Highlights

How good are TSO load and renewable generation forecasts: learning curves, challenges, and the road ahead

Hussain Kazmi, Zhenmin Tao

- Load, solar and wind forecasts from 16 European TSOs are analysed
- Forecast errors scale linearly with increasing demand or generation
- Overall renewable forecast errors have nearly doubled in the last five years
- Forecasts tend to outperform naïve baselines but errors remain highly autocorrelated
- Considerable room for improvement remains in most forecasts

How good are TSO load and renewable generation forecasts: learning curves, challenges, and the road ahead

Hussain Kazmi[a], Zhenmin Tao[a]

[a]*KU Leuven, Belgium*

Abstract

Transmission system operators (TSOs) forecast load and renewable energy generation to maintain smooth functioning of the grid by contracting sufficient generation and reserve capacity. These forecasts are also utilized by third parties, such as energy generators and demand aggregators, in their own forecasting and decision-making pipelines e.g. to determine suitable trading strategies. Inaccurate forecasts by the TSOs can therefore lead to increased balancing needs as well as elevated societal and market costs. The situation is further exacerbated by the challenges arising due to rapidly increasing renewable generation and the effects of the post-covid era. In this paper, we analyse five years of TSO forecasts for load, wind and solar generation for 16 European countries. More concretely, using a comprehensive set of metrics, we explore relevant questions such as whether there are TSO specific differences in forecast accuracy, and how forecast errors have changed over time and if they can be reduced further. Our results show that while errors tend to increase linearly with demand or renewable generation, most TSOs still have considerable room for improvement in terms of accuracy. The paper concludes with a set of recommendations for TSOs to improve their forecasts, as well as the ENTSO-E transparency platform where we obtained the data used in this study.

Keywords: Forecasting, electricity demand, renewable energy generation, accuracy, learning curves

1. Introduction

Forecasting future energy demand and generation lies at the heart of planning and decision-making in the modern electric power sector. In fact, it is widely seen as the cornerstone that underpins stable grid operation by matching production and consumption in real time [1]. Over time, forecasting the future has gained in importance with energy market liberalization and increasing proliferation of intermittent renewable energy sources [2].

Today, supply-side volatility is a growing concern in many countries across the world with inherently variable renewable generation displacing (largely) predictable thermal generation [3]. In the future, the adoption of electric vehicles (EVs) and electric heating [4], the formation of consumer-centric energy systems [5], and the introduction of virtual power plants (VPPs) [6] are expected to further complicate the forecasting tasks necessary for stable grid operation [7].

These challenges arise in part due to the non-stationary character inherent of the time series under consideration. Both electricity demand and generation through renewables exhibit strong seasonalities and trends, which are in continuous flux. For instance, increasing the installed base of renewable energy sources means that historic data should be used with caution while training forecast models that predict the future. Likewise, changing electricity demand trends - for instance due to electrification of transportation - also pose the danger to make established load forecasts less accurate than they used to be. To better understand the scale of change over the past few years, Fig. 1 shows how the demand and renewable generation in several European countries has evolved since 2015 and 2017 respectively. It is obvious that, for many countries, even though the overall demand levels have remained largely stable, renewable generation has increased rapidly during this time. The figure also identifies how different European countries have extremely diverse demand and generation profiles.

1.1. Forecasting in power grids

In the context of electric power systems, there are numerous ways to categorize forecasting techniques [7]. Depending on the time frame and purposes, forecasting models can be classified into long-term and short-term forecasting techniques. Likewise, forecast models can also be classified based on how the model is constructured, including white- and black-box models (also referred to as judgment-based and empirical or data-driven models respectively in forecasting literature). As the name implies, a white-box forecast utilizes an energy conversion model that mimics the physical processes in the real world [8, 9]. In contrast, black-box forecasting techniques are typically data-driven and are driven by statistical and machine learning theory [7, 10]. Therefore, white-box forecasting methods typically enjoy better interpretability due to the existence of a physical model [11], while black-box forecasting techniques tend to generalize better by continuously adjusting model parameters to the latest observation data [12]. Another emerging, yet important, distinction in forecasting methods is that of global and local forecast models [13]. Local forecast models predict the future for a single time series, while global models can predict several time series simultaneously, thereby potentially improving generalization by leveraging cross-learning across time series [14].

