International Journal of Forecasting 36 (2020) 191–194

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Discussion

Forecasting with high frequency data: M4 competition and beyond

Tao Hong

Systems Engineering and Engineering Management, University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, United States

A B S T R A C T

The M4 competition included 100,000 time series, with the frequencies ranging from yearly to hourly. The team rankings differ notably across frequencies for both point and probabilistic forecasting. I discuss the performances of these methods, with an emphasis on the hourly series of the M4 competition. I also discuss forecasting with high-frequency data in general.

1. High-frequency data

In this era of big data, computing, sensor and communication technologies have enabled the collection and transmission of data at higher frequencies than ever before. Many business decisions nowadays are based heavily on forecasts that use data collected in real-time. In financial markets, high-frequency trading relies on high performance computers to execute models and algorithms that are based on high-frequency financial data. In high-tech industry, marketers and website owners monitor web traffic in an endeavor to understand customers' responses to various marketing campaigns, as well as their purchasing behaviors. Department stores and retail shops analyze video camera recordings in order to model customer flow and to optimize display items and labor resources. In the energy industry, power companies measure the transmission grid dozens of times per second using phasor measurement units to help prevent blackouts. In the sporting industry, professional basketball teams optimize training plans for their athletes using the high-frequency data collected during games and training sessions. In practice, transactional data and video streams are often converted to high-frequency time series data, commonly hourly and sub-hourly series, for analysis and modeling.

2. Another look at M4 competition performance

Of the 100,000 time series released as part of the M4 competition, 414 are hourly series (Makridakis, Spiliotis, & Assimakopoulos, 2020). Tables 1 and 2 depict heat maps of the rankings, where a cooler (greener) color indicates a higher rank. Overall, the rankings differ notably across different frequencies. For instance, the No. 1 team on hourly series for point forecasts is ranked No. 9 overall. When considering prediction intervals, the ARIMA benchmark is ranked No. 2 on hourly series, but No. 9 overall.

As is shown in Table 1, the bottom six benchmarks are ranked 15 or worse across all frequencies. While only two teams (namely Pawlikowski et al. and Jaganathan & Prakash) are ranked in the top 10 across all frequencies, Pawlikowski et al. is the only team that beats all 12 benchmarks at every frequency, offering the most robust forecasts across different frequencies. When evaluating prediction intervals, none of the teams dominate the ARIMA benchmark.

Compared with the time series at daily and lower frequencies, hourly time series usually present an additional seasonal pattern: time of day. This requires the forecasting methodology to be able to capture multiple seasonalities. Typically, weaknesses in capturing seasonalities are exposed more severely in hourly series than in lower-frequency series. While most methods in the M4

E-mail address: hongtao01@gmail.com.

https://doi.org/10.1016/j.ijforecast.2019.03.013
0169-2070/

Table 1
Rankings of teams across different frequencies in terms of point forecasts.

	Hourly	Daily	Weekly	Monthly	Quarterly	Yearly
Smyl	2	22	8	1	2	1
Montero-Manso et al.	5	20	7	3	1	3
Pawlikowski et al.	3	1	3	4	5	10
Jaganathan & Prakash	4	9	6	2	6	8
Fiorucci & Louzada	15	3	11	5	4	5
Petropoulos & Svetunkov	14	5	2	6	3	7
Shaub	13	24	19	9	20	4
Legaki & Koutsouri	24	10	22	13	18	2
Doornik et al.	1	19	5	8	7	13
Pedregal et al.	6	6	16	11	9	12
Spiliotis & Assimakopoulos	18	7	17	14	12	11
Roubinchtein	10	25	9	7	10	14
Ibrahim	21	26	20	19	14	6
Tartu M4 seminar	11	2	10	22	23	9
Waheeb	8	13	4	21	8	19
Darin & Stellwagen	7	27	1	10	11	18
Dantas& Cyrino Oliveira	16	23	18	16	16	15
Theta - Benchmark	23	14	23	15	21	17
Comb - Benchmark	27	4	13	18	13	16
ARIMA - Standard for comparison	9	21	15	12	19	21
Damped - Benchmark	26	12	12	20	17	20
ETS - Standard for comparison	17	11	14	17	15	22
Holt - Benchmark	28	8	21	24	22	23
SES - Benchmark	20	18	24	23	24	27
Naive2 - Benchmark	22	15	25	25	25	24
Naive - Benchmark	29	15	25	26	26	24
sNaive - Benchmark	12	15	25	27	27	24
RNN - Benchmark	25	28	28	28	28	29
MLP - Benchmark	19	29	29	29	29	28

competition have some ability to capture seasonality, it is hard to perform an apple-to-apple comparison among them, due largely to the fact that the other components are not identical. However, two methods, namely the naïve benchmark and the seasonal naïve benchmark, differ only in their treatment of seasonality. Thus, they are a good example for showing the contrast in their performances for hourly series and others. As Table 1 shows, the naïve benchmark (Naïve) and the seasonal naïve benchmark (sNaïve) are fairly close for the non-hourly frequencies, with ranking positions within two places. At the hourly frequency, though, the seasonal naïve benchmark is ranked No. 12, while the naïve benchmark is ranked at the very bottom, No. 29.

