RAINFALL RUNOFF MODELLING USING ARTIFICIAL NEURAL NETWORK

ZHU MORAN

DEPARTMENT OF CIVIL & ENVIRONMENTAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2018/2019

RAINFALL RUNOFF MODELLING
USING ARTIFICIAL NEURAL NETWORK

ZHU MORAN

A THESIS SUBMITTED

FOR THE DEGREE OF BACHELOR OF ENGINEERING

DEPARTMENT OF CIVIL & ENVIRONMENTAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

Acknowledgement

I would like to express my sincere gratitude to my supervisor Associate Professor Vladan Babovic for for his patience, enthusiasm, insightful comments, invaluable suggestions, helpful information and practical advice which have helped me tremendously during the course of this project. I benefitted extensively from our discussions about this project in using Artificial Neural Network.

I also wish to express my thanks to Desmoitier Nicolas Louis Roger for his assistance in lab sessions.

I would like to thank my parents and friends. Their unwavering support and encouragement are my source of strength.

Contents

Summary

Artificial Neural Network (ANN) is a complex mathematical model which is capable in predicting relationships between input and output data. The rainfall runoff of Kent Ridge Catchment is predicted using Artificial Neural Network as a function of lead time. The accuracy of prediction is investigated using testing data sets. Several data pre-processing techniques are implemented on training and cross-validation data sets. The techniques include taking difference, Simple Moving Average and Box Cox Transformation. The prediction results are compared against naïve forecast. Sophisticated methods provide better results in various length of lead time. Different methods have different advantages and disadvantages. Satisfactory results show that ANN could be used as an effective approach in predicting rainfall-runoff relationship for catchment areas.

List of Figures

List of Tables

1. Introduction

1.1 Project Background

In order to derive at a more accurate description of rainfall runoff processes in Singapore, a monitoring programme at pilot catchment (Kent Ridge Catchment at campus of National University of Singapore) was established to collect dense hydrological data over a representative area. The monitoring program collects hydrological data during entire year. The data are used to provide background information for development of an accurate hydrological distributed model for the pilot catchment.

Sensors including level meters and velocity meters were implemented in main drains and channels in the catchment area. Fig 0 is the map of the catchment area with sensor locations.

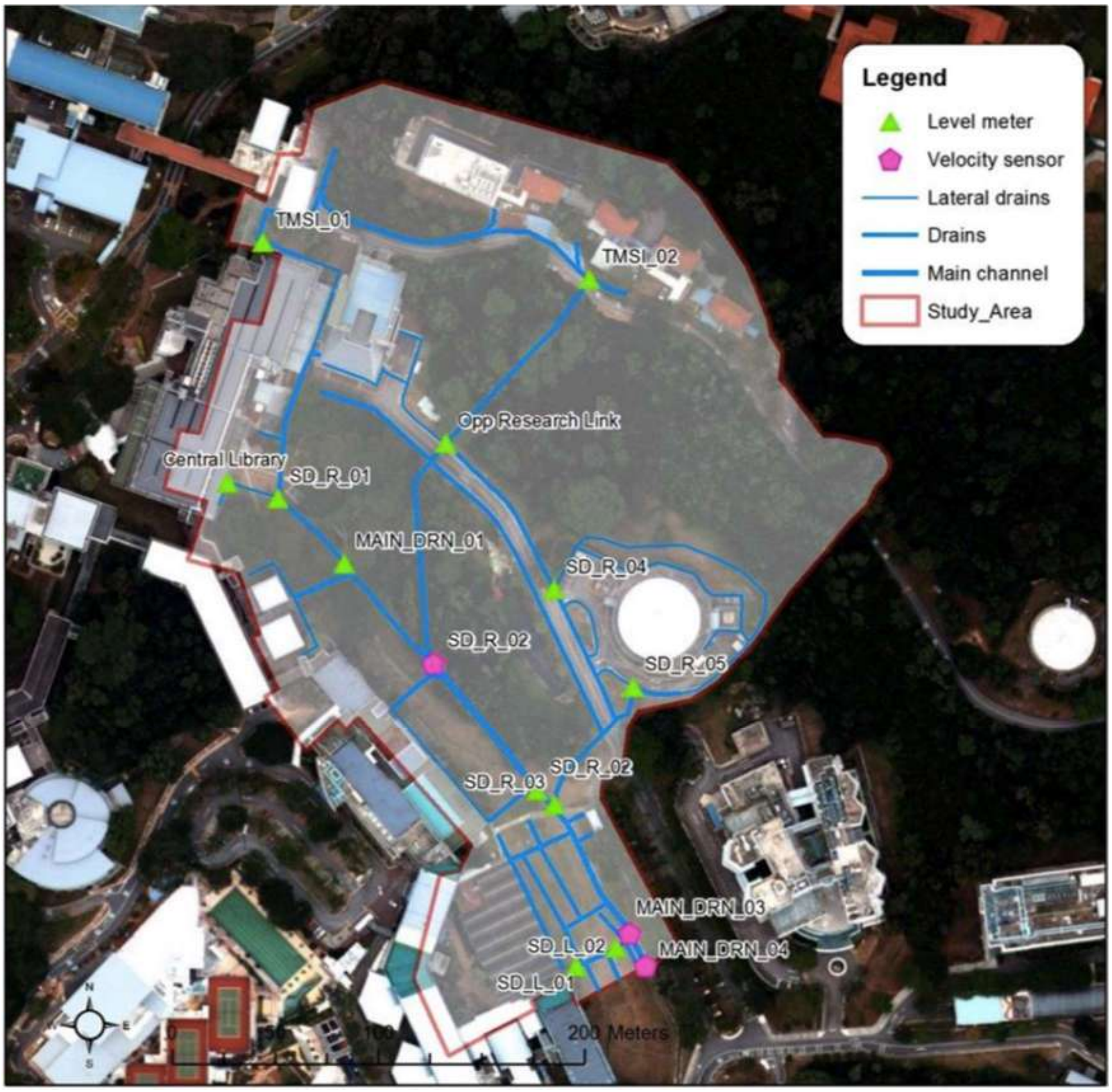


Fig 0. Map of Kent Ridge Catchment with sensor locations

A sample of data consisting of 41 events collected throughout year 2011 is used for this project. The data consists of time series at resolution of 1 minute of the following quantities.

- Catchment averaged rainfall intensities. (unit: millimetre)
- Catchment cumulative rainfall amount. (unit: millimetre)
- Duration of individual rainfall event. (unit: minute)
- Time series of discharges at Central Library. (unit: litre per second l/s)
- Time series of discharges at OPPRLink. (unit: litre per second l/s)
- Time series of discharges at Main_Drain_01. (unit: litre per second l/s)
- Time series of discharges at Main_Drain_02. (unit: litre per second l/s)
- Time series of discharges at Main_DRN_04. (unit: litre per second l/s)

The main objective of the project is to use suitable Artificial Neural Network (ANN) architecture to produce 10, 20 and 60 minutes forecast of flow rates at location Main Drain_04.

1.2 Artificial Neural Network

Learning from the past to understand the present and predict the future has been the motivation for researchers in various academic fields. Many mathematical models were created and refined over the years to find the pattern or correlation of data sets. However, most of the modelling methods suffer from imprecision and uncertainty which lead to unsatisfactory prediction of results.

Artificial Neural Networks (ANN) become popular since the 1980s. ANNs have the ability to solve complicated problems like pattern recognition and nonlinear modelling. It could be a more accurate prediction tool if carefully trained.

Structure of ANNs

ANNs process information through nodes and links. The link connecting nodes in each layer has weights which represent the connection strength. Mathematically speaking, the weight acts as the coefficient of functions in calculation of the output values. The activation functions are often sigmoid functions. Sigmoid functions are easy to use due to its simple derivative equation. It could also model non-linear process.

Training

Training, also referred to as learning, is the process of obtaining reasonable weights and bias factors for the specific ANN using experiment data. Most of the hydrologic applications use supervised learning, a process where the external operator adjusts the weights and bias for nodes of ANN based on the difference between output data and actual data, Overfitting, or over-training refers to the situation where the ANN is only tuned to the training data sets. The operator should also understand when to terminate training.

Issues in Modelling

Although ANNs are powerful tools in modelling complex problems, it would still be beneficial for the users to consider several issues before training the ANNs. First of all, selecting appropriate variables is crucial in modelling ANNs. There could be many different variables available such as temperature, water level, evaporation rate, tidal flows and so on. Instead of blindly including all related variables and create a large network, the user could study the problem and find the most relevant variables. Not only to produce a succinct network, but also saves time for training by reducing the number of nodes and links. At the same time, the output variables should also be carefully checked to find out probable problems with the existing ANN.

Also, the pre-processing of data plays a significant role in creating a good ANN. The reliability of relevant data should be checked before inputting into ANN, especially if the data is obtained remotely or by external parties. Sieving out anomaly could greatly improve the speed of training. In common practices, the first-handed data needs to be normalised or de-trended before feeding to ANN. The suitable and reasonable method to pre-process data is important. Inappropriate ways to transform data could result in unnecessary difficulties during the training step.

1.3 Project Tools

The main tools used for this project including Microsoft Excel and NeuroSolutions 7.

NeuroSolutions 7 is a neural network development environment developed by NeuroDimension. It combines a modular, icon-based (component-based) network design interface with an implementation of advanced learning procedures, such as conjugate gradients, Levenberg-Marquardt and backpropagation through time. The software is used to design, train and deploy neural network (supervised learning and unsupervised learning) models to perform a wide variety of tasks such as data mining, classification, function approximation, multivariate regression and time-series prediction.

2. Selection of parameters

2.1 Selection of input and output data

Artificial Neural Network relies on well-processed input data sets to provide decent output results. Therefore, the input and output data sets need to be chosen carefully.

A correlation analysis was performed to determine which data sets to be used as input for the ANN. Correlation coefficients are numerical factors used to describe the strength of the relationship between two variables. The type of correlation coefficient used here is Pearson correlation. The value measures the strength of linear relationship between two variables. For example, a value of 1.0 indicates that there exists a perfect positive linear relationship and a value of -1.0 indicates that the relationship is perfectly negative. Since rainfall is the independent factor among all data sets, the correlation analysis is performed with respect of rainfall. Table 1 is the computed correlation coefficient of main data sets. As it can be seen from the table, the first four data sets show relatively strong linear relationship with rainfall.

	Correlation Coefficient					
	Q_MD01	Q_MD02	Q_CTRLIB	Q_OPPRL	Raincum	Duration
Rainfall	0.707215	0.795895	0.884045	0.750784	-0.12677	0.216596

Table 1. Correlation coefficient of main data sets

Thus, there will be the following five inputs:

1. Catchment averaged rainfall intensities. (unit: millimetre)
2. Discharges at Central Library. (unit: litre per second l/s)
3. Discharges at OPPRLink. (unit: litre per second l/s)
4. Discharges at Main_Drain_01. (unit: litre per second l/s)
5. Discharges at Main_Drain_02. (unit: litre per second l/s)

The output will be:

1. Discharges at Main_Drain_04. (unit: litre per second l/s)

2.2 Selection of testing, training and cross-validation data sets

The data time series of discharges at Main_Drain_04 is plotted and shown in Fig 1. The largest value of discharge, i.e., the most extreme case, occurs between event No. 8037 and No.8611. According to the principles in ANN training suggest by J.Zupan (1994), the training data should include the most extreme case. Therefore, the last 60% of data (approx. No.4001 to No.10000) will be used as training data sets. The first 20% (approx. No.1 to No.2000) will be used as testing data sets and the remaining 20% (approx. No.2001 to No.4000) will be used as cross-validation data sets.

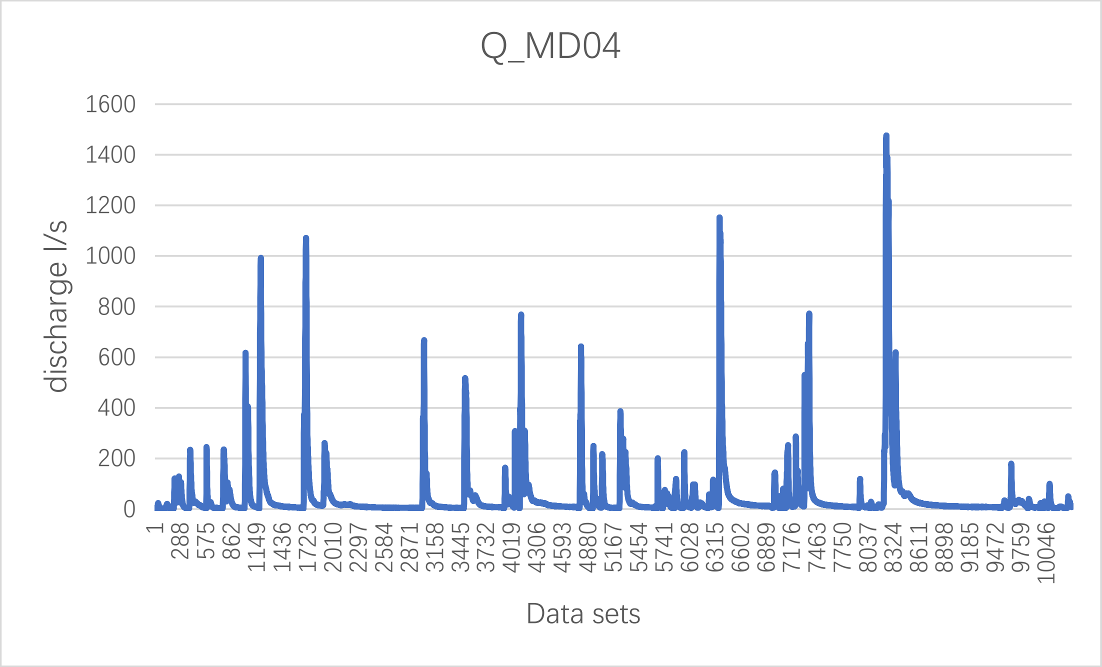


Fig 1. Discharge distribution of MD04

2.3 Selection of ANN structure

Time-Delay Neural Network (TDNN) would be the most appropriate network structure for this project. Since the data obtained is in time sequence at resolution of 1 minute, TDNN will be able to recognize this time difference and select a certain window of events for training. Unlike traditional Multi-layer network (MLP) which considers all inputs at the same time, TDNN will not destroy the time series signal. To be specific, the option to be chosen in NeuroSolutions 7 is *Regression TDNN (Time-Dependent data)*.

3. Training methods.

3.1 Naïve forecast

Naïve forecasting refers to the forecast technique in which the last period's trends are used as the next period's forecast, without adjusting the data or trying to process the data in advance. It is commonly treated as the most cost-effective forecasting model or method because little effort needs to be made in order to produce a naïve forecast. This method could work quite well for data with strong patterns such as economic and financial time series.

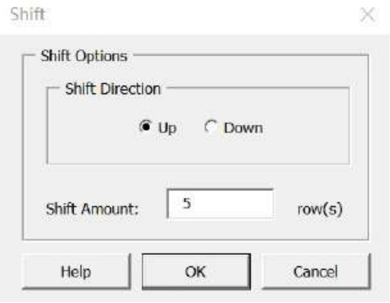
The first training method to be adopted in this project is naïve forecast. The results obtained from naïve forecast could provide a benchmark against which more sophisticated methods can be compared. It will also provide the opportunity to get familiar with the software in order to improve efficiency for future training attempts.

3.1.1 5-minute naïve forecast

Lead time could be a critical factor in affecting the quality of prediction. In this project, forecast results will be generated from short lead time to long lead time in order to provide analysis on quality of prediction against the function of lead time. The lead time for analysis will be 5 minutes, 10 minutes, 20 minutes, 30 minutes and 60 minutes.

The detailed training procedure for 5-minute naïve forecast is as follows:

1. Using the preprocess function in NeuroSolutions 7 to shift the column Q_MD04 up by 5 rows. The step is to produce data of lead time 5 minutes.


Shift

Shift Options

Shift Direction

● Up ○ Down

Shift Amount: 5 row(s)

Fig 2. Shift option in NeuroSolutions 7

2. Delete the last 5 rows for each individual event in Excel Spreadsheet. This step is to eliminate redundant data to achieve more accurate results. Taking event No.1 as an example, there are 99 data sets in the original spreadsheet. After shifting data up by 5 rows, there are only 94 data sets remain useful. In other words, there is no information about Q_MD04 of lead time 5 minutes for the last five minutes of event No.1.

3. Tag inputs and desired columns. Using the embedded NeuroSolutions in Excel, Tag rainfall, Q_MD01, Q_MD02, Q_Cntrlib and Q_Opprlink as Input. Tag Q_MD04 (shifted by 5 rows) as Desired.

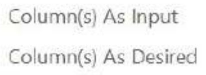


Fig 3. Tag columns in NeuroSolutions 7

4. Tag the first 2000 rows as Testing, the next 2000 rows as Cross-Validation and the rest rows as Training.

Row(s) As Training

Row(s) As Cross Validation

Row(s) As Testing

Fig 4. Tag rows in NeuroSolutions 7

5. Build Regression TDNN using NeuroSolutions 7.

Regression TDNN (Time-Dependent Data)

Fig 5. Build TDNN in NeuroSolutions 7

Fig 6 shows the visualisation of created neural network. Although the actual neural network structure could be more complicated, the software still provides a platform where users can make necessary changes to the default neural network structure.

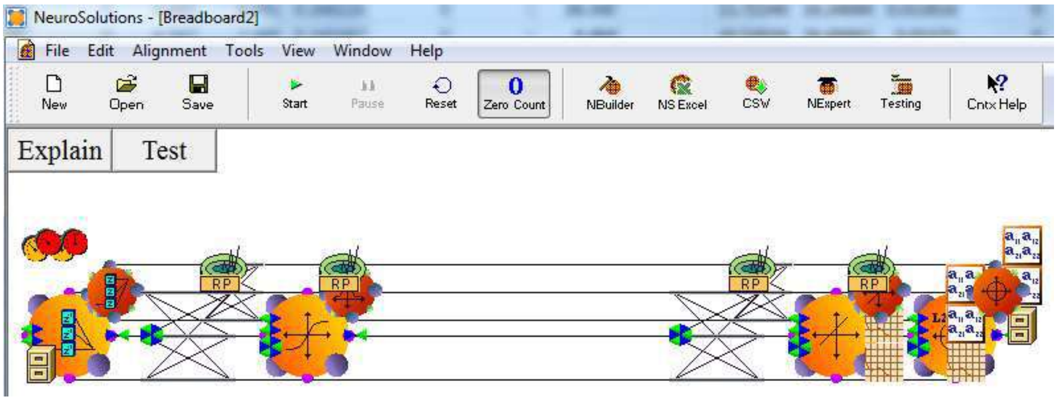


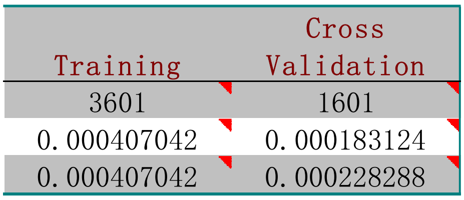
Fig 6. NeuroSolutions 7 user interface

6. Train TDNN.

Select the Train Network option and set the number of epochs to be 10,000. Set the terminating criteria at 2000 epochs without improvement. This training epoch number is more than sufficient and most of the time training would end at 3000 to 4000 epochs.

7. Check training and cross-validation results.

After training process is completed, NeuroSolutions 7 will generate a training result report including the minimum mean-square error and final mean-square error. Mean-square error (MSE) is an estimator measures the averaged square of error between estimated values and original values. In neural network training, the MSE measures the training out come for a neural network. In principle, a well-trained ANN should have a very low MSE at the end of training and cross-validation phase. For example, the final MSE for 5-minute naïve forecast training is 0.00040704, indicating that the ANN outputs and original data have very small difference. The MSE for cross-validation phase is also very small, which is desirable.

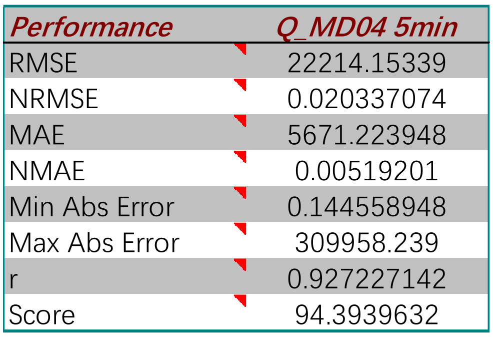


Training	Cross Validation
3601	1601
0.000407042	0.000183124
0.000407042	0.000228288

Fig 7. training reports for 5-minute naïve forecast

8. Test neural network

After ensuring the training and cross-validation phase is valid, next step will be testing the created neural network using the testing data sets. Select the Test Network option in NeuroSolutions to create testing result. The software will generate a testing report containing a graph comparing all original data against the neural network output. A table of various parameters will also be produced. Fig 8 shows a typical table in the testing report.



Performance	*Q_MD04 5min*
RMSE	22214.15339
NRMSE	0.020337074
MAE	5671.223948
NMAE	0.00519201
Min Abs Error	0.144558948
Max Abs Error	309958.239
r	0.927227142
Score	94.3939632

Fig 8. Testing report generated by NeuroSolutions 7

The parameters listed are:

- RMSE: Root-Mean-Squared Error

- NRMESE: Normalized Root-Mean-Square Error

- MAE: Mean Absolute Error
- NMAE: Normalized Mean Absolute Error
- Min. Abs. Error: Indicating the smallest error between two outputs
- Max. Abs. Error: Indicating the largest error between two outputs
- r: correlation coefficient between two outputs
- Score: An auto-generated score to evaluate the testing results.

9. Although the score provided by NeuroSolutions 7 could help to evaluate the performance, the more appropriate method would be comparing the final results of ANN output and the original data. Fig 9 shows the comparison of Q_MD04 5-minute naïve forecast with original data. As it can be seen from the graph, the predicted results are relatively accurate.

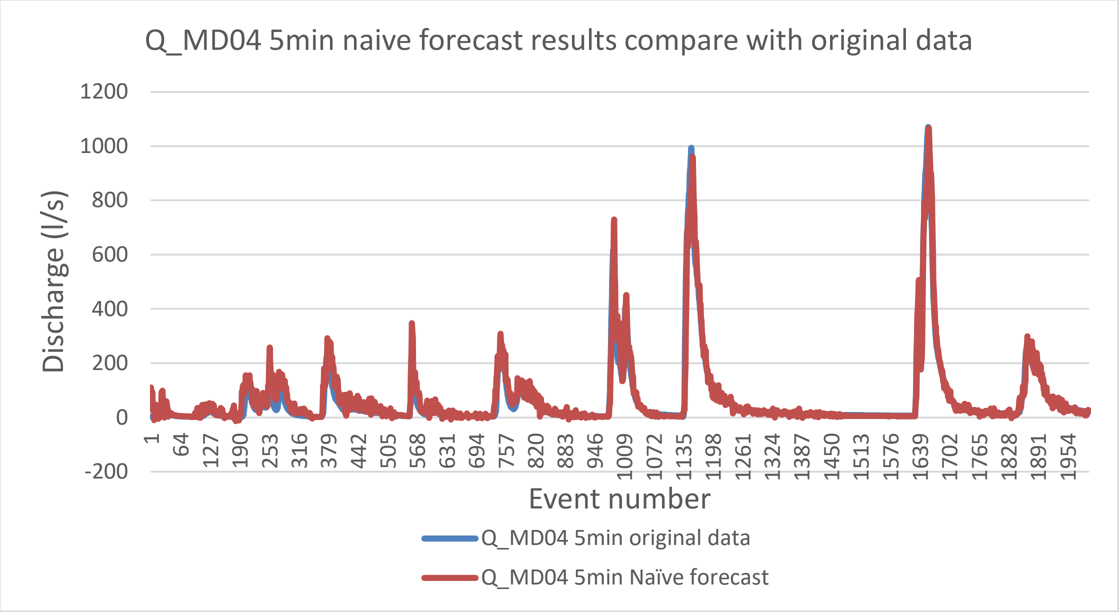


Fig 9. Q_MD04 5min naive forecast results compare with original data

3.1.2 10-minute naïve forecast

Following the similar procedures as above, the 10-minute naïve forecast is produced. Note that in this case, event no.22 is removed because it has only seven data sets. Fig 10 shows the comparison of Q_MD04 10-minute forecast with original data.

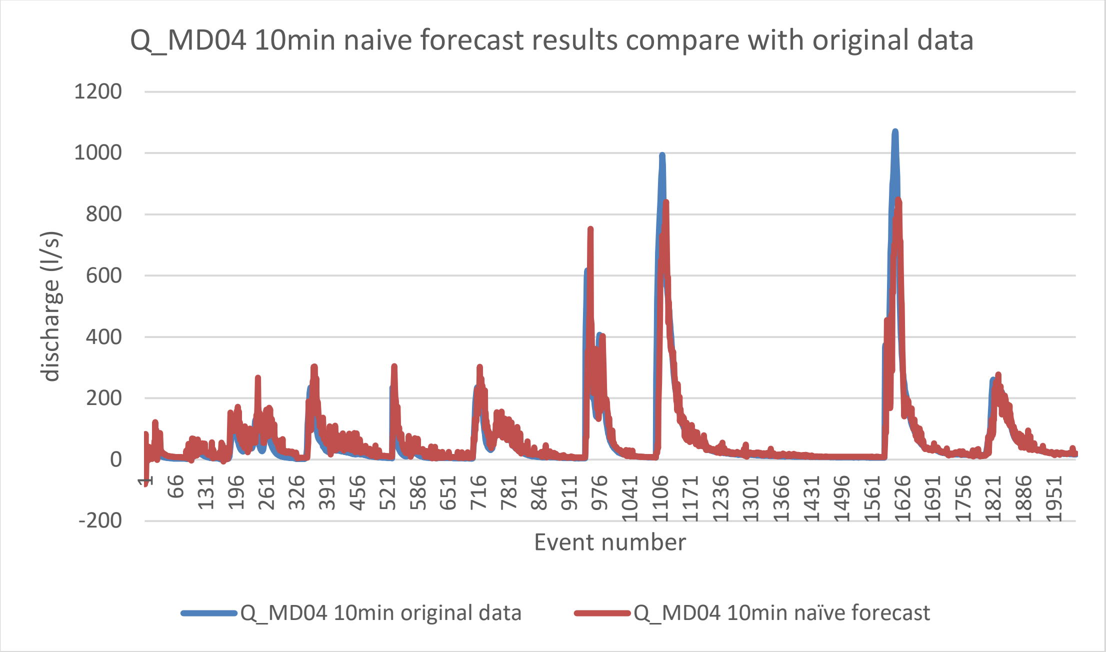


Fig 10. Q_MD04 10min naive forecast results compare with original data

3.1.3 20-minute naïve forecast

Following the similar procedures as above, the 20-minute naïve forecast is produced. Note that in this case, event no.22 is removed because it has fewer than 20 data sets. Fig 11 shows the comparison of Q_MD04 20-minute forecast with original data.

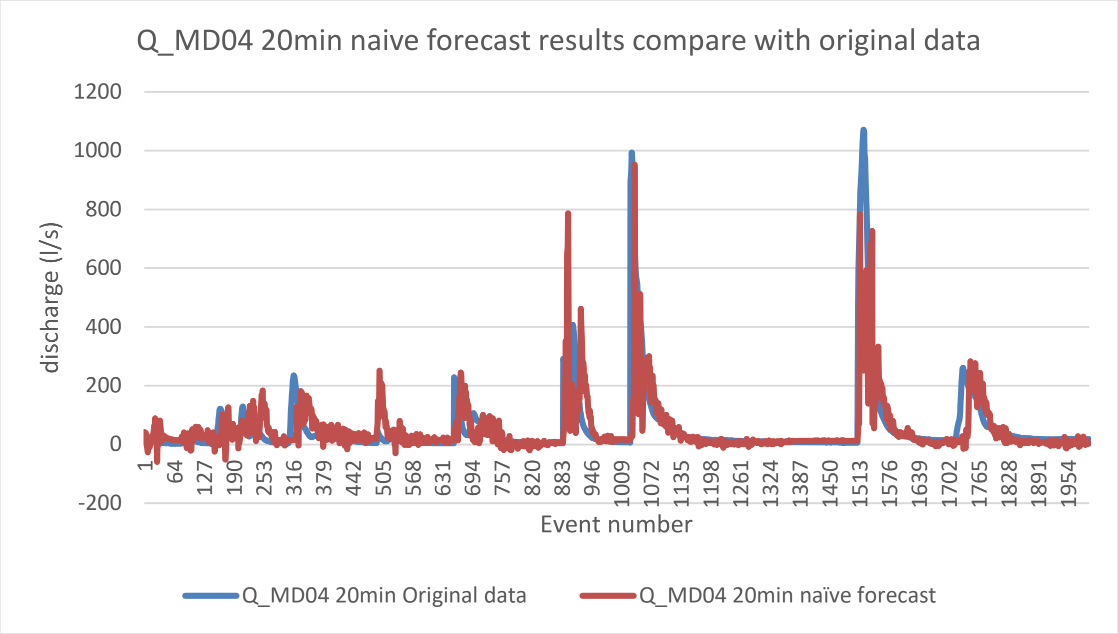


Fig 11. Q_MD04 20min naive forecast results compare with original data

3.1.4 30-minute naïve forecast

Following the similar procedures as above, the 30-minute naïve forecast is produced. Note that in this case, event no.22 and no.24 are removed because they have fewer than 30 data sets. Fig 12 shows the comparison of Q_MD04 30-minute forecast with original data.

