Day-Ahead Electricity Demand Forecasting Competition: Post-COVID Paradigm

MOSTAFA FARROKHABADI[1] **(Member, IEEE),**
JETHRO BROWELL[2] **(Senior Member, IEEE), YI WANG**[3] **(Member, IEEE),**
STEPHEN MAKONIN[4] **(Senior Member, IEEE),**
WENCONG SU[5] **(Senior Member, IEEE),**
AND HAMIDREZA ZAREIPOUR[6] **(Fellow, IEEE)**

[1]BluWave-ai, Ottawa, ON K2K 2A9, Canada
[2]School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, U.K.
[3]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, SAR, China
[4]Engineering Science Department, Simon Fraser University (SFU), Burnaby, BC V5A 1S6, Canada
[5]Department of Electrical and Computer Engineering, University of Michigan-Dearborn, Dearborn, MI 48128 USA
[6]Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
CORRESPONDING AUTHOR: J. BROWELL (jethro.browell@glasgow.ac.uk)

ABSTRACT The COVID-19 related shutdowns have made significant impacts on the electric grid operation worldwide. The global electrical demand plummeted around the planet in 2020 continuing into 2021. Moreover, demand shape has been profoundly altered as a result of industry shutdowns, business closures, and people working from home. In view of such massive electric demand changes, energy forecasting systems struggle to provide an accurate demand prediction, exposing operators to technical and financial risks, and further reinforcing the adverse economic impacts of the pandemic. In this context, the "IEEE DataPort Day-Ahead Electricity Demand Forecasting Competition: Post-COVID Paradigm" was organized to support the development and dissemination state-of-the-art load forecasting techniques that can mitigate the adverse impact of pandemic-related demand uncertainties. This paper presents the findings of this competition from the technical and organizational perspectives. The competition structure and participation statistics are provided, and the winning methods are summarized. Furthermore, the competition dataset and problem formulation is discussed in detail. Finally, the dataset is published along with this paper for reproducibility and further research.

INDEX TERMS Electricity demand forecasting, forecasting competition, COVID-19, electricity demand uncertainty.

I. INTRODUCTION

ACCURATE electricity demand forecasting is an essential component of decision-making in power systems operation. Forecasts are used in control rooms as well as in processes such as unit commitment and economic dispatch [1]. Improving the load forecast accuracy is an effective way to reduce the operational costs of the system through reducing the need for reserves and adjusting generators' output to more economic schedules [2].

Demand (or 'load') forecasting has attracted wide attention and extensive efforts have been devoted to developing new tools and techniques. These include statistical-based methods such as multiple linear regression, time-series analysis-based methods such as auto-regressive integrated moving average (ARIMA), and machine learning-based methods such as artificial neural networks and support vector machines. With the advancement of artificial intelligence, deep neural network (DNN) based methods have been applied to load forecasting [3]. Furthermore, so-called ensemble learning models have been developed to combine the advantages of various base models to further improve the forecasting performance [4], [5]. In addition to traditional deterministic load

forecasting, various probabilistic forecasting algorithms have also been proposed to assist the system operator's decisions in an uncertain environment.

As shown in [1], there is a rapid growth in energy forecasting literature during the last decade (2010–2019), where load forecasting papers account for about half of the energy forecasting literature. Different papers demonstrated the effectiveness and superiority of their proposed methods on different datasets (open datasets, private datasets) or/and with different settings (data partition, forecasting time horizon, etc.). Many of them cannot be replicated because the data is not published or the experimental settings are not provided. In this situation, several questions may be raised. For example, does the superiority of the proposed methods in the literature review still exist on a different dataset? In which condition do DNN-based methods perform worse than transitional statistics-based methods even though they are powerful tools for regression? To what extent do the ensemble learning methods improve the forecasting accuracy compared to the best individual forecasting model? Hosting forecasting competitions seems to partially answer these questions because different forecasting methods/algorithms can be tested on the same platform, i.e., the same datasets, time horizon, and evaluation metrics.