3. Beyond the M4 competition

The five main categories of M4 competition data are from micro, macro, industry, finance, and demographic applications, though none of these include hourly data (Makridakis et al., 2020). Although the M4 competition data are more diverse than those of the previous M-competitions, there are still many important applications that are being missed. In the forecasting literature, a major source of hourly data is the energy industry, which has been using hourly load, price and generation data for decades. In addition, the transportation sector has also been working with hourly data, such as traffic load and accidents. Likewise, the retail and healthcare industries have also been building forecasting models using hourly demand.

One barrier to conducting forecasting research on high frequency data is the problem of data accessibility, because many data sources are not available publicly. Recently, though, several sets of hourly data have been made available publicly through the Global Energy Forecasting Competitions (Hong, Pinson, & Fan, 2014; Hong, Pinson, Fan, Zareipour, Troccoli, & Hyndman, 2016; Hong, Xie, & Black, in press), allowing researchers to benchmark their results against the winning forecasts. Many electricity market operators, such as PJM, Nord Pool, and the

Table 2
Rankings of teams across different frequencies in terms of prediction intervals.

	Hourly	Daily	Weekly	Monthly	Quarterly	Yearly
Smyl	3	1	11	1	1	1
Montero-Manso et al.	11	11	10	6	2	2
Doornik et al.	1	3	2	9	5	3
ETS - Standard for comparison	10	5	8	5	4	4
Fiorucci & Louzada	5	4	5	2	3	5
Petropoulos & Svetunkov	8	6	1	3	6	6
Roubinchtein	6	9	6	4	7	7
Talagala et al.	13	12	9	11	9	8
ARIMA - Standard for comparison	2	7	7	7	8	9
Ibrahim	12	15	13	8	10	10
Iqbal et al.	15	2	4	13	11	12
Reilly	7	14	3	12	13	13
Wainwright et al.	4	10	15	10	12	15
Segura-Heras et al.	9	13	12	14	15	11
Naive - Benchmark	14	8	14	15	14	14

Australian Energy Market Operator, also publish hourly or sub-hourly load and price data, while some, such as ISO New England, also publish the corresponding weather data. In addition, the energy community has published several datasets of smart meter measurements (Wang, Chen, Hong, & Kang, 2019). Since these energy-related dependent variables are also driven strongly by other explanatory variables, these datasets are great resources for developing and testing causal models.

In practice, the frequencies of causal variables may not be the same as those of the dependent variables. As a consequence, forecasting with data of mixed frequencies is an important research topic. While many investigations of forecasting with mixed-frequency data have been conducted in the area of economic forecasting (Armesto, Engemann, & Owyang, 2010), the research in other domains and for high frequencies is still limited.

Computing is another aspect of forecasting research that is important but often down-played. When the data frequency is high, the dataset will be large relative to a low-frequency dataset for the same period. Moreover, since many business decisions that rely on high-frequency data are made in real-time or nearly real-time, the forecasts have to be generated in a timely manner. One step that was taken in the M4 Competition was to assess the computing time required for each of the methods submitted, as the competition organizing team reproduced all of the methods. Makridakis et al. (2020) presented a clear relationship showing that the computational time grows exponentially with reductions in the forecast error. This implies several valuable research questions, such as how to implement the forecasting methodologies to make them most efficient, how to design forecasting methodologies to take advantage of modern computing power, and how to optimize the tradeoff between computational complexity and forecast accuracy.

Last but not least, I would like to recognize a field that is underrepresented in the International Institute of Forecasters and the *International Journal of Forecasting*, namely weather forecasting. Weather forecasters have been using large amounts of high-frequency and high-resolution measurements of meteorological variables, and making use of high-performance computing technologies to run their numerical weather prediction models. Traditionally, high-quality weather forecasting research has been published by journals in meteorological science. On the other hand, the statistical models and machine learning models that dominate the forecasting literature are not considered as the state-of-the-art in weather forecasting. Apparently, weather data can form good test cases for statistical and machine learning models. Meanwhile, improvements in statistical and machine learning models may offer a credible alternative to numerical weather prediction. Combining the two approaches may lead to better forecasts than can be obtained from either individually. In future, I hope that meteorologists and the forecasting community can take an interdisciplinary approach to advancing both fields.

Acknowledgment

I would like to acknowledge the review performed by Rafal Weron, who offered constructive comments that helped to improve the quality of this paper.

References

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 521–536.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*(3), 896–913.

Hong, T., Xie, J., & Black, J. Global energy forecasting competition 2017: hierarchical probabilistic load forecasting. *International Journal of Forecasting* (in press) https://doi.org/10.1016/j.ijforecast.2019.02.006.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.

Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). Review of smart meter data analytics: applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, *10*(3), 3125–3148.

Tao Hong is Associate Professor and Research Director of Systems Engineering and Engineering Management Department, Director of BigDEAL (Big Data Energy Analytics Laboratory), NCEMC Faculty Fellow of Energy Analytics, and associate of Energy Production and Infrastructure Center at University of North Carolina at Charlotte.