In this regard, long-term forecasting techniques, which are used to guide investments in new capacity and even policy-making, are usually designed and implemented using physical principles. This is due primarily to their huge business impact and the resulting requirements on interpretability [10, 15]. Short-term forecasting needs, on the other hand, arise commonly for operational generation and transmission planning, such as day-ahead market clearing. In this regard, one can find both white- and black-box forecasting methods. A comprehensive review of different forecasting techniques deployed in the electric power sector can be found in [2]. In the remainder of this study, we focus exclusively on short-term, day-ahead forecasts.

1.2. Impact of forecast errors

As highlighted earlier, transmission system operators (TSOs) need to ensure that electricity demand and generation is balanced at all times on the transmission grid level [16].

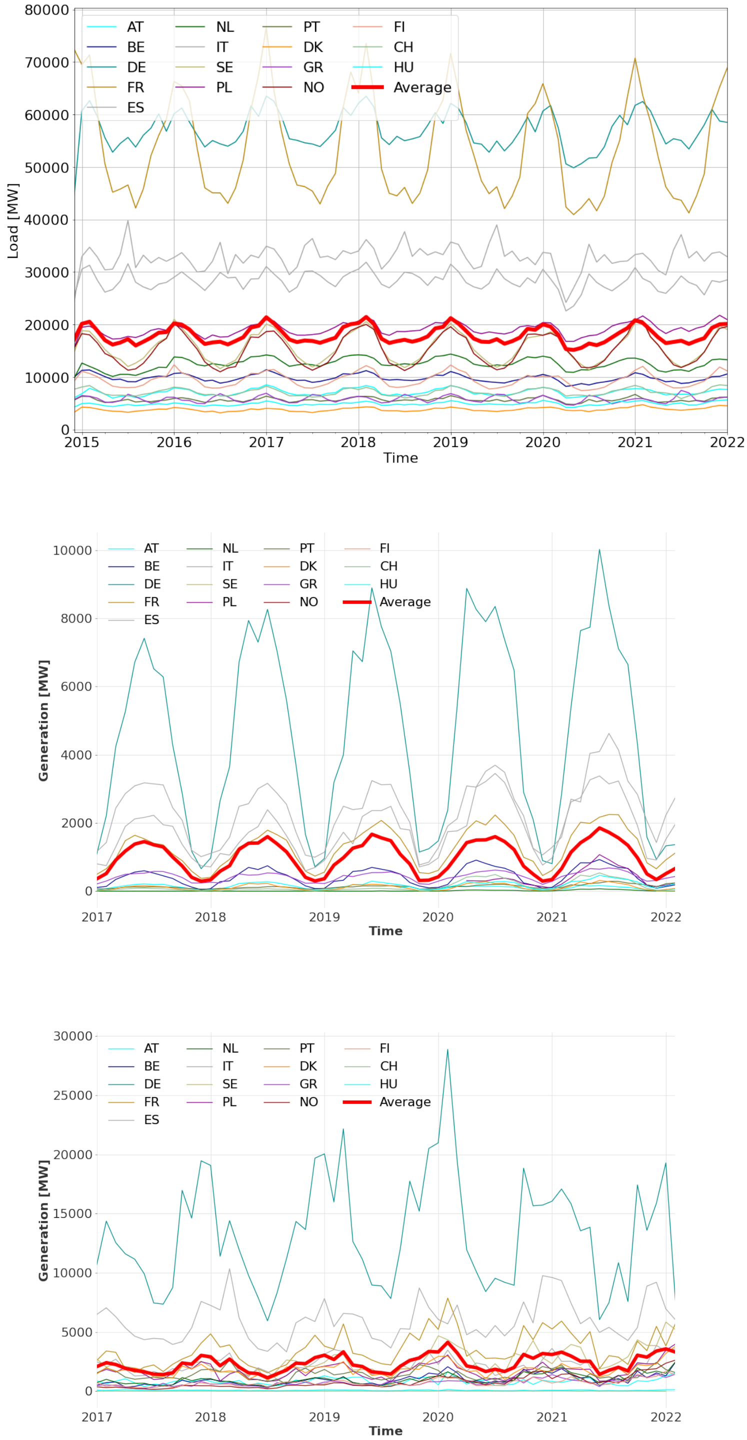


Figure 1: Evolution of load (top), solar generation (middle), and wind generation (bottom) in 16 different European countries; the average demand and renewable generation is also included to provide an overall indication of the overall trend and seasonality. The annual variations are caused by both the weather conditions and an increasing installed renewable base.

This is achieved by forecasting future demand, and then ensuring that sufficient capacity is available to meet it at all times. Therefore, the accuracy of these forecasts is key to ensuring the technical stability and economical efficiency. Technically, supply-demand imbalances cause frequency deviations on the grid from its nominal frequency [17]. Left untreated, such imbalances can lead to large scale disruption in the form of grid outages and equipment failures. Moreover, TSOs typically contract and operate a frequency reserve pool, which can be activated to ensure that imbalance-driven frequency deviations do not exceed certain thresholds [18]. Large forecast errors consequently increase the absolute values of imbalance, influence subsequent market prices, and necessitate larger reserves [19]. Additionally, in some cases, these forecast errors can even lead to operational curtailment of renewable generation (i.e. when there is insufficient demand in the system), which can reduce the willingness of operators to accept these technologies in their area of control.

From an economic perspective, such forecasts are used by power generation companies to guide both the bidding and the optimal operation of their power generation assets. In the presence of demand response, these forecasts also serve as a preliminary reference for imbalance service optimal bidding and operation. The economic impact of (in)accurate short-term forecasting has been quantified and discussed in [20], in which the authors concluded that even simple electricity trading strategies can lead to substantial economic impact if combined with a decent forecasting technique. Likewise, the actual impact of market forecast errors on revenue generated by battery storage is explored in [21]. Existing research also indicates that underforecasting wind generation at the TSO level leads power producers to make decisions that favor social welfare (rather than corporate benefit), and vice versa [22].

1.3. Sources of forecast error

There are numerous sources of forecast errors. As alluded to earlier, this can be due to the changing character of the time series being forecast (e.g. due to increasing installed base of renewables or caused by pandemic-induced disruptions [23]). Several other sources of forecast error exist as well. These include predominantly the forecast time horizon - demand and generation in the near future tends to be easier to forecast accurately than that in the distant future [24]. Likewise, errors can also arise due to mis-specification of the forecaster, e.g. due to choices made during the modelling process [25], and can manifest as bias (i.e. systematically over- or under-predicting), and variance. Moreover, the incorrect deployment of forecasting models or provision of incorrect input data to the forecasters can also lead to significant errors [26].

Finally, forecast accuracy can also degrade due to upstream forecast errors. An example of this is in the use of meteorological forecasts as input features to predict electricity demand and generation. These coupled inputs also greatly increase the risk of correlated errors, e.g. across TSOs in several neighboring countries [27]. In fact, while the aggregated variability of wind and solar is in general less than that of each resource alone, it is extreme weather conditions, such as storms, that often lead to the largest forecast errors [28]. A recent study concludes that, over the last decade, increasingly accurate weather forecasts have saved consumers over $150 million in annual energy savings [29].

1.4. Contributions

Over the past decades, TSOs have actively adapted and improved the capability of their forecasting techniques so as to tackle the challenges and impacts discussed above. During this period, the field of forecasting has also evolved considerably. In practice, however, since the TSOs do not publicly publish their methodology, it is difficult to anticipate when forecast accuracy should have improved. Likewise, this lack of transparency also limits cross-TSO comparisons, since it is unclear whether several TSOs utilize the same methodology and input data streams, or if they implement their own. Furthermore, cross-comparisons over different forecasting techniques have remained limited and are often restricted to theoretical settings ([30, 31]). This is mainly due to the fact that the performance of forecasting techniques is highly context-dependent. Intuitively, deploying identical forecasting techniques to different regions will lead to different performances.

Nevertherless, even though the forecast models cannot be directly evaluated, publicly available forecasts can be. This is enabled by the fact that TSO forecasts for both load and renewable generation are available for most European countries over the past few years. To evaluate these forecasts, several metrics have been developed in literature, which are well-aligned in both academia and industry [32]. Following the methodology formalized in [33], this paper makes several important contributions to address the issues highlighted above:

1. It establishes the forecast accuracy of several European TSOs' load and renewable generation forecasts, based on their published day-ahead forecasting data between year 2015-2021 for load, and 2017-2021 for renewable generation.
2. It evaluates the individual TSO forecast errors and compares them against baselines to help understand whether the existing forecasts can be further improved upon, and provides specific recommendations where this is the case. This is important for TSOs to improve their own forecasts, but can also be utilized by other market agents to improve the openly accessible TSO forecasts.
3. It creates learning curves for forecasting as a function of varying electricity demand and generation. These help us better understand the drivers of forecast accuracy and yield insights into whether there are inherent differences between time series forecastability in different countries.

The rest of this paper is organized as follows: Section 2 highlights the methodology and data used in this paper, while Section 3 provides a description of the most important results, including error metrics for load, solar and wind generation in several countries. Section 4 discusses the most important insights and recommendations resulting from the analysis. Section 5 concludes the paper.

2. Methodology and data

The day-ahead forecasts considered in this paper are made available by European TSOs through their own web interfaces or dashboards, and are aggregated by the European Network of Transmission System Operators for Electricity (ENTSO-E) Transparency Platform

(TP) for easier, unified access [34]. In this section, we first describe the ENTSO-E TP from which we obtained the observed and forecast load and generation values used in this paper. Next, we describe the methodology we followed to evaluate these forecasts.

2.1. The ENTSO-E transparency platform (TP)

The past decade has seen increasing integration of the European electrical grids and markets. With rising cross-border power flows, it is critical to record and process this information in a centralized manner. Consequently, these data or information flows transcend the responsibility or jurisdiction of any individual transmission system operator (TSO). While several papers point out a general lack of open data in the energy sector [35], in this case the transparency platform from ENTSO-E fills this gap and has been widely utilized in research [36]. Several papers have also appeared discussing its extensions [37] and limitations [38].

The TP publishes and updates around 50 data items distributed in six different categories, which include load, generation, transmission, balancing, outages and congestion management. The reporting duration and resolution are dependent on the data item and the TSO. The TP also distinguishes between a number of zonal aggregations, including country, bidding zone, control areas, and market balance areas. Subsequently, in this paper, we focus exclusively on the country-level aggregation for easier interpretability of results. This does occasionally lead to some challenges as a single country may have multiple TSOs and the datasets or forecasts provided by each one of these may differ in quality. The TP makes it possible to access this data in several ways, including using the website's graphical user interface (GUI), an application programming interface (API) and a file transfer protocol (FTP) service. We used the API to retrieve the data used in this study.

2.2. Data

It is important to note that while ENTSO-E makes an enormous amount of data available through its API (over ten thousand time series for several years according to some estimates [38]), we utilize only a small subset of these in our analysis. More specifically, since we are primarily interested in evaluating the day-ahead forecast accuracy of load and generation, we work with (1) actual and day-ahead forecasts of load, and (2) actual and day-ahead forecasts of renewable sources (wind, solar).

We focused on a group of 16 countries that make the bulk of energy demand and generation in the European Union (EU). These include Austria (AT), Belgium (BE), Germany (DE), France (FR), Spain (ES), the Netherlands (NL), Italy (IT), Sweden (SE), Poland (PL), Portugal (PT), Denmark (DK), Greece (GR), Norway (NO), Finland (FI), Switzerland (CH), and Hungary (HU). However, the sheer size of the dataset makes understanding and communicating results challenging: not all results for all countries can be visualized and presented in a single paper. Furthermore, it is unclear whether the data used in the analysis (i.e. downloaded from the TP) has been post-processed in the time since it was first made available (which is when the forecast would have been originally used).

It is pertinent to point out some issues with the data collection phase here. Not every country uses the same time resolution, therefore some measurements were on an hourly interval while others were on a half- or quarter-hourly interval. To keep cross-country comparisons fair, we sub-sampled data from all countries to an hourly interval. A more serious issue arose due to API calls that did not always yield usable data, and occasionally crashed. This was particularly an issue for renewable generation, where data for a number of countries was only available since the beginning of 2017. We consequently limit most of the analysis to the period of 2017-2021 (inclusive). In line with earlier findings [38], we also experienced several time-out errors while retrieving data using the API. This led to roughly 5-10% of renewable energy sources values to be missing in the final analysed dataset (depending on country), which were not used in subsequent forecast evaluations. In addition, we also observed a few anomalous measurements for load and generation time series for some countries, which were filtered out. We highlight these findings in subsequent sections.

2.3. Evaluating forecast errors

It is important to note that, pandemic-induced disruptions notwithstanding, the European electricity demand has not seen a level shift over the past few years (Fig. 1). On the other hand, renewable generation has expanded considerably in several countries during this time. All three time series are inherently seasonal, with solar energy generation being arguably the most pronounced example of this. This existing trend and seasonality makes the evaluation of forecast errors inherently tricky as any change in error over time can be attributed to changes in the time series to be forecast or the forecasting algorithm (or both). Unfortunately, the algorithms used by the various TSOs to make the forecasts remain black-box, and to the best of our knowledge, there is no documentation to study whether they have changed over time or not. It is also unclear, though unlikely, whether different TSOs coordinate while making forecasts and what meteorological (or other) data sources they ingest as input features while making their forecasts. Consequently, this section focuses on forecast evaluation metrics for the different countries considered in the study for all three forecasts (load, solar, and wind).

The forecast errors are calculated for each individual country using several metrics. There are two practical reasons for this. First, it is unknown which error metrics the different TSOs are utilizing. Second, good forecast methods should theoretically perform well on multiple metrics, rather than on only the one they are trained to optimize for [33]. Consequently, we consider (1) the coefficient of determination (R^2), (2) mean error (ME), (3) mean absolute error (MAE), (4) weighted mean absolute percentage error (WAPE), (5) relative mean absolute error (rMAE), (6) autocorrelation function (ACF) of residuals, and (7) a non-stationarity of residuals test. Taken together, the metrics provide a rather holistic view of the forecast performance. The different error metrics are defined as follows:

$$R^2 = \frac{(\sum(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}))^2}{\sum((y_t - \bar{y})^2(\hat{y}_t - \bar{\hat{y}})^2)} \tag{1}$$

$$ME = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t) \tag{2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{3}$$

$$WAPE = 100 \cdot \frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{\sum_{t=1}^{n}|y_t|} \tag{4}$$

$$rMAE = \frac{MAE_{TSO}}{MAE_{Baseline}} \tag{5}$$

$$r_k = \frac{\sum_{t=k+1}^{n}(e_t - \bar{e})(e_{t-k} - \bar{e})}{\sum_{t=1}^{n}(e_t - \bar{e})^2} \tag{6}$$

Here, y and $\hat{y}$ are the observed and forecast time series (load, wind and solar for all sixteen countries). Likewise, $\bar{y}$ and $\bar{\hat{y}}$ are the mean values for these observations and forecasts, and n is the extent of the time series (in this case, mostly five years). e_t is the error at time t (i.e. $y_t - \hat{y}_t$), and $\bar{e}$ is the mean error. r_k is the autocorrelation function at lag k.

2.3.1. Scale-preserving metrics

Of these metrics, only the ME and MAE follow the same units as the observed time series. The ME provides a measure of whether the forecaster is biased or not, while the MAE provides insights into the extent the forecast deviates from the observations. These metrics, ME and MAE, however do not let us compare the forecast errors between countries with very different overall demand or generation (such as Germany and Austria or Belgium).

2.3.2. Relative and scaled metrics

R^2, WAPE and rMAE can be used for this purpose instead. R^2 is considered a measure of the amount of variation in the observed time series that is explained by the prediction. Another way to think about it is the strength of correlation between the observed and predicted time series. Instead of the standard MAPE, we use WAPE to prevent divisions by zero, something which is especially an issue for solar forecasts. Due to the trend and seasonality in the data, the average also provides a slightly more indicative denominator of how the forecast can be expected to behave. Finally, the rMAE compares the forecast models' performance against a baseline model. In this case, a daily persistence model is used for both demand and generation. The accuracy from such a baseline model may slightly exaggerate its skill since in practice the day-ahead forecast is sometimes created 36 to 40 hours in advance. This is balