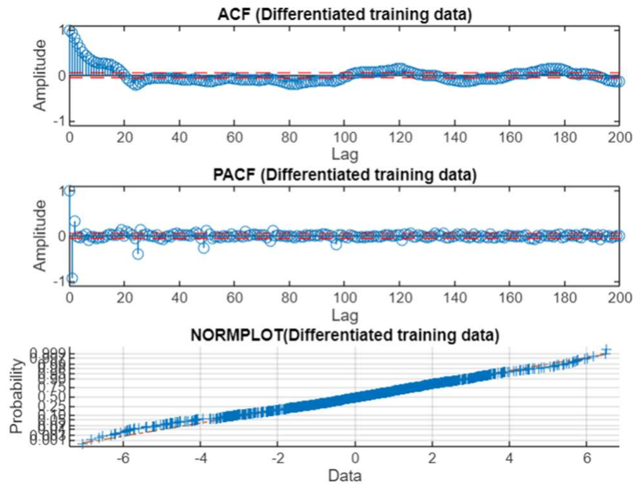


Figure 5: ACF, PACF and NORMPLOT of Differenced Data

The model is as follows.

```
model_max =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 1.254 (+/- 0.02588) z^-1 + 0.3454 (+/- 0.02591) z^-2
```

All the parameters are significant. We can move on to residual analysis.

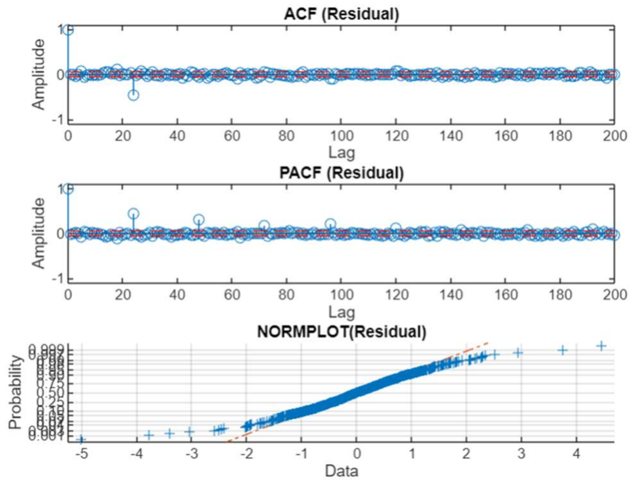


Figure 6: Residual analysis of AR(2)

We have already put two parameters from PACF, so we need to check ACF this time. We observe that 24 is outside the confidence interval, so naturally we include C_{24} into the model, and the model becomes ARMA(2,24) model. But the only parameter in the MA part is C_{24}. The estimated ARMA model is as follows:

```
model_max =
Discrete-time ARMA model: A(z)y(t) = C(z)e(t)
  A(z) = 1 - 1.282 (+/- 0.02642) z^-1 + 0.3255 (+/- 0.02641) z^-2

  C(z) = 1 - 0.8562 (+/- 0.01479) z^-24
```

All the parameters are significant, now we move on to the residual analysis. We plot the ACF, PACF and Normplot of the residual.

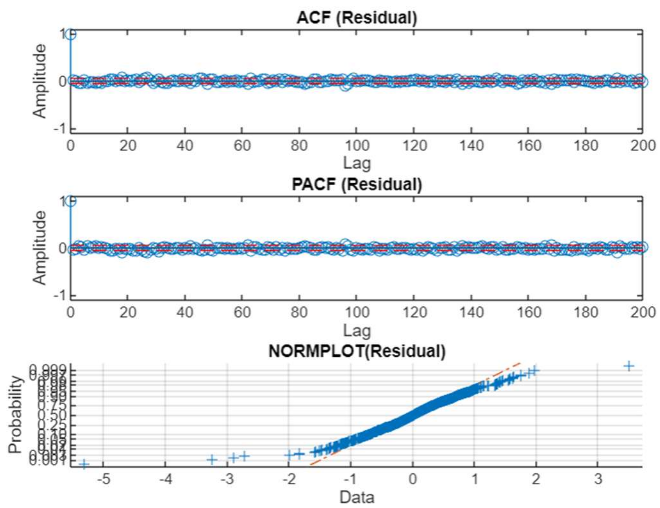


Figure 7: Residual analysis of ARMA model

The residual looks to be white in the ACF and PACF, and in the Normplot, the residual looks close to be normal. We can move on to the residual test.

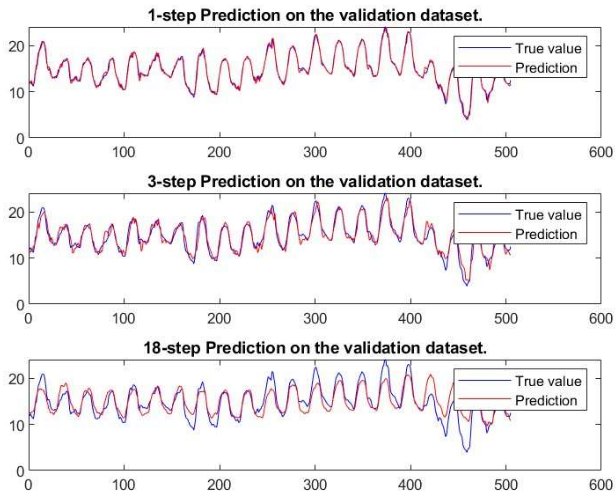
```
The residual is deemed to be WHITE according to the Monti-test (as 26.75 < 31.41).
ans = logical
```

```
The D'Agostino-Pearson's K2 test indicates that the Agostino-Pearson is NOT normal distributed.
```

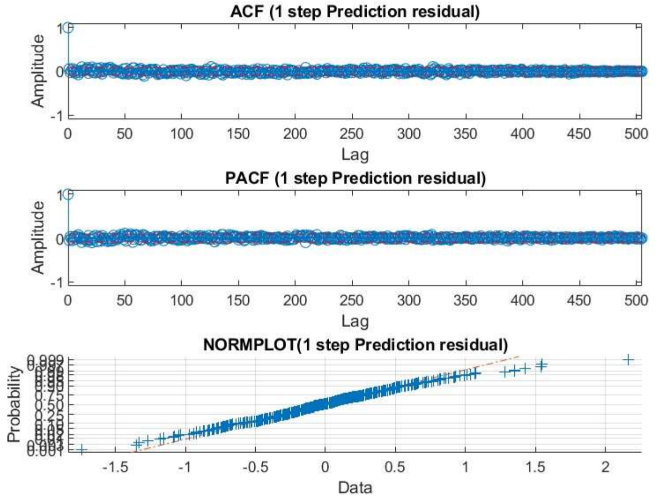
The test shows the residual is white if it is normal distributed but the residual can not pass the normal test. But we still take the residual as normal distributed, however, if something goes wrong in the prediction part, this might be the problem.

3. Validation

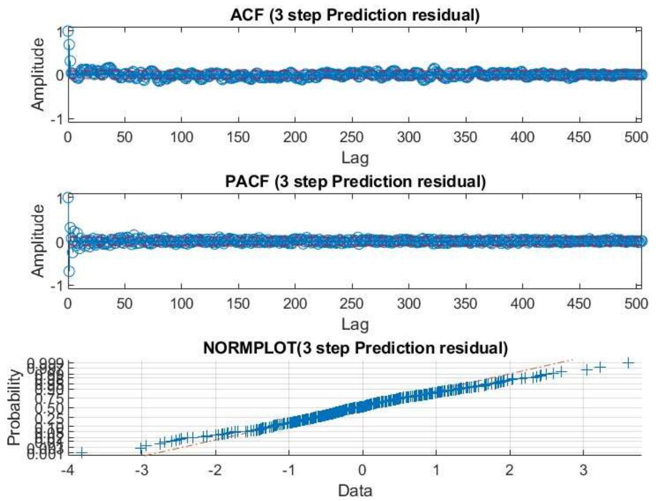
Now we can form prediction on the validation dataset and see if it works.




```
Prediction results for the 1-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    0.23. The normalized variance is 0.0190.
  The variance of the naive predictor is    0.81. The normalized variance is 0.0675.
The residual is deemed to be WHITE according to the Monti-test (as 19.74 < 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```




```
Prediction results for the 3-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    1.05. The normalized variance is 0.0878.
  The variance of the naive predictor is    5.54. The normalized variance is 0.4611.
The residual is NOT deemed to be white according to the Monti-test (as 412.27 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```




```
Prediction results for the 18-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    3.96. The normalized variance is 0.3295.
  The variance of the naive predictor is    4.37. The normalized variance is 0.3639.
The residual is NOT deemed to be white according to the Monti-test (as 596.17 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

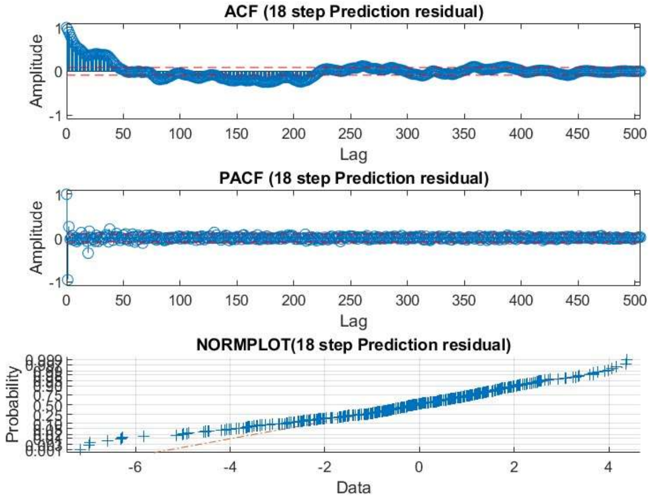


Figure 8: Predictions on the validation dataset

We can see from Fig.8, the 1-step and 3-step predictions work good, but in the 18-step prediction from 400th to 460th point, the prediction can not catch the decrease. But this is reasonable as this is a 18-step prediction, at the first period where decrease shows, all the information we have is the information up to last period where there is no information about the decrease. So even if this is not so good, we can accept that and move on to the prediction on the test dataset.

Overall, our predictor performs significantly better than the naïve one on validation data for 1step and 3step prediction, however, for 18 step prediction, our predictor is just a little better. And the 1 step prediction residual is white as supposed, 3 step prediction looks like a MA(3), however, 18 step prediction seems not a MA(18).

4. Prediction

For test dataset 1, which is close to the training dataset, it is supposed to be the easiest prediction for the predictor, we can check how it works.

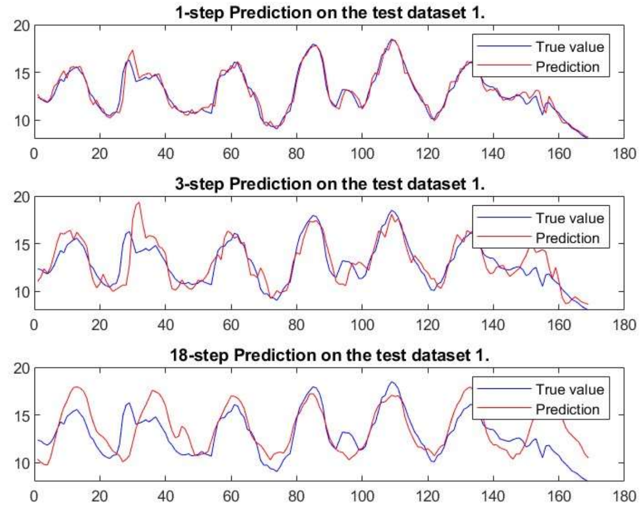


Figure 9: Predictions on the test dataset 1

Surprisingly, the prediction works not so good as in the validation dataset. But then we find this dataset is actually not easy to predict, as this set seems to be not so periodic, especially in the 90^{th} to the 100^{th} part and 140^{th} to the end. Now we use the variance and the normalized variance to evaluate this predictor's performance as well as naïve predictor on this dataset.

```
Prediction results for the 1-step prediction:
  The variance of the validations data is      5.32
  The variance of the proposed predictor is    0.31. The normalized variance is 0.0588.
  The variance of the naive predictor is    0.51. The normalized variance is 0.0958.
The residual is deemed to be WHITE according to the Monti-test (as 30.39 < 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result shows the predictor works better than naïve predictor on this dataset's 1-step prediction. And the prediction residual is white as it is supposed to be for it is a 1-step prediction. The ACF and PACF of prediction residual is in Fig.10, we can see some points a little bit outside the confidence interval, but it is acceptable as we can not find a true model.

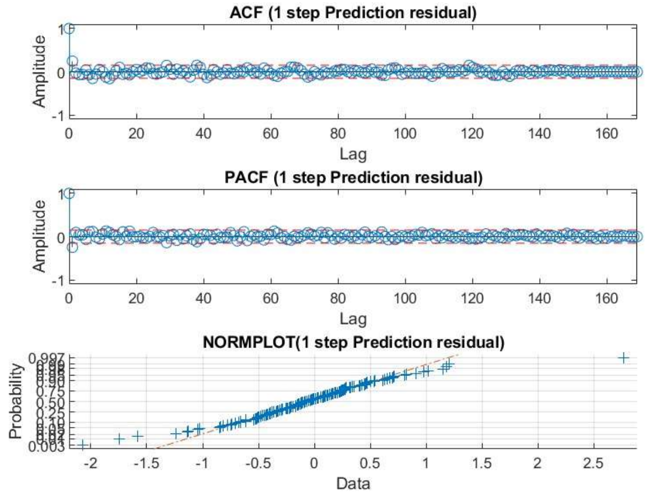


Figure 10: ACF and PACF of the 1-step prediction

We can move on to the 3-step prediction now.

```
Prediction results for the 3-step prediction:
  The variance of the validations data is      5.32
  The variance of the proposed predictor is    1.82. The normalized variance is 0.3428.
  The variance of the naive predictor is    3.18. The normalized variance is 0.5972.
The residual is NOT deemed to be white according to the Monti-test (as 176.73 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result shows the predictor works better than naïve predictor on this dataset's 3-step prediction. And the prediction residual seems to be a MA(3) as it is supposed to be. The ACF and PACF of prediction residual is in Fig.11.

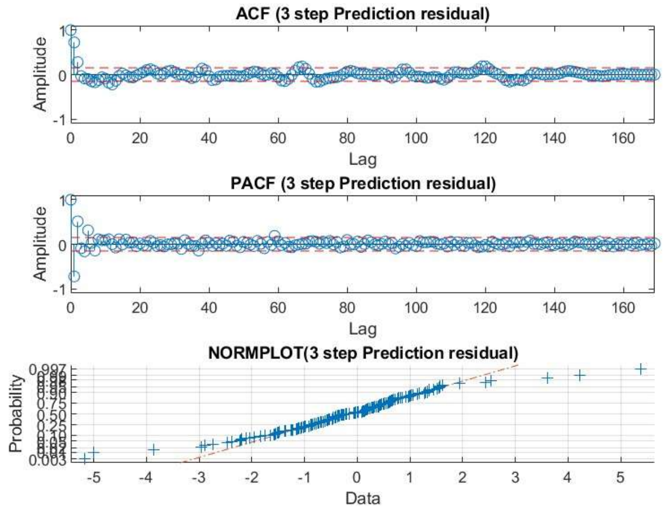


Figure 11: ACF and PACF of the 3-step prediction

We can move on to the 18-step prediction now.

```
Prediction results for the 18-step prediction:
  The variance of the validations data is      5.32
  The variance of the proposed predictor is    3.96. The normalized variance is 0.7443.
  The variance of the naive predictor is    4.09. The normalized variance is 0.7697.
The residual is NOT deemed to be white according to the Monti-test (as 217.71 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result shows the predictor works a little bit better than naive predictor on this dataset's 18-step prediction. And we can not really tell if this is MA(18), it seems to have strong AR part. The ACF and PACF of prediction residual is in Fig.12.

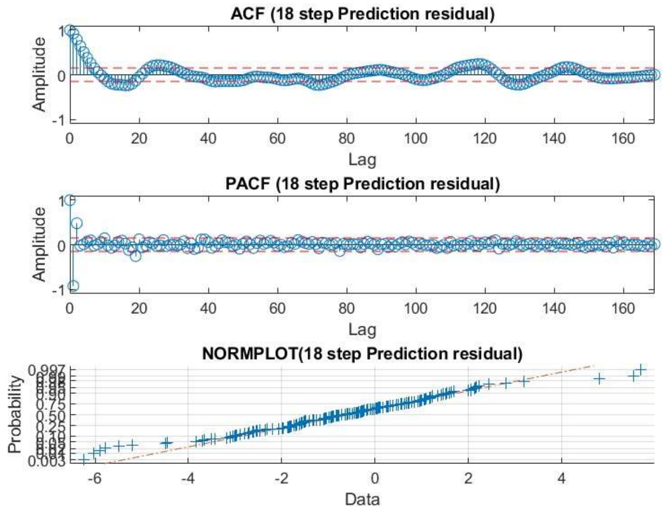


Figure 12: ACF and PACF of the 18-step prediction

For test dataset 2, which is in winter, we have the following result. This is an interesting test dataset, actually this might be one of the hardest weeks to predict, as the pattern are quite different from the training dataset. In training dataset, validation dataset and test dataset 1, the period is strong, hence the naïve predictor for 18-step prediction still works fine, however, the period is weak in this week, the periodic part of both our predictor and naïve predictor go wrong, so both of them perform bad. We can check this with the variance of the prediction error.

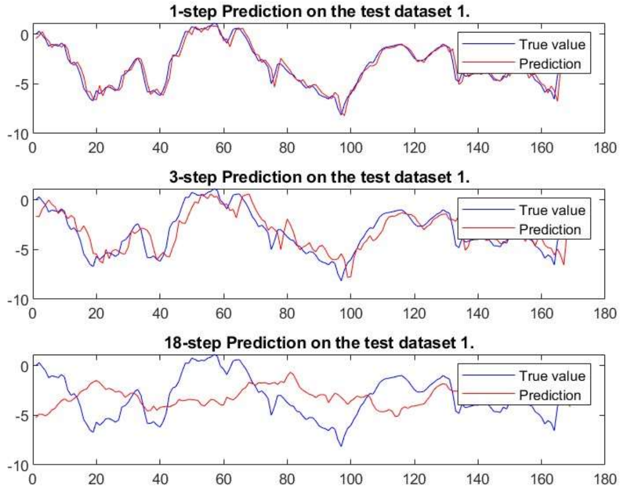


Figure 13: Predictions on the test dataset 2

The 1-step prediction is as follows, the proposed predictor works better than the naïve one. And the prediction residual seems to be white.

```
Prediction results for the 1-step prediction:
   The variance of the validations data is      4.64
   The variance of the proposed predictor is    0.30. The normalized variance is 0.0637.
   The variance of the naive predictor is    0.35. The normalized variance is 0.0763.
 The residual is deemed to be WHITE according to the Monti-test (as 15.60 < 31.41).
 The D'Agostino-Pearson's K2 test indicates that the PACF is NORMAL distributed.
```

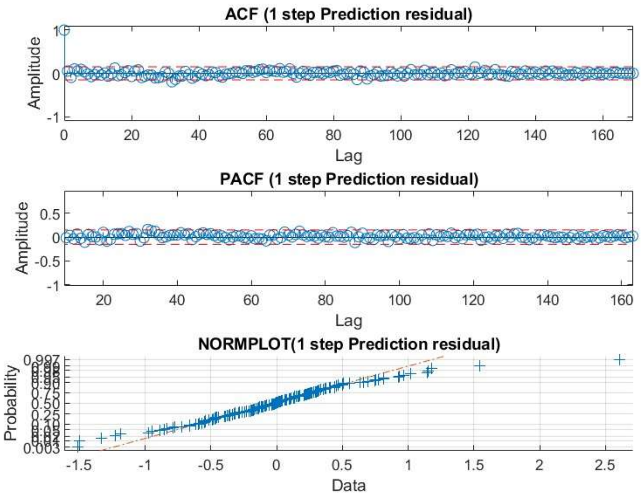


Figure 14: ACF and PACF of the 1-step prediction

We can move on to the 3-step prediction now.

```
Prediction results for the 3-step prediction:
   The variance of the validations data is      4.64
   The variance of the proposed predictor is    1.35. The normalized variance is 0.2908.
   The variance of the naive predictor is    1.64. The normalized variance is 0.3534.
 The residual is NOT deemed to be white according to the Monti-test (as 137.99 > 31.41).
 The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result shows the predictor works better than naïve predictor on this dataset's 3-step prediction. And the prediction residual might be a MA(3). The ACF and PACF of prediction residual is in Fig.15.

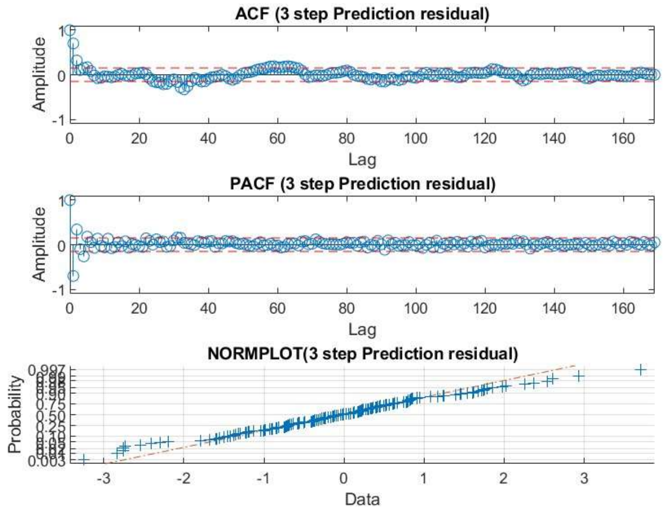


Figure 15: ACF and PACF of the 3-step prediction

We can move on to the 18-step prediction now.

```
Prediction results for the 18-step prediction:
   The variance of the validations data is      4.64
   The variance of the proposed predictor is    6.56. The normalized variance is 1.4130.
   The variance of the naive predictor is   16.97. The normalized variance is 3.6562.
 The residual is NOT deemed to be white according to the Monti-test (as 191.97 > 31.41).
 The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result is interesting as the variance of both proposed predictor and naïve predictor is very large. For long term prediction, period is important, we include the period in our model, but the period is too weak here, that may be why the 18-step prediction is that bad. Besides, as our model is far from the data here, the prediction residual should be something other than MA(18).

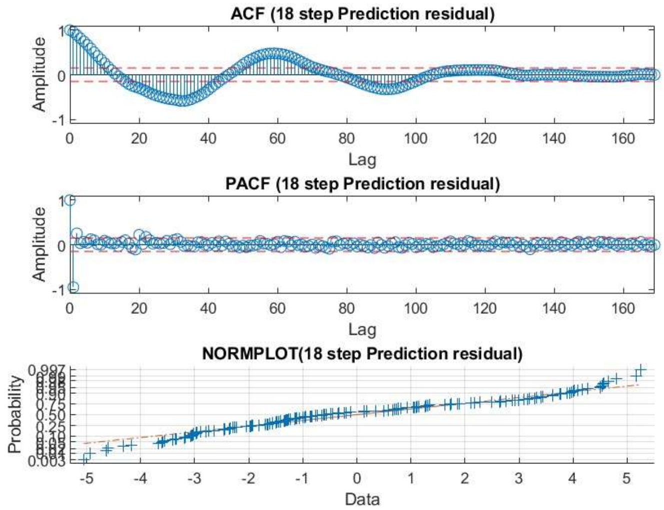


Figure 16: ACF and PACF of the 18-step prediction

The evaluation depends much on the test dataset we choose, in order to evaluate the predictor from a more objective way, we choose a third test dataset as we said in the Introduction part. From the figure 17, the predictor seems to be good.

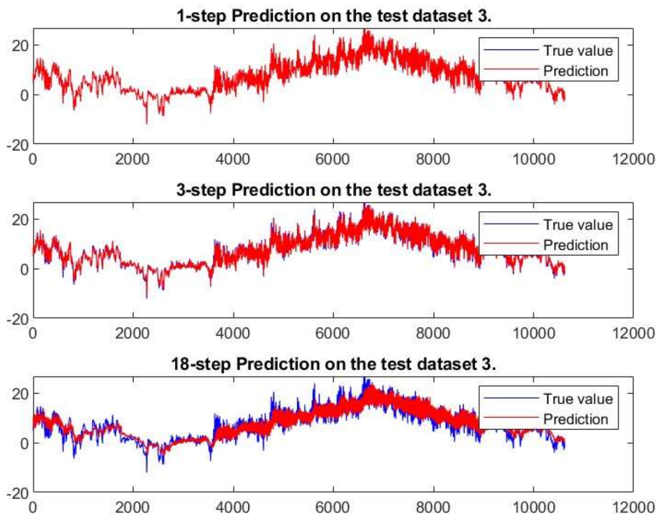


Figure 17: Predictions on the test dataset 3

The 1-step prediction is as follows, the proposed predictor works better than the naïve one. But the prediction residual seems to be white, actually we train the model at 30025^{th} point, and this test dataset begins from 50001, so it is not surprising that the model actually changed.

```
Prediction results for the 1-step prediction:
  The variance of the validations data is     37.62
  The variance of the proposed predictor is    0.24. The normalized variance is 0.0063.
  The variance of the naive predictor is    0.44. The normalized variance is 0.0118.
The residual is NOT deemed to be white according to the Monti-test (as 163.16 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

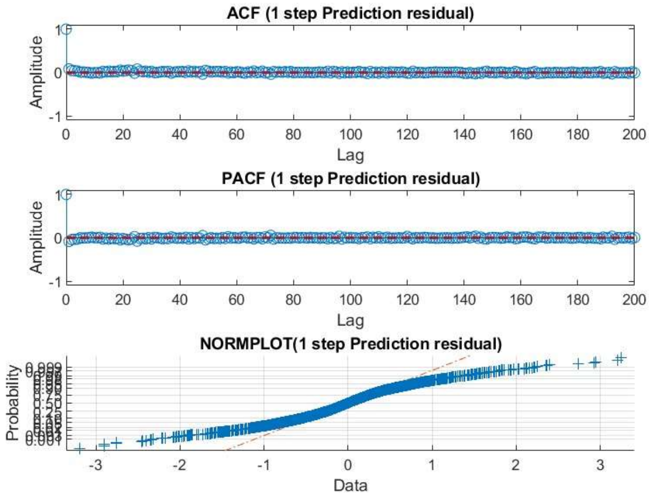


Figure 18: ACF and PACF of the 1-step prediction

For the 3-step prediction.

```
Prediction results for the 3-step prediction:
  The variance of the validations data is     37.62
  The variance of the proposed predictor is    1.19. The normalized variance is 0.0316.
  The variance of the naive predictor is    2.72. The normalized variance is 0.0722.
The residual is NOT deemed to be white according to the Monti-test (as 8854.03 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result shows the predictor works better than naïve predictor on this dataset's 3-step prediction. And the prediction residual might be a MA(3). The ACF and PACF of prediction residual is in Fig.15.

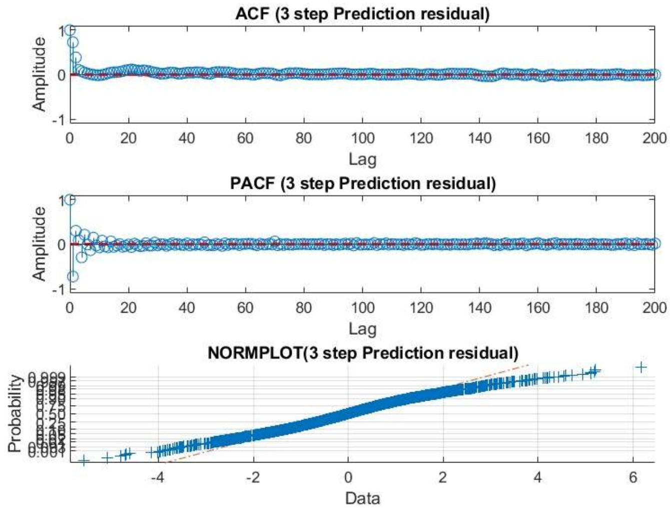


Figure 19: ACF and PACF of the 3-step prediction

We can move on to the 18-step prediction now.

```
Prediction results for the 18-step prediction:
  The variance of the validations data is     37.62
  The variance of the proposed predictor is    5.22. The normalized variance is 0.1389.
  The variance of the naive predictor is    6.05. The normalized variance is 0.1608.
The residual is NOT deemed to be white according to the Monti-test (as 12962.02 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

The result is both proposed predictor and naïve predictor seem to work better than in test dataset 2. That is because the period comes back. And our predictor is better than the naïve one. Besides, the prediction residual is not MA(18), which implies the model might have changed.

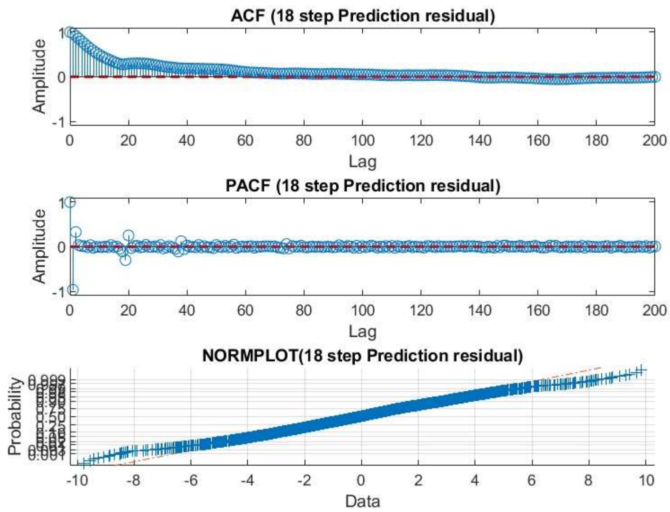


Figure 20: ACF and PACF of the 18-step prediction

Part B: Box-Jenkins Model

1. Model the input data

With the previous information from part A, to account for daily periodicity, the training data and external input data (Net radiation) are differentiated with 24. First, we plot the ACF and PACF of the data and by PACF, A_1 should be included in the model.

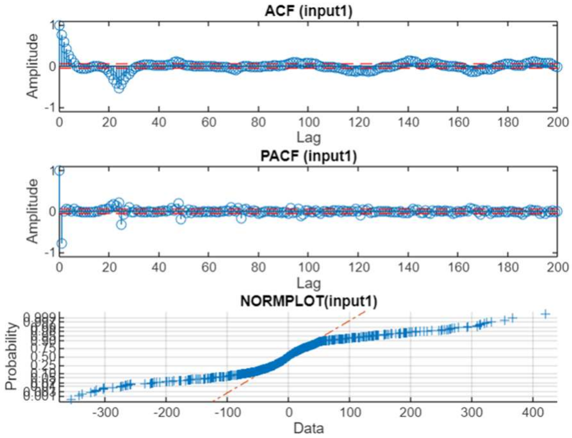


Figure 21: ACF, PACF and NORMPLOT of Differenced Input Data

```
m1im =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 0.7705 (+/- 0.01756) z^-1
```

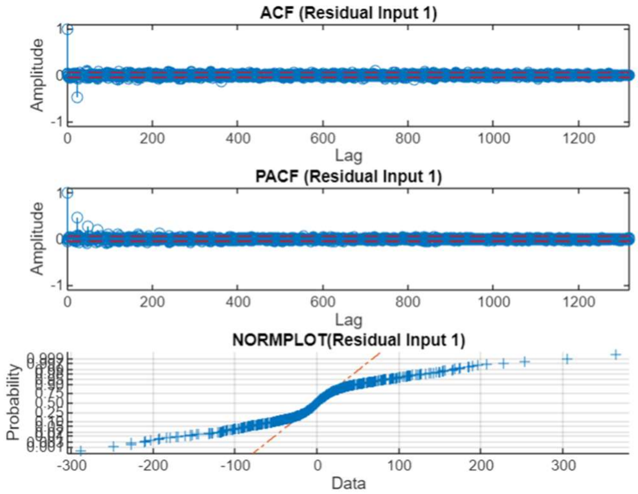


Figure 12: Residual analysis of AR(1)

By the ACF of the residual, we observed that 24 is outside the confidence interval, so we include C_{24}.

Our new model for input data is below:

```
m1im =
Discrete-time ARMA model: A(z)y(t) = C(z)e(t)
  A(z) = 1 - 0.7444 (+/- 0.01891) z^-1

  C(z) = 1 - 0.8188 (+/- 0.01613) z^-24
```

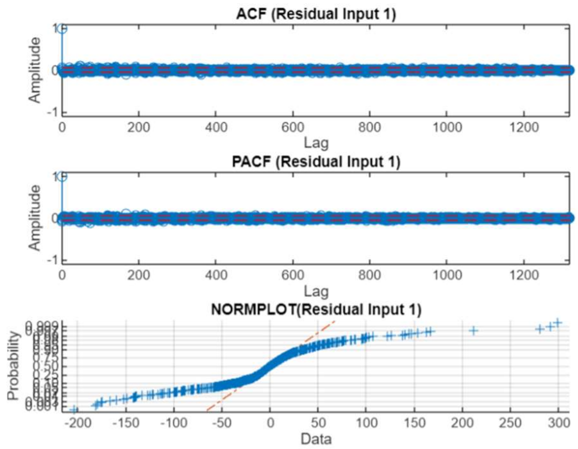


Figure 13: Residual analysis for refined input model

All the parameters are significant and the residual seems white, then we check it with Monti test.

```
The residual is deemed to be WHITE according to the Monti-test (as 23.47 < 31.41).
```

So far we finished the input model.

2. Cross-Correlation Analysis

To model the relationship between the external input and air temperature, we need to analyze the cross-correlation function (CCF) between the residuals of the input model and the temperature data.

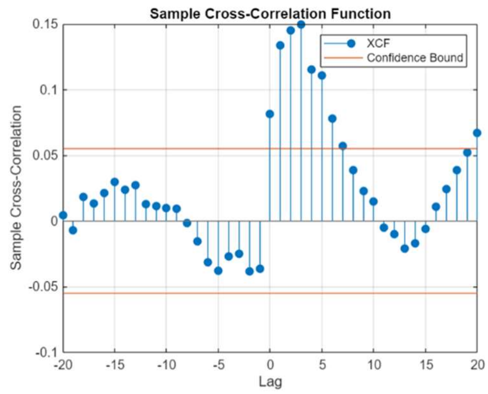


Figure 14: Cross correlation between residuals and output

We observe that when lag is 1 it becomes significant and when lag is 3 it reaches its top, so we try d=0, r=2, s=3 first.

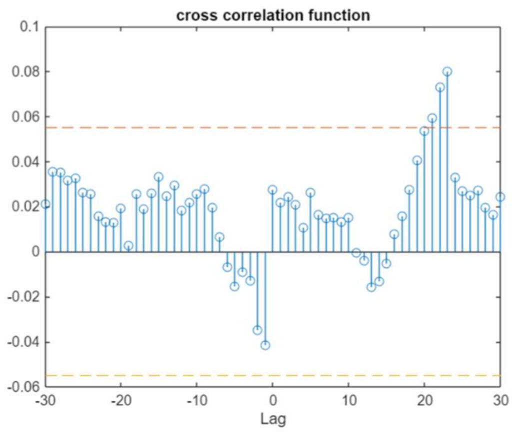


Figure 15: Cross correlation of residuals and output

There are still some significant values in CCF here, which means the model does not fully capture the impact of the input on the output. It seems not so good, we then change s from 3 to 4 and then it becomes acceptable.

```
m1bm =
Discrete-time OE model: y(t) = [B(z)/F(z)]u(t) + e(t)
  B(z) = 0.003679 (+/- 0.001159) + 0.002849 (+/- 0.001745) z^-1 - 0.001637 (+/- 0.0008208) z^-2 - 0.001647 (+/-

  F(z) = 1 - 0.3399 (+/- 0.4437) z^-1 - 0.6277 (+/- 0.4089) z^-2
```

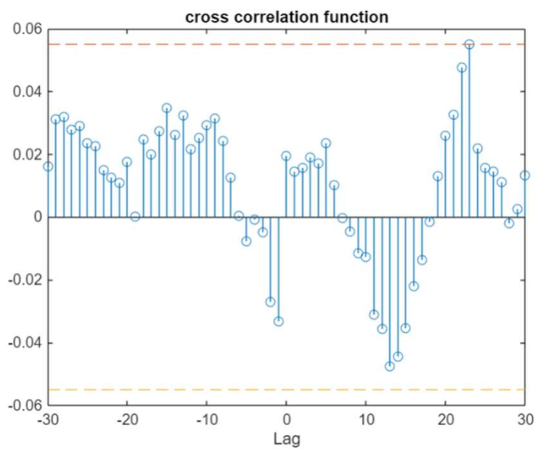


Figure 16: Cross correlation of residuals and output (refined model)

We can see all the values in CCF are zero, which means the residual of the model contains only random noise that is unrelated to the input.

Now, we get the etilda and it is not correlated with input.

3. Modeling etilda

We use the etilda from previous step, and try to extract the information from it. As usual we use ACF and PACF to identify the model.

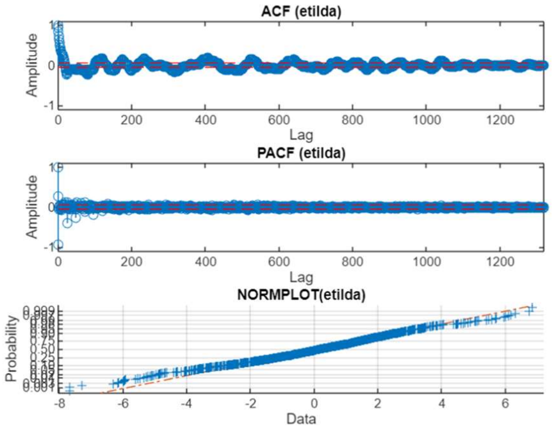


Figure 17: ACF, PACF and NORMPLOT of etilda

By observing the ACF and PACF, with the previous experience, we try the ARMA model which contains a seasonal MA term. We put it below along with its residual analysis:

```
m1om =
Discrete-time ARMA model: A(z)y(t) = C(z)e(t)
  A(z) = 1 - 1.248 (+/- 0.0266) z^-1 + 0.2921 (+/- 0.0266) z^-2

  C(z) = 1 - 0.8814 (+/- 0.01344) z^-24
```

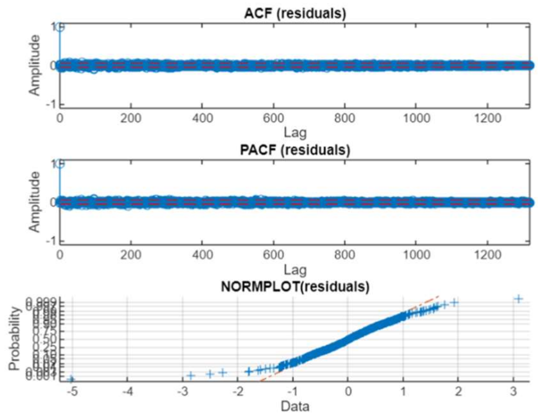


Figure 17: Residual analysis for residual model

```
The residual is deemed to be WHITE according to the Monti-test (as 24.34 < 31.41).
```

After the residual analysis we found the residual is significant to be white, which means we extract the information successfully.

So far we finish the we have completed the modeling of each part of the BJ model, and then we need to integrate the model.

4. Re-estimate the whole model

Keep the model structure and re-estimate all the parameters at the same time, we get the Box-Jenkins model for this dataset.

```
Mbj =
Discrete-time BJ model: y(t) = [B(z)/F(z)]u(t) + [C(z)/D(z)]e(t)
  B(z) = 0.006345 (+/- 0.0003722) + 0.001092 (+/- 0.001492) z^-1 - 0.002509 (+/- 0.0006055) z^-2 - 0.0007746 (+/- 0.0006055) z^-3 - 0.001394 (+/- 0.0003964) z^-4

  C(z) = 1 - 0.8644 (+/- 0.01429) z^-24

  D(z) = 1 - 1.262 (+/- 0.02664) z^-1 + 0.3072 (+/- 0.02664) z^-2

  F(z) = 1 - 0.5091 (+/- 0.2271) z^-1 - 0.4448 (+/- 0.2159) z^-2
```

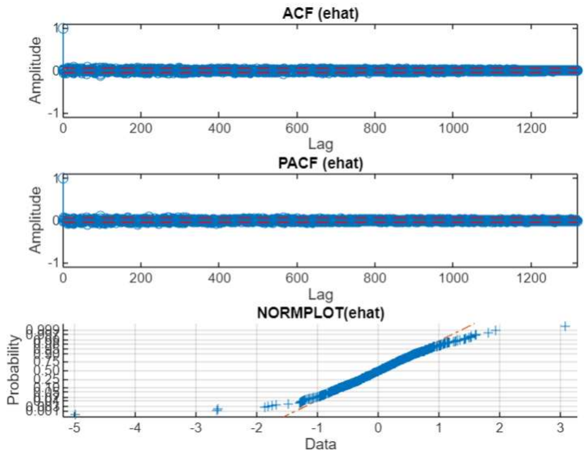
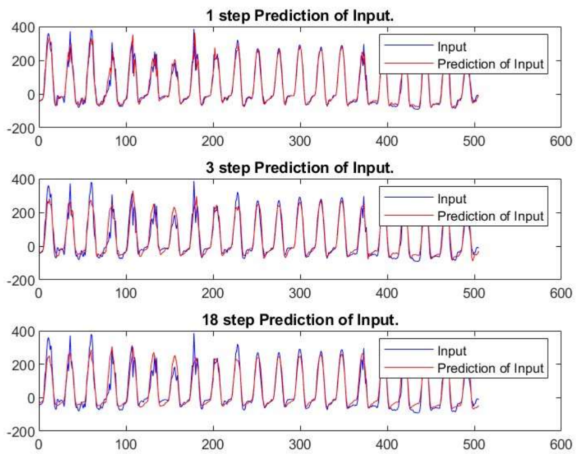


Figure 17: Residual analysis for integrated model

```
The residual is deemed to be WHITE according to the Monti-test (as 25.75 < 31.41).
```

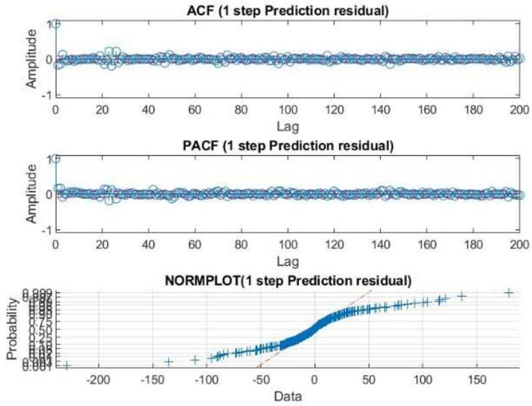
All the parameters are significant and the residual seems to be white, we can now move on to the validation dataset.

5. Validation



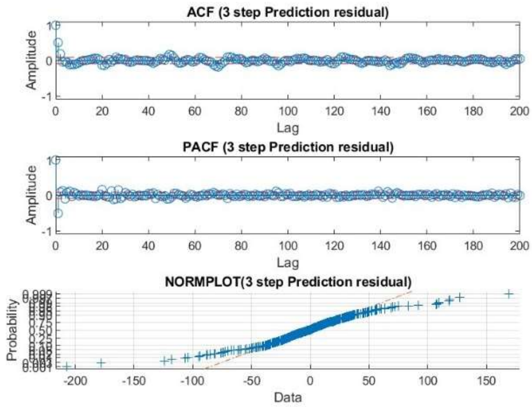
For 1 step prediction.

```
Prediction results for the 1-step prediction:
  The variance of the validations data is   14753.56
  The variance of the proposed predictor is 1006.39. The normalized variance is 0.0682.
  The variance of the naive predictor is 2464.70. The normalized variance is 0.1671.
The residual is NOT deemed to be white according to the Monti-test (as 42.04 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```



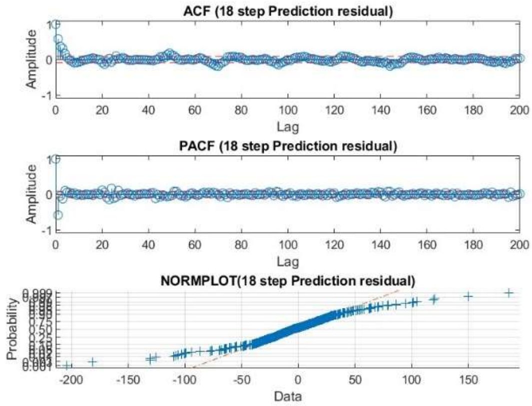
For 3 step prediction.

```
Prediction results for the 3-step prediction:
  The variance of the validations data is   14753.56
  The variance of the proposed predictor is 1341.16. The normalized variance is 0.0909.
  The variance of the naive predictor is 12813.80. The normalized variance is 0.8685.
The residual is NOT deemed to be white according to the Monti-test (as 179.15 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```

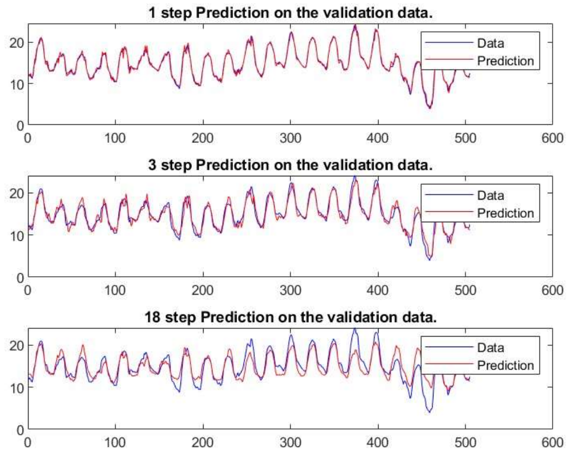


For 18 step prediction.

```
Prediction results for the 18-step prediction:
  The variance of the validations data is   14753.56
  The variance of the proposed predictor is 1474.11. The normalized variance is 0.0999.
  The variance of the naive predictor is 2515.73. The normalized variance is 0.1705.
The residual is NOT deemed to be white according to the Monti-test (as 201.54 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NOT normal distributed.
```



For the input validations, our input predictor is always better than naïve one and the residuals are as it should be. We move on to the output.



For 1 step prediction.

```
Prediction results for the 1-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    0.27. The normalized variance is 0.0224.
  The variance of the naive predictor is    0.81. The normalized variance is 0.0675.
The residual is deemed to be WHITE according to the Monti-test (as 16.10 < 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NORMAL distributed.
```

For 3 step prediction.

```
Prediction results for the 3-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    1.04. The normalized variance is 0.0867.
  The variance of the naive predictor is    5.54. The normalized variance is 0.4611.
The residual is NOT deemed to be white according to the Monti-test (as 379.61 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NORMAL distributed.
```

For 18 step prediction.

