Supplementary document for

A framework for timely and accessible long-term forecasting of shale gas production based on time series pattern matching

submitted to

International Journal of Forecasting

Author information concealed for double-blind review

June 24, 2024

Contents

1 Supplementary document for the beginning of Section 3 in the original article

Table 1: Hyperparameter optimisation results.

Split	# LSTM neurons	# of FC neurons	Learning rate
1	200	50	0.001
2	175	40	0.001
3	150	50	0.01
4	75	20	0.001
5	175	40	0.001
6	125	30	0.01
7	75	10	0.01
8	200	10	0.001
9	125	20	0.001
10	175	20	0.01

2 Supplementary document for Section 3.3 in the original article

Figure 1 and Figure 2 demonstrate the comparison of OMAPE, OME, and OMR between the proposed method and the baseline methods for different M and w_3 settings. These figures expand Figure 9 in the original article. Figure 3 and Figure 4 demonstrate the distributions of the percentage that the proposed method has lower MAPE, MAE, and MR than the baseline methods for different M and w_3 settings. These figures expand Figure 10 in the original article.

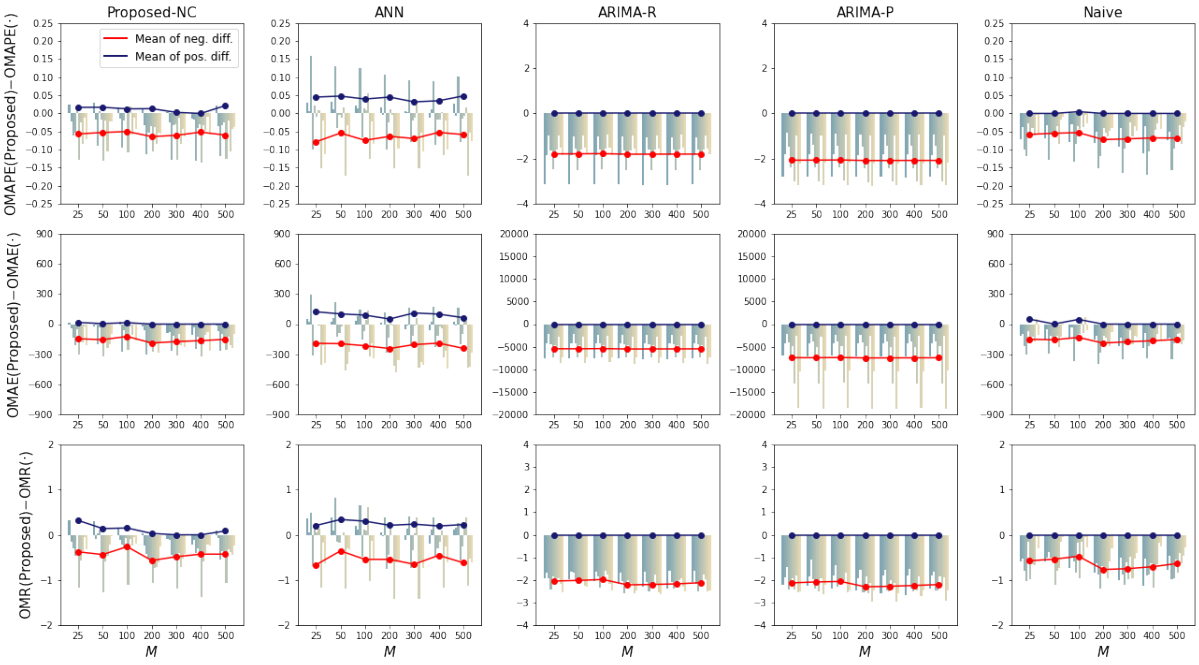

Figure 1: The differences in OMAPE, OMAE, and OMR between the proposed method and the baseline methods for each data split when changing M. Each bar represents a data split. The height of the bar represents the magnitude of the difference. The negative difference indicates the proposed method is better.

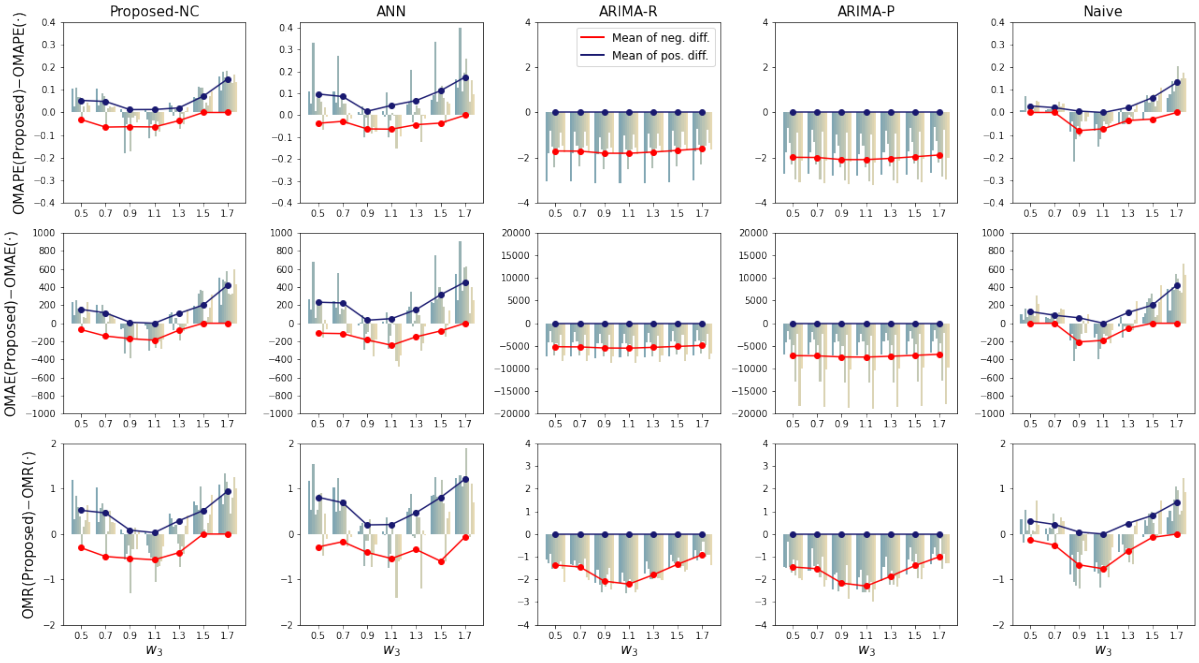

Figure 2: The differences in OMAPE, OMAE, and OMR between the proposed method and the baseline methods for each data split when changing w_3. Each bar represents a data split. The height of the bar represents the magnitude of the difference. The negative difference indicates the proposed method is better.

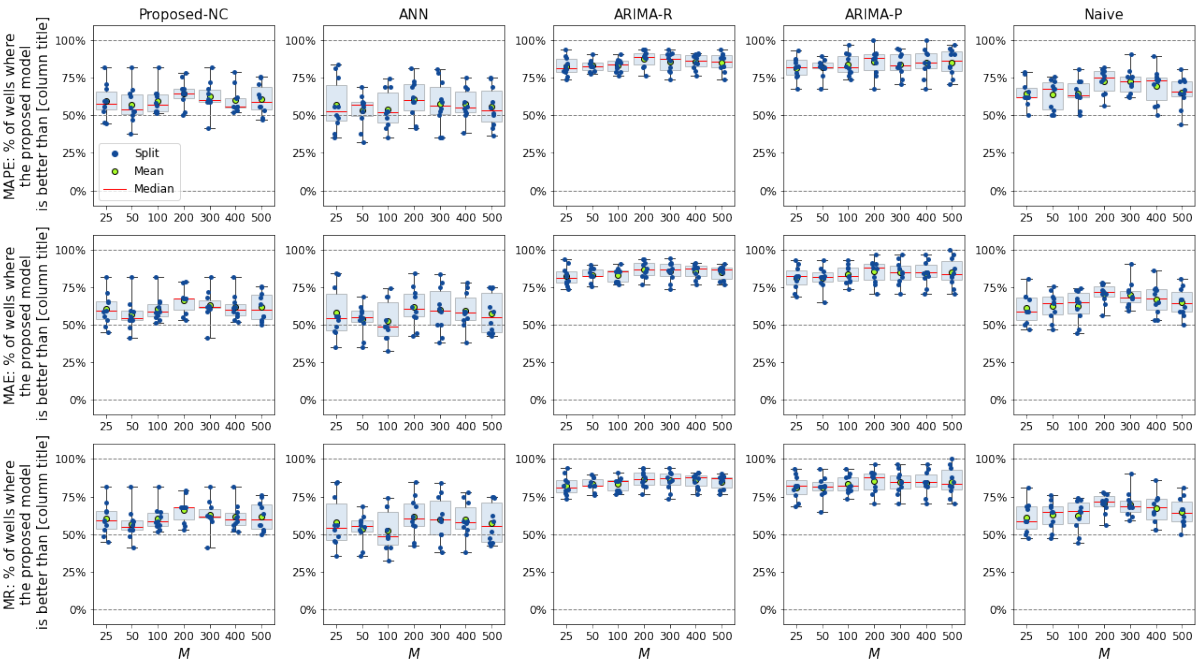

Figure 3: Distributions of the percentage that the proposed method has lower MAPE, MAE, or MR than the baseline methods for each data split when changing M.

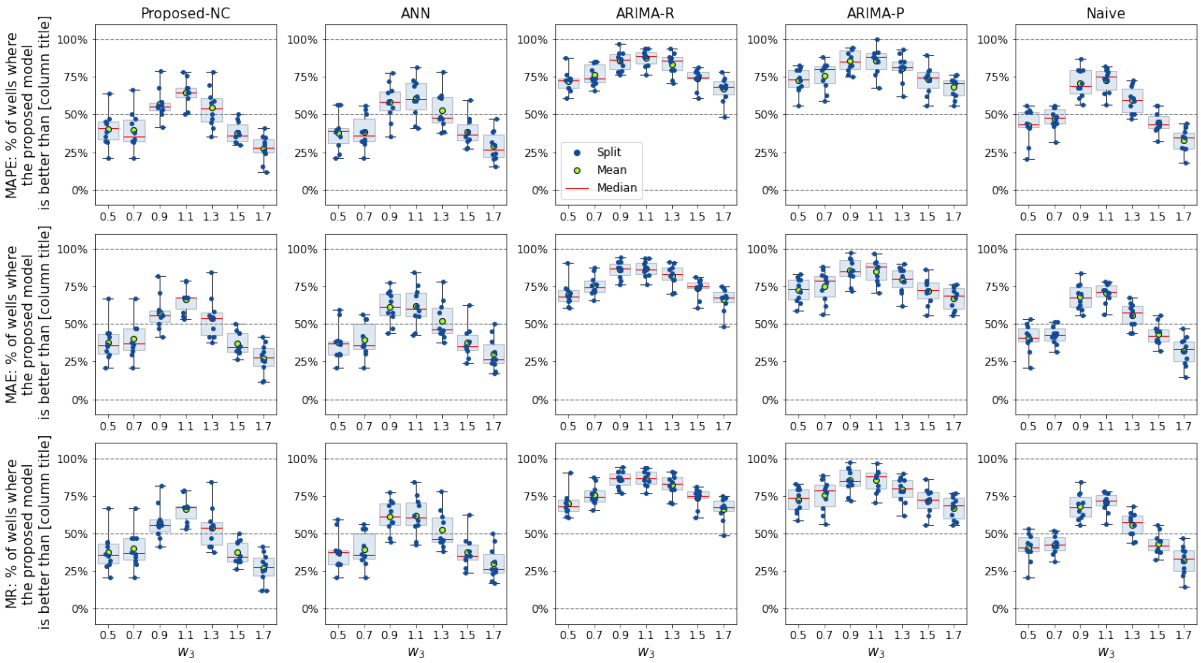

Figure 4: Distributions of the percentage that the proposed method has lower MAPE, MAE, or MR than the baseline methods for each data split when changing w_3.

3 Supplementary document for Section 3.4 in the original article

The following four figures displays the comparison of each method listed in Table 4 in the original article with the proposed method with $l_q = 21$. All OMR and MR values are calculated from the 14 methods listed in Table 4 in the original article.

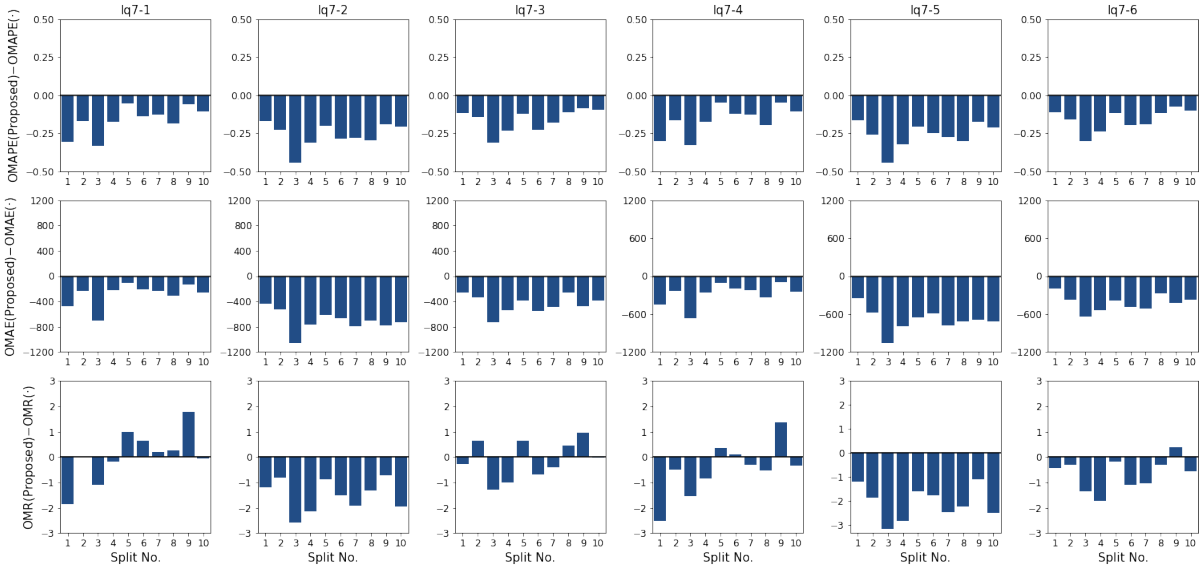

Figure 5: The differences in OMAPE, OMAE, and OMR between the proposed method with $l_q = 21$ and the proposed method with $l_q = 7$. Each bar represents a data split. The height of the bar represents the magnitude of the difference. The negative difference indicates the proposed method with $l_q = 21$ is better.

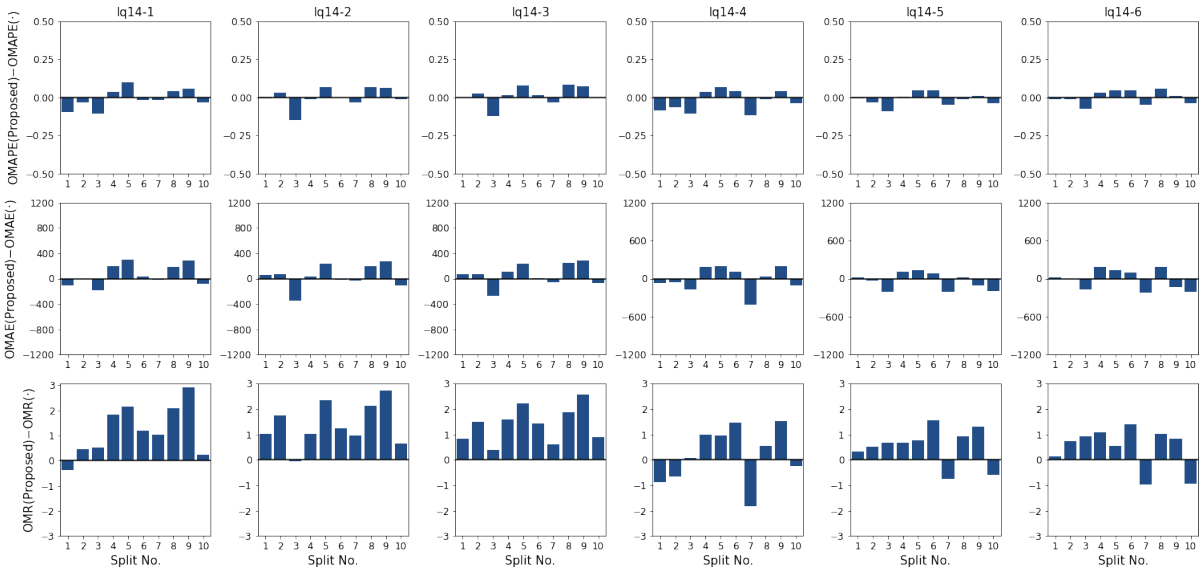

Figure 6: The differences in OMAPE, OMAE, and OMR between the proposed method with $l_q = 21$ and the proposed method with $l_q = 14$. Each bar represents a data split. The height of the bar represents the magnitude of the difference. The negative difference indicates the proposed method with $l_q = 21$ is better.

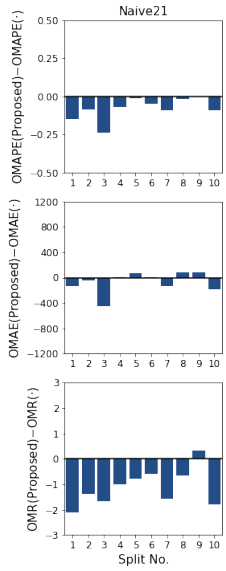

Figure 7: The differences in OMAPE, OMAE, and OMR between the proposed method with $l_q = 21$ and Naive21. Each bar represents a data split. The height of the bar represents the magnitude of the difference. The negative difference indicates the proposed method with $l_q = 21$ is better.

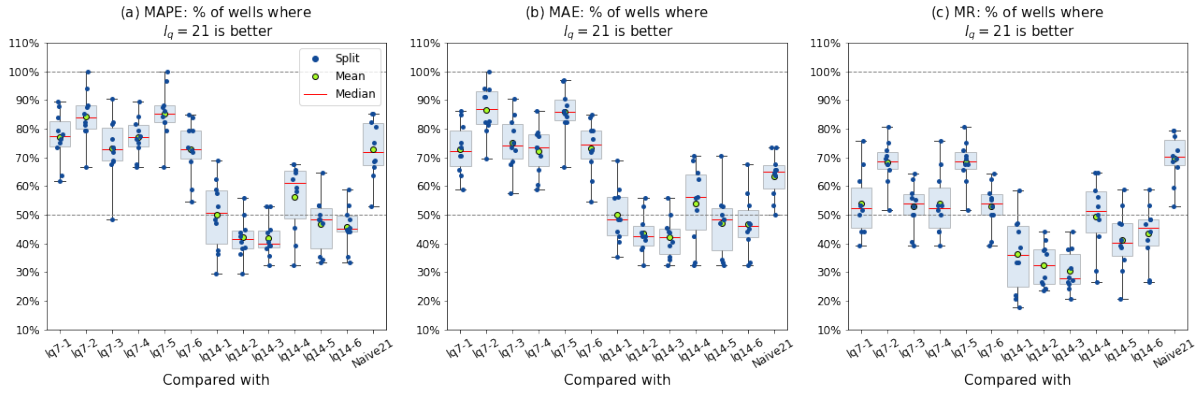

Figure 8: Distributions of the percentage that the proposed method with $l_q = 21$ has lower MAPE, MAE, or MR than the proposed method with $l_q = 7$, the proposed method with $l_q = 14$, and Naive21.

4 Supplementary document for Section 3.6.1 in the original article

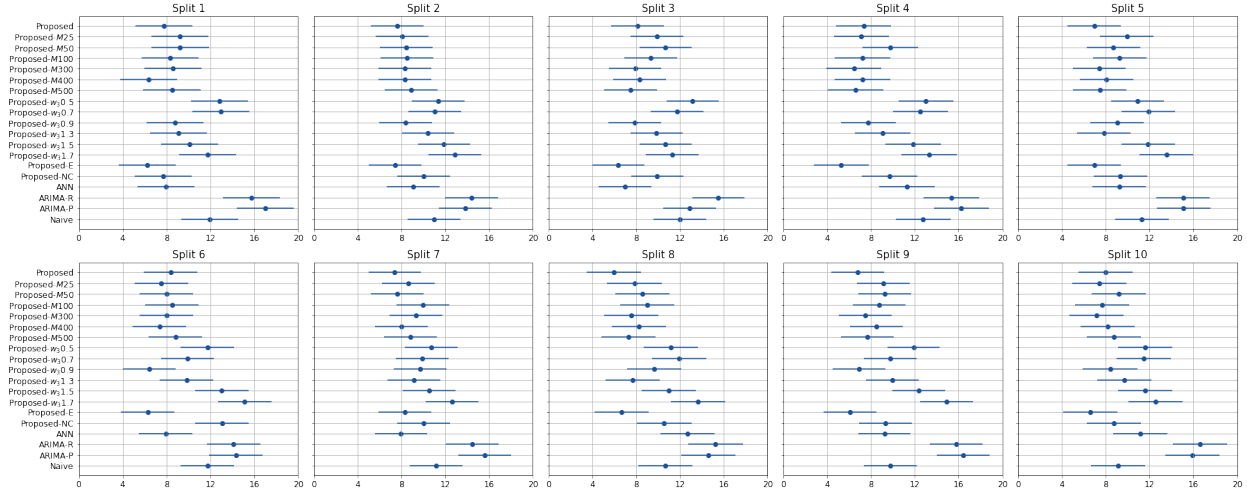

Figure 9: Nemenyi test results for the 19 methods listed in Table 6 in the original article with regard to MAPE under different data splits. Scatters stand for the average rank of the respective method with regard to MAPE. Lines stand for CD intervals.

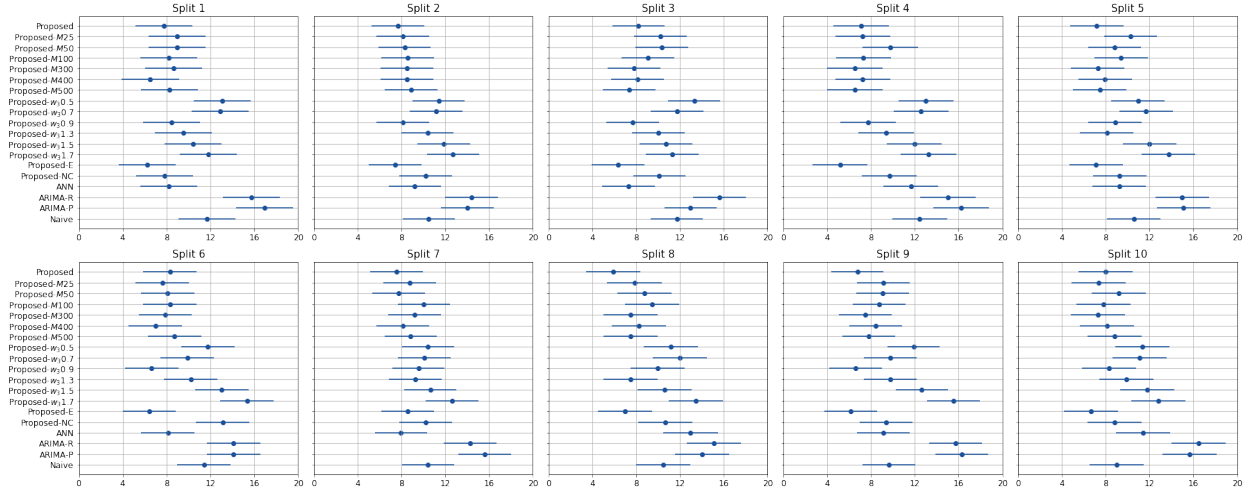

Figure 10: Nemenyi test results for the 19 methods listed in Table 6 in the original article with regard to MAE under different data splits. Scatters stand for the average rank of the respective method with regard to MAE. Lines stand for CD intervals.