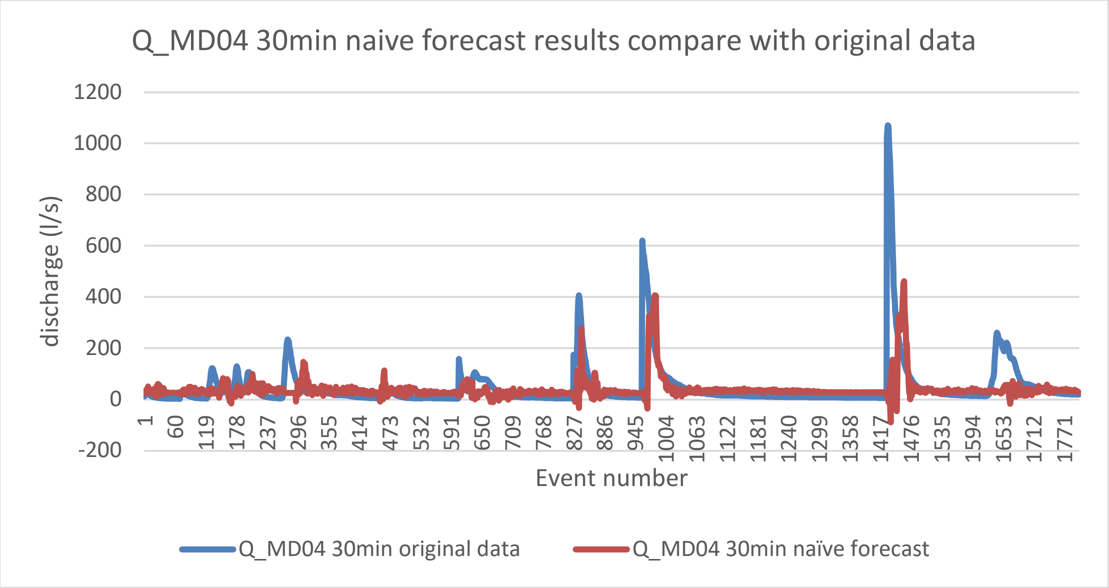


Fig 12. Q_MD04 30min naive forecast results compare with original data

3.1.5 60-minute naïve forecast

Following the similar procedures as above, the 60-minute naïve forecast is produced. Note that in this case, event no.20, no.22, no.24, no.36, no.64, no.65 and no.66 are removed because they have fewer than 60 data sets. Fig 13 shows the comparison of Q_MD04 60-minute forecast with original data.

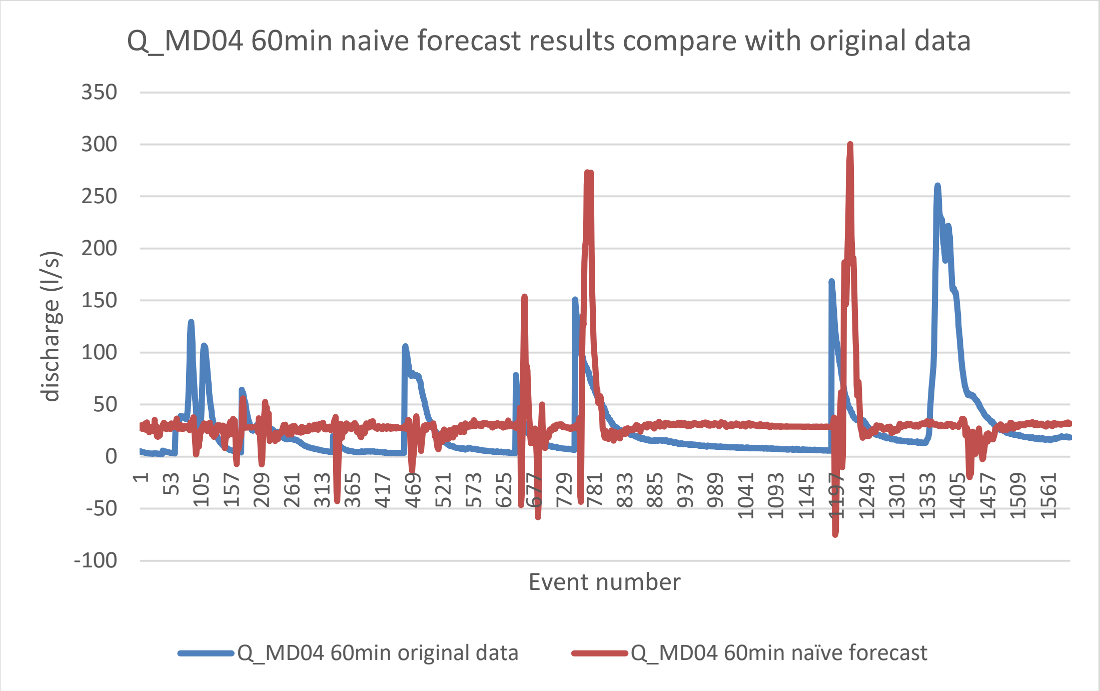


Fig 13. Q_MD04 60min naive forecast results compare with original data

3.2 Predicting the difference of flow

Although the naïve forecast could give relatively good results for short lead time, it can be anticipated that the accuracy of forecast would decrease by a large extent for long lead time. Measures should be taken to increase the accuracy for long time prediction. Additionally, even if naïve forecast could produce good results, naïve forecast actually does not have any forecast ability. This part would be further discussed in the results and discussion section. Thus, more sophisticated methods should be implemented to produce useful forecasts.

Since the rainfall amount for each minute is given, the difference in water flow in the drains will be more correlated to the rainfall. Therefore, instead of predicting the flow in MD04 directly, predicting the difference of flow could possibly achieve better results.

A relevant mathematical equation is:

$$Q(t+n) = Qt + \widehat{dQ} \tag{3.2.1}$$

Assuming the available data is flow rate at t=0 and the aim is to predict the flow at 5 minutes later. If the difference of flow between t=5 and t=0 can be obtained (dQ), the flow at t=5 will be calculated directly.

3.2.1 5-minute dQ forecast

The detailed training procedures for 5-minute dQ forecast will be explained in this section.

1. Calculating the difference of flow for each individual drain.

   Using the flow at 5 minutes later to minus the flow at the current point of time. Similar to 5-minute naïve forecast, there will be redundant data for each event after this step.

2. Shift the computed dQ_Q_MD04 up by 5 rows. This step is crucial as it produces the data for lead time of 5 minutes.

   Delete the last 10 rows of each individual event. After data processing in step 1 and 2, the last

10 data sets of each event become unneeded. Taking the set of data at 12:05 as an example:

DateTime	Rainffall	Q_MD01	Q_MD02	Q_CNTRLIB	Q_OPPRLINK		Q_MD04
19/9/2011 12:00	0.05	15.493	37.756	1.97181075	5.93246513		59.657
19/9/2011 12:01	0.05	14.317	32.401	1.37095584	5.12466441		53.801
19/9/2011 12:02	0.05	12.602	29.882	1.37095584	4.39510829		48.399
19/9/2011 12:03	0.05	10.813	25.549	1.49626639	3.78154021		44.345
19/9/2011 12:04	0	9.679	23.351	2.06711797	3.24348804		39.053
19/9/2011 12:05	0.2	8.547	21.702	2.57458706	2.74316506		36.336
19/9/2011 12:06	0	7.953	25.277	2.41539755	2.44510035		33.282
19/9/2011 12:07	0.15	7.819	23.623	2.41539755	2.24529806		30.57
19/9/2011 12:08	0.1	7.684	21.152	2.19342465	2.08149251		28.872
19/9/2011 12:09	0.05	7.415	21.974	2.44739755	1.99095365		27.516
19/9/2011 12:10	0.1	7.953	20.875	2.06711797	2.24529806		26.502

Table 2. 11 consecutive data sets in event no.1

DateTime	dQ_Q_MD01	dQ_Q_MD02	dQ_Q_CNTRLIB	dQ_Q_OPPRLINK		dQ_Q_MD04
19/9/2011 12:05	-6.946	-16.054	0.602776312	-3.189300068		-9.834

Table 3. Training data at 12:05

Table 3 shows the data to be used in neural network training. The first 4 inputs are differences of flow between 12:05 and 12:00, the last column is the difference of flow in MD04 between 12:10 and 12:05.

3. The rest of the steps are similar with the procedure stated in section 3.1.1.

4. Results of 5-minute dQ forecast

Adding the difference of flow in MD_04 produced by Artificial Neural Network back to the original data in order to check the quality of forecast. Fig 14 shows the comparison between ANN forecast and the original data.

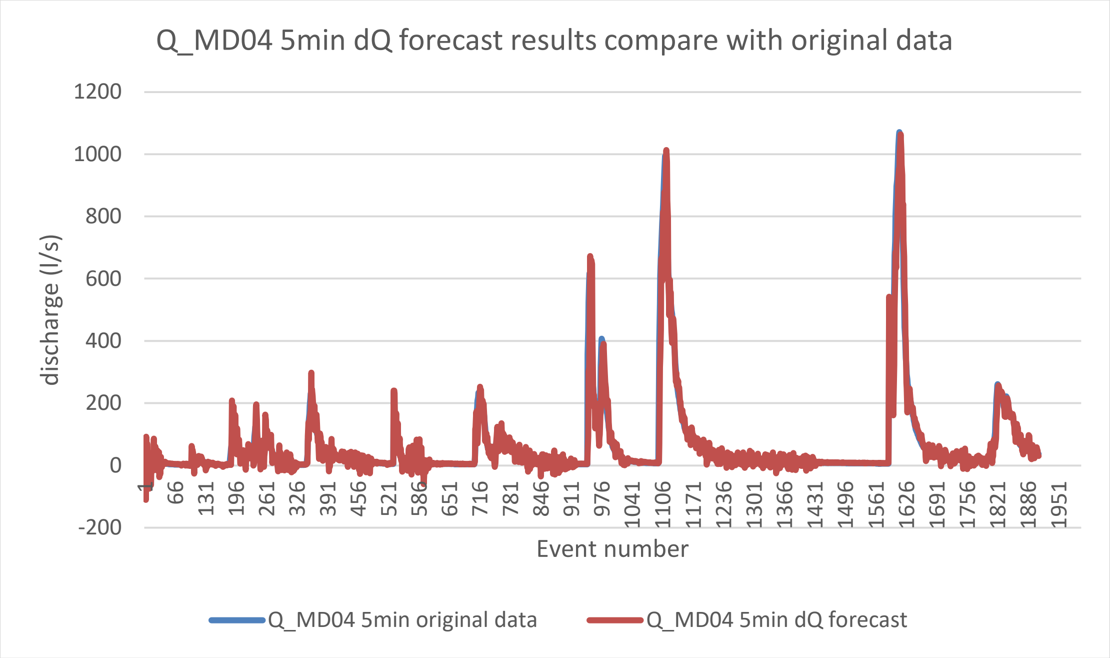


Fig 14. Q_MD04 5min dQ forecast results compare with original data

3.2.2 10-minute dQ forecast

Following the similar procedures as above, 10-minute dQ forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 15 shows the comparison between ANN forecast and the original data.

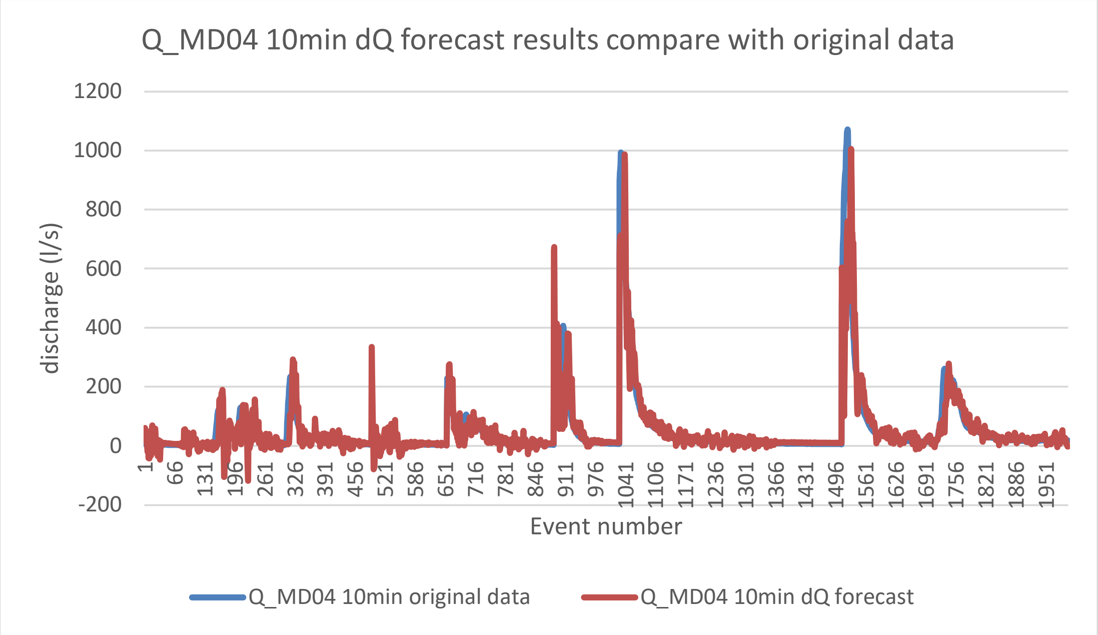


Fig 15. Q_MD04 10min dQ forecast results compare with original data

3.2.3 20-minute dQ forecast

Following the similar procedures as above, 20-minute dQ forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 16 shows the comparison between ANN forecast and the original data.

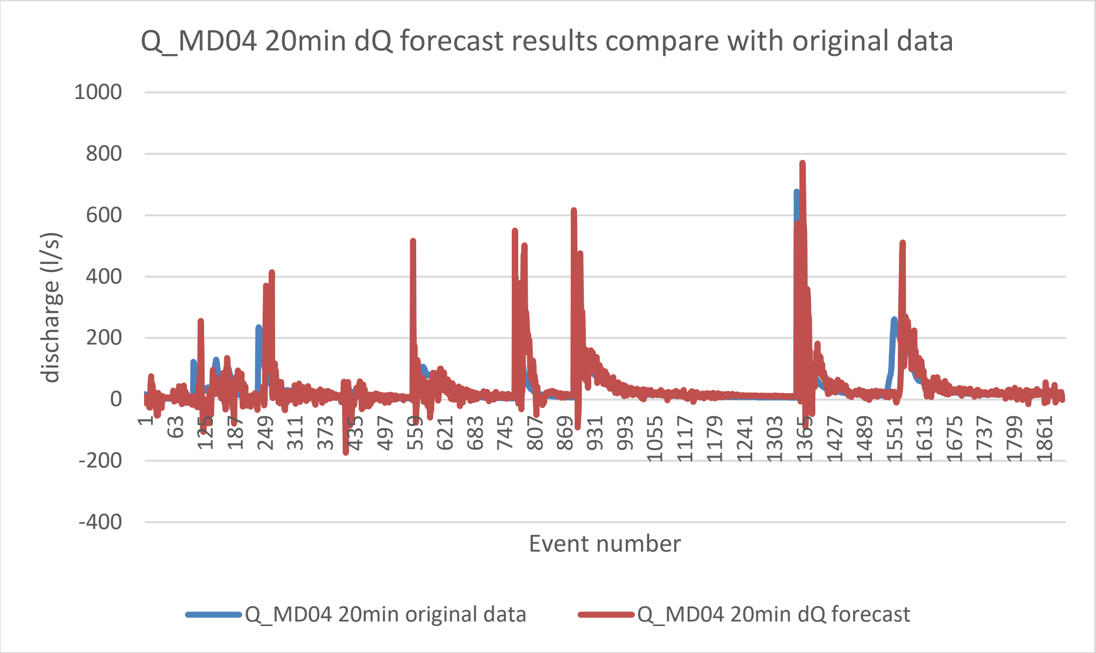


Fig 16. Q_MD04 20min dQ forecast results compare with original data

3.2.4 30-minute dQ forecast

Following the similar procedures as above, 30-minute dQ forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 17 shows the comparison between ANN forecast and the original data.

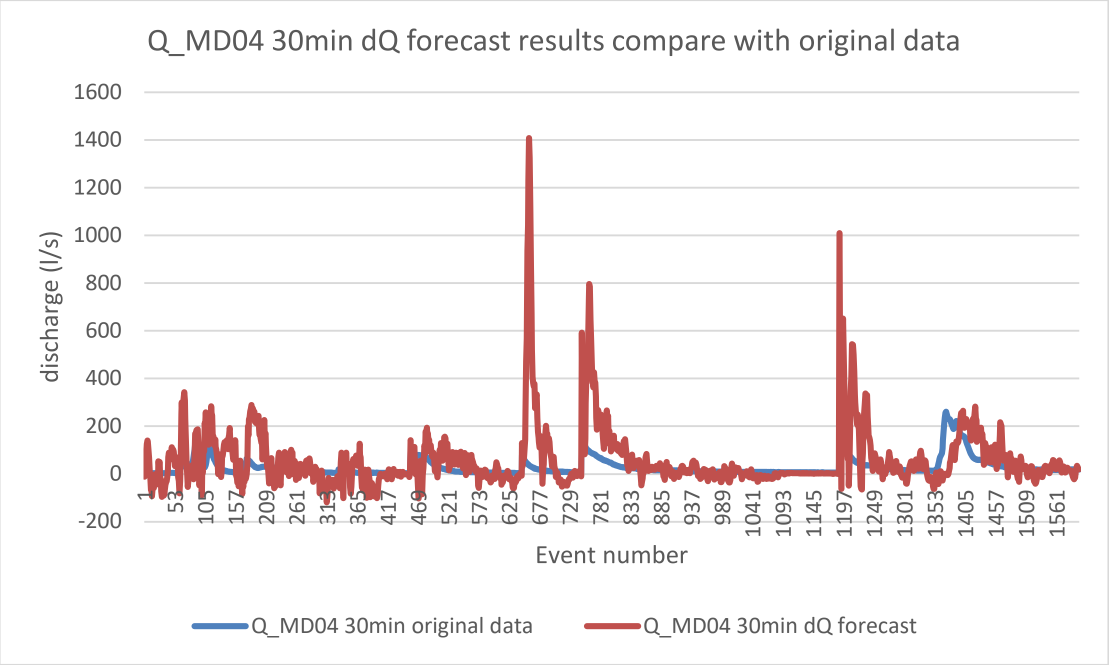


Fig 17. Q_MD04 30min dQ forecast results compare with original data

3.2.5 60-minute dQ forecast

Following the similar procedures as above, 60-minute dQ forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 18 shows the comparison between ANN forecast and the original data.

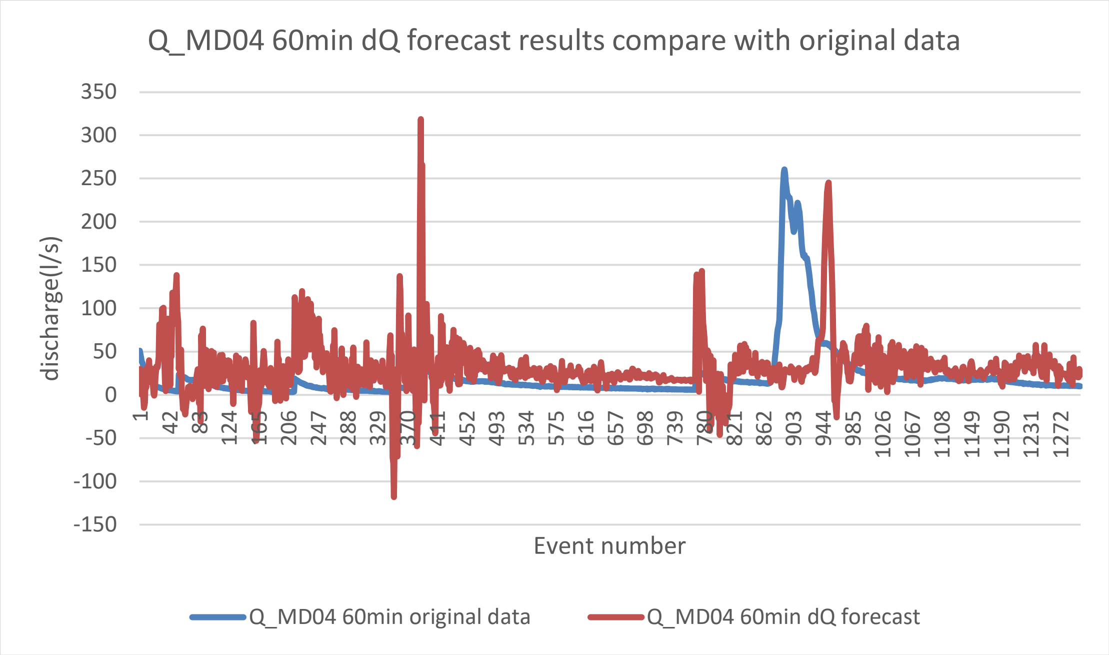


Fig 18. Q_MD04 60min dQ forecast results compare with original data

3.3 Simple Moving Average

The concept of moving average (rolling average or running average) is a calculation to analyse data by computing a series of averaged data sets from the original data sets. A moving average is usually used in time-series data to smooth fluctuations in the short term and accentuate general trends in the long term. In other words, it could help to filter out the “noise” from a large amount of data. There are several variations of moving average including simple, cumulative, exponential and weighted.

The type of moving average used in this project is simple moving average. It is an arithmetic moving average calculated by adding value of adjacent data points then dividing that by the total number of data points. The formula for SMA is:

$$\mathrm{SMA} = \frac{A_1+A_2+A_3+\cdots+A_n}{n} \tag{3.3.1}$$

Where

$A_n =$ the data at time n

$n =$ the number of total data points

One major factor that must be taken into consideration is the lag resulting from computing simple moving average. Simple moving average lag the actual data change because it is based on past data sets. The more data points to use, the greater the lag. Fig 19 shows a 16-week simple moving average throughout the period of 4 years on a products’ sales data. As it can be seen from the graph, there exists a significant lag after calculating the simple moving average despite the fact that simple moving average does reflect the general trend of the data. Therefore, the lag of simple moving average needs to be carefully treated and dealt with to minimize its effect in the prediction.

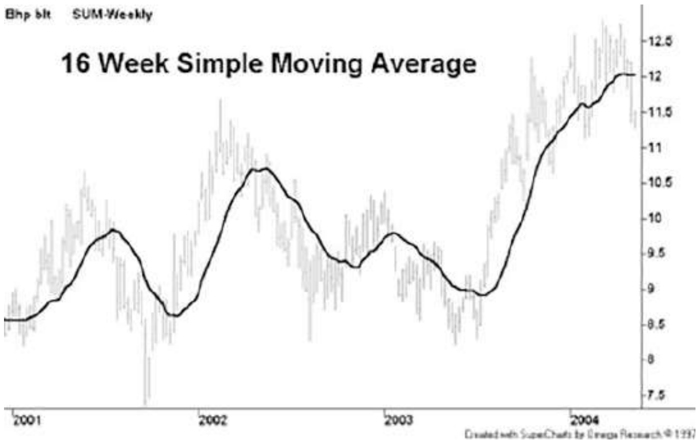
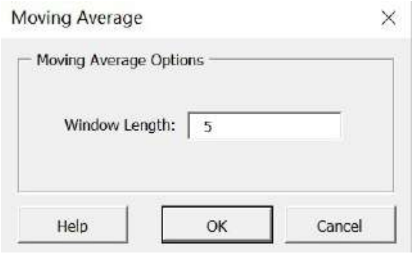


Fig 19. Example of lag in SMA

Moreover, dQ will still be the aim for prediction after applying simple moving average. This series of prediction is represented as dQ+SMA prediction.

3.3.1 dQ+SMA 5-minute forecast

1. Using the preprocess function in NeuroSolutions 7, compute the moving average for each input. Set the window length to be 5, which means that the generated data will be the average of 5 adjacent original data points.



Window Length: 5

Fig 20. Simple Moving Average option in NeuroSolutions 7

2. Shift the moving average data up by 3 rows.

   This step attempts to eliminate the lag resulting from taking moving average. Fig 21 shows the comparison between original data and moving average. The first 200 data points of Q_MD01 are compared. It can be seen that there exists a noticeable lag, especially for the peak values. Fig 22 shows the comparison between original data and shifted moving average. The lag now has been reduced. The peak values for both data sets are aligned.

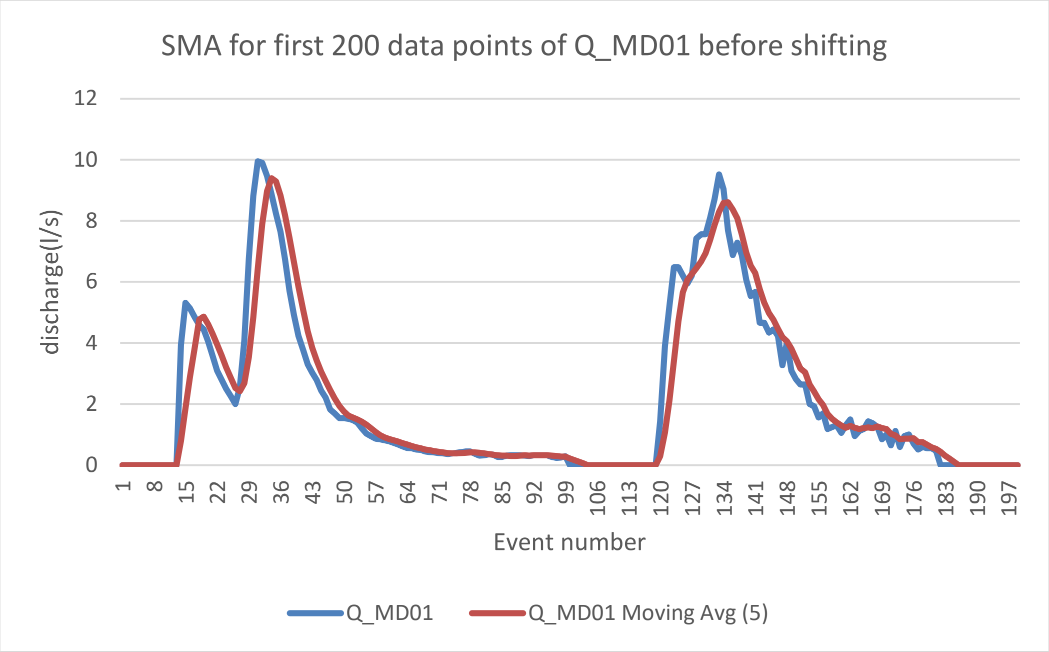


Fig 21. SMA for first 200 data points of Q_MD01 before shifting

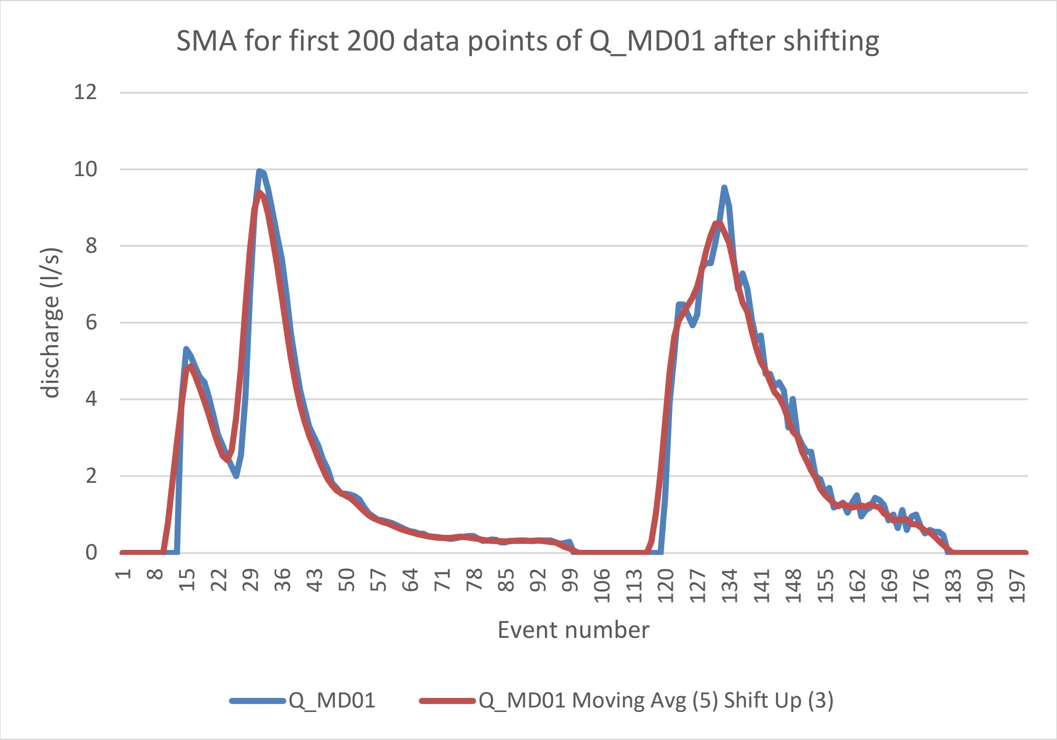


Fig 22. SMA for first 200 data points of Q_MD01 after shifting

3. Calculating the flow difference for each drain.

   The calculations are performed based on the moving average.

4. The rest of the steps are similar with procedures stated in section 3.2.1.

5. Results of dQ+SMA 5-minute forecast

   The forecast results produced by Artificial Neural Network is added back to the original data for comparison. Fig 23 shows the comparison of ANN forecast and the original data.

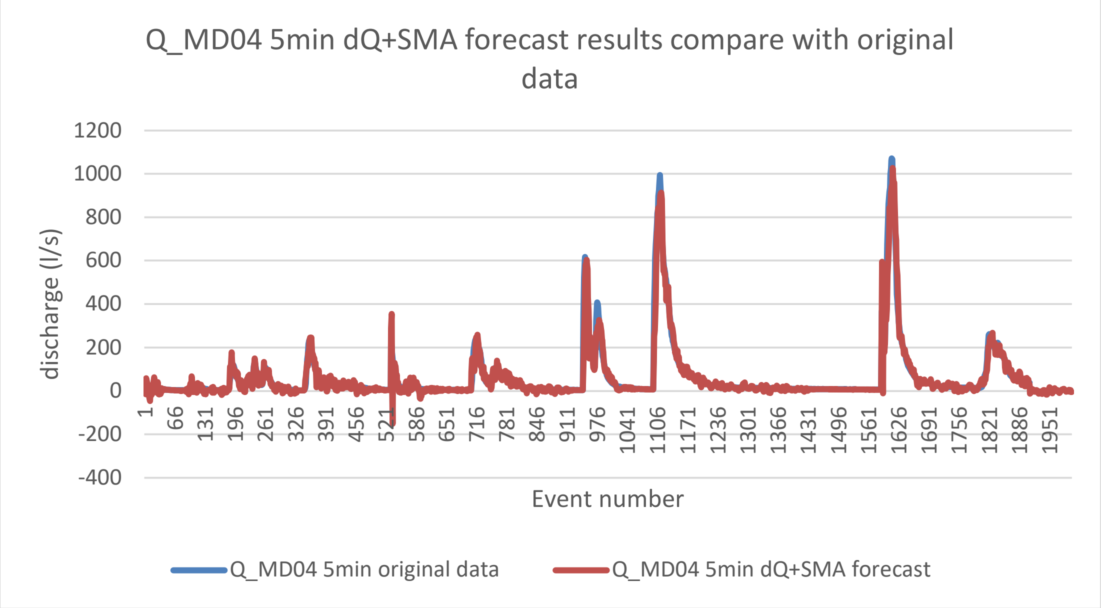


Fig 23. Q_MD04 5min dQ+SMA forecast results compare with original data

3.3.2 dQ+SMA 10-minute forecast

Following the similar procedures as above, 10-minute dQ+SMA forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 24 shows the comparison between ANN forecast and the original data.

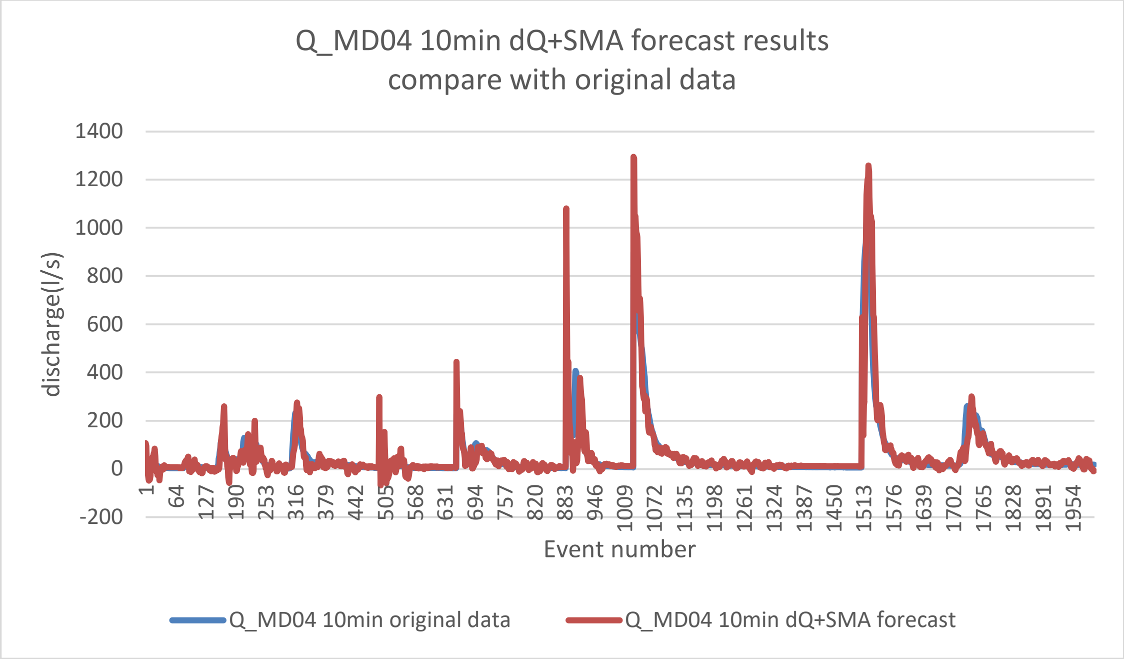


Fig 24. Q_MD04 10min dQ+SMA forecast results compare with original data

3.3.3 dQ+SMA 20-minute forecast

Following the similar procedures as above, 20-minute dQ+SMA forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 25 shows the comparison between ANN forecast and the original data.

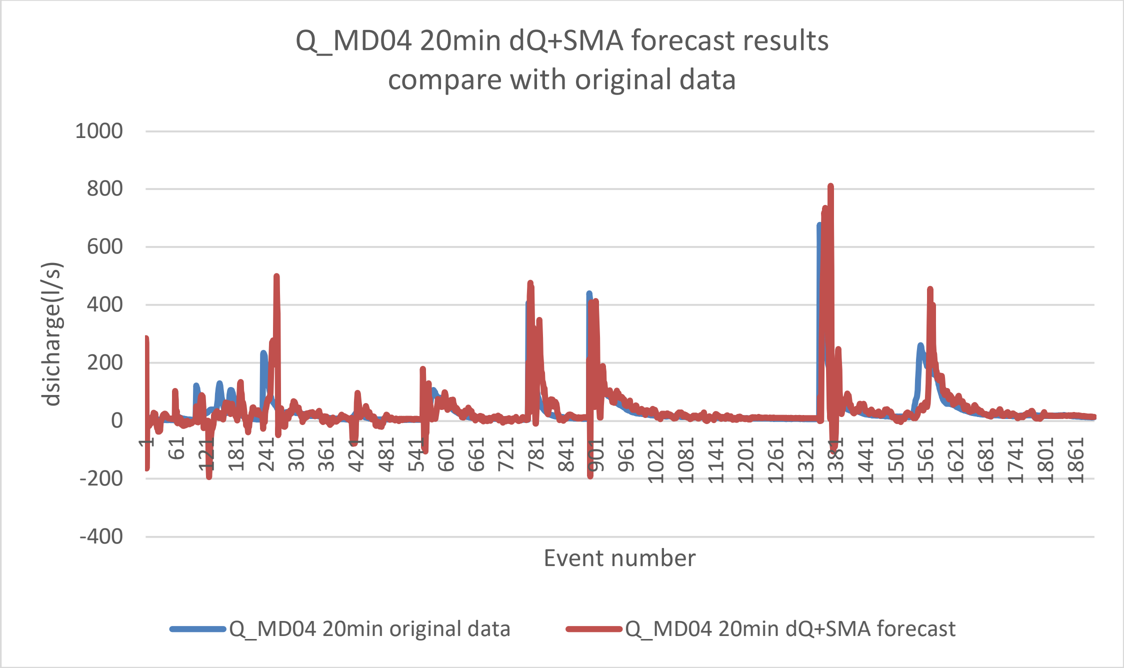


Fig 25. Q_MD04 20min dQ+SMA forecast results compare with original data

3.3.4 dQ+SMA 30-minute forecast

Following the similar procedures as above, 30-minute dQ+SMA forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 26 shows the comparison between ANN forecast and the original data.

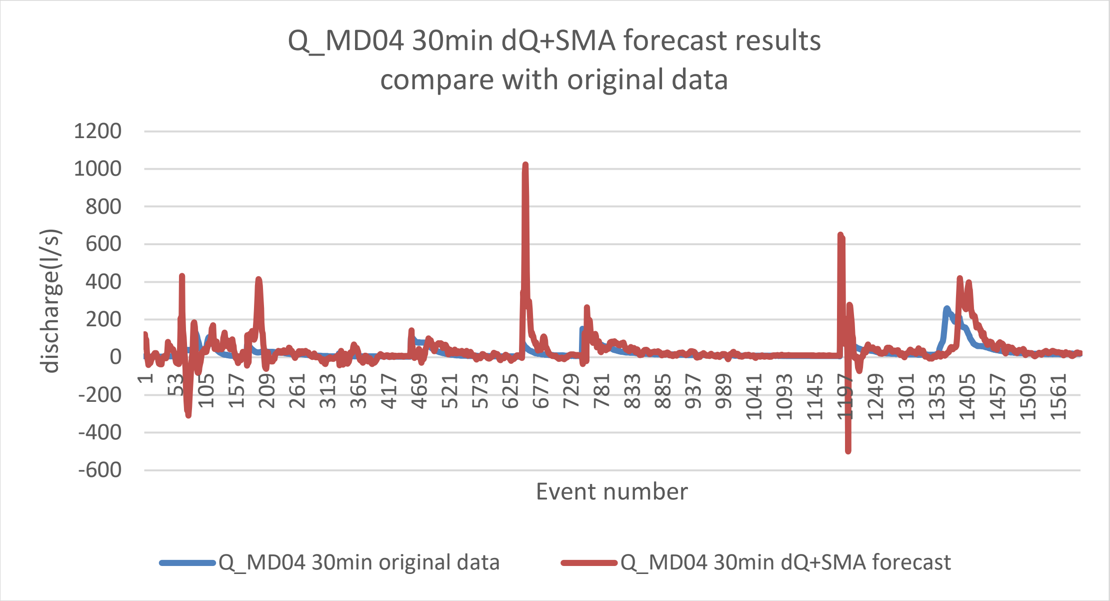


Fig 26. Q_MD04 30min dQ+SMA forecast results compare with original data

3.3.5 dQ+SMA 60-minute forecast

Following the similar procedures as above, 60-minute dQ+SMA forecast is produced. Events with insufficient numbers of events have been eliminated before computing the difference of flow. Fig 27 shows the comparison between ANN forecast and the original data.

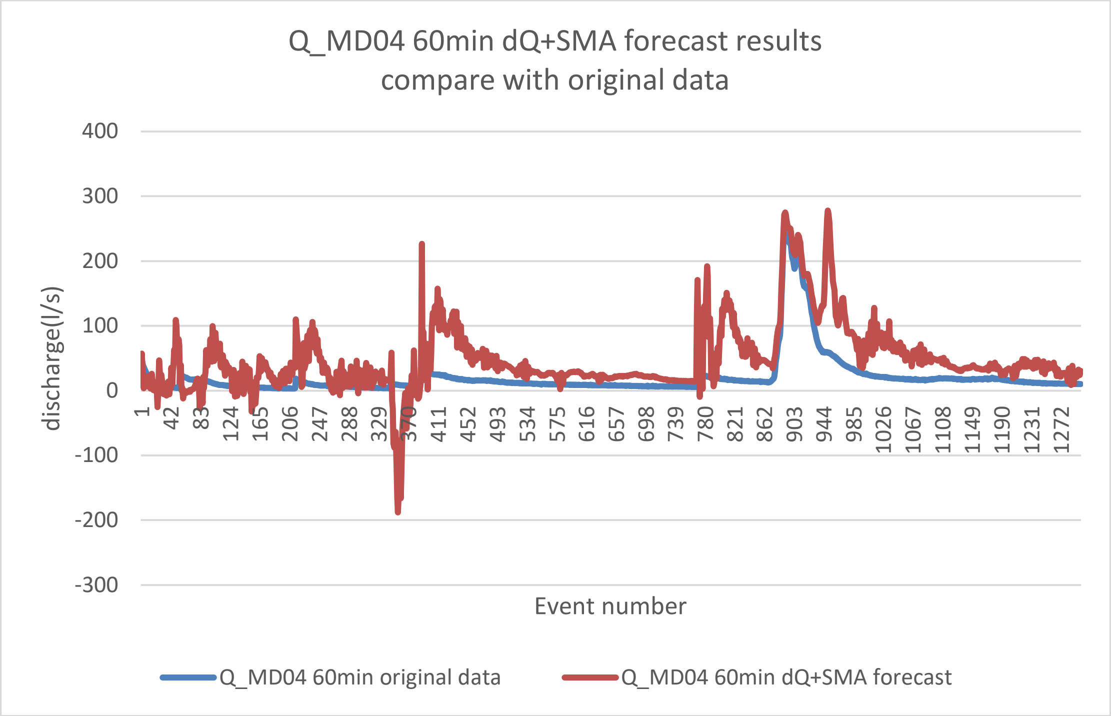


Fig 27. Q_MD04 60min dQ+SMA forecast results compare with original data

3.4 Box Cox Transformation

Artificial Neural Network responses well against data at a comparable range and data normalisation would improve the quality of forecast in training of ANN (J.Liu et al., 2003). The are many techniques available to normalise data such as transforming data using a z-score or t-score and rescaling data to have values between 0 and 1. The method used to normalise existing data in this project is Box Cox transformation.

Box Cox transformation is named after statisticians George Box and Sir David Roxbee Cox who worked together and developed this technique in year 1964. The main concept of Box Cox transformation is a parameter called lambda (λ) which its value could vary from -5 to 5. The optimal value of λ depends on the one which results in the best approximation of a normal distribution. The transformation involves the following mathematical calculation:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, if\ \lambda \neq 0 \\ log\, y, if\ \lambda = 0 \end{cases} \quad (3.4.1.1)$$

Where:

y represents the original data

This transformation works only for data with positive values. Nevertheless, Box and Cox provided a second formula which can be used for negative values:

$$y(\lambda) = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1}, if\ \lambda_1 \neq 0 \\ log\ (y+\lambda_2), if\ \lambda_1 = 0 \end{cases} \quad (3.4.1.2)$$

Where:

y represents the original data;

λ_1 represents the transformation coefficient;

λ_2 represents the numerical value to add on to make negative values positive.

In this project, λ values of 2, 1, 0.5, -0.5 and -1 are used for Box Cox transformation. Analysis will be performed to show how quality of forecast varies with different λ value.

When λ = 2,

$$y'(2) = \frac{y^2-1}{2} \quad (3.4.1.3)$$

When λ = 1,

$$y'(1) = y - 1 \quad (3.4.1.4)$$

When λ = 0.5,

$$y'(0.5) = \frac{\sqrt{y}-1}{0.5} = 2\sqrt{y} - 2 \quad (3.4.1.5)$$

When λ = -0.5

$$y'(-0.5) = \frac{\frac{1}{\sqrt{y}}-1}{-0.5} = 2 - \frac{2}{\sqrt{y}} \qquad (3.4.1.6)$$

When λ = -1

$$y'(-1) = \frac{\frac{1}{y}-1}{-1} = 1 - \frac{1}{y} \qquad (3.4.1.7)$$

Where y' represents the transformed value and y represents the original value.

3.4.1 Box Cox transformation for flow

The first training attempt is to perform Box Cox transformation on flow data directly followed by training of Artificial Neural Network on the transformed data. After transforming the data with various λ value, forecast of lead time 5, 10, 20, 30 and 60 minutes are produced.

3.4.1.1 5-minute Box Cox transformed flow forecast

The neural network training procedure for 5-minute Box Cox transformed flow forecast as follows:

1. Calculate new inputs using different λ values.
2. Shift the new Q_MD04 up by 5 rows.
3. Delete the last 5 rows of each individual event.
4. The next few steps are similar with the procedure stated in section 3.1.1
5. Reverse Box Cox transformation.

   At this point of time, the result obtained is in transformed scale. One more step needs to be performed which is reverse Box Cox transformation in order to make the ANN results and original data comparable. The mathematical calculations for each λ value are:

When λ = 2,

$$y(2) = \sqrt{2y' + 1} \tag{3.4.2.1}$$

When λ = 1,

$$y(1) = y' + 1 \tag{3.4.2.2}$$

When λ = 0.5,

$$y(0.5) = (\frac{y'+2}{2})^2 \tag{3.4.2.3}$$

When λ = -0.5,

$$y(-0.5) = (\frac{2}{2-y\prime})^2 \tag{3.4.2.4}$$

When λ = -1,

$$y(-0.5) = \frac{1}{1-y\prime} \tag{3.4.2.5}$$

Where y' represents the transformed value and y represents the original value.

Fig 28, 29, 30, 31 and 32 show the comparison between ANN forecast and original value for lead time of 5 minutes against various λ values.

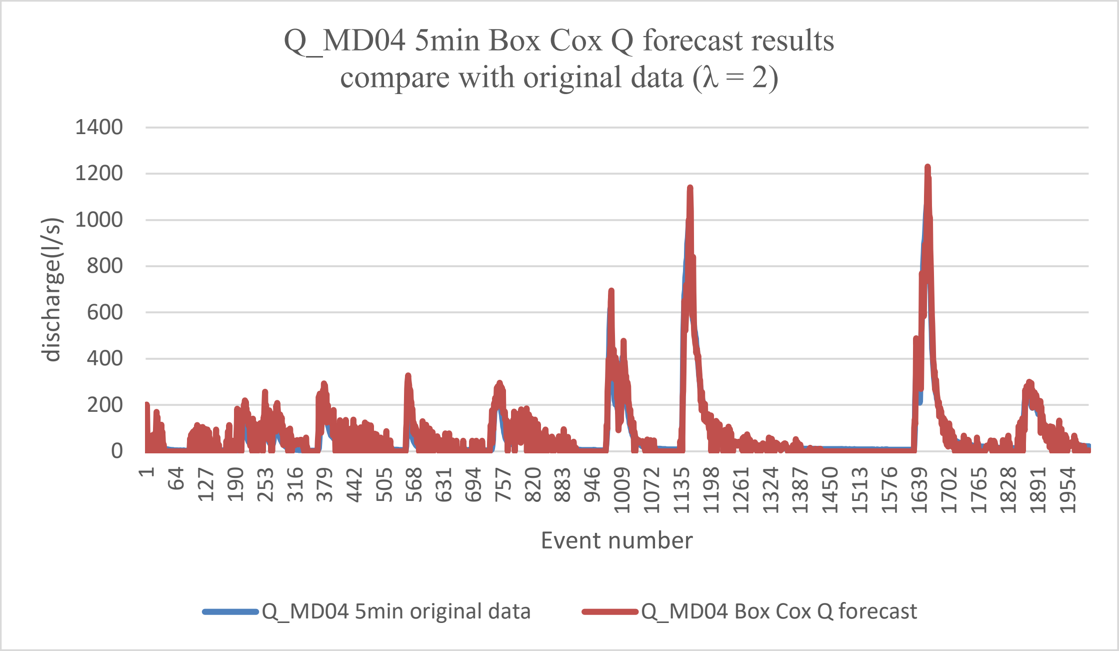


Fig 28. Q_MD04 5min Box Cox Q forecast results compare with original data (λ = 2)

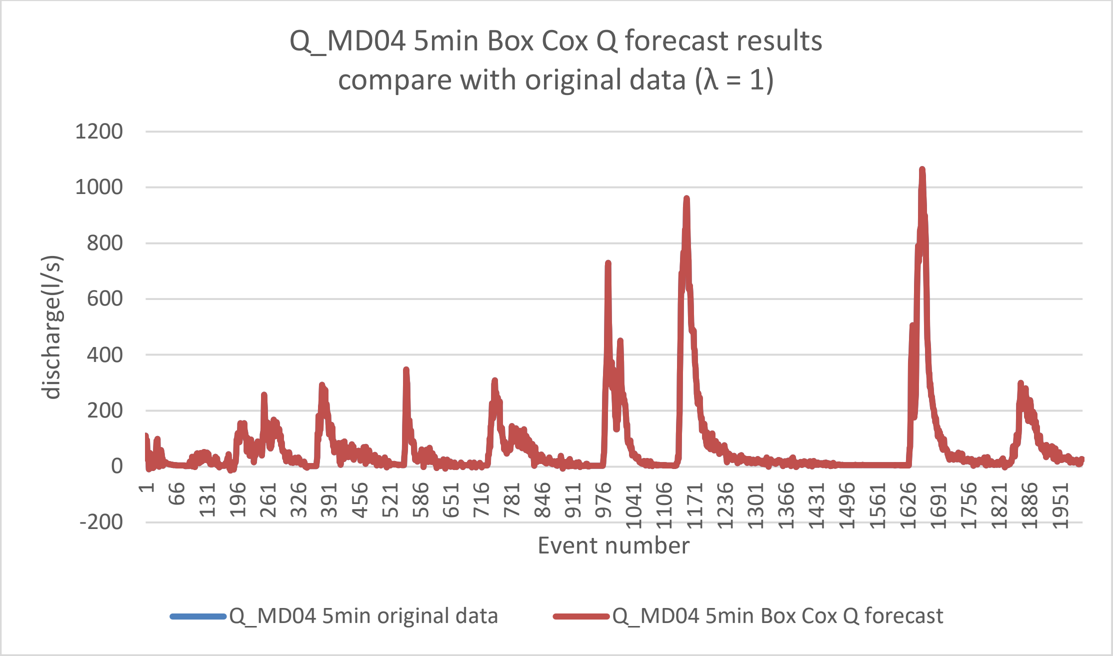


Fig 29. Q_MD04 5min Box Cox Q forecast results compare with original data (λ = 1)

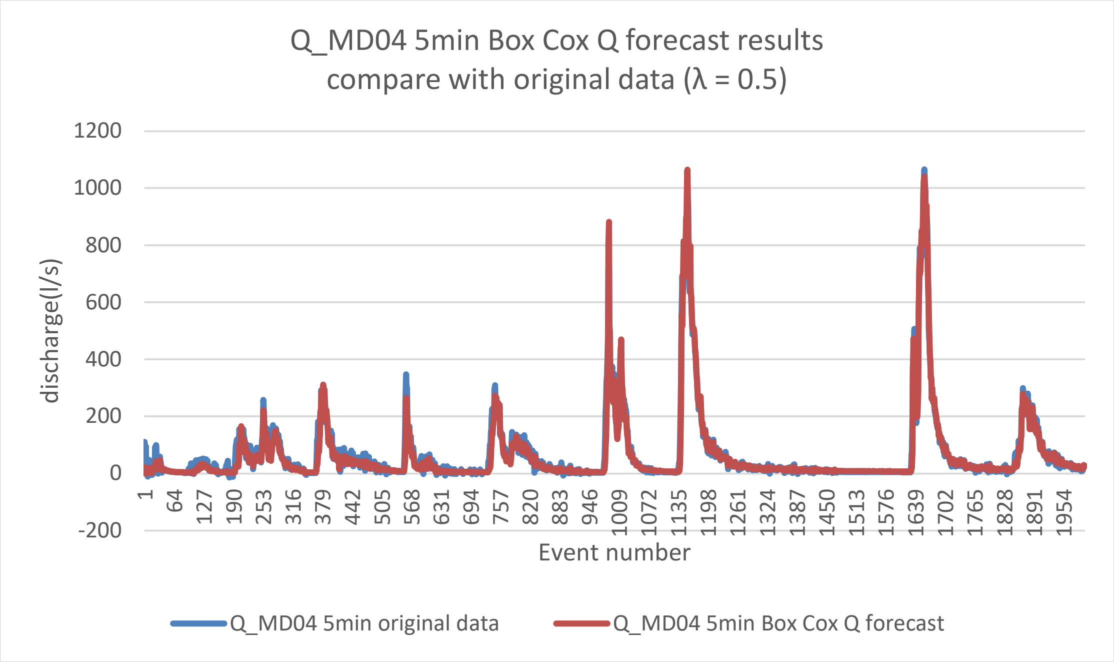


Fig 30. Q_MD04 5min Box Cox Q forecast results compare with original data (λ = 0.5)

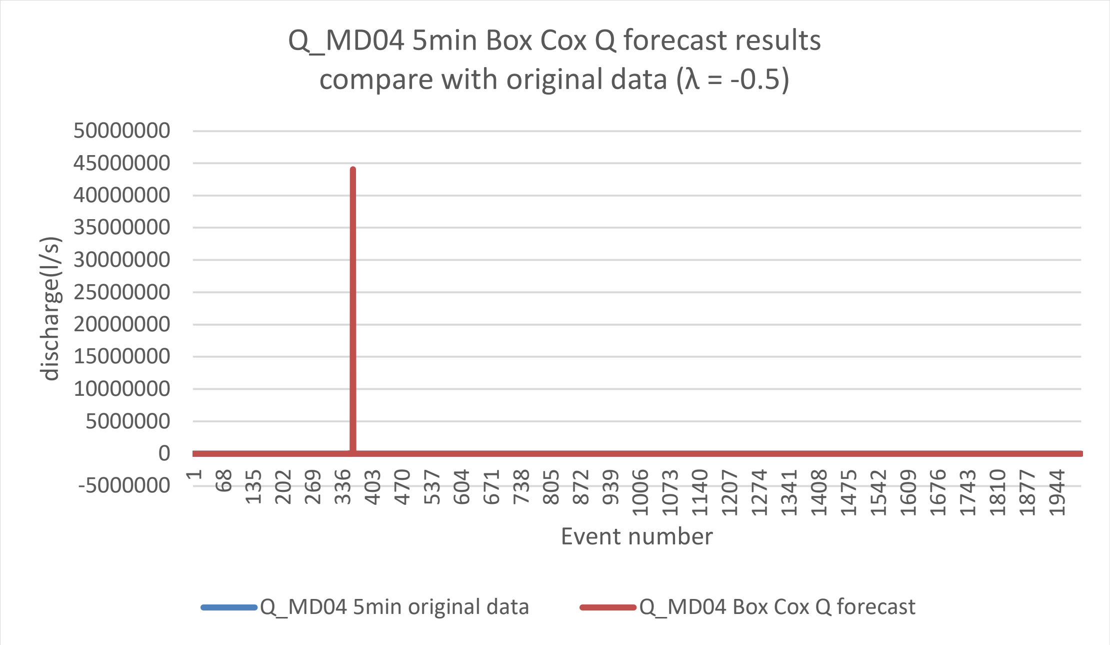


Fig 31. Q_MD04 5min Box Cox Q forecast results compare with original data (λ = -0.5)

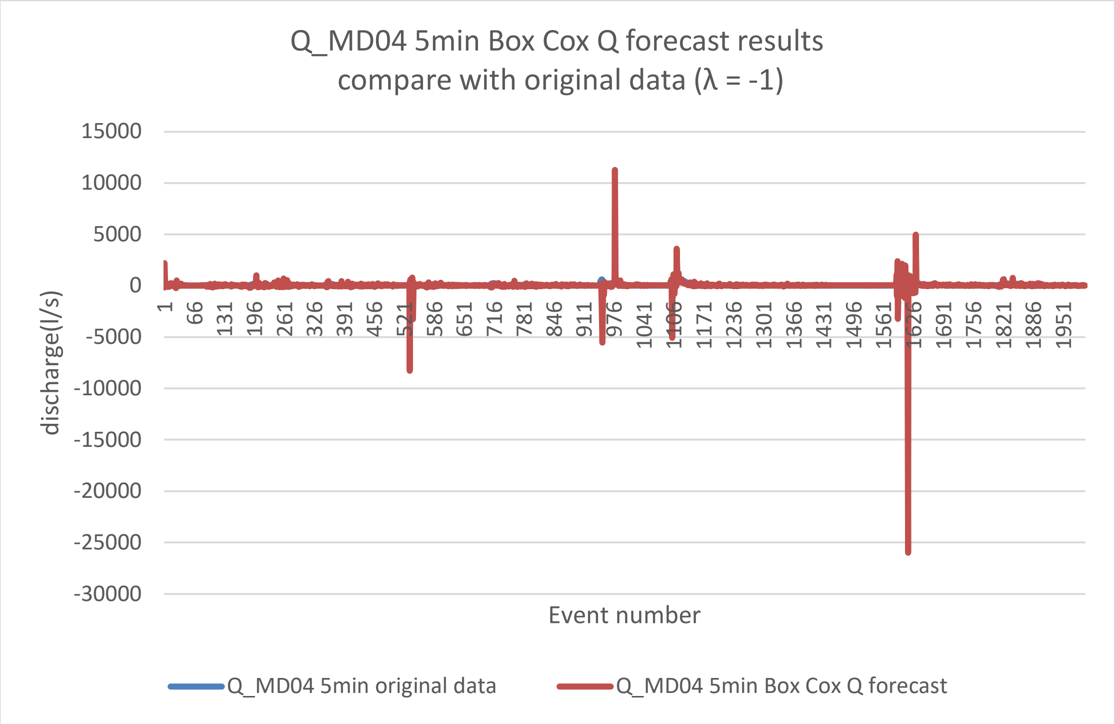


Fig 32. Q_MD04 5min Box Cox Q forecast results compare with original data (λ = -1)

3.4.1.2 10-minute Box Cox transformed flow forecast

Following the similar procedure as above, 10-minute Box Cox transformed flow forecast is produced. Fig 33, 34, 35, 36 and 37 shows the comparison between ANN forecast and original value for lead time of 10 minutes against various λ values.

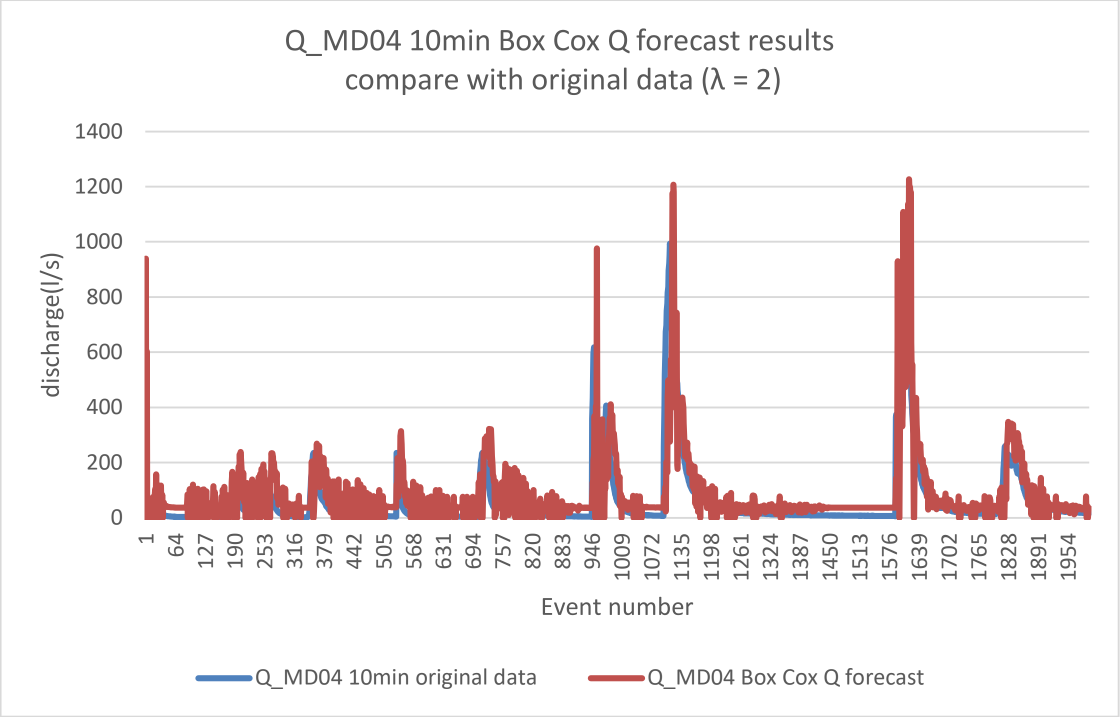


Fig 33. Q_MD04 10min Box Cox Q forecast results compare with original data (λ = 2)

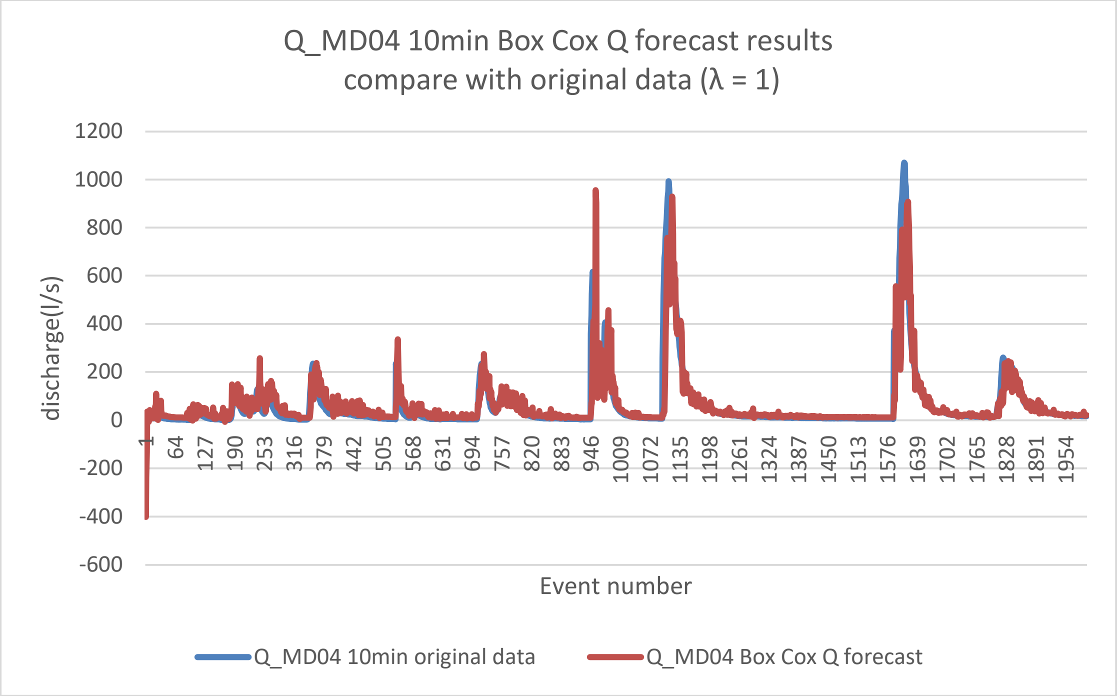


Fig 34. Q_MD04 10min Box Cox Q forecast results compare with original data (λ = 1)

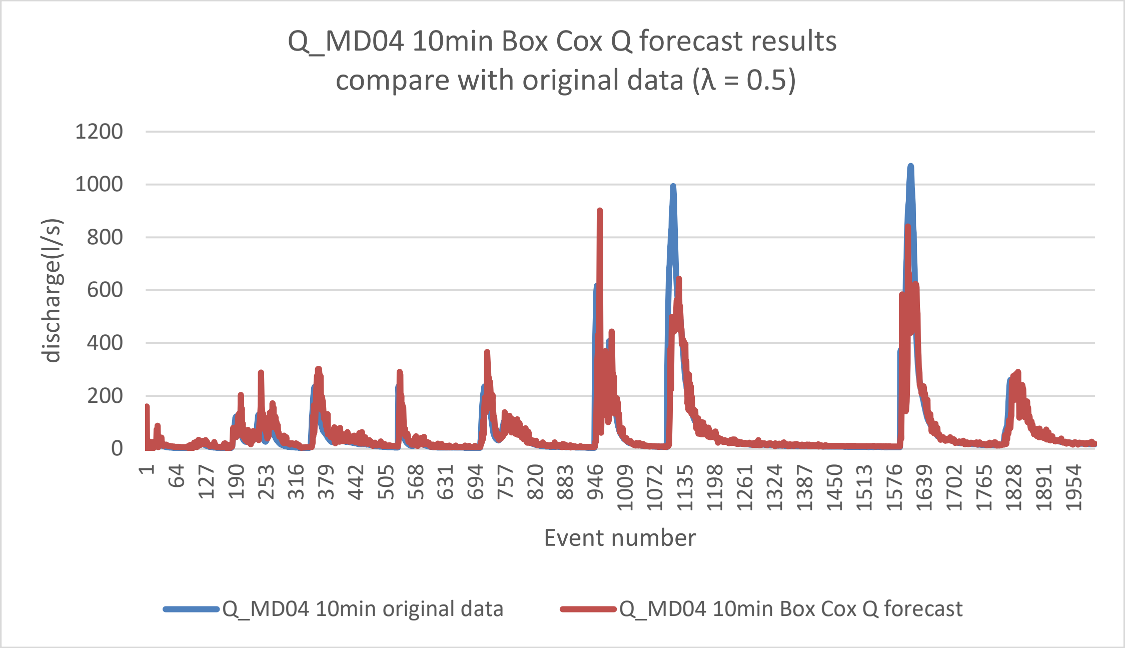


Fig 35. Q_MD04 10min Box Cox Q forecast results compare with original data (λ = 0.5)

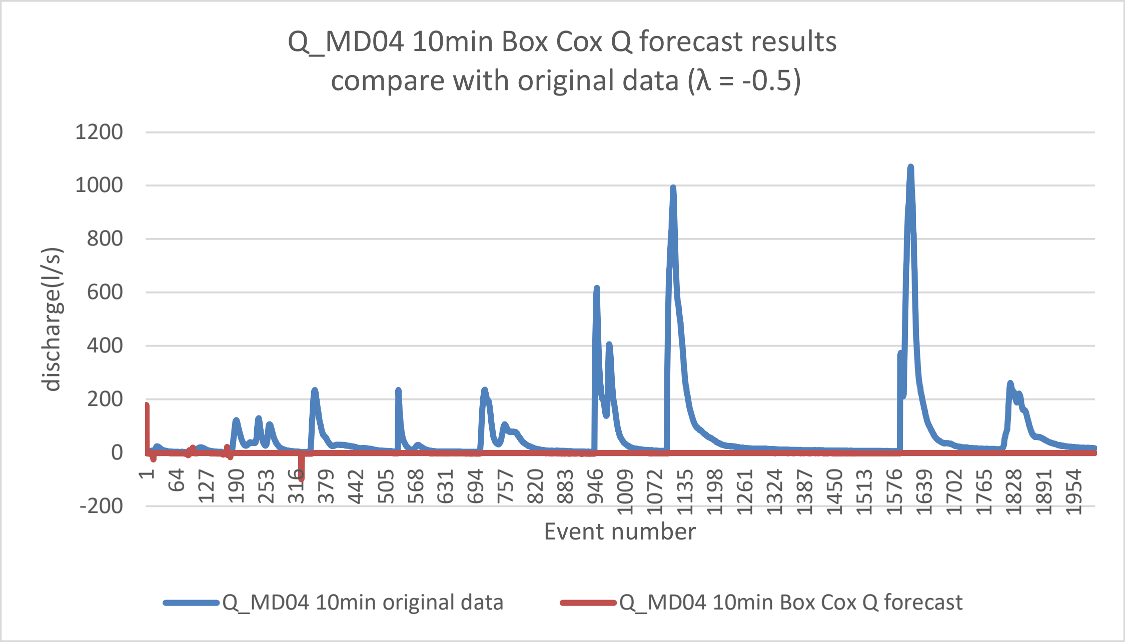


Fig 36. Q_MD04 10min Box Cox Q forecast results compare with original data (λ = -0.5)

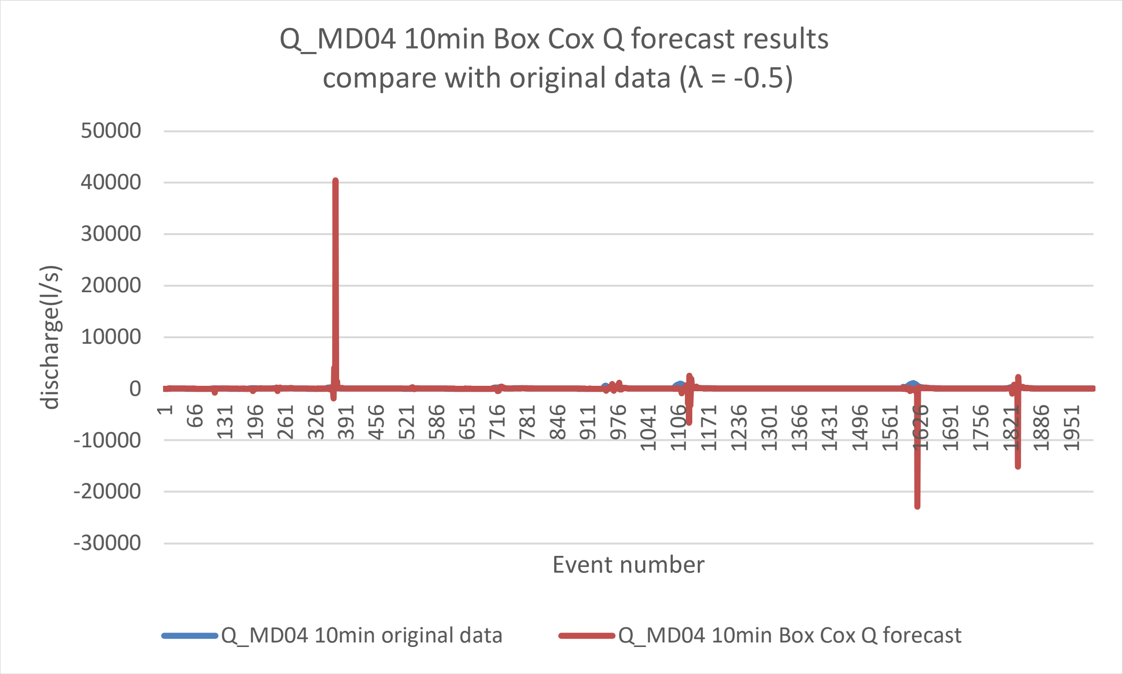


Fig 37. Q_MD04 10min Box Cox Q forecast results compare with original data (λ = -1)

3.4.1.3 20-minute Box Cox transformed flow forecast

Following the similar procedure as above, 20-minute Box Cox transformed flow forecast is produced. Fig 38, 39 and 40 shows the comparison between ANN forecast and original value for lead time of 20 minutes against various λ values. As it can be seen from the last 2 series of training attempts, λ with negative values produce poor quality forecasts. Therefore, only positive λ values are considered for the future training attempts.

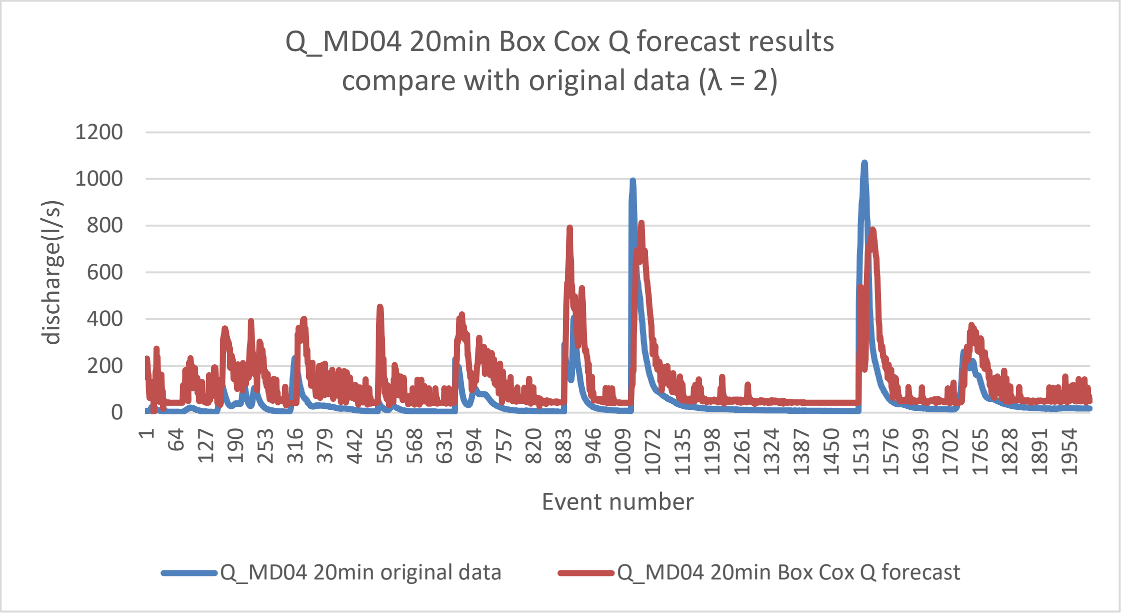


Fig 38. Q_MD04 20min Box Cox Q forecast results compare with original data (λ = 2)

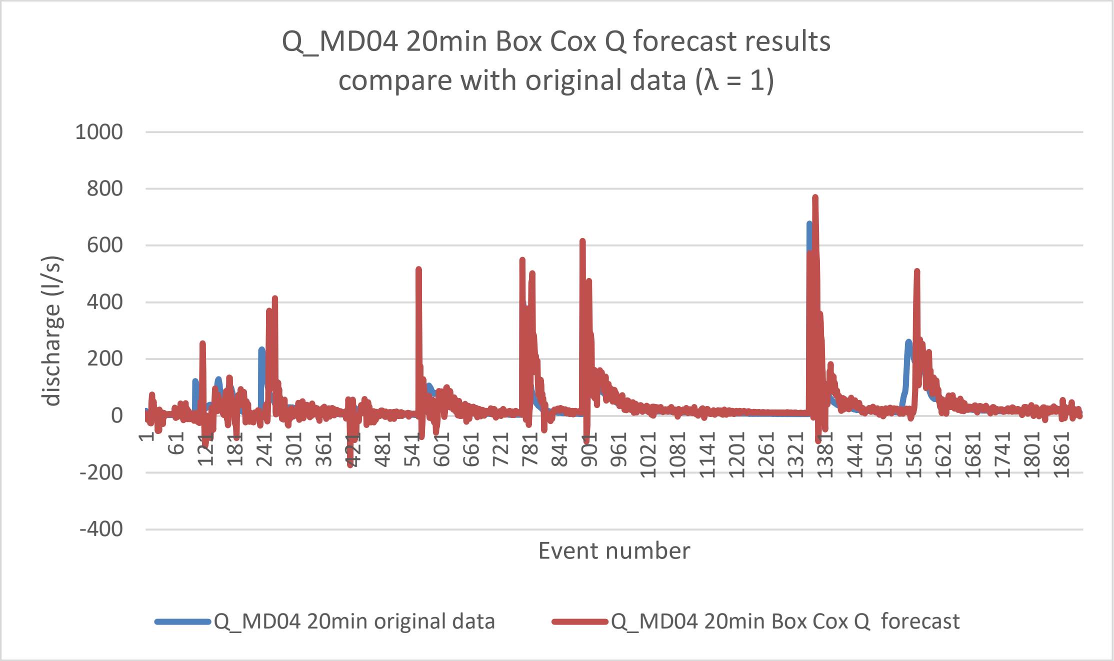


Fig 39. Q_MD04 20min Box Cox Q forecast results compare with original data (λ = 1)

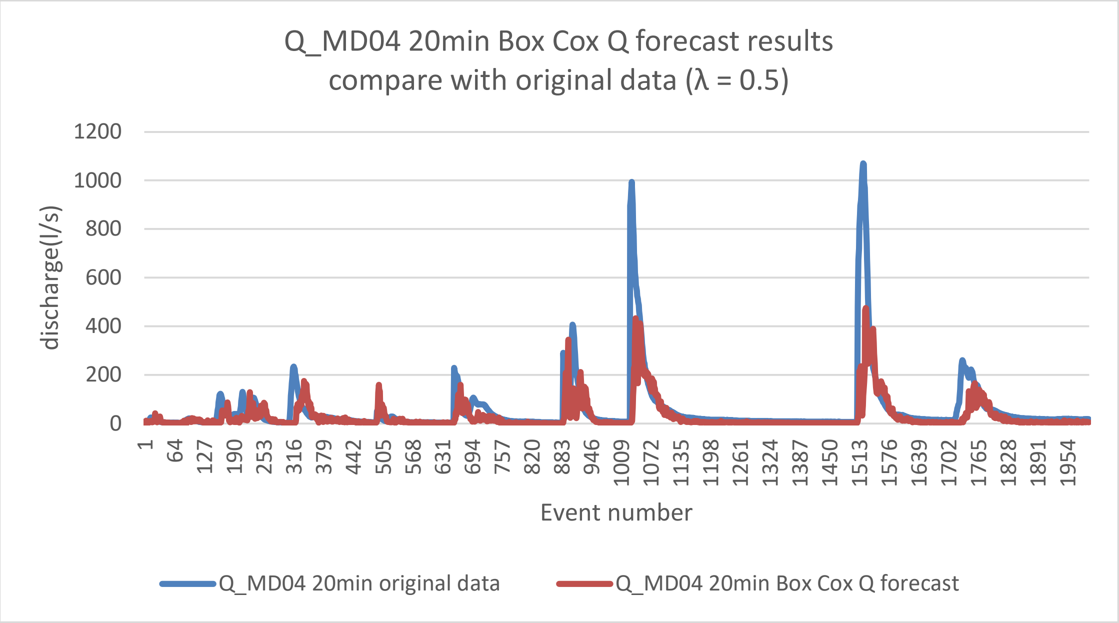


Fig 40. Q_MD04 20min Box Cox Q forecast results compare with original data (λ = 0.5)

3.4.1.4 30-minute Box Cox transformed flow forecast

Following the similar procedure as above, 30-minute Box Cox transformed flow forecast is produced. Fig 41, 42 and 43 show the comparison between ANN forecast and original value for lead time of 30 minutes against various λ values.

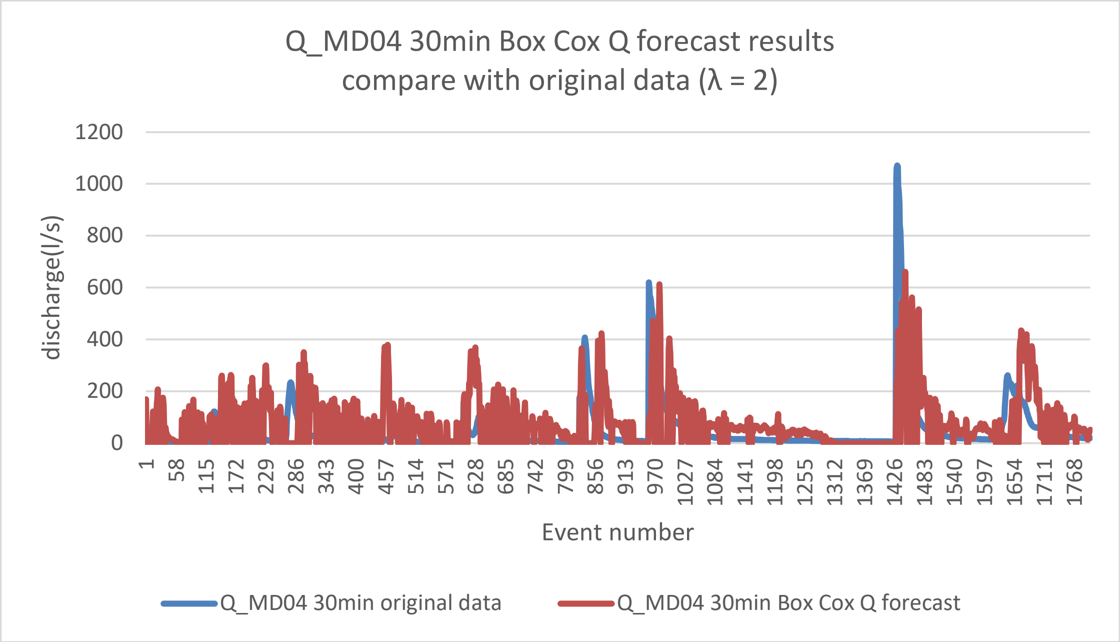


Fig 41. Q_MD04 30min Box Cox Q forecast results compare with original data (λ = 2)

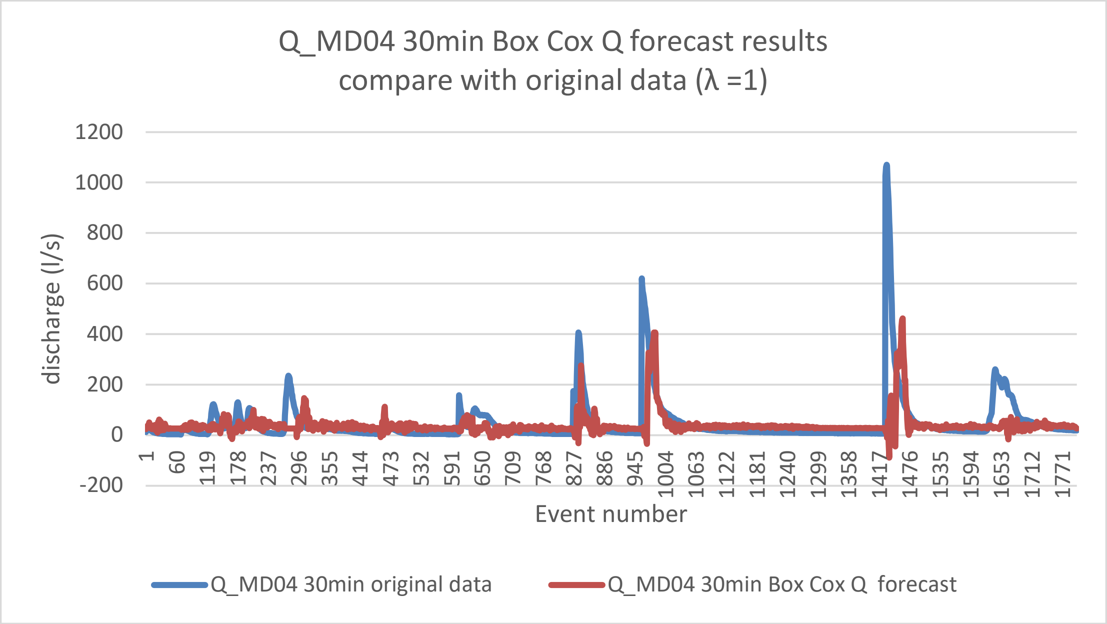


Fig 42. Q_MD04 30min Box Cox Q forecast results compare with original data (λ = 1)

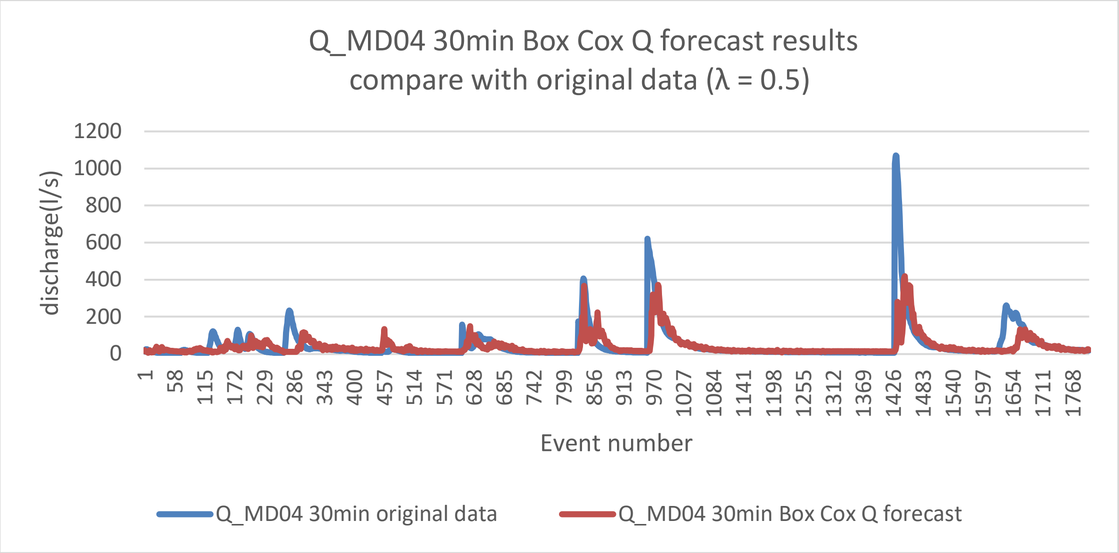


Fig 43. Q_MD04 30min Box Cox Q forecast results compare with original data (λ = 0.5)

3.4.1.5 60-minute Box Cox transformed flow forecast

Following the similar procedure as above, 60-minute Box Cox transformed flow forecast is produced. Fig 44, 45 and 46 show the comparison between ANN forecast and original value for lead time of 60 minutes against various λ values.

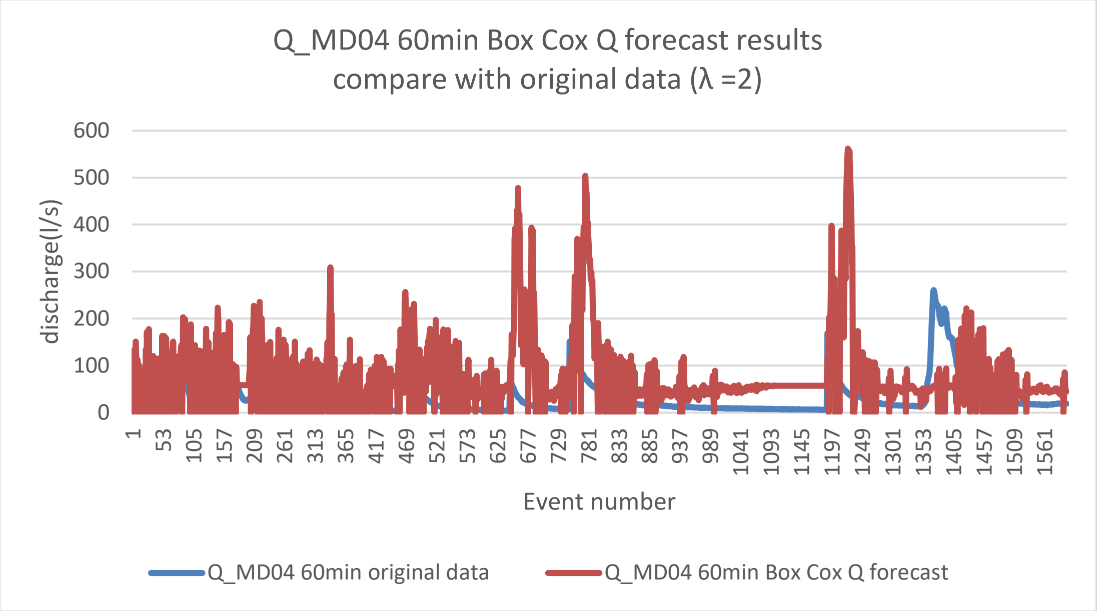


Fig 44. Q_MD04 60min Box Cox Q forecast results compare with original data (λ =2)

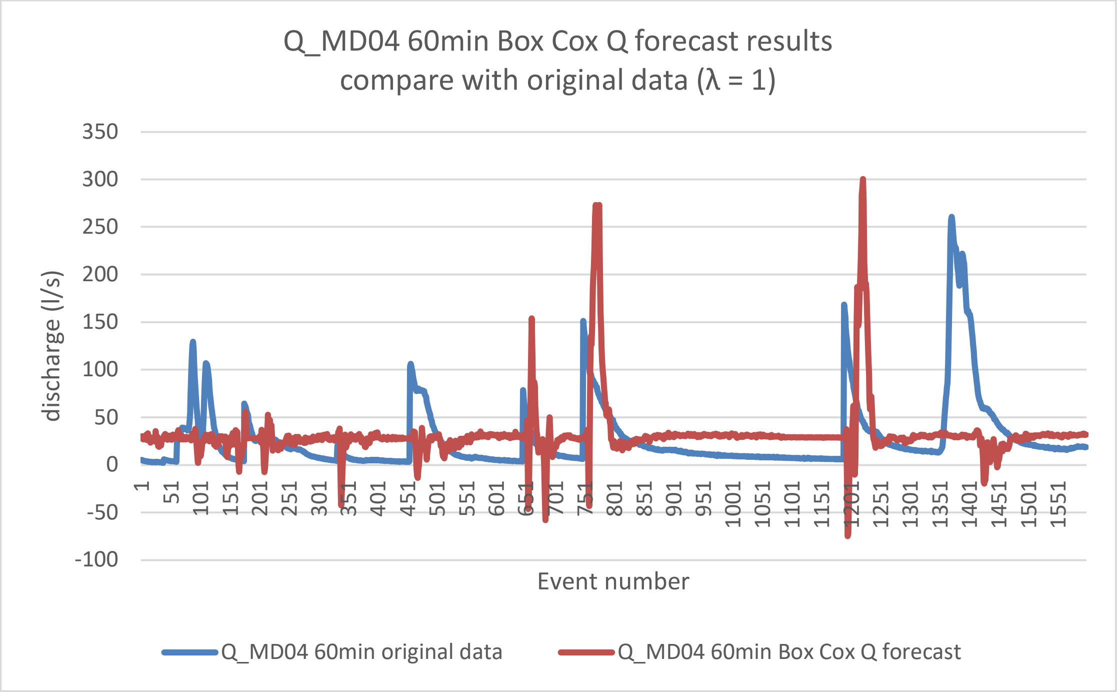


Fig 45. Q_MD04 60min Box Cox Q forecast results compare with original data (λ = 1)

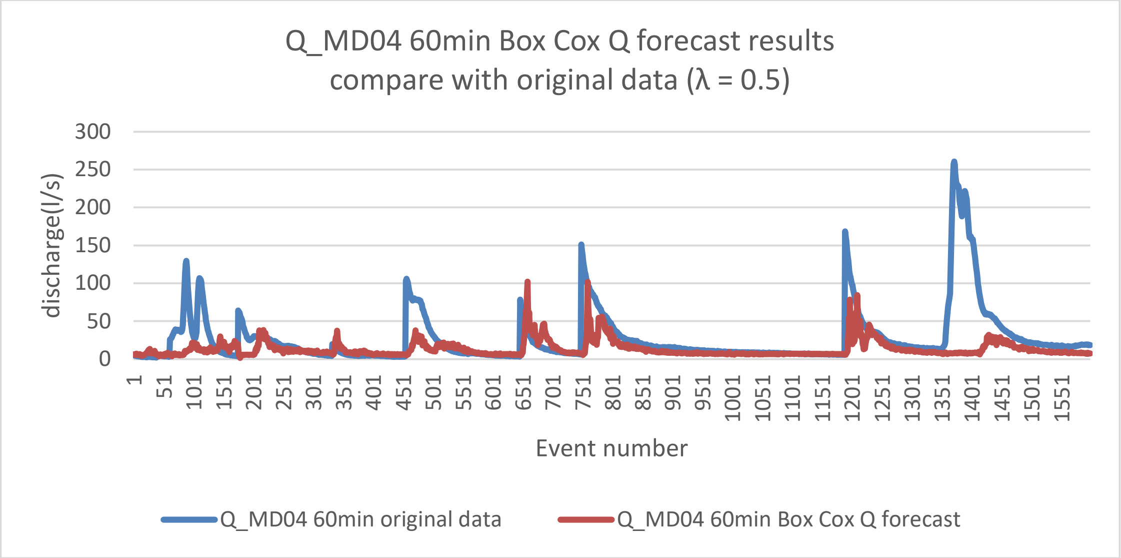


Fig 46. Q_MD04 60min Box Cox Q forecast results compare with original data (λ = 0.5)

3.4.2 Box Cox transformation for difference of flow (dQ)

The next training attempt is based on Box Cox transformation for difference of flow followed by neural network training. After transforming the data (dQ) with various λ value, forecast of lead time 5, 10, 20, 30 and 60 minutes are produced.

Dealing with negative values

After taking difference of flow of the drains, it can be noticed that there are many negative data points. The second equation suggested by Box and Cox becomes applicable. Taking dQ of 5 minutes as an example, the minimum value of dQ is -469.85 using Microsoft Excel. Therefore, the λ_2 value can be set to 470. Adding a value of 470 to all data points to obtain positive values. The reverse Box Cox transformation step needs to minus 470 from all data sets in order to get the correct values.

3.4.2.1 5-minute Box Cox transformed dQ forecast

The training procedure for box-cox transformed 5-minute dQ forecast is similar as steps mentioned in section 3.2.1. Fig 47, 48 and 49 show the comparison between ANN forecast and original value for lead time of 5 minutes against various λ values.

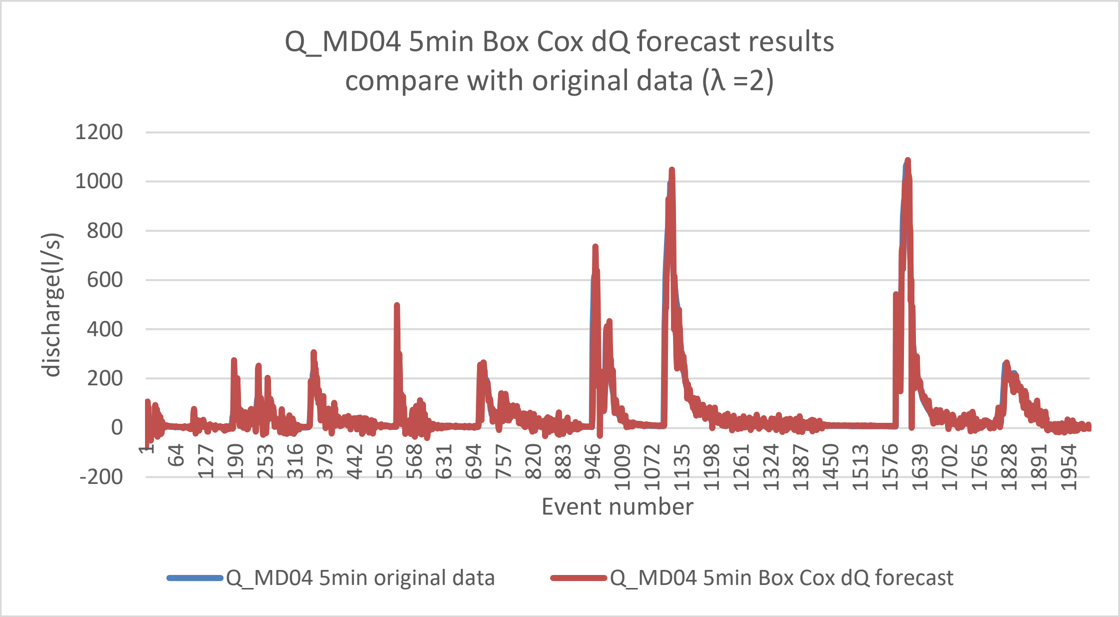


Fig 47. Q_MD04 5min Box Cox dQ forecast results compare with original data (λ =2)

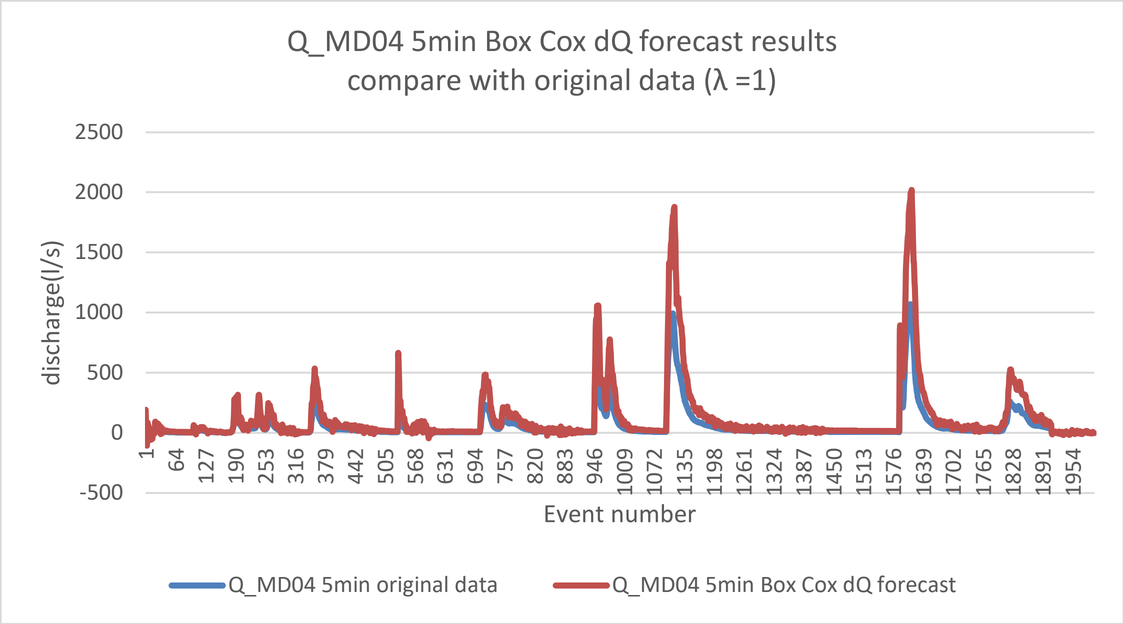


Fig 48. Q_MD04 5min Box Cox dQ forecast results compare with original data (λ =1)

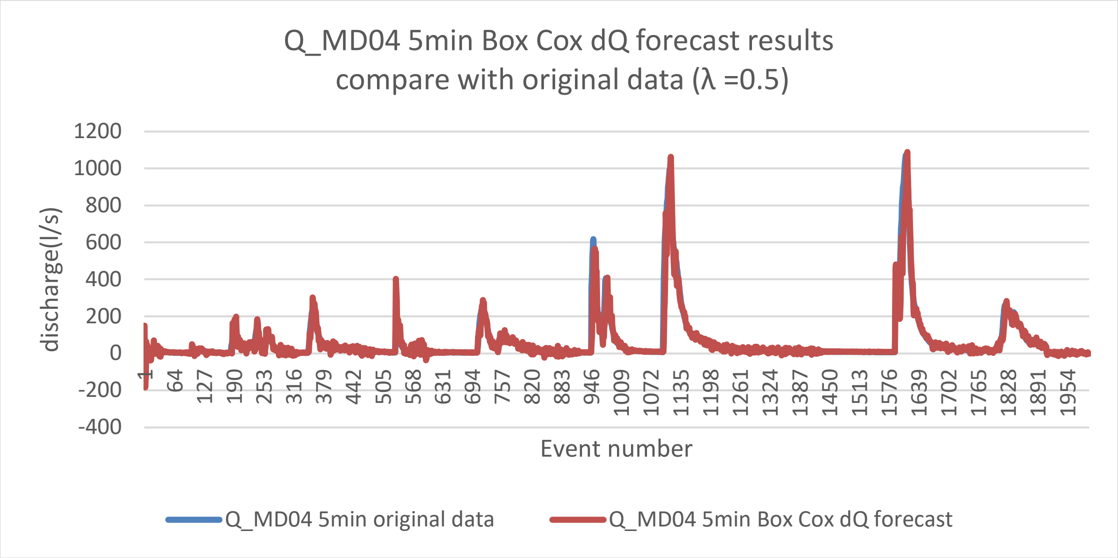


Fig 49. Q_MD04 5min Box Cox dQ forecast results compare with original data (λ =0.5)

3.4.2.2 10-minute Box Cox transformed dQ forecast

Fig 50, 51 and 52 show the comparison between ANN forecast and original value for lead time of 10 minutes against various λ values.

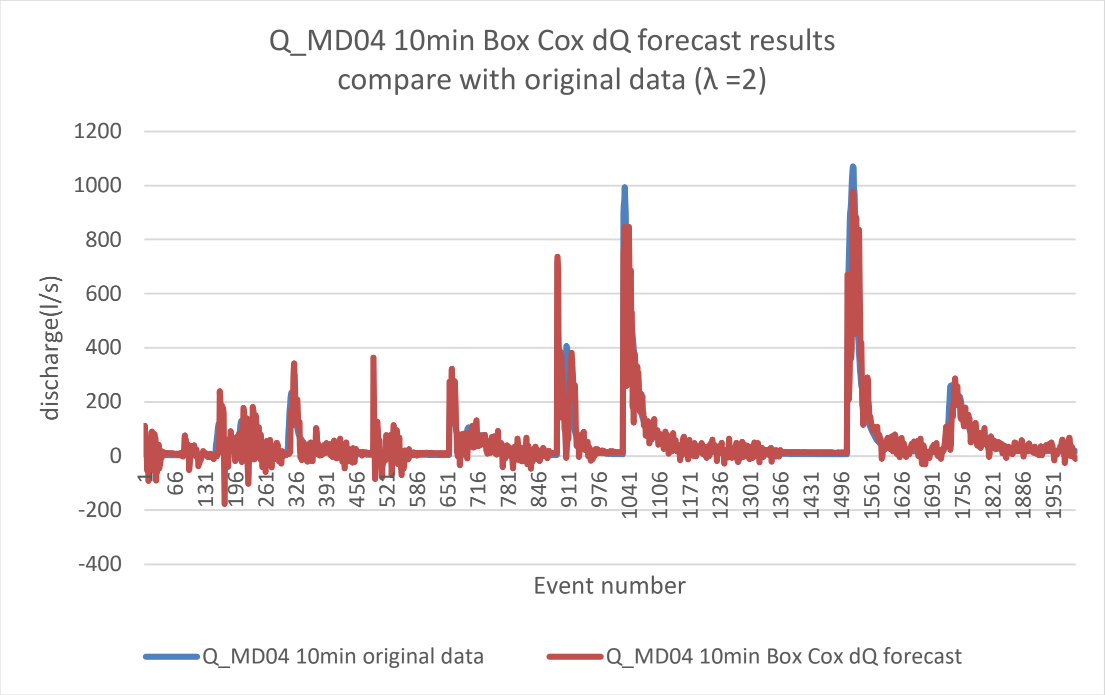


Fig 50. Q_MD04 10min Box Cox dQ forecast results compare with original data (λ =2)

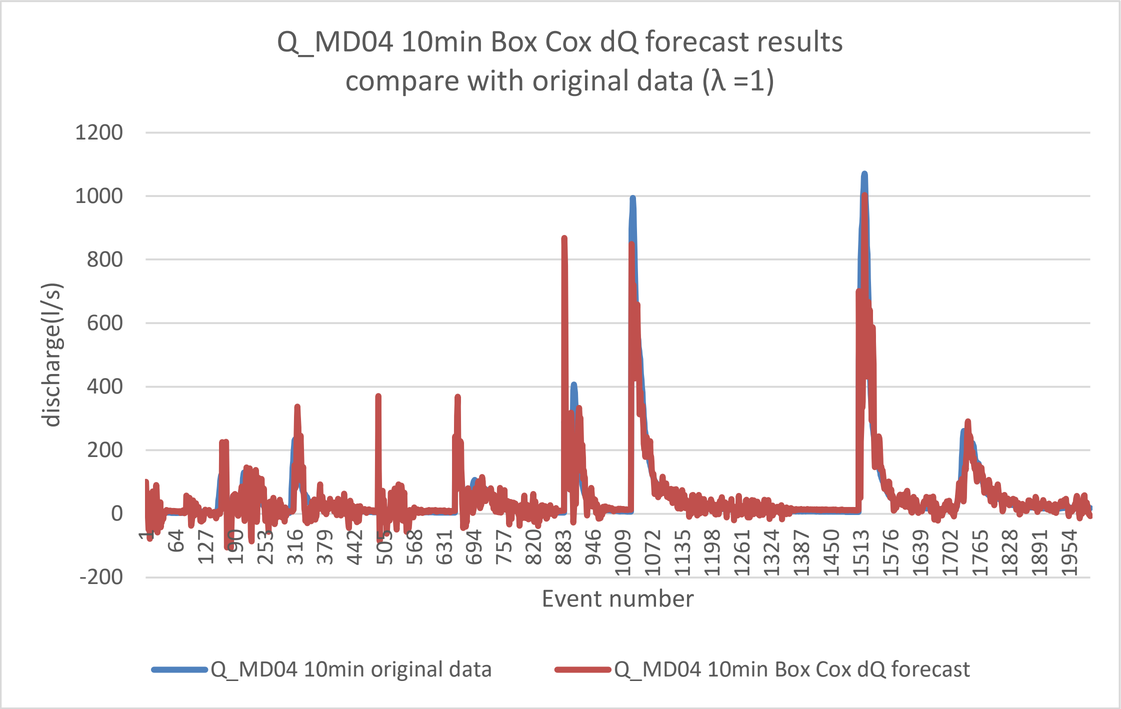


Fig 51. Q_MD04 10min Box Cox dQ forecast results compare with original data (λ =1)

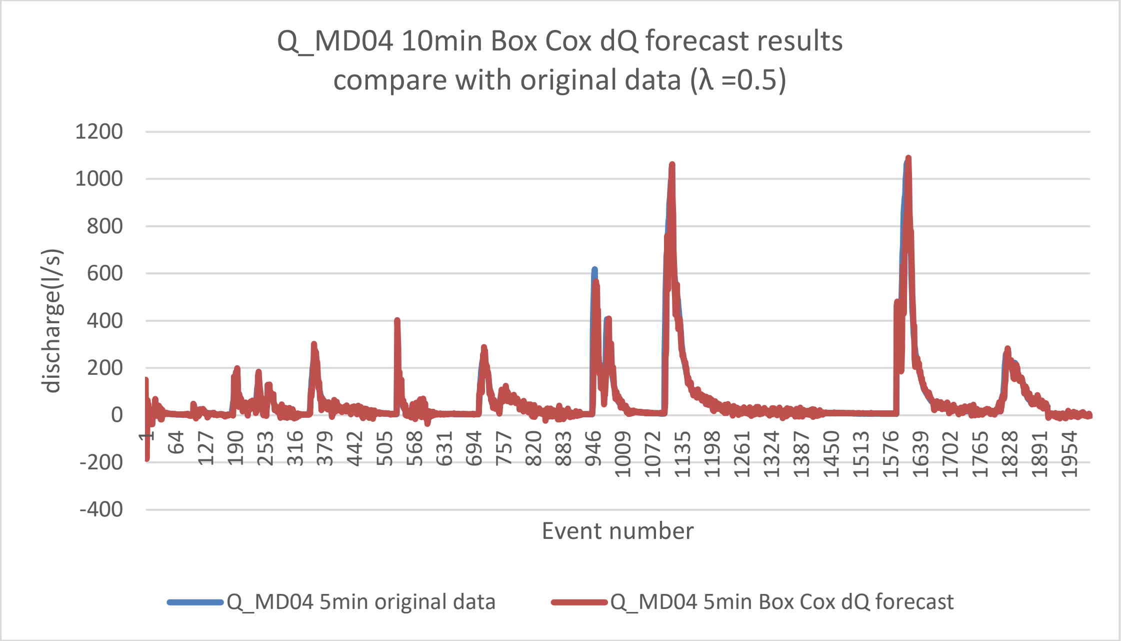


Fig 52. Q_MD04 10min Box Cox dQ forecast results compare with original data (λ =0.5)

3.4.2.3 20-minute Box Cox transformed dQ forecast

Fig 53, 54 and 55 show the comparison between ANN forecast and original value for lead time of 20 minutes against various λ values.

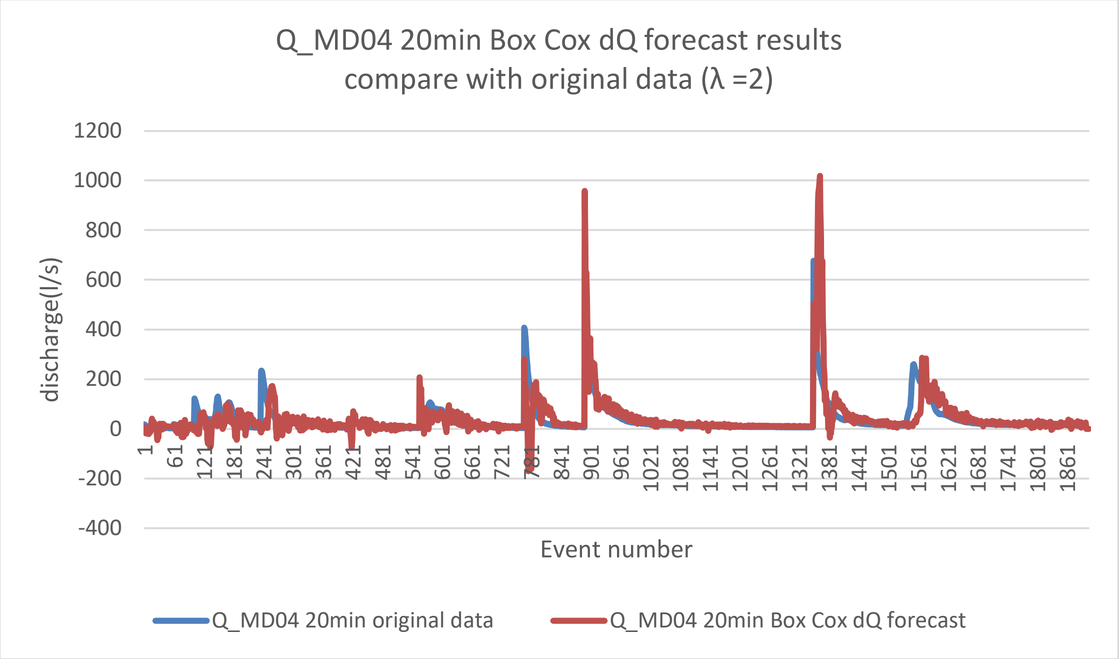


Fig 53. Q_MD04 20min Box Cox dQ forecast results compare with original data (λ =2)

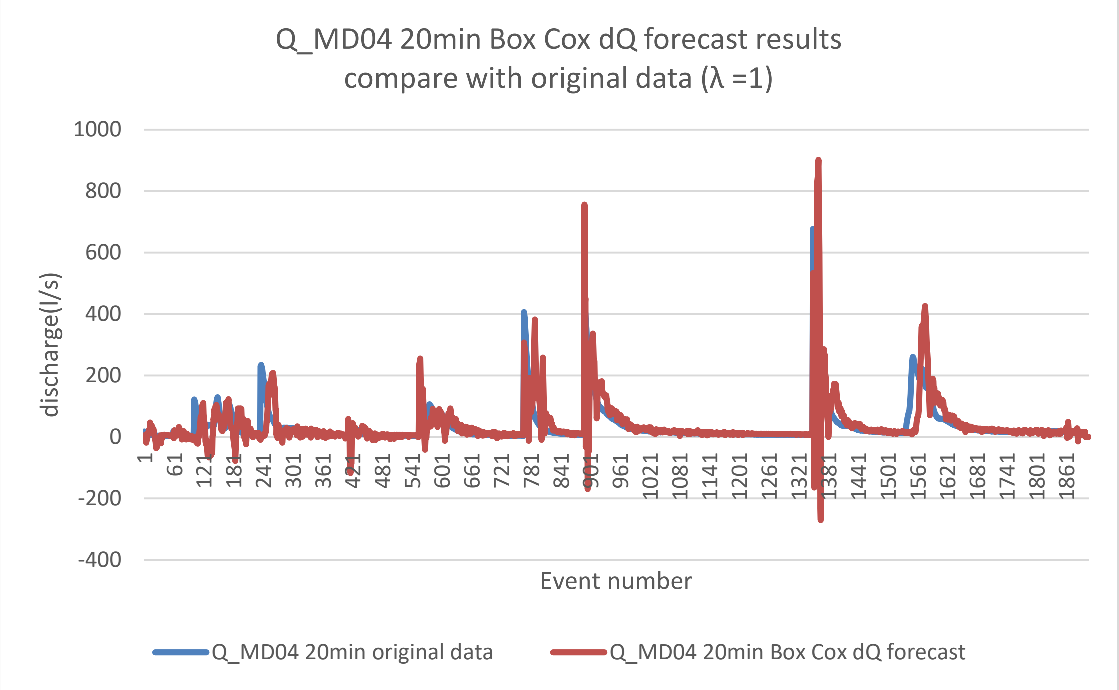


Fig 54. Q_MD04 20min Box Cox dQ forecast results compare with original data (λ =1)

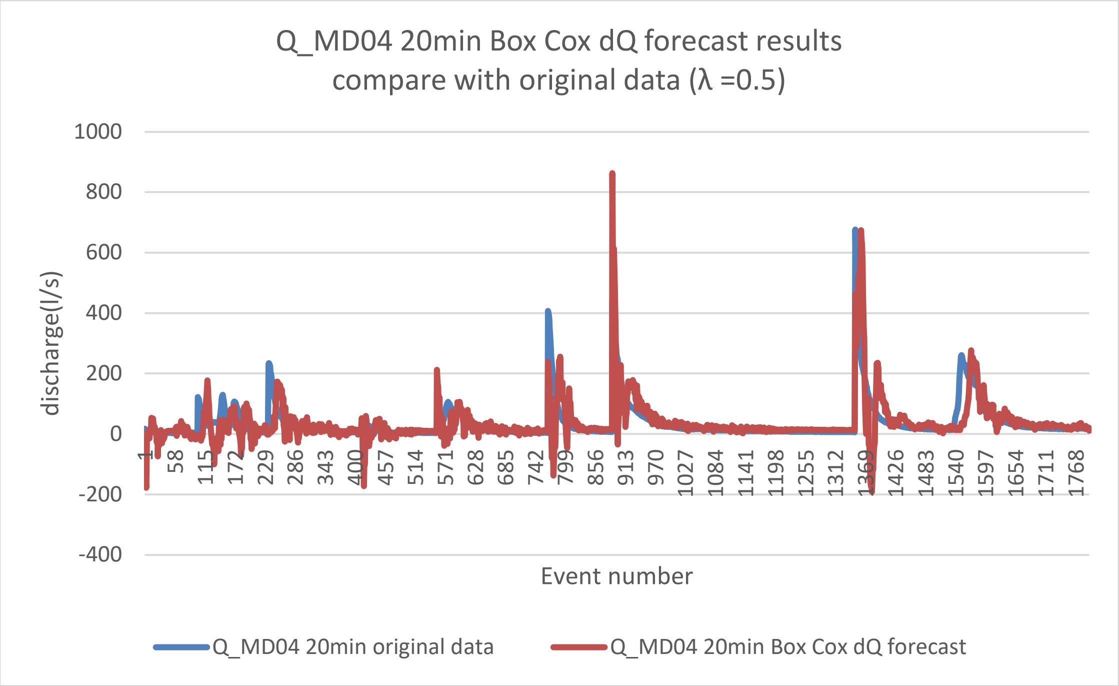


Fig 55. Q_MD04 20min Box Cox dQ forecast results compare with original data (λ =0.5)

3.4.2.4 30-minute Box Cox transformed dQ forecast

Following the similar procedure as above, 30-minute Box Cox transformed dQ forecast is produced. Fig 56, 57 and 58 show the comparison between ANN forecast and original value for lead time of 30 minutes against various λ values.

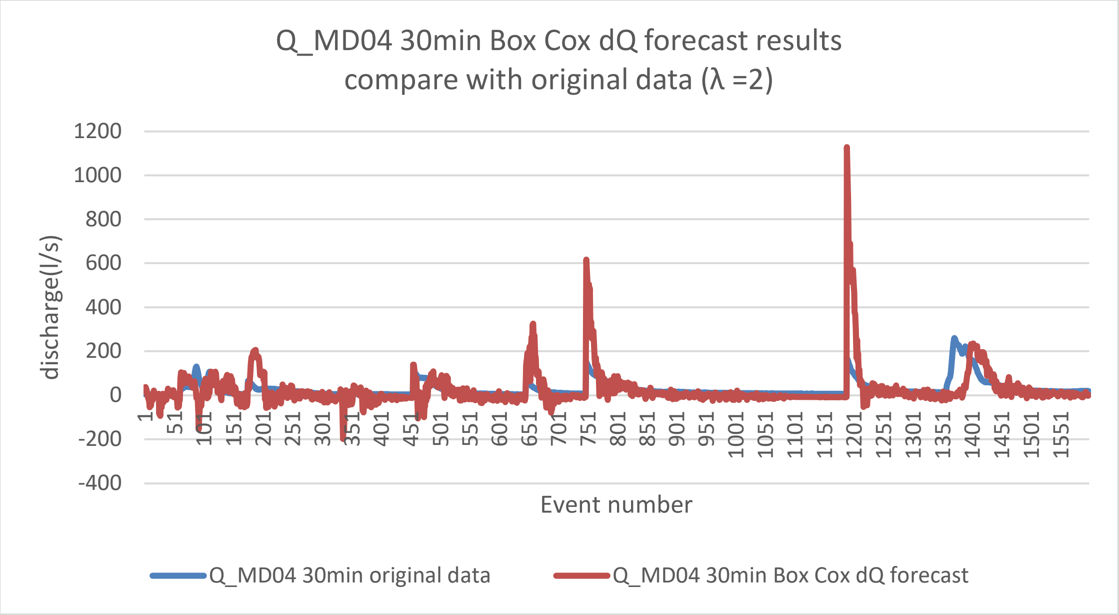


Fig 56. Q_MD04 30min Box Cox dQ forecast results compare with original data (λ =2)

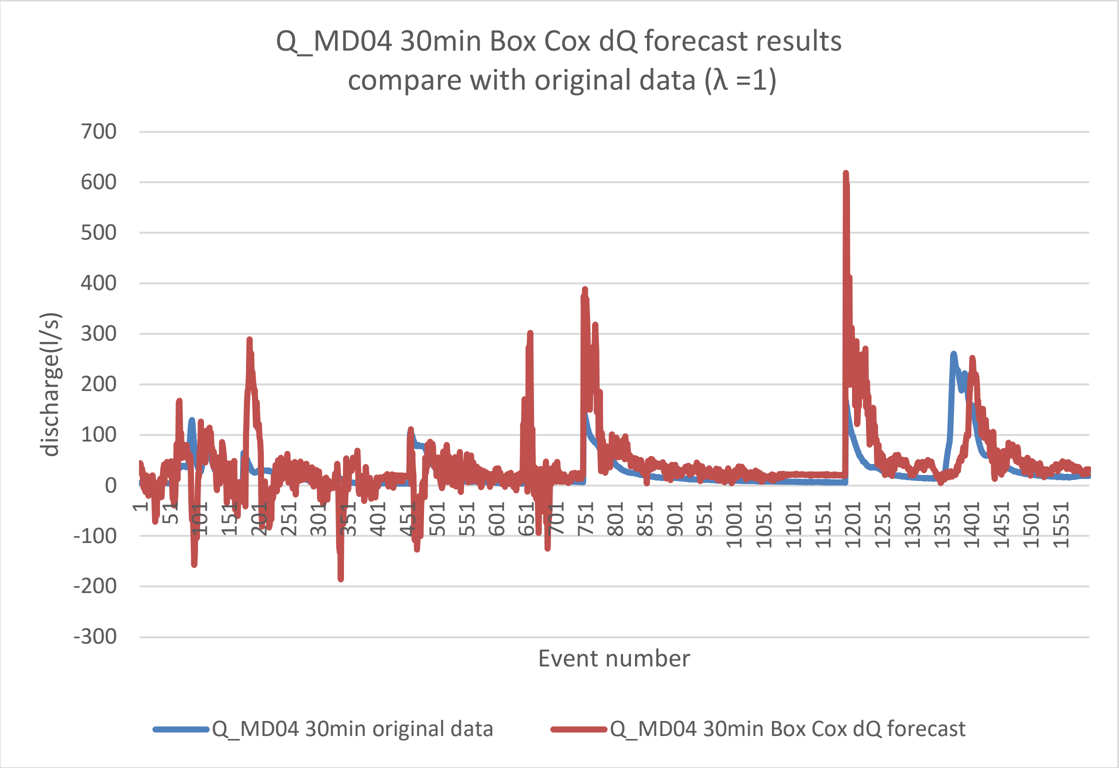


Fig 57. Q_MD04 30min Box Cox dQ forecast results compare with original data (λ =1)

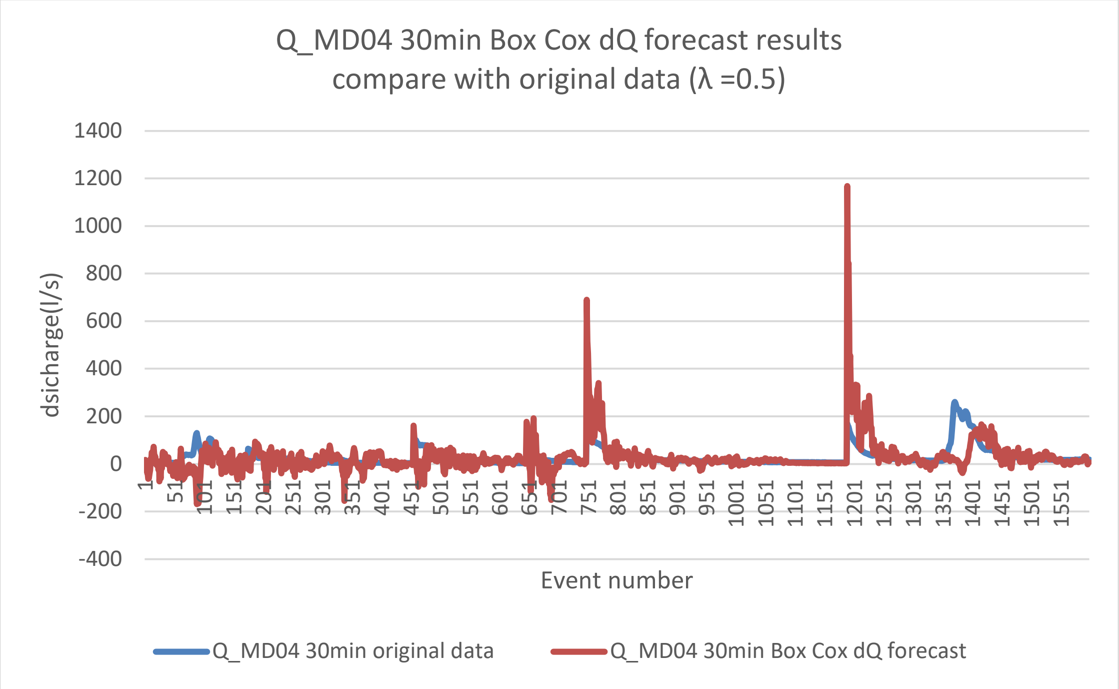


Fig 58. Q_MD04 30min Box Cox dQ forecast results compare with original data (λ = 0.5)

3.4.2.5 60-minute Box Cox transformed dQ forecast

Fig 59, 60 and 61 show the comparison between ANN forecast and original value for lead time of 60 minutes against various λ values.

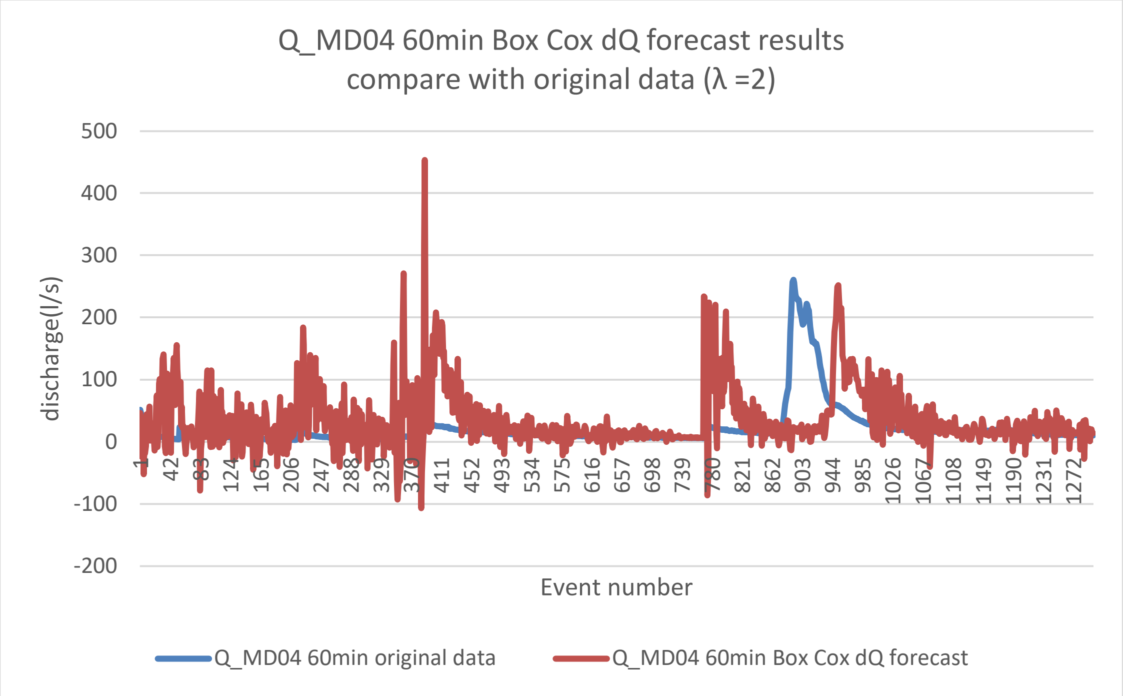


Fig 59. Q_MD04 60min Box Cox dQ forecast results compare with original data (λ =2)

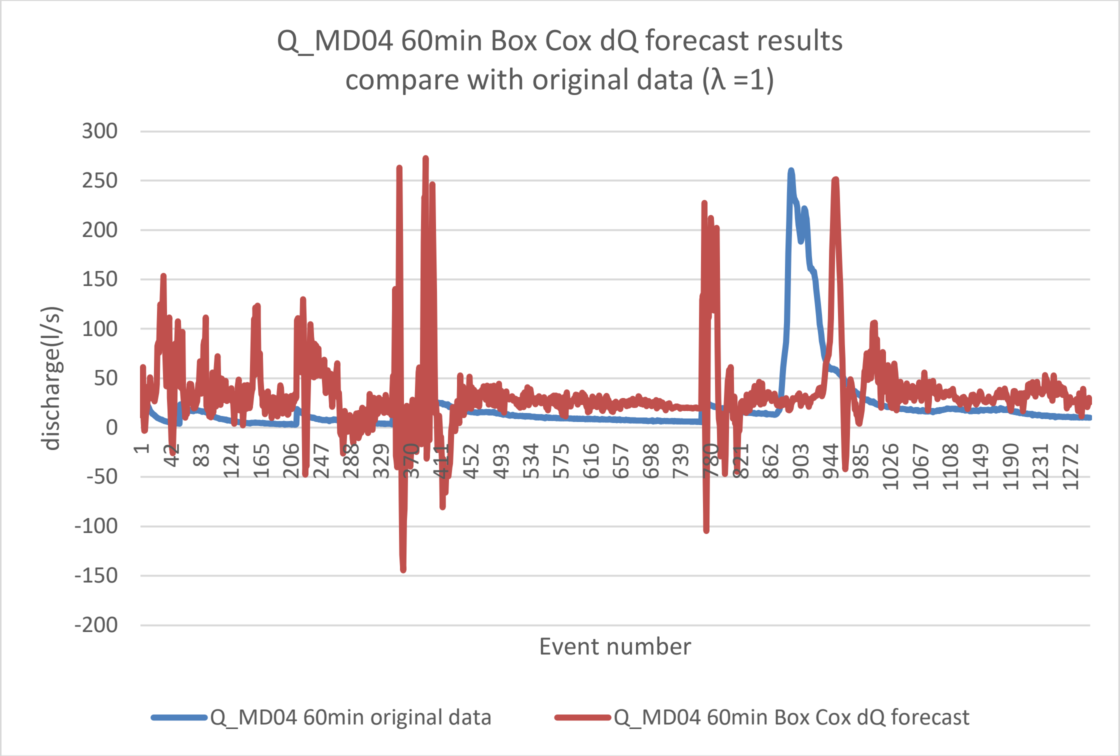


Fig 60. Q_MD04 60min Box Cox dQ forecast results compare with original data (λ =1)

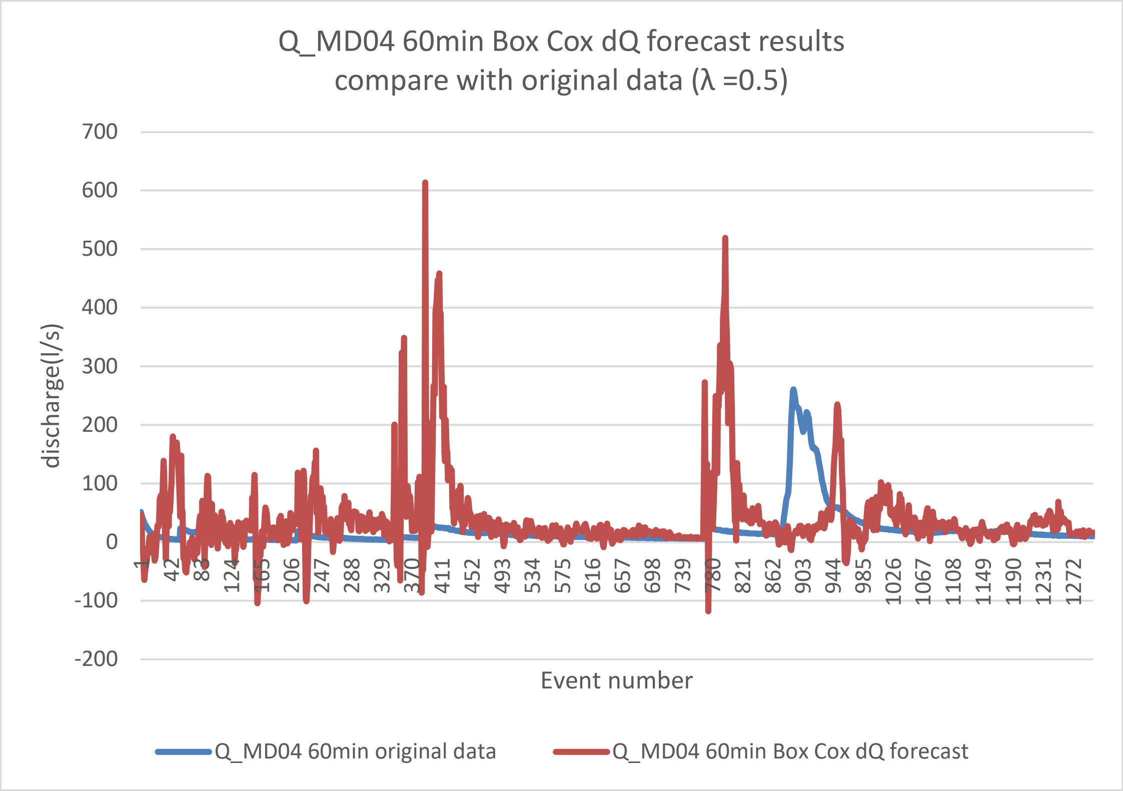


Fig 61. Q_MD04 60min Box Cox dQ forecast results compare with original data (λ =0.5)

4. Results and discussion

The quality of forecast is determined using correlation coefficient between ANN outputs and original data. The value of correlation coefficient varies from 0 to 1. A large correlation coefficient value indicates that there is a strong linear relationship between ANN outputs and the actual data, which suggests that the quality of forecast is satisfactory and vice versa. In this section, comparison will be made between different training methods and discuss the results.

4.1 Comparing naïve forecast with dQ forecast

Correlation coefficient between ANN results and original data

naïve forecast		dQ forecast	
Lead time	r	Lead time	r
5	0.974	5	0.963
10	0.875	10	0.857
20	0.593	20	0.502
30	0.243	30	0.383
60	0.122	60	0.107

Table 4. Correlation coefficient for naïve forecast and dQ forecast

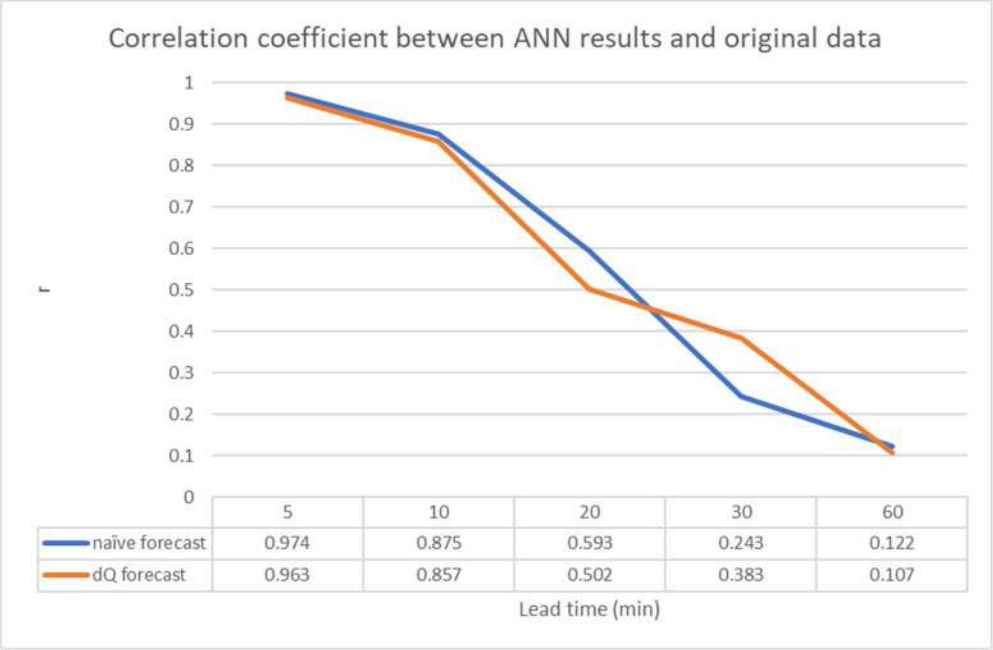


Fig 62. Correlation coefficient vs. lead time for naïve forecast and dQ forecast

Generally, both forecast methods have decent quality of forecast for lead time from 5 minutes to 20 minutes. dQ forecast method has relatively better results for lead time of 30 minutes and both methods perform poorly for lead time of 60 minutes.

4.2 Comparing naïve forecast with dQ+SMA forecast

Correlation coefficient between ANN results and original data

naïve forecast		dQ+SMA forecast	
Lead time	r	Lead time	r
5	0.974	5	0.971
10	0.875	10	0.916
20	0.593	20	0.545
30	0.243	30	0.414
60	0.122	60	0.101

Table 5. Correlation coefficient for naïve forecast and dQ+SMA forecast

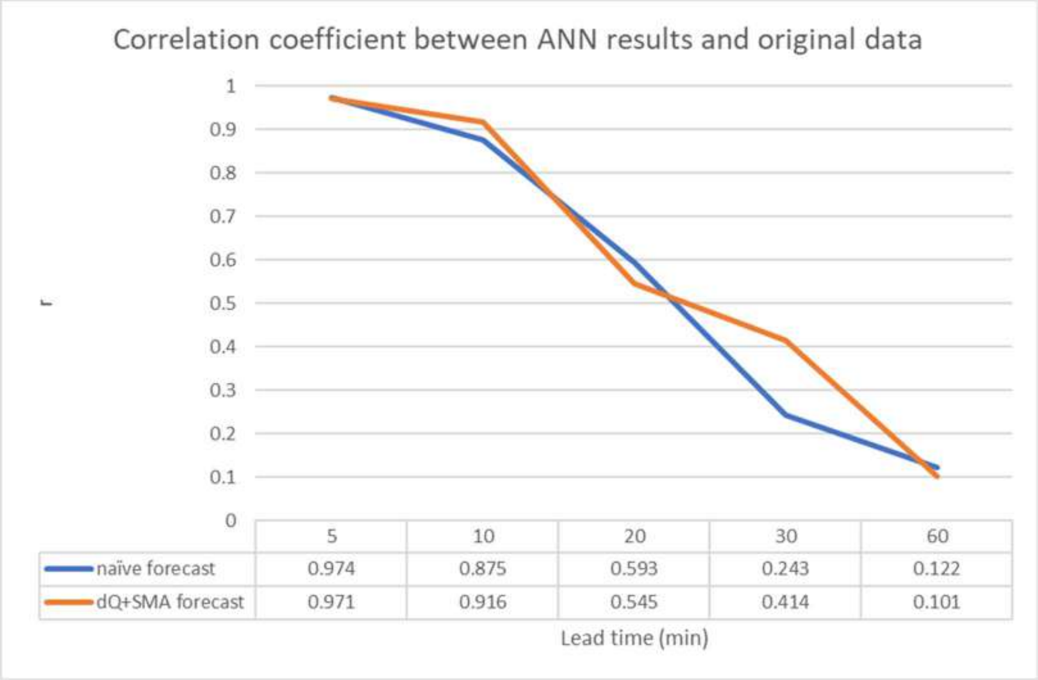


Fig 63. Correlation coefficient vs. lead time for naïve forecast and dQ+SMA forecast

While the results are similar with dQ method, introducing Simple Moving Average does show improvement in quality of forecast in terms of short lead time. It reflects excellent results for lead time of 5 and 10 minutes. dQ+SMA method also shows considerably good results for lead time of

30 minutes comparing with naïve forecast. Simple Moving Average helps to smooth the data and captures the trend in data sets. Referring to Fig 23 to 25, it is evident that the general trend in the original data are well-captured by results produced by ANN. Especially for data with relatively smaller values, ANN performs well in the prediction.

4.3 Limitations of naïve forecast

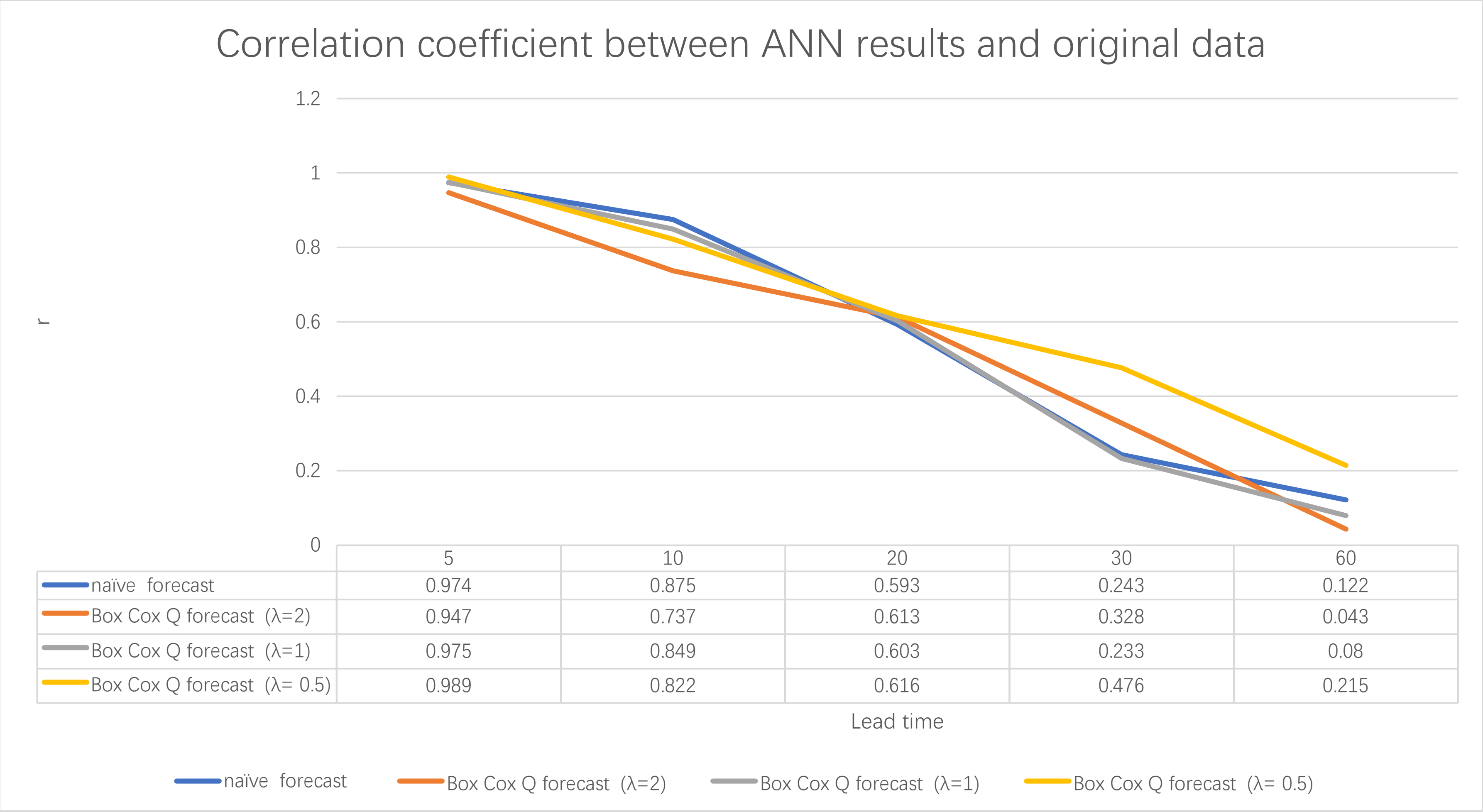
After the two comparisons, it can be concluded that naïve forecast performs well for relatively short lead time and sometimes naïve forecast could outperform more sophisticated methods to a certain extent. However, the limitation of naïve forecast is that it does not have any real forecast ability. The main concept of naïve forecast is using previous trend as next period's forecast. It fully relies on historical data and does not respond to any random variations. Additionally, this forecast method does not reflect any knowledge of the person who performed the forecast. In fact, naïve forecast should be used as a reference in comparing with more complex forecast techniques and should not be used alone.

4.4 Comparing naïve forecast with Box Cox Q forecast

Correlation coefficient between ANN results and original data

naïve forecast		Box Cox Q forecast (λ=2)		Box Cox Q forecast (λ=1)		Box Cox Q forecast (λ= 0.5)	
Lead time	r	Lead time	r	Lead time	r	Lead time	r
5	0.974	5	0.947	5	0.975	5	0.989
10	0.875	10	0.737	10	0.849	10	0.822
20	0.593	20	0.613	20	0.603	20	0.616
30	0.243	30	0.328	30	0.233	30	0.476
60	0.122	60	0.043	60	0.08	60	0.215

Table 6. Correlation coefficient for naïve forecast and Box Cox Q forecast



	5	10	20	30	60
naïve forecast	0.974	0.875	0.593	0.243	0.122
Box Cox Q forecast (λ=2)	0.947	0.737	0.613	0.328	0.043
Box Cox Q forecast (λ=1)	0.975	0.849	0.603	0.233	0.08
Box Cox Q forecast (λ= 0.5)	0.989	0.822	0.616	0.476	0.215

Fig 64. Correlation coefficient vs. lead time for naïve forecast and Box Cox Q forecast

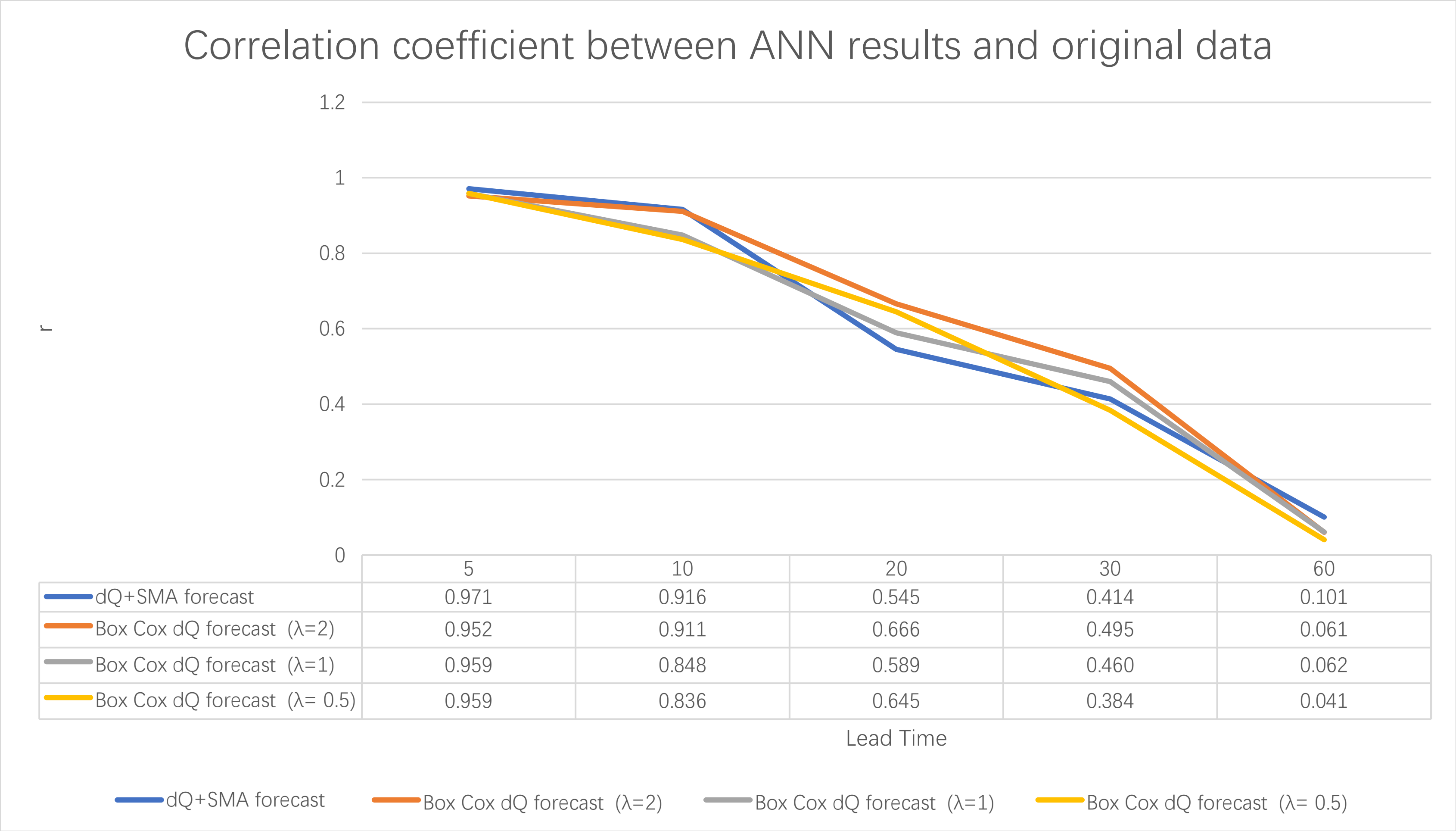
Comparing with naïve forecast, training results after performing Box Cox transformation does have its advantages to a certain extent. Using λ = 1 the neural network shows similar results with naïve forecast while using λ = 2 shows less satisfactory results. The reason could be that the optimal λ value for Box Cox transformation is not close to either 1 or 2. Box Cox transformation using λ = 0.5 shows significant improvement in lead time of 30 and 60 minutes. It suggests that the optimal λ value could possibly be around 0.5. The effort made for Box Cox transformation results in improvement of forecast quality.

4.5 Comparing dQ+SMA forecast with Box Cox dQ forecast

Correlation coefficient between ANN results and original data

dQ+SMA forecast		Box Cox dQ forecast (λ=2)		Box Cox dQ forecast (λ=1)		Box Cox dQ forecast (λ= 0.5)	
Lead time	r	Lead time	r	Lead time	r	Lead time	R
5	0.971	5	0.982	5	0.959	5	0.959
10	0.916	10	0.851	10	0.848	10	0.836
20	0.545	20	0.666	20	0.589	20	0.645
30	0.414	30	0.495	30	0.460	30	0.384
60	0.101	60	0.061	60	0.062	60	0.041

Table 7. Correlation coefficient for naïve forecast and Box Cox dQ forecast



	5	10	20	30	60
dQ+SMA forecast	0.971	0.916	0.545	0.414	0.101
Box Cox dQ forecast (λ=2)	0.952	0.911	0.666	0.495	0.061
Box Cox dQ forecast (λ=1)	0.959	0.848	0.589	0.460	0.062
Box Cox dQ forecast (λ= 0.5)	0.959	0.836	0.645	0.384	0.041

Fig 65. Correlation coefficient vs. lead time for naïve forecast and Box Cox dQ forecast

It can be noticed from Fig 65 that while dQ+SMA forecast does provide good results for lead time of 5 minutes, Box Cox dQ forecast performs better than dQ+SMA forecast in longer lead time. In forecast lead time of 20 minutes, all Box Cox dQ forecast results are better than dQ+SMA forecast.

Taking a closer examination at the individual results, it can be concluded that dQ+SMA and Box Cox dQ methods both have advantages and limitations. For predictions involving SMA, the trend is captured at high accuracy. SMA highlights the trend in data by taking average value from adjacent data points. The trend could be detected by Artificial Neural Network and ANN can make accurate predictions based on the data. However, the limitation of applying SMA is about deciding the window period. In this project, relatively short window periods are used and there is no training attempts made for longer window periods. While using too short window periods could result in losing general trend of the data, using too long window periods could lead to increase in lag of SMA. Therefore, more training attempts could be made in future to determine the suitable window period for SMA in order to achieve more accurate results.

The advantage of using Box-Cox transformation focuses on capturing peak values of discharge. It can be noticed that while dQ+SMA method sometimes fails to capture the peak values of discharge, Box Cox dQ method could detect the peak values at most of the time. Even if at long lead time such as 30 and 60 minutes where dQ+SMA method show overall unsatisfactory prediction results, Box Cox dQ method still has the ability to identify the time periods that the peak values may occur. Identifying peak values is important in flooding prevention as well as in environment protection. If potential flooding could be predicted based on hydrological data, necessary actions could be taken in advance. However, the limitation of using Box Cox transformation is the difficulty in deciding optimal λ value. There could be different λ values for different series of data and the results could be improved if the ANN is trained based on well-normalized data.

The forecast for lead time of 60 minutes is unsatisfactory in all training methods. The reason is that the lead time exceeds the capacity of prediction for this catchment area. The limitation of forecast depends on catchment concentration time, which is the time needed for water to flow from the most remote point from a watershed to the watershed outlet. It is obvious that 60 minutes is larger than catchment concentration time for Kent Ridge Catchment.

5. Summary and future work

Using Artificial Neural Network (ANN) to predict rainfall runoff is introduced in this project. The training of ANN uses hydrological data collected from Kent Ridge Catchment in National University of Singapore. Several data pre-processing techniques are applied on discharge data in various ANN training attempts. Naïve forecast is conducted for reference and comparison. Various training attempts are made and forecasts with relatively satisfactory quality with respect to lead time are achieved. The predicting ability of ANN is limited by catchment concentration time. ANN exhibits better performance when trained with sophisticated data sets such as Simple Moving Average (SMA) and Box Cox Transformation. The comparison made between dQ+SMA method and Box Cox dQ method suggests that each training method has advantages and disadvantages. The ANN could possibly perform better if combination of different methods is made for training, which is not presented in this project.

To conclude, ANN modelling is promising and could be used to construct hydrological models to make accurate predictions. It is widely applicable to catchments with different sizes and climate conditions. Forecast using ANN could help in preventing flooding and protecting ecological environment. The future work will focus on testing the methodology on larger catchment area such as Kallang River in Singapore. Effort will also be made to combine ANN forecasting with linear prediction to obtain accurate forecast results.

6. References

Algorithms for Designing Feedforward Networks. (2015). Feedforward Neural Network Methodology Information Science and Statistics, 129-202. doi:10.1007/0-387-22649-4_5

Babovic, V., & Keijzer, M. (2002). Rainfall Runoff Modelling Based on Genetic Programming. Hydrology Research, 33(5), 331-346. doi:10.2166/nh.2002.0012

Chadalawada, J. & Havlicek, V. & Babovic, V. (2017). A Genetic Programming Approach to System Identification of Rainfall-Runoff Models. Water Resources Management. 10.1007/s11269-017-1719-1.

Cheng, R. (2017). Box-Cox Transformations. Oxford Scholarship Online. doi:10.1093/oso/9780198505044.003.0010

Govindaraju, R. S., & Rao, A. R. (2000). Artificial neural networks in hydrology. Dordrecht: Kluwer Academic.

Hall, S. (2016). Modeling Rainfall Runoff. Eos. doi:10.1029/2016eo062109

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688. doi:10.1016/j.ijforecast.2006.03.001

Liu, J., Savenije, H. H., & Xu, J. (2003). Forecast of water demand in Weinan City in China using WDF-ANN model. Physics and Chemistry of the Earth, Parts A/B/C, 28(4-5), 219-224. doi:10.1016/s1474-7065(03)00026-3

Moving Average. (2005). Dictionary of Statistics & Methodology. doi:10.4135/9781412983907.n1213

Pattie, D. C., & Snyder, J. (1996). Using a neural network to forecast visitor behavior. Annals of Tourism Research, 23(1), 151-164. doi:10.1016/0160-7383(95)00052-6

Rainfall Runoff Modelling. (2012). Flood Risk Assessment and Management, 80-86. doi:10.2174/978160805047511101010080

Reeves, C. R. (1995). Training Set Selection in Neural Network Applications. Artificial Neural Nets and Genetic Algorithms, 476-478. doi:10.1007/978-3-7091-7535-4_123

Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. The Statistician, 41(2), 169. doi:10.2307/2348250

Sun, Y. & Babovic, V. & Soon C, E. (2012). Artificial neural networks as routine for error correction with an application in Singapore regional model. Ocean Dynamics. 62. 10.1007/s10236-012-0524-x.

Selection of Initial Conditions During Neural Network Adjustment — Typical Neural Network Input Signals. (2007). Neural Networks Theory, 207-222. doi:10.1007/978-3-540-48125-6_12

Zhang, T., & Yang, B. (2017). Box–Cox Transformation in Big Data. Technometrics, 59(2), 189-201. doi:10.1080/00401706.2016.1156025

Appendix

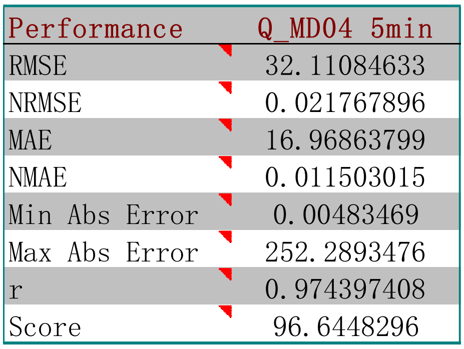
Appendix A: Number of data sets of events

Appendix B: Training results

Appendix A: Number of data sets of events

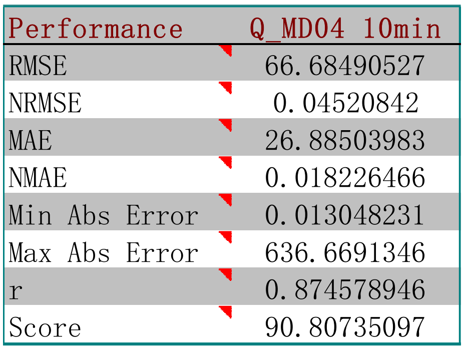
Event number	Number of data sets
1	99
2	82
7	174
8	217
12	183
13	251
14	162
15	501
16	1313
17	483
18	451
19	129
20	59
21	662
22	7
23	125
24	315
25	416
26	24
27	62
28	125
29	94
30	175
31	115
32	120
33	604
34	79
35	36
36	43
37	816
38	128
39	61
40	1403
61	87
62	230
63	135
64	38
65	58
66	56
67	141
75	71

Appendix B: Testing results



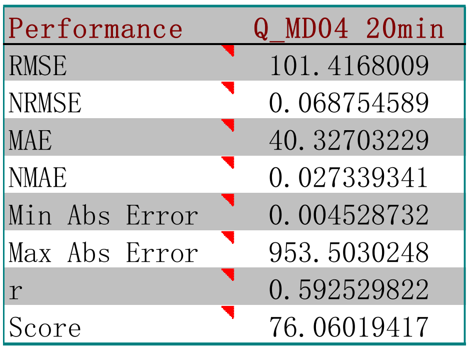
Performance	Q_MD04 5min
RMSE	32.11084633
NRMSE	0.021767896
MAE	16.96863799
NMAE	0.011503015
Min Abs Error	0.00483469
Max Abs Error	252.2893476
r	0.974397408
Score	96.6448296

Q_MD04 5min naive forecast testing results



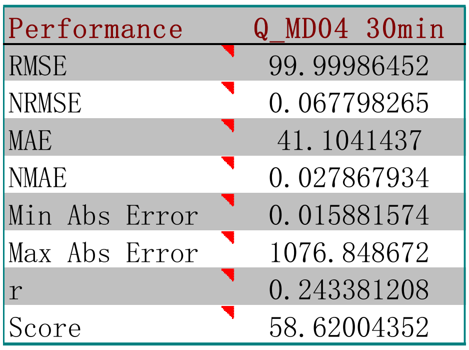
Performance	Q_MD04 10min
RMSE	66.68490527
NRMSE	0.04520842
MAE	26.88503983
NMAE	0.018226466
Min Abs Error	0.013048231
Max Abs Error	636.6691346
r	0.874578946
Score	90.80735097

Q_MD04 10min naive forecast testing results



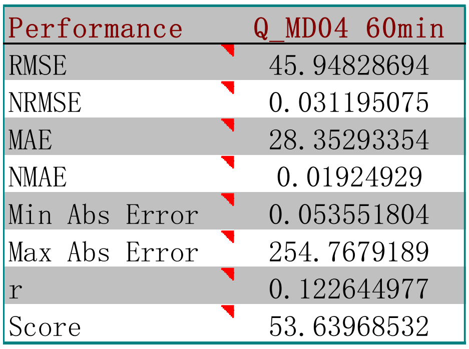
Performance	Q_MD04 20min
RMSE	101.4168009
NRMSE	0.068754589
MAE	40.32703229
NMAE	0.027339341
Min Abs Error	0.004528732
Max Abs Error	953.5030248
r	0.592529822
Score	76.06019417

Q_MD04 20min naive forecast testing results



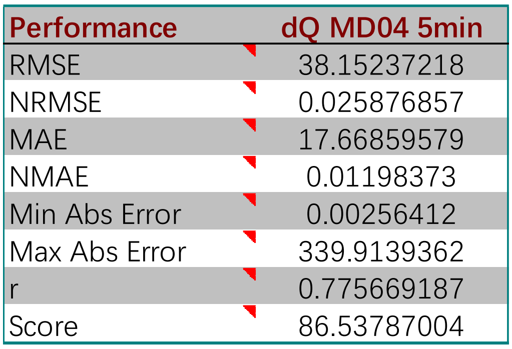
Performance	Q_MD04 30min
RMSE	99.99986452
NRMSE	0.067798265
MAE	41.1041437
NMAE	0.027867934
Min Abs Error	0.015881574
Max Abs Error	1076.848672
r	0.243381208
Score	58.62004352

Q_MD04 30min naive forecast testing results



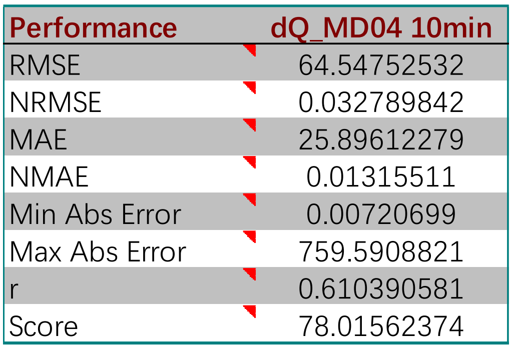
Performance	Q_MD04 60min
RMSE	45. 94828694
NRMSE	0. 031195075
MAE	28. 35293354
NMAE	0. 01924929
Min Abs Error	0. 053551804
Max Abs Error	254. 7679189
r	0. 122644977
Score	53. 63968532

Q_MD04 60min naive forecast testing results



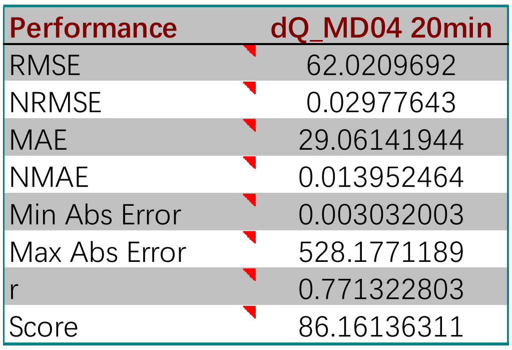
Performance	dQ MD04 5min
RMSE	38.15237218
NRMSE	0.025876857
MAE	17.66859579
NMAE	0.01198373
Min Abs Error	0.00256412
Max Abs Error	339.9139362
r	0.775669187
Score	86.53787004

Q_MD04 5min dQ forecast testing results



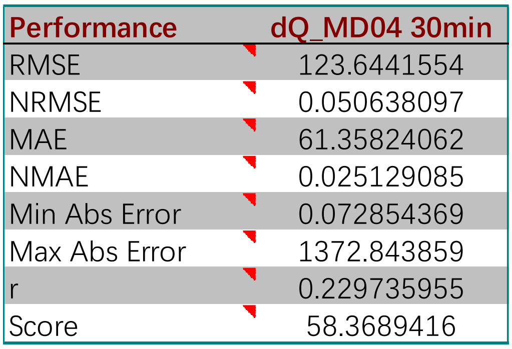
Performance	dQ_MD04 10min
RMSE	64.54752532
NRMSE	0.032789842
MAE	25.89612279
NMAE	0.01315511
Min Abs Error	0.00720699
Max Abs Error	759.5908821
r	0.610390581
Score	78.01562374

Q_MD04 10min dQ forecast testing results



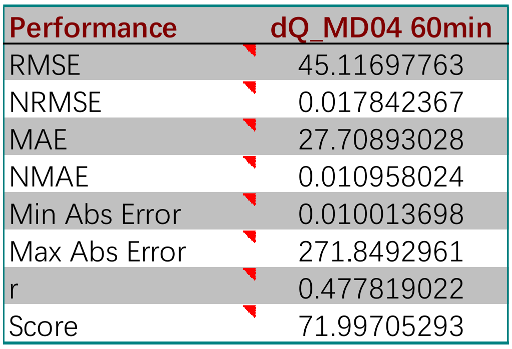
Performance	dQ_MD04 20min
RMSE	62.0209692
NRMSE	0.02977643
MAE	29.06141944
NMAE	0.013952464
Min Abs Error	0.003032003
Max Abs Error	528.1771189
r	0.771322803
Score	86.16136311

Q_MD04 20min dQ forecast testing results



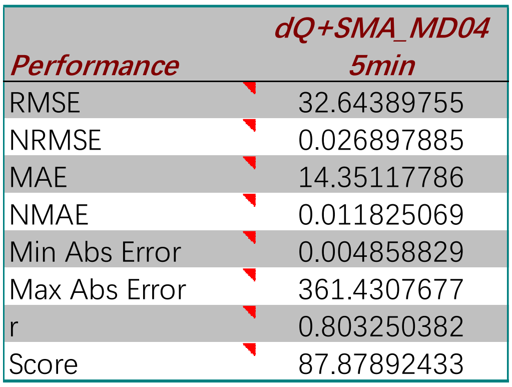
Performance	dQ_MD04 30min
RMSE	123.6441554
NRMSE	0.050638097
MAE	61.35824062
NMAE	0.025129085
Min Abs Error	0.072854369
Max Abs Error	1372.843859
r	0.229735955
Score	58.3689416

Q_MD04 30min dQ forecast testing results



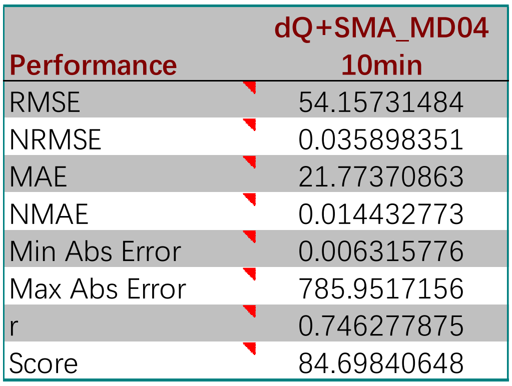
Performance	dQ_MD04 60min
RMSE	45.11697763
NRMSE	0.017842367
MAE	27.70893028
NMAE	0.010958024
Min Abs Error	0.010013698
Max Abs Error	271.8492961
r	0.477819022
Score	71.99705293

Q_MD04 60min dQ forecast testing results



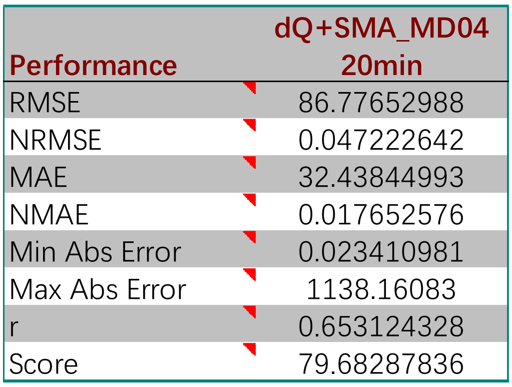
Performance	*dQ+SMA_MD04 5min*
RMSE	32.64389755
NRMSE	0.026897885
MAE	14.35117786
NMAE	0.011825069
Min Abs Error	0.004858829
Max Abs Error	361.4307677
r	0.803250382
Score	87.87892433

Q_MD04 5min dQ+SMA forecast testing results



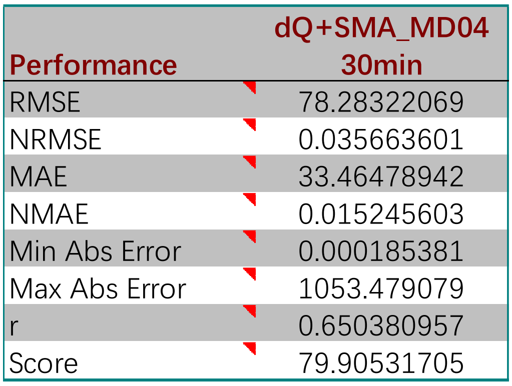
Performance	dQ+SMA_MD04 10min
RMSE	54.15731484
NRMSE	0.035898351
MAE	21.77370863
NMAE	0.014432773
Min Abs Error	0.006315776
Max Abs Error	785.9517156
r	0.746277875
Score	84.69840648

Q_MD04 10min dQ+SMA forecast testing results



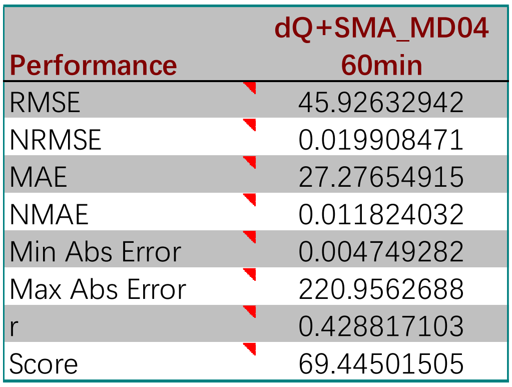
Performance	dQ+SMA_MD04 20min
RMSE	86.77652988
NRMSE	0.047222642
MAE	32.43844993
NMAE	0.017652576
Min Abs Error	0.023410981
Max Abs Error	1138.16083
r	0.653124328
Score	79.68287836

Q_MD04 20min dQ+SMA forecast testing results



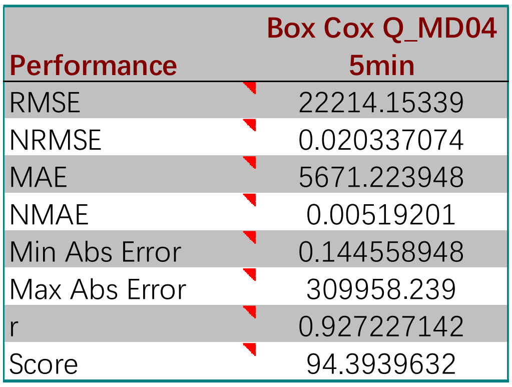
Performance	dQ+SMA_MD04 30min
RMSE	78.28322069
NRMSE	0.035663601
MAE	33.46478942
NMAE	0.015245603
Min Abs Error	0.000185381
Max Abs Error	1053.479079
r	0.650380957
Score	79.90531705

Q_MD04 30min dQ+SMA forecast testing results



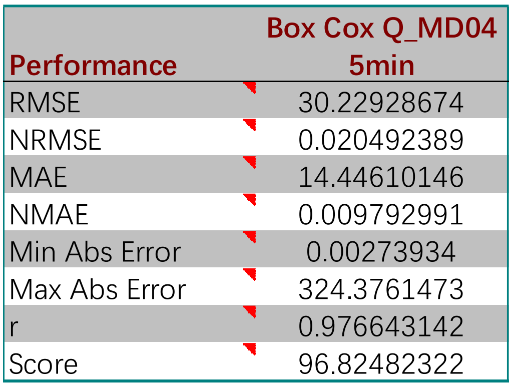
Performance	dQ+SMA_MD04 60min
RMSE	45.92632942
NRMSE	0.019908471
MAE	27.27654915
NMAE	0.011824032
Min Abs Error	0.004749282
Max Abs Error	220.9562688
r	0.428817103
Score	69.44501505

Q_MD04 60min dQ+SMA forecast testing results



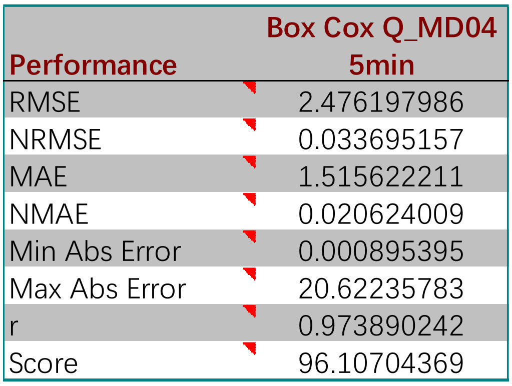
Performance	Box Cox Q_MD04 5min
RMSE	22214.15339
NRMSE	0.020337074
MAE	5671.223948
NMAE	0.00519201
Min Abs Error	0.144558948
Max Abs Error	309958.239
r	0.927227142
Score	94.3939632

Q_MD04 5min Box Cox Q forecast results (λ = 2)



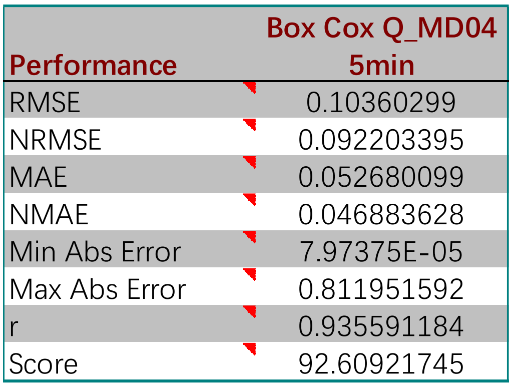
Performance	Box Cox Q_MD04 5min
RMSE	30.22928674
NRMSE	0.020492389
MAE	14.44610146
NMAE	0.009792991
Min Abs Error	0.00273934
Max Abs Error	324.3761473
r	0.976643142
Score	96.82482322

Q_MD04 5min Box Cox Q forecast results (λ = 1)



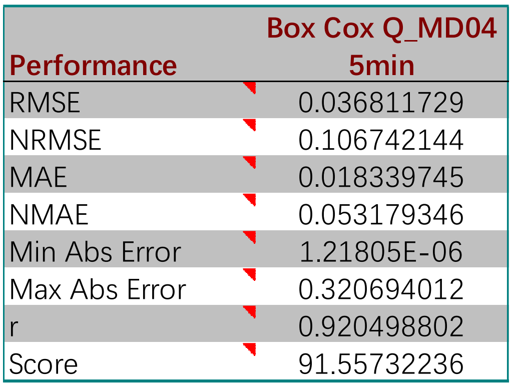
Performance	Box Cox Q_MD04 5min
RMSE	2.476197986
NRMSE	0.033695157
MAE	1.515622211
NMAE	0.020624009
Min Abs Error	0.000895395
Max Abs Error	20.62235783
r	0.973890242
Score	96.10704369

Q_MD04 5min Box Cox Q forecast results (λ = 0.5)