```
Prediction results for the 18-step prediction:
  The variance of the validations data is     12.01
  The variance of the proposed predictor is    3.60. The normalized variance is 0.2999.
  The variance of the naive predictor is    4.37. The normalized variance is 0.3639.
The residual is NOT deemed to be white according to the Monti-test (as 528.39 > 31.41).
The D'Agostino-Pearson's K2 test indicates that the PACF is NORMAL distributed.
```

By add an extra input, the variance becomes lower, which means it works.

6. Prediction

Same as part A, we are going to use this predictor on the three test dataset, for test dataset 1. The left are the prediction figures of input, and the right are the output, it works better than part A, but still can’t catch the decrease for 3 and 18 step predictions.

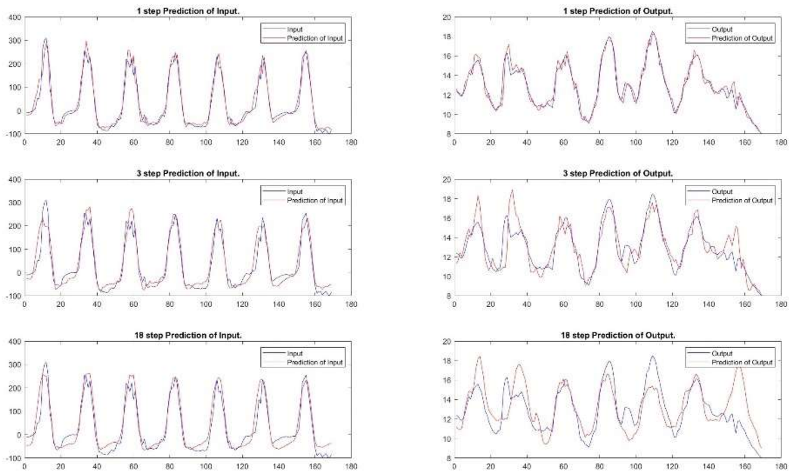


Figure 21: Prediction of test dataset 1

We put all the normalized variance in the following table.

	1-step	3-step	18-step
Input naïve	0.1717	0.9703	0.1881
Input proposed	0.0662	0.1165	0.1151
Output naive	0.0958	0.5972	0.7697
Output proposed	0.0610	0.3208	0.6602

(the variance of test dataset1 is 5.32)

For test dataset 2.

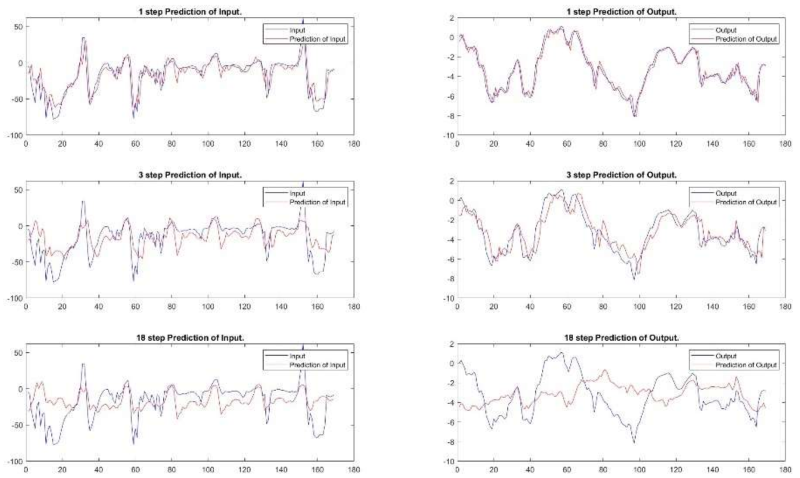


Figure 22: Prediction of test dataset 2

We put all the normalized variance in the following table.

	1-step	3-step	18-step
Input naïve	0.3233	1.0830	0.8025
Input proposed	0.2902	0.6605	0.9429
Output naive	0.0763	0.3534	3.6562
Output proposed	0.0598	0.2355	1.2778

(the variance of test dataset 2 is 4.64)

For test dataset 3.

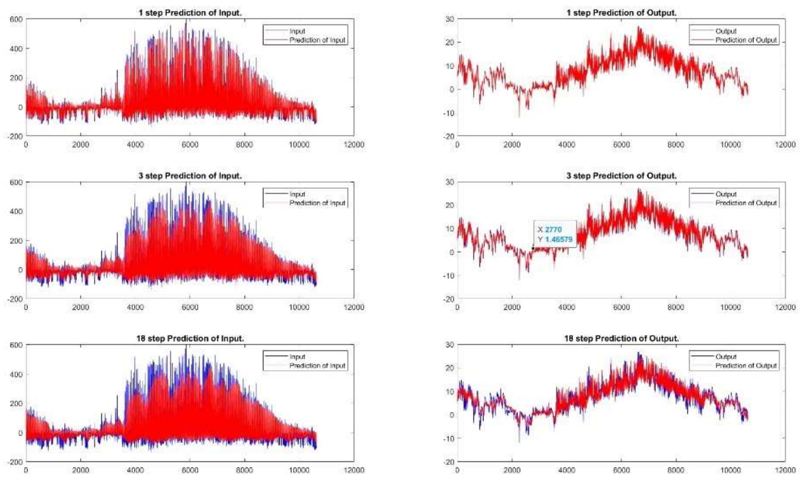


Figure 23: Prediction of test dataset 3

We put all the normalized variance in the following table.

	1-step	3-step	18-step
Input naïve	0.1335	0.6690	0.3555
Input proposed	0.0766	0.1726	0.2231
Output naive	0.0118	0.0722	0.1608
Output proposed	0.0069	0.0295	0.1210

(the variance of test dataset 3 is 37.62)

The predictor works better than naïve predictor and the predictor in Part A.

Part C: Kalman filter

1. State Variables

As we want to use the result from Part B, we build the state with all the non-zero parameters in the ARMAX model transferred from Box-Jenkins model.

The state variable is

$[\boldsymbol{KA_1}\ \boldsymbol{KA_2}\ \boldsymbol{KA_3}\ \boldsymbol{KA_4}\ \boldsymbol{KB_1}\ \boldsymbol{KB_2}\ \boldsymbol{KB_3}\ \boldsymbol{KB_4}\ \boldsymbol{KB_5}\ \boldsymbol{KB_6}\ \boldsymbol{KB_7}\ \boldsymbol{KC_1}\ \boldsymbol{KC_2}\ \boldsymbol{KC_{24}}\ \boldsymbol{KC_{25}}\ \boldsymbol{KC_{26}}]$.

And the corresponding C_t is

$$[-y_{t-1}\ -y_{t-2}\ -y_{t-3}$$
$$-y_{t-4}\ u_t\ u_{t-1}\ u_{t-2}\ u_{t-3}\ u_{t-4}\ u_{t-5}\ u_{t-6}\ \hat{e}_{t-1}\ \hat{e}_{t-2}\ \hat{e}_{t-24}\ \hat{e}_{t-25}\ \hat{e}_{t-26}].$$

Where y_t is the differentiated Air temperature and u_t is the differentiated Net radiation.

While in the prediction part, we can only use the data up to time $t-k$, then for the y and u we can not use, we need to use the estimation instead, and for $\hat{e}$ from time $t-k+1$, we need to use 0 instead.

For the estimated u, we can form a matrix where the ith row and jth col means i-step prediction for the u_t. This can be done with both Kalman filter or polynomial division method, we choose the polynomial method.

2. Initial State and Parameters

We use the model from Part B to form the initial state by convolution,

$KA = conv(D, F), \quad KB = conv(D, B), \quad KC = conv(F, C).$

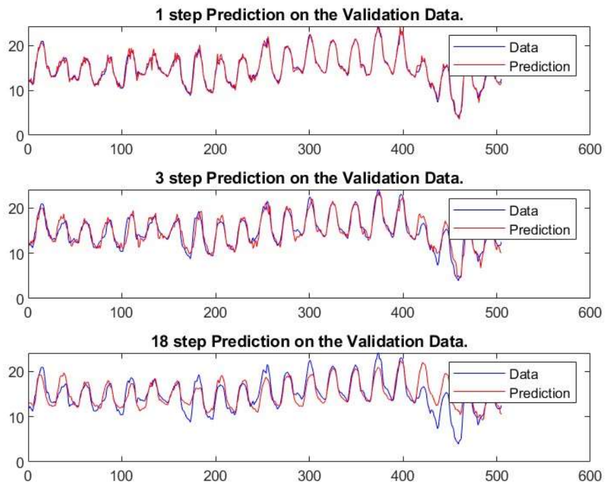
The notation is the same as the matlab program.

As the Part B prediction is good, we can trust much on this, so we choose a small $R_e = 10^{-6}$ to keep it stable. After trying different R_w, we choose 200.

3. Validation

We put the estimation part into the loop, and start the loop at $t = 27$, and then we need to transform the prediction back to the original domain by adding y_{t-24}, as we only make up to 18-step prediction, y_{t-24} is always available.

As the estimated input we use is from polynomial division, it is the same as the Part B, so we will not show this again.



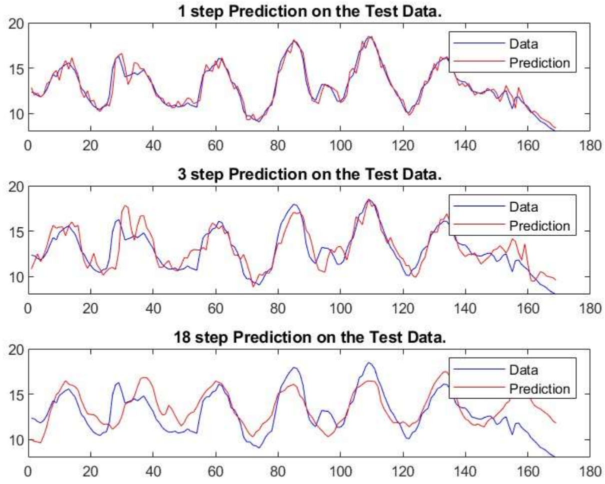
Also, we put the normalized variance in the following table.

	1 step	3 steps	18 steps
Naïve	0.0675	0.4611	0.3639
Kalman	0.0297	0.1064	0.3419

This predictor is better than the naïve predictor, but not significant better from BJ model. This may imply that the dataset is stable.

4. Prediction

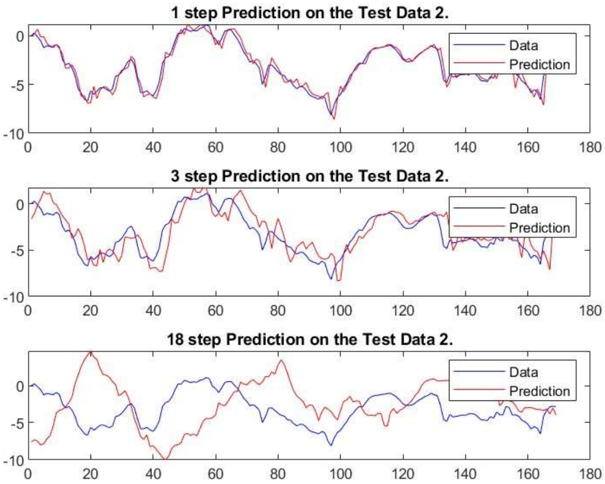
For test dataset 1.



Also, we put the normalized variance in the following table.

	1 step	3 steps	18 steps
Naïve	0.0958	0.5972	0.7697
Kalman	0.0643	0.3051	0.6311

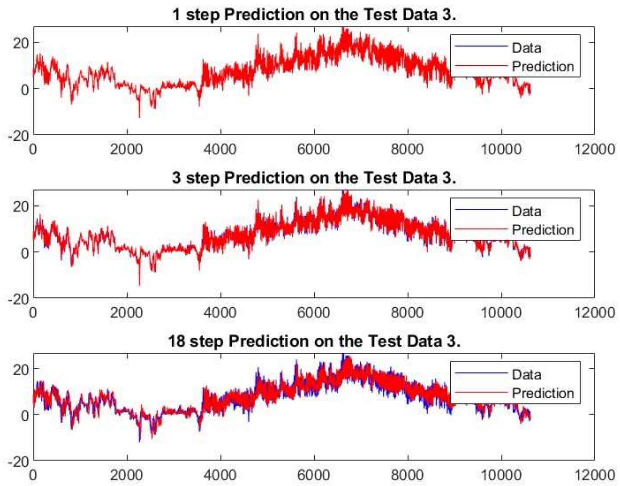
For test dataset 2.



Also, we put the normalized variance in the following table.

	1 step	3 steps	18 steps
Naïve	0.0763	0.3534	3.6562
Kalman	0.0776	0.4038	4.1008

For test dataset 3.



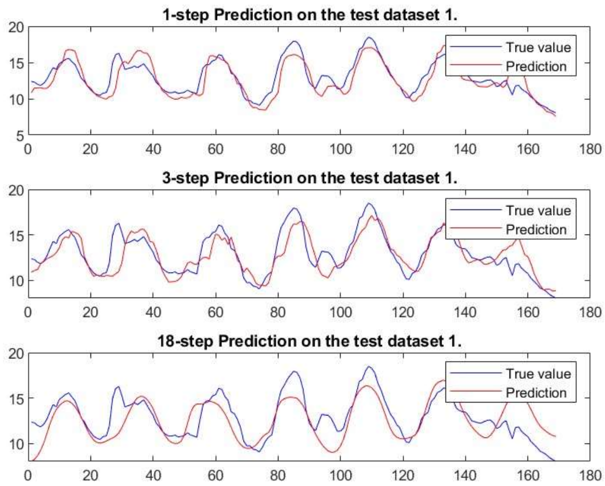
Also, we put the normalized variance in the following table.

	1 step	3 steps	18 steps
Naïve	0.0118	0.0722	0.1608
Kalman	0.0097	0.0432	0.1481

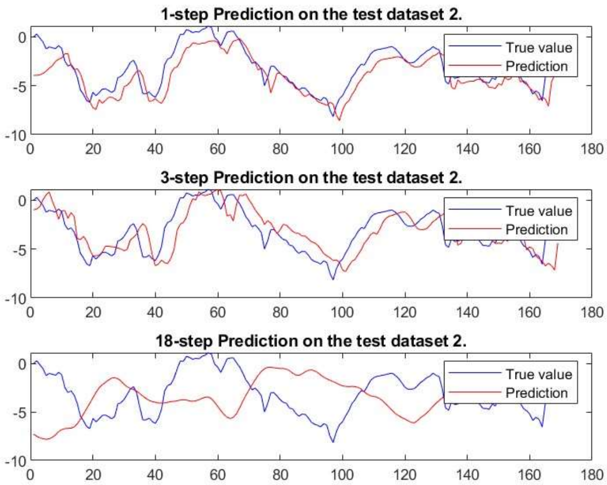
One thing surprising is, in the test dataset 2, Kalman filter even does worse, it is completely meaningless on this particular dataset. So is it because this dataset is really impossible to predict?

Part D: Transformer

Transformer is a sequence to sequence deep learning model, we keep the training data of this predictor is the same as the Kalman filter. We can see the short-term prediction is not as precise as the other model, the reason might be in this structure, the model cannot get the distance from the output and input, only correspondence. But for long term prediction, this works almost as good as the Kalman filter. Besides, in the test dataset 2, this model works a little better than the naïve predictor but still not ideally. This might imply that there was something unusual happening, and the pattern changed, the period became weak and the correspondence changed. So all the predictors go wrong.

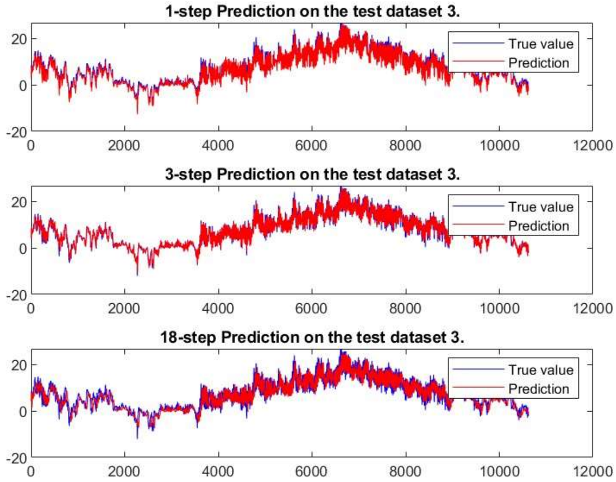


	1 step	3 steps	18 steps
Naïve	0.0958	0.5972	0.7697
Transformer	0.3615	0.3663	0.6902



	1 step	3 steps	18 steps
Naïve	0.0763	0.3534	3.6562
Transformer	0.2683	0.5472	2.4921

For test dataset 3.



	1 step	3 steps	18 steps
Naïve	0.0118	0.0722	0.1608
Transformer	0.0408	0.0558	0.1408

Conclusion

This dataset does not change much, but there are some part where the change happens quickly and change back to the pattern before quickly, which cause a lot of trouble to the recursive algorithm. So in this dataset, the BJ model perform best and then the ARMA model and Kalman filter. The deep learning based transformer performs worst, but it has some advantages such as most part of this predictor can be automatically done by the program itself. Also, the other three models are essentially linear models, the transformer can build some non-linear models.