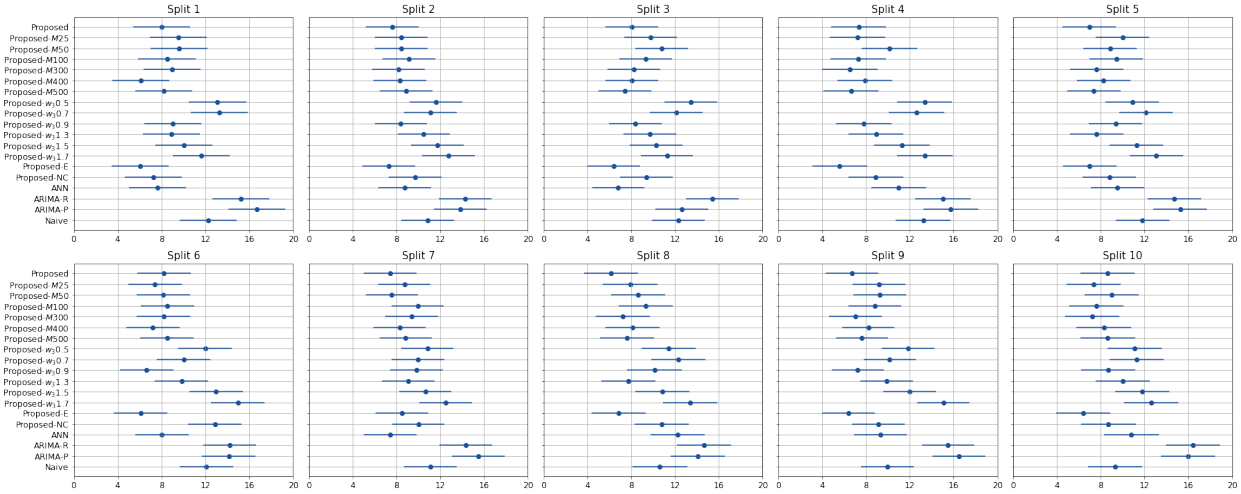

Figure 11: Nemenyi test results for the 19 methods listed in Table 6 in the original article with regard to MR under different data splits. Scatters stand for the average rank of the respective method with regard to MR. Lines stand for CD intervals.

5 Supplementary document for Section 3.6.2 in the original article

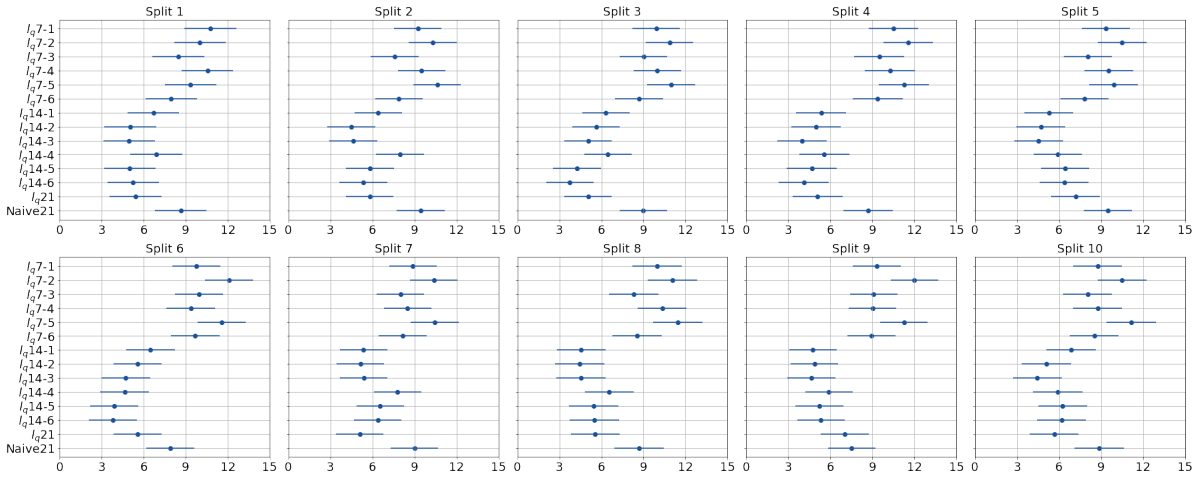

Figure 12: Nemenyi test results for the proposed method with different l_q values with regard to MAPE under different data splits. Scatters stand for the average rank of the respective method with regard to MAPE. Lines stand for CD intervals.

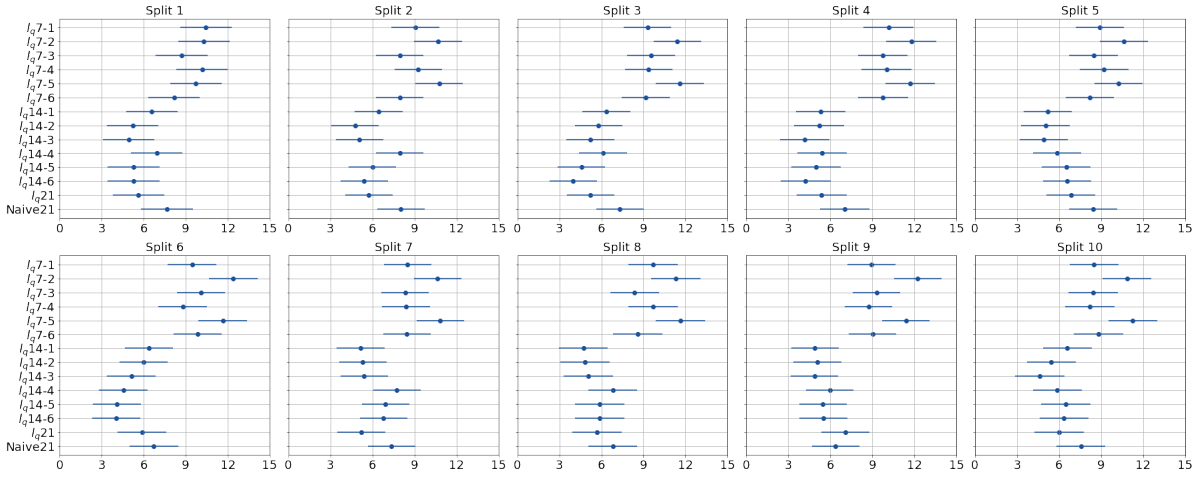

Figure 13: Nemenyi test results for the proposed method with different l_q values and Naive21 with regard to MAE under different data splits. Scatters stand for the average rank of the respective method with regard to MAE. Lines stand for CD intervals.

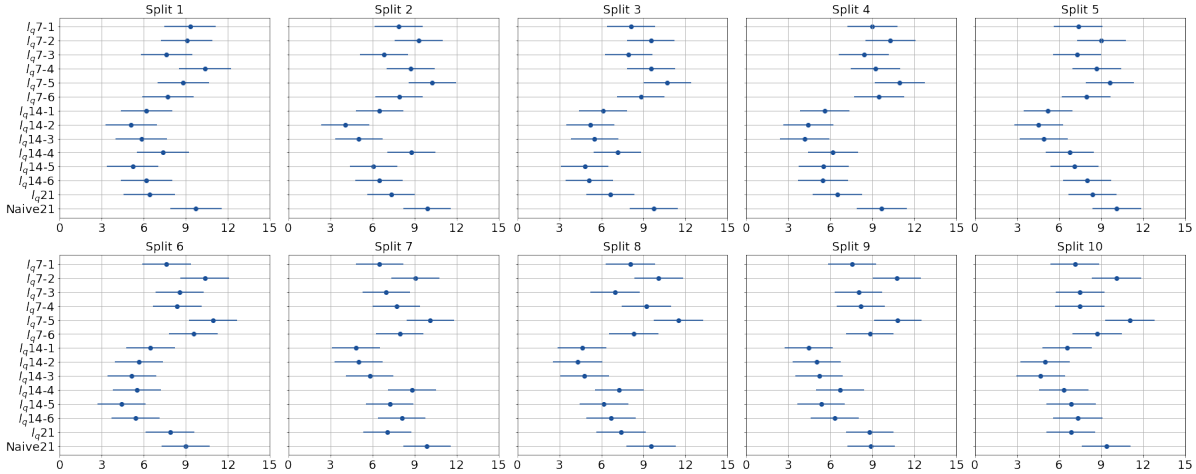

Figure 14: Nemenyi test results for the proposed method with different l_q values and Naive21 with regard to MR under different data splits. Scatters stand for the average rank of the respective method with regard to MR. Lines stand for CD intervals.

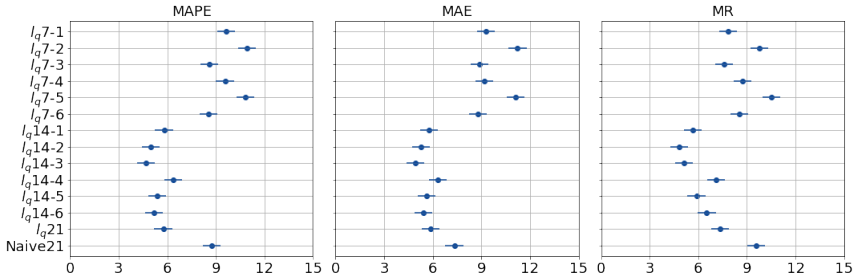

Figure 15: Nemenyi test results for the proposed method with different l_q values and Naive21 with regard to MAPE, MAE, and MR for all test data. Scatters stand for the average rank of the respective method with regard to the respective metric. Lines stand for CD intervals.