The time series forecasting competitions can be traced back to the 1970s and have promoted the development of forecasting research and applications [6]. In the electrical load forecasting area, Hong and his collaborators organized a series of Global Energy Forecasting Competitions, a.k.a. GEFCom2012 [7], GEFCom2014 [8] and GEFCom2017 [9]. The main focuses of these three competitions are hierarchical load forecasting, probabilistic load forecasting, and hierarchical probabilistic load forecasting, respectively. The probabilistic forecasts were evaluated by pinball loss. In addition to system-level load forecasting, a competition on building energy consumption forecasting was jointly organized by the IEEE PES AMPS/ISS [10]. The forecasts were evaluated using a comprehensive metric by combining absolute errors, standard deviation (SD) of errors, etc. The wide installation of smart meters makes it possible to conduct forecasting on the individual consumer level [11]. IEEE Computational Intelligence Society (IEEE-CIS) partnered up with one international energy provider, E.ON SE, and held a competition on smart meter data [12], which focused not only on the accuracy but also on the explainability of the predictions. These load forecasting competitions covered wide topics from deterministic forecasting to probabilistic forecasting, from system level to individual consumer level, and also attracted participants from both academia and industry.

At the beginning of 2020, the novel coronavirus disease (COVID-19) has rapidly spread worldwide. The ongoing COVID-19 related shutdowns have had a profound impact on the electric demand profiles and power systems operation all around the world, as governments put strict mitigation and suppression measures in place [13], [14]. The global electrical demand plummeted around the planet in March,

April, and May 2020, with countries such as Spain and Italy experiencing more than 20% decrease in their usual electric consumption. In view of such massive electric demand changes, electricity network operators are facing unprecedented challenges in scheduling energy resources, as energy forecasting systems struggle to provide an accurate demand prediction [15], [16]. In fact, power systems' operational reliability highly depends on an accurate projection of the future demand and scheduling an appropriate mixture of generation resources accordingly. Particularly, day-ahead forecasts are critical in managing market operation uncertainty. Thus, recent changes expose operators to technical and financial risks, further reinforcing the adverse economic impacts of the pandemic.

Since COVID-19 has largely changed the electricity consumption behavior of consumers, including households, commercial buildings, and industrial plants, forecasting models trained before COVID-19 are unlikely to correctly capture the characteristics of load profiles in the post-COVID paradigm. How to provide accurate forecasts in this situation is a challenging issue. This paper presents a day-ahead electricity demand forecasting competition that was established to motivate experts worldwide to tackle this issue and share their learning. This paper introduces the competition set-up, and publicizes the load data used in the competition. In addition, the top-ranking methods are summarized, and the future of load forecasting under similar conditions is discussed.

The rest of this paper is organized as follows: Section II provides basic information of the competition in post-COVID paradigm; Section III introduces the data and forecasting methods that have been used in the competition; Section IV summarizes important findings and makes recommendations on future competitions.

II. COMPETITION GENERAL INFORMATION
A. ORGANIZERS, STRUCTURE, AND DATES
This day-ahead demand forecasting competition technical committee was led by Dr. Mostafa Farrokhabadi (BluWave-ai, Canada), and other committee members included Dr. Jethro Browell (University of Glasgow, UK), Dr. Yi Wang (The University of Hong Kong, Hong Kong), Dr. Wencong Su (University of Michigan-Dearborn, USA), and Dr. Stephen Makonin (Simon Fraser University, Canada). The committee was advised by Dr. Hamidreza Zareipour (University of Calgary, Canada).

This competition aimed at a detailed analysis of the impacts of the COVID-19 related measures on electricity demand, calling for strategies to mitigate the impact on day-ahead forecasting techniques' performance. In particular, the competition was focused on day-ahead prediction of city-wide demand. The competition included one-track only, deterministic forecasting of hourly load, 16 to 40 hours ahead. The competition simulated operational forecasting by requiring participants to submit forecasts on a daily basis and providing them with actual demand data after submission.

Historical data was released on December 14, 2020. The registration portal was open until March 1, 2021. The evaluation period runs from March 15 to April 13 for a consecutive period of 30 days. The final report and code submission was due on April 19, 2021. The final competition results and winners were announced in early May 2021. Supported by the IEEE Foundation Donor Supported Program, the top three participants received a prize of 5,000, 3,500, and 1,500 USD, respectively.

B. EVALUATION AND RANKING METHOD

Forecasts were evaluated using the Mean Absolute Error (MAE) with final ranking based on the teams' MAE of all 30 days of the competition period. The MAE for forecasts $\hat{y}_t$ of y_t for time period $t = 1, \ldots, T$ is given by

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t|. \tag{1}$$

In instances where a team missed a submission, forecasts from the benchmark method, described in Section III-C, were used in their place to ensure that the evaluation period was exactly the same for all teams. Teams with more than 5 missing submissions were disqualified.

C. WINNING TEAMS

There were a total of 239 unique registrations; 37 teams entered the evaluation period, out of which 20 teams successfully finished the competition. The final leader board can be found on the competition website [17]. The top three winners of the competition are as follows:

- **First Place**: Joseph de Vilmarest (Électricité de France R&D and Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, France) and Yannig Goude (Électricité de France R&D and Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, France);
- **Second Place**: Hongqiao Peng (Guangdong Power Grid Co., Ltd., China);
- **Third Place**: Florian Ziel (Universität Duisburg-Essen, Germany).

III. COMPETITION DESCRIPTION, DATA, AND METHOD
A. PROBLEM DESCRIPTION

The competition included one-track only, deterministic forecasting of hourly load, 16 to 40 hours ahead. Thus, participants had to submit 24 predictions for 24 hourly intervals of a full test day, based on data up to 8 AM of the previous day. The focus was specifically on the day-ahead utility-scale load prediction.

B. DATA

The competition data belonged to a metropolitan electric utility and represent the total system load for the metropolitan area. Throughout the competition, the teams were provided with approximately four years of data, spanning from March 18, 2017 to February 16, 2021. Data was provided to participants in CSV files, including the following hourly measurements and day-ahead forecasts:

- Electricity demand in kW,
- Air Pressure observation and forecast in kPa,
- Cloud cover observation and forecast in %,
- Humidity observation and forecast in %,
- Temperature observation and forecast in °C,
- Wind direction observation and forecast in degrees,
- Wind speed observation and forecast in km/h.

Data was split into training set and test set. The training set was provided in two different releases, first one in December 2020 including data up to November 6, 2020, and second one in March 2021 including data up to January 16, 2021. It was assumed that the weather forecasts are available 48 hours in advance and remain the same throughout the 48-hour period. The test data was released in daily batches for 30 consecutive days from March 15 to April 13, 2021. Each batch contained 24 hours of daily measurements up to 8 AM as well as weather forecasts for the next 24-48 hours. The first batch of test data included demand measurement up to 8 AM of January 17, 2021; the last batch of test data included demand measurement up to 8 AM of February 15, 2021. Participants had a period of approximately 24 hours to submit their day-ahead prediction; each submission consisted 24 predictions for 24 hourly intervals 16 to 40 hours ahead, i.e., beginning to end of the next day.

Figure 1 demonstrates the load and temperature data used in this competition. As seen in this figure, the COVID-19-related shutdowns have had a significant impact on the load profile, with a drastic decrease in load average, peak, and variance observed around February to June 2020. Figure 1d demonstrates the load during the first week of June in 2019 and 2020; observe the significant difference in both load shape and magnitude.

C. BENCHMARK

A persistence-based method was implemented as a benchmark and included in the competition leader board to provide a common reference for participants, and to fill missing submissions in the event that a team failed to submit a forecast. Benchmark forecasts were issued at 8am for midnight-to-midnight of the next day. The load measured on most recent complete day of the same type was used as a forecast for the target day. Summary of the benchmark is provided in Table 1. Ultimately, this simple benchmark proved challenging to beat with only nine teams having a significantly lower MAE than the benchmark over the evaluation period, as discussed in the next section.

D. TOP PERFORMING METHODS

Participants provided the computer code and a summary report to complete the competition. These were reviewed to ensure the rules of the competition were followed, and to facilitate dissemination of learnings from the competition. The approaches taken by several top-placed teams are
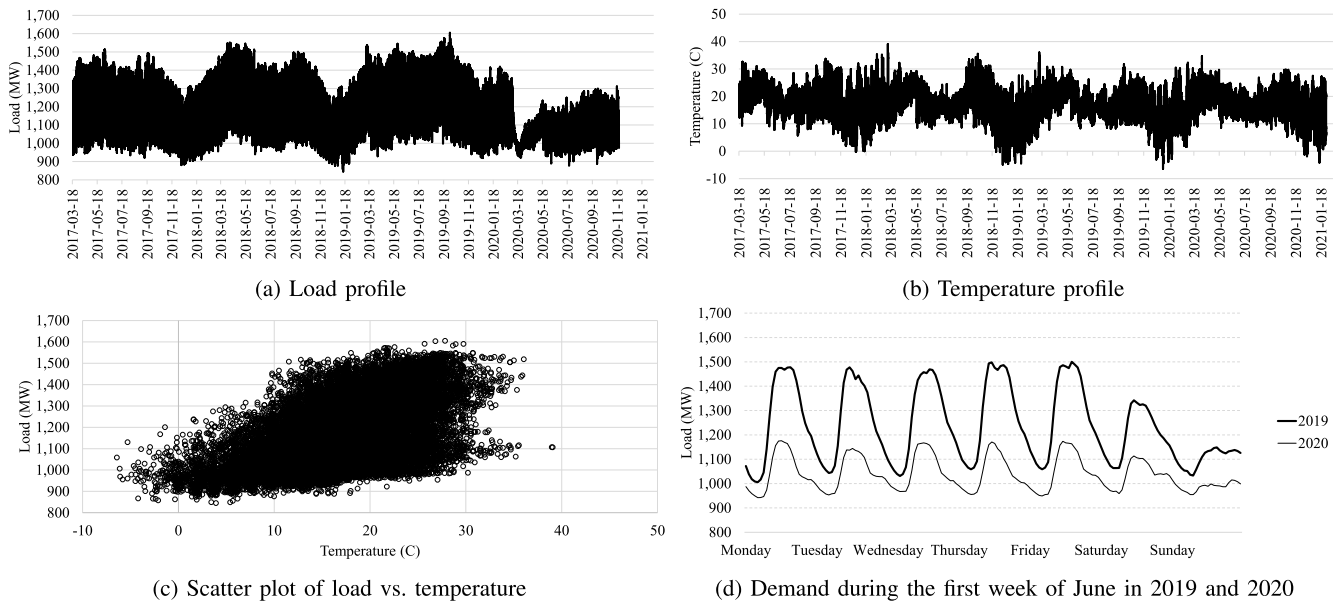
(a) Load profile

(b) Temperature profile

(c) Scatter plot of load vs. temperature

(d) Demand during the first week of June in 2019 and 2020

FIGURE 1. Load and temperature data used for the competition.

TABLE 1. Persistence-based benchmark load forecasts.

Target Day	Data used as Forecast
Sunday	Previous Sunday's load
Monday, Tuesday	Previous Friday's load
Wednesday–Friday	Load from two days ago
Saturday	Previous Saturday's load

summarised in Table 2 based on the details provided in these reports.

All teams used information from recent days as an input to re-train or update their models. This is a marked contrast to successful methods from previous load forecasting competitions (including GEFcom 12, 14 and 17 [7]–[9] and all other competitions cited in the Introduction) where data has been released in blocks making the use of lagged observations impossible. Several teams combined forecasts from large pools of models, including those placing 1st, 3rd and 4th. All three of these teams' ensemble methods were distinct but all had a time-varying component with final combination based on the recent performance of individual models.

The impact of having a relatively short evaluation period and a competitive field of entrants warrants consideration as apparent difference in performance may be the result of sample variation rather than superior forecast performance. Therefore, in addition to calculating the MAE for each participant, we have investigated the impact of sample variation using bootstrapped skill scores and the Diebold-Mariano (DM) test.

A skill score is given by

$$\text{Skill} = \frac{M_{\text{ref}} - M}{M_{\text{ref}} - M_{\text{perf}}} \qquad (2)$$

for a metric with value M for the candidate forecast, M_{ref} for a reference forecast, in this case the benchmark, and where M_{perf} is the metrics value for a perfect forecast, so $M_{\text{perf}} = 0$ in the case of MAE. We further perform block bootstrap re-sampling in order to estimate the scale of sample variation while controlling for auto-correlation with a block size of 24h. We have bootstrapped skill scores as this provides greater discrimination than bootstrapping metrics directly [20]. The results of this analysis are shown in Figure 2 for the top ten performing teams.

Only the top nine teams are providing positive skill that can be discriminated from the benchmark. The teams ranked 1st to 5th have skill scores between 20% and 30%, and perform significantly better than those ranked 6th to 9th with a skill of 10% and lower. Distinguishing within these groups is more challenging. We can be fairly confident that difference between Team 4's skill score and that of Teams 14 and 7 is not the result of sample variation, although evidence of Teams 14's superiority over Team 7 is tenuous. It is not possible to distinguish between the skill scores of Teams 36 and 19 ranked 4th and 5th, respectively, or Teams 23, 9, 25 and 13 ranked 6th to 9th.

We have also calculated the skill score for each of the top 10 teams relative to one another and performed the DM test [21] to assess the significance of apparent differences in performance between all pairs of forecasts. However, we note that this test is likely to be conservative given the auto-correlation observed in forecast errors and modest size of the evaluation period. These results are presented in Figure 3. The DM test confirms the superiority of Team 4's forecasts over all others and inability to separate other teams in the top 10. Notably, the DM test provides evidence against separating the performance of Teams 14 and 7. Furthermore,

TABLE 2. Summary of methodologies based on reports submitted by participants. Due to an error there was a problem with the first three submissions from team 6, who would have finished in 8th position if submissions 1–3 were discounted for all teams.

| Team	Rank	Data Preparation	Techniques
4	1	Correction to meteorological forecasts using linear model on recent errors and observations. Feature engineering: exponentially smoothed temperature with smoothing factor 0.95 and 0.99	Multi-model ensemble (72 models): Auto-regression, linear regression, generalised additive models, random forest (quantiles 0.4, 0.5 and 0.6), random forest for GAM residuals, multi-layer perceptron. Kalman filter adaptation. Intraday correction (AR model). Ensemble using 'online robust aggregation of experts'. Based on [18]
14	2	Normalisation of load by daily peak	Combination similar days from training data using day type and peak temperature. Adjustment based on peak load forecast recent profiles from days of the same type
7	3	Visual outlier detection, filled with interpolation. Creation of north/south and east/west wind components, daily rolling means of weather features, multiple ReLU transformed weather features and interactions. Creation of holiday-adjusted training data.	Multi-model ensemble (674 models): STL-decomposed exponential smoothing, AR(p), generalized additive models, and lasso-estimated high-dimensional linear models. Separate models for each lead-time. Combination by smoothed Bernstein Online Aggregation.
36	4	Data normalized to [0,1]	Ensemble of random forest, gradient boosted machine and XGBoost models with 1) recursive multi-step set-up using lagged load, time of day and day of week only, 2) models for each hour of the day with addition of temperature forecasts. Weighted average combination based on recent performance.
19	5	None	Deep residual networks [19]
25	8	Feature engineering: sin/cos of hour of day and day of year, averages of recent load, averages of recent weather forecast error	Gradient boosted regression tree with quantile loss, separate models for 0–8h and 9–24h due to availability of 24h lag, tuning via grid search
13	9	None	Load from the same day in the last three weeks is averaged and then multiplied 1.02
6	22	None	Multi-model ensemble (400 models): Facebook *prophet* to forecast temperature, cloud cover and load, XGBoost to forecast temperature and load using forecasts from prophet as inputs. 100 XGBoost models fit for 4 training/validation splits resulting in 400 models and forecasts which were combined using simple linear regression.
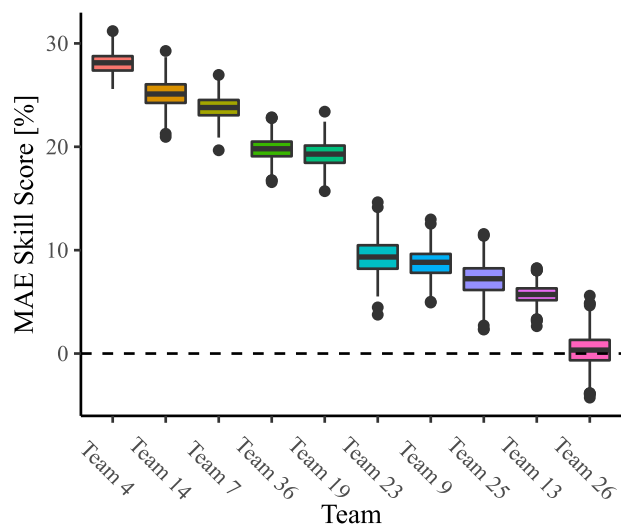
FIGURE 2. MAE skill scores relative to the benchmark. Box plots indicate sample variation produced via block bootstrap re-sampling with 500 samples and blocks of length 24. Overlapping boxes, and whiskers to a lesser extent, suggest a lack of statistically significant difference in skill.

the test for equal performance of multiple forecasts proposed in [22] at a significance level of 5% suggests that teams ranked 2nd to 5th have equal predictive ability, as do those ranked 6th to 10th plus the benchmark, also illustrated in Figure 3.

IV. FINDINGS AND RECOMMENDATIONS FOR FUTURE COMPETITIONS

This competition has provided, for the first time, the opportunity for competitors to test methods in an on-line fashion with daily feed-back and the availability of recent load observations as an input to forecasting models, similar to an actual operational setting. The value of recent data is clear as all top performing teams used it as an input, and most also used it to update or adapt their forecasting models. Given both the value and relevance of the on-line set-up it would be a positive step for it to be the future of demand forecasting competitions. However, it places an additional burden on competition organisers and participants by increasing the frequency of data release, forecast submission, and evaluation.

Three of the top four teams combined forecasts from multiple models. While multi-model approaches are not new to forecasting, this is the first forecasting competition where their dominance has been so pronounced. A similar trend has been observed in the M-series of competitions with the top performing teams in M5 all combining forecasts from multiple models [23]. This trend is expected to continue as large computational resources become more accessible and
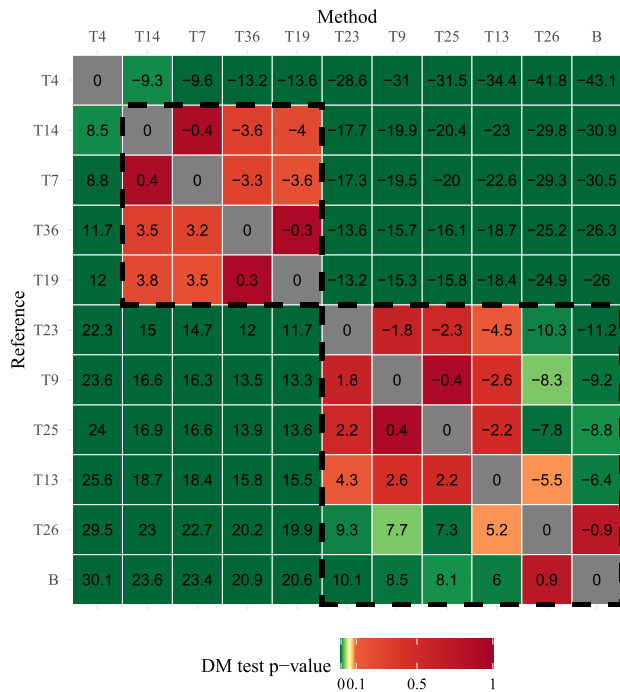
FIGURE 3. MAE skill score (numbers, %) and Diebold-Mariano (DM) test p-values (color) for all combinations of the top 10 entries and benchmark. Dashed boxes indicate groups of forecasts that have equal predictive ability following the multivariate DM test.

tools for automatic model selection and tuning, so called *autoML*, continue to improve.

The on-line format of the evaluation period required a strategy for missing submissions; the technical committee's consensus was to replace the missing submissions by benchmark. The benchmark performed relatively well in this competition, and thus the decision could have been leading to a conflict if any of the top teams were to benefit from it. Hence, future organizers of similar competitions are advised to take extra care in dealing with missing submissions. In addition, the platform on which the on-line format is being organized must be reliable and easy to interact with, given the tight deadline for submissions each day. Furthermore, since each participant submitted a new file each day, a naming convention and standard format was needed to facilitate the storage and analysis of results. While the organizers announced the naming convention and sent multiple reminders, some participants continued to ignore the defined naming format throughout the evaluation period. A possible remedy is an online submission portal that automatically verifies formatting.

In this competition, the organisers made a trade-off between the length of evaluation period and the burden of running and participating in a 'live' competition for an extended period of time. As a result, dissemination between the performance of closely matched teams was challenging or impossible. Future competitions in a similar positions

should consider this aspect of competition design carefully to ensure they are able to draw meaningful conclusions from competition results and the fairness of final rankings. For instance, any procedure to award joint rankings if scores are statistically indistinguishable could be part of the competition design. In addition to statistical tests, additional criteria could be employed, such as judging teams based on the explainability of their methods, as in [12].

The competition ran as an academic exercise in contrast to commercial data science competitions where findings are often kept private for commercial exploitation. Publication of results and methods is therefore critical to maximize the benefit to society from this activity. We hope the dataset set and forecast will serve as a useful test case and benchmark for academics and practitioners working on electricity load forecasting, and time series forecasting in general.

V. CONCLUSION

"Day-Ahead Electricity Demand Forecasting Competition: Post-Covid Paradigm" was well-received among the forecasting community, with approximately 250 unique registrations and 40 teams entering the competition. The competition on-line evaluation period received positive feedback from the participants, providing a benchmark format for future energy forecasting competitions. Multi-model ensembles were used by two of the top three winners of the competition. We published the competition data alongside the predictions submitted by top three teams on the competition official web page [17].

One future direction is analyzing the practicality and generality of the submitted methods. First, submitted methods used a considerable portion of post-COVID-19 data to train their models; it is not clear if these models are robust to load profile changes caused by other similar global or local events. Second, large multi-model ensembles are costly to productize and maintain as each member of the ensemble places demands on staff and computational effort for marginal forecast improvement, raising questions on whether system operators will be able to justify the additional overhead of using such models. However, they are adaptable to sudden and unexpected changes in conditions, as imposed by COVID-19, so may be justified on the grounds of improving resiliency to future disruptive events. Finally, it is not clear if the winning methods would perform well in other jurisdictions. A similar competition could be organized using data from around the globe to test the robustness of the methods to location, demand rating, etc.

REFERENCES

[1] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376–388, 2020.

[2] D. K. Ranaweera, G. G. Karady, and R. G. Farmer, "Economic impact analysis of load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1388–1392, Aug. 1997.

[3] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 511–516.

[4] L. Wang, S. Mao, B. M. Wilamowski, and R. M. Nelms, "Ensemble learning for load forecasting," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 616–628, Apr. 2020.

[5] L. Von Krannichfeldt, Y. Wang, and G. Hug, "Online ensemble learning for load forecasting," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 545–548, Jan. 2021.

[6] R. J. Hyndman, "A brief history of forecasting competitions," *Int. J. Forecasting*, vol. 36, no. 1, pp. 7–14, Jan. 2020.

[7] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecasting*, vol. 30, no. 2, pp. 357–363, Apr./Jun. 2014.

[8] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.

[9] T. Hong, J. Xie, and J. Black, "Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1389–1399, 2019.

[10] *Smart Grid Competitions*. Accessed: Sep. 30, 2021. [Online]. Available: http://www.gecad.isep.ipp.pt/smartgridcompetitions

[11] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[12] (May 2021). *Fuzz-IEEE Competition on Explainable Energy Prediction*. Accessed: Sep. 30, 2021. [Online]. Available: https://ieee-dataport.org/competitions/fuzz-ieee-competition-explainable-energy-prediction

[13] G. Ruan *et al.*, "A cross-domain approach to analyzing the short-run impact of COVID-19 on the U.S. electricity sector," *Joule*, vol. 4, no. 11, pp. 2322–2337, Nov. 2020.

[14] G. Ruan, J. Wu, H. Zhong, Q. Xia, and L. Xie, "Quantitative assessment of U.S. bulk power systems and market operations during the COVID-19 pandemic," *Appl. Energy*, vol. 286, Mar. 2021, Art. no. 116354.

[15] P. Jiang, Y. V. Fan, and J. J. Klemeš, "Impacts of COVID-19 on energy demand and consumption: Challenges, lessons and emerging opportunities," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116441.

[16] J. López Prol and O. Sungmin, "Impact of COVID-19 measures on short-term electricity consumption in the most affected EU countries and USA states," *iScience*, vol. 23, no. 10, Oct. 2020, Art. no. 101639.

[17] (May 2021). *Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*. [Online]. Available: https://ieee-dataport.org/competitions/day-ahead-electricity-demand-forecasting-post-covid-paradigm

[18] D. Obst, J. de Vilmarest, and Y. Goude, "Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4754–4763, Sep. 2021.

[19] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[20] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, and I. Schicker, "Evaluation of wind power forecasts-an up-to-date view," *Wind Energy*, vol. 23, no. 6, pp. 1461–1481, 2020.

[21] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, 1995.

[22] R. S. Mariano and D. Preve, "Statistical tests for multiple forecast comparison," *J. Econometrics*, vol. 169, no. 1, pp. 123–130, Jul. 2012.

[23] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 competition: Background, organization, and implementation," *Int. J. Forecasting*, 2021, doi: 10.1016/j.ijforecast.2021.07.007.

• • •