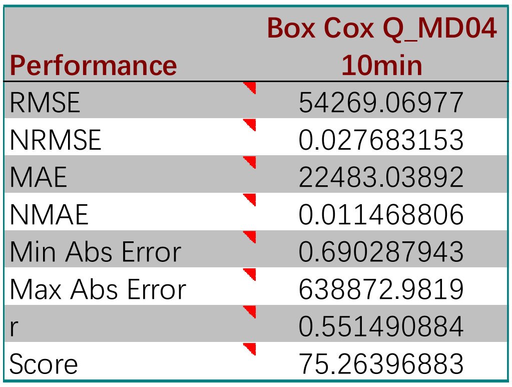
Performance	Box Cox Q_MD04 5min
RMSE	0.10360299
NRMSE	0.092203395
MAE	0.052680099
NMAE	0.046883628
Min Abs Error	7.97375E-05
Max Abs Error	0.811951592
r	0.935591184
Score	92.60921745

Q_MD04 5min Box Cox Q forecast results (λ = -0.5)



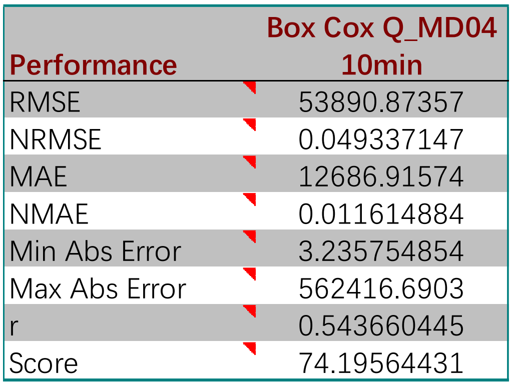
Performance	Box Cox Q_MD04 5min
RMSE	0.036811729
NRMSE	0.106742144
MAE	0.018339745
NMAE	0.053179346
Min Abs Error	1.21805E-06
Max Abs Error	0.320694012
r	0.920498802
Score	91.55732236

Q_MD04 5min Box Cox Q forecast results (λ = -1)



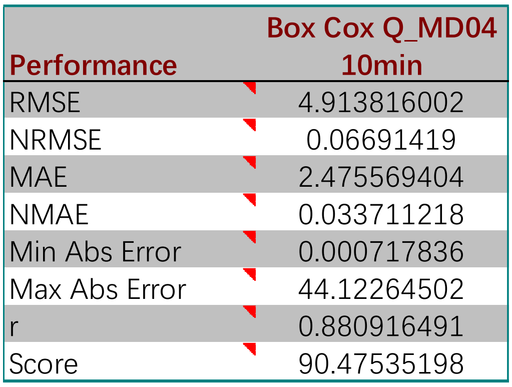
Performance	Box Cox Q_MD04 10min
RMSE	54269.06977
NRMSE	0.027683153
MAE	22483.03892
NMAE	0.011468806
Min Abs Error	0.690287943
Max Abs Error	638872.9819
r	0.551490884
Score	75.26396883

Q_MD04 10min Box Cox Q forecast testing results (λ = 2)



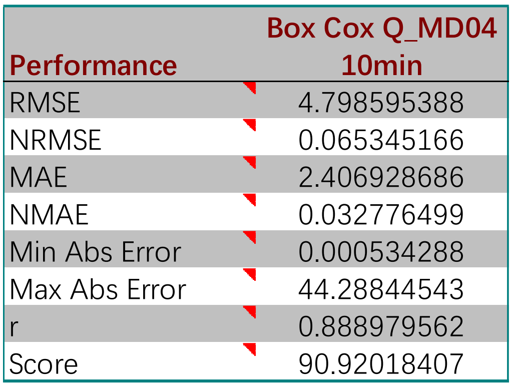
Performance	Box Cox Q_MD04 10min
RMSE	53890.87357
NRMSE	0.049337147
MAE	12686.91574
NMAE	0.011614884
Min Abs Error	3.235754854
Max Abs Error	562416.6903
r	0.543660445
Score	74.19564431

Q_MD04 10min Box Cox Q forecast testing results (λ = 1)



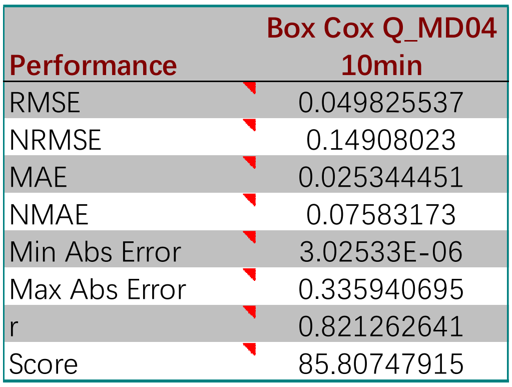
Performance	Box Cox Q_MD04 10min
RMSE	4.913816002
NRMSE	0.06691419
MAE	2.475569404
NMAE	0.033711218
Min Abs Error	0.000717836
Max Abs Error	44.12264502
r	0.880916491
Score	90.47535198

Q_MD04 10min Box Cox Q forecast testing results (λ = 0.5)



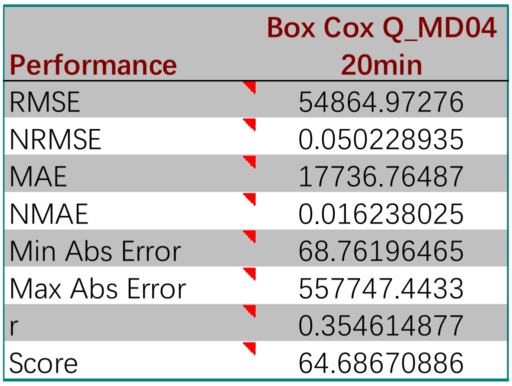
Performance	Box Cox Q_MD04 10min
RMSE	4.798595388
NRMSE	0.065345166
MAE	2.406928686
NMAE	0.032776499
Min Abs Error	0.000534288
Max Abs Error	44.28844543
r	0.888979562
Score	90.92018407

Q_MD04 10min Box Cox Q forecast testing results (λ = -0.5)



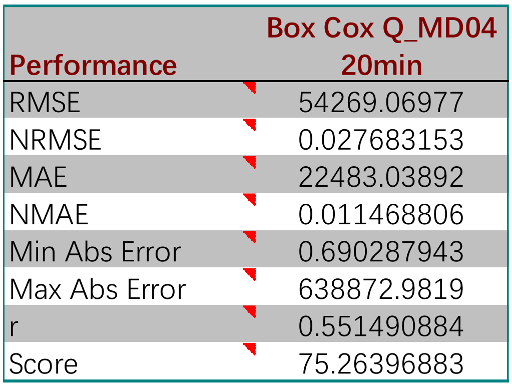
Performance	Box Cox Q_MD04 10min
RMSE	0.049825537
NRMSE	0.14908023
MAE	0.025344451
NMAE	0.07583173
Min Abs Error	3.02533E-06
Max Abs Error	0.335940695
r	0.821262641
Score	85.80747915

Q_MD04 10min Box Cox Q forecast testing results (λ = -1)



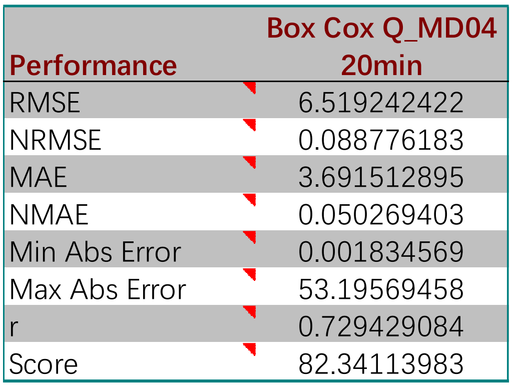
Performance	Box Cox Q_MD04 20min
RMSE	54864.97276
NRMSE	0.050228935
MAE	17736.76487
NMAE	0.016238025
Min Abs Error	68.76196465
Max Abs Error	557747.4433
r	0.354614877
Score	64.68670886

Q_MD04 20min Box Cox Q forecast testing results (λ = 2)



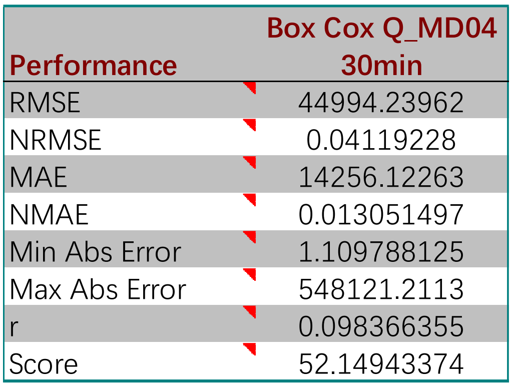
Performance	Box Cox Q_MD04 20min
RMSE	54269.06977
NRMSE	0.027683153
MAE	22483.03892
NMAE	0.011468806
Min Abs Error	0.690287943
Max Abs Error	638872.9819
r	0.551490884
Score	75.26396883

Q_MD04 20min Box Cox Q forecast testing results (λ = 1)



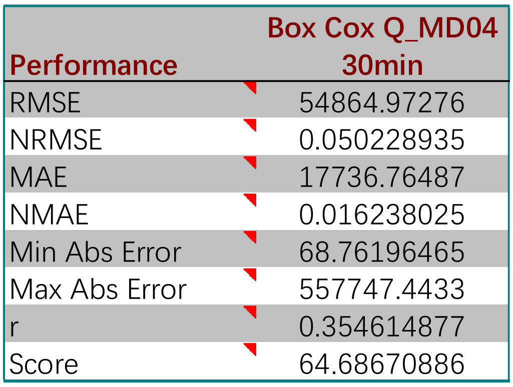
Performance	Box Cox Q_MD04 20min
RMSE	6.519242422
NRMSE	0.088776183
MAE	3.691512895
NMAE	0.050269403
Min Abs Error	0.001834569
Max Abs Error	53.19569458
r	0.729429084
Score	82.34113983

Q_MD04 20min Box Cox Q forecast testing results (λ = 0.5)



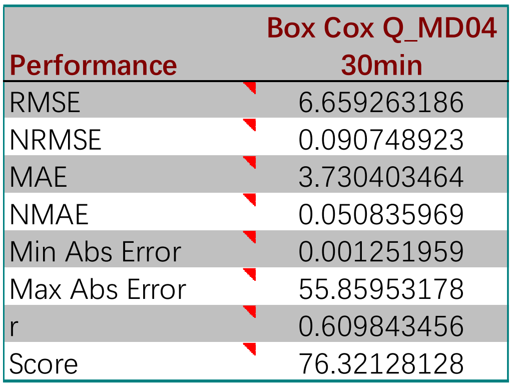
Performance	Box Cox Q_MD04 30min
RMSE	44994.23962
NRMSE	0.04119228
MAE	14256.12263
NMAE	0.013051497
Min Abs Error	1.109788125
Max Abs Error	548121.2113
r	0.098366355
Score	52.14943374

Q_MD04 30min Box Cox Q forecast testing results (λ = 2)



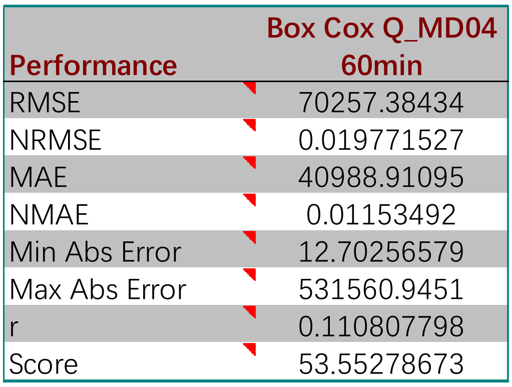
Performance	Box Cox Q_MD04 30min
RMSE	54864.97276
NRMSE	0.050228935
MAE	17736.76487
NMAE	0.016238025
Min Abs Error	68.76196465
Max Abs Error	557747.4433
r	0.354614877
Score	64.68670886

Q_MD04 30min Box Cox Q forecast testing results (λ = 1)



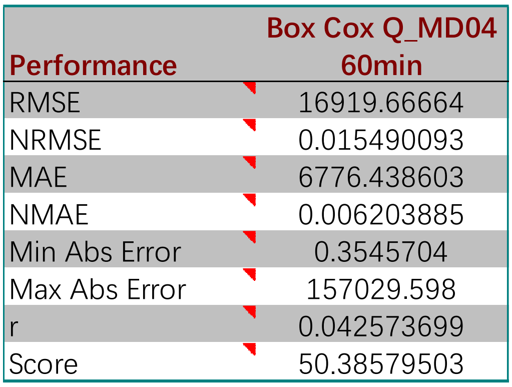
Performance	Box Cox Q_MD04 30min
RMSE	6.659263186
NRMSE	0.090748923
MAE	3.730403464
NMAE	0.050835969
Min Abs Error	0.001251959
Max Abs Error	55.85953178
r	0.609843456
Score	76.32128128

Q_MD04 30min Box Cox Q forecast testing results (λ = 0.5)



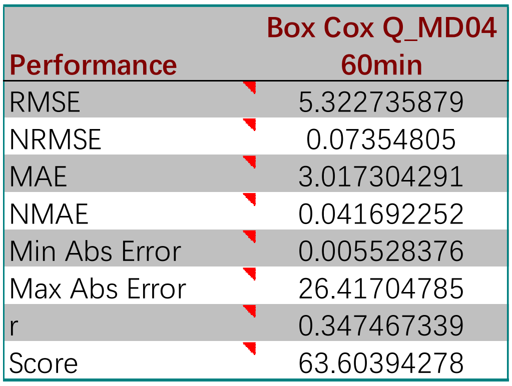
Performance	Box Cox Q_MD04 60min
RMSE	70257.38434
NRMSE	0.019771527
MAE	40988.91095
NMAE	0.01153492
Min Abs Error	12.70256579
Max Abs Error	531560.9451
r	0.110807798
Score	53.55278673

Q_MD04 60min Box Cox Q forecast testing results (λ =2)



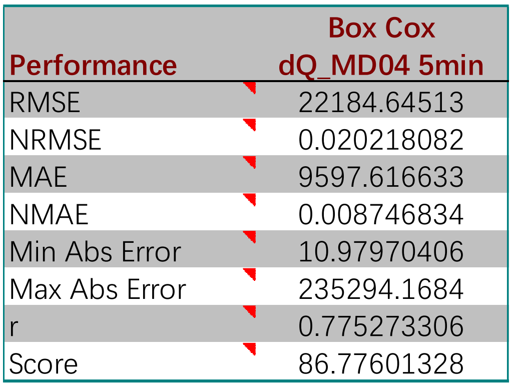
Performance	Box Cox Q_MD04 60min
RMSE	16919.66664
NRMSE	0.015490093
MAE	6776.438603
NMAE	0.006203885
Min Abs Error	0.3545704
Max Abs Error	157029.598
r	0.042573699
Score	50.38579503

Q_MD04 60min Box Cox Q forecast testing results (λ = 1)



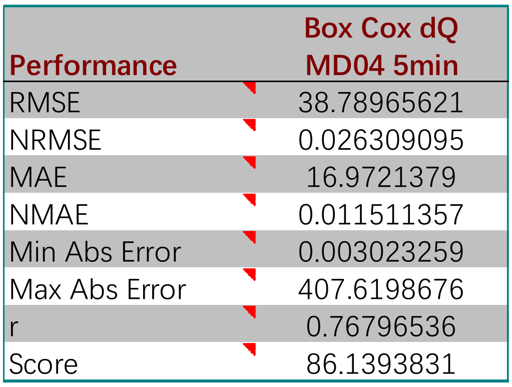
Performance	Box Cox Q_MD04 60min
RMSE	5.322735879
NRMSE	0.07354805
MAE	3.017304291
NMAE	0.041692252
Min Abs Error	0.005528376
Max Abs Error	26.41704785
r	0.347467339
Score	63.60394278

Q_MD04 60min Box Cox Q forecast testing results (λ = 0.5)



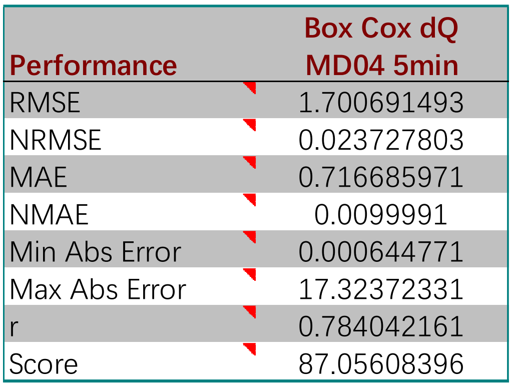
Performance	Box Cox dQ_MD04 5min
RMSE	22184.64513
NRMSE	0.020218082
MAE	9597.616633
NMAE	0.008746834
Min Abs Error	10.97970406
Max Abs Error	235294.1684
r	0.775273306
Score	86.77601328

Q_MD04 5min Box Cox dQ forecast testing results (λ =2)



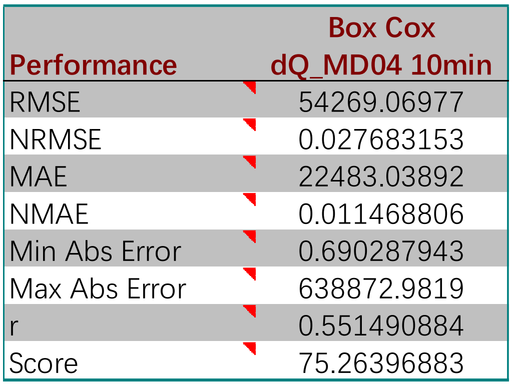
Performance	Box Cox dQ MD04 5min
RMSE	38.78965621
NRMSE	0.026309095
MAE	16.9721379
NMAE	0.011511357
Min Abs Error	0.003023259
Max Abs Error	407.6198676
r	0.76796536
Score	86.1393831

Q_MD04 5min Box Cox dQ forecast testing results (λ =1)



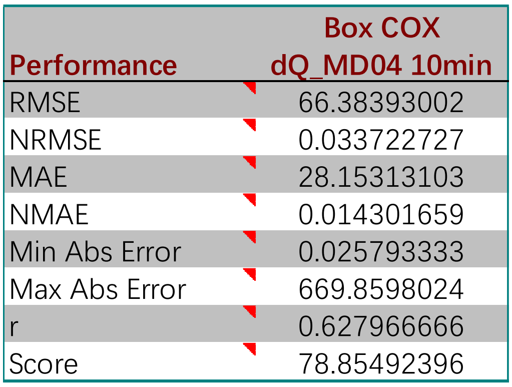
Performance	Box Cox dQ MD04 5min
RMSE	1.700691493
NRMSE	0.023727803
MAE	0.716685971
NMAE	0.0099991
Min Abs Error	0.000644771
Max Abs Error	17.32372331
r	0.784042161
Score	87.05608396

Q_MD04 5min Box Cox dQ forecast testing results (λ =0.5)



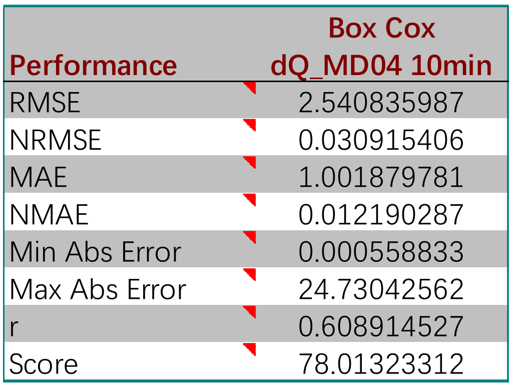
Performance	Box Cox dQ_MD04 10min
RMSE	54269.06977
NRMSE	0.027683153
MAE	22483.03892
NMAE	0.011468806
Min Abs Error	0.690287943
Max Abs Error	638872.9819
r	0.551490884
Score	75.26396883

Q_MD04 10min Box Cox dQ forecast testing results (λ =2)



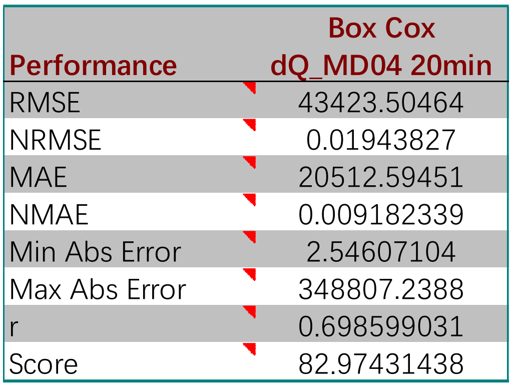
Performance	Box COX dQ_MD04 10min
RMSE	66.38393002
NRMSE	0.033722727
MAE	28.15313103
NMAE	0.014301659
Min Abs Error	0.025793333
Max Abs Error	669.8598024
r	0.627966666
Score	78.85492396

Q_MD04 10min Box Cox dQ forecast testing results (λ =1)



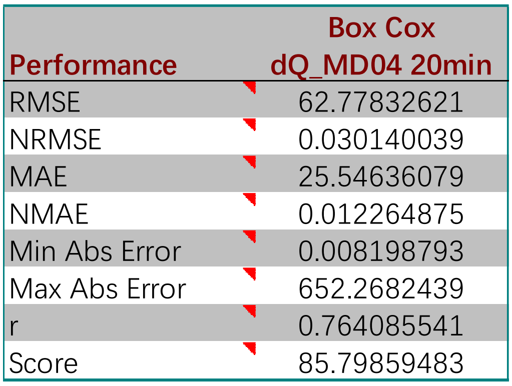
Performance	Box Cox dQ_MD04 10min
RMSE	2.540835987
NRMSE	0.030915406
MAE	1.001879781
NMAE	0.012190287
Min Abs Error	0.000558833
Max Abs Error	24.73042562
r	0.608914527
Score	78.01323312

Q_MD04 10min Box Cox dQ forecast testing results (λ =0.5)



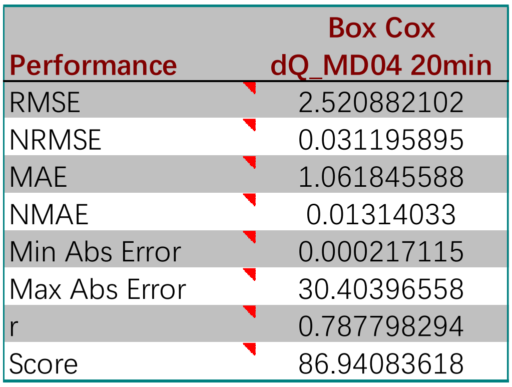
Performance	Box Cox dQ_MD04 20min
RMSE	43423.50464
NRMSE	0.01943827
MAE	20512.59451
NMAE	0.009182339
Min Abs Error	2.54607104
Max Abs Error	348807.2388
r	0.698599031
Score	82.97431438

Q_MD04 20min Box Cox dQ forecast testing results (λ =2)



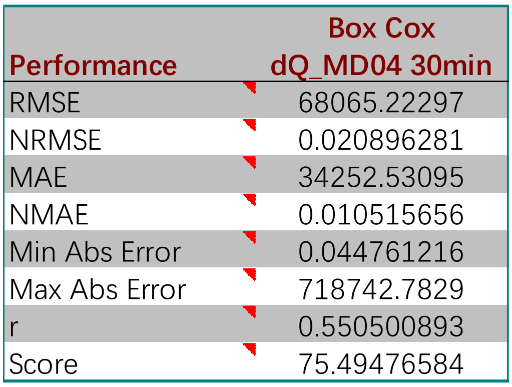
Performance	Box Cox dQ_MD04 20min
RMSE	62.77832621
NRMSE	0.030140039
MAE	25.54636079
NMAE	0.012264875
Min Abs Error	0.008198793
Max Abs Error	652.2682439
r	0.764085541
Score	85.79859483

Q_MD04 20min Box Cox dQ forecast testing results (λ =1)



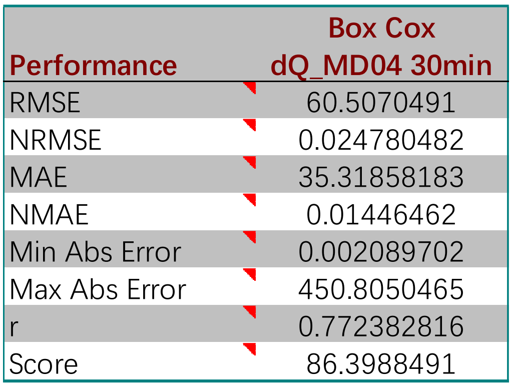
Performance	Box Cox dQ_MD04 20min
RMSE	2.520882102
NRMSE	0.031195895
MAE	1.061845588
NMAE	0.01314033
Min Abs Error	0.000217115
Max Abs Error	30.40396558
r	0.787798294
Score	86.94083618

Q_MD04 20min Box Cox dQ forecast testing results (λ =0.5)



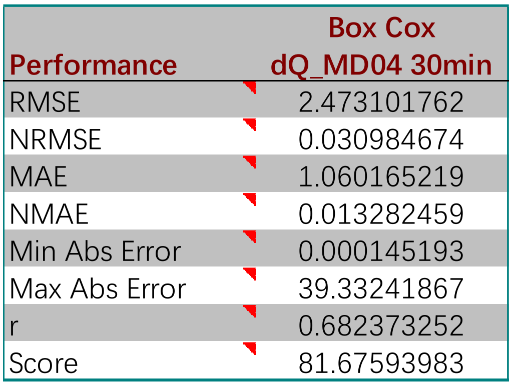
Performance	Box Cox dQ_MD04 30min
RMSE	68065.22297
NRMSE	0.020896281
MAE	34252.53095
NMAE	0.010515656
Min Abs Error	0.044761216
Max Abs Error	718742.7829
r	0.550500893
Score	75.49476584

Q_MD04 30min Box Cox dQ forecast testing results (λ =2)



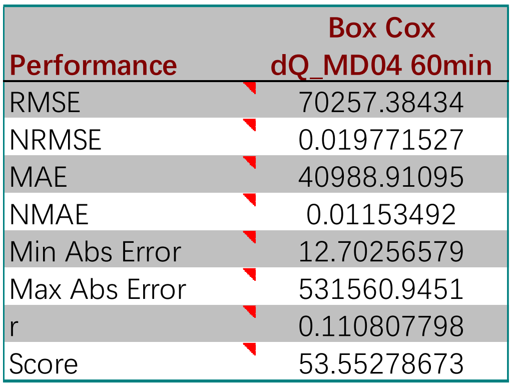
Performance	Box Cox dQ_MD04 30min
RMSE	60.5070491
NRMSE	0.024780482
MAE	35.31858183
NMAE	0.01446462
Min Abs Error	0.002089702
Max Abs Error	450.8050465
r	0.772382816
Score	86.3988491

Q_MD04 30min Box Cox dQ forecast testing results (λ =1)



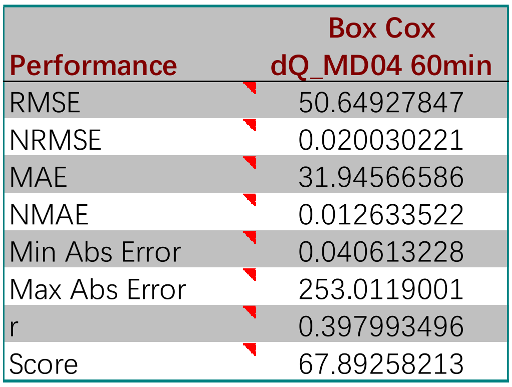
Performance	Box Cox dQ_MD04 30min
RMSE	2.473101762
NRMSE	0.030984674
MAE	1.060165219
NMAE	0.013282459
Min Abs Error	0.000145193
Max Abs Error	39.33241867
r	0.682373252
Score	81.67593983

Q_MD04 30min Box Cox dQ forecast testing results (λ = 0.5)



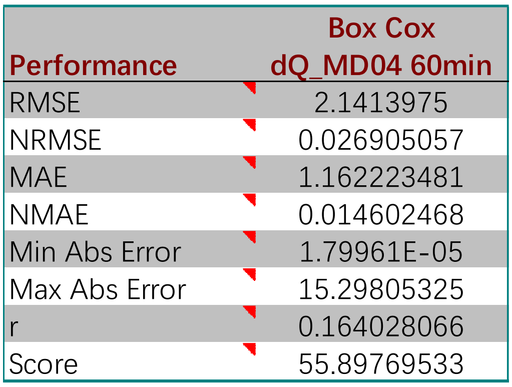
Performance	Box Cox dQ_MD04 60min
RMSE	70257.38434
NRMSE	0.019771527
MAE	40988.91095
NMAE	0.01153492
Min Abs Error	12.70256579
Max Abs Error	531560.9451
r	0.110807798
Score	53.55278673

Q_MD04 60min Box Cox dQ forecast testing results (λ =2)



Performance	Box Cox dQ_MD04 60min
RMSE	50.64927847
NRMSE	0.020030221
MAE	31.94566586
NMAE	0.012633522
Min Abs Error	0.040613228
Max Abs Error	253.0119001
r	0.397993496
Score	67.89258213

Q_MD04 60min Box Cox dQ forecast testing results (λ =1)



	Box Cox
Performance	dQ_MD04 60min
RMSE	2.1413975
NRMSE	0.026905057
MAE	1.162223481
NMAE	0.014602468
Min Abs Error	1.79961E-05
Max Abs Error	15.29805325
r	0.164028066
Score	55.89769533

Q_MD04 60min Box Cox dQ forecast testing results (λ =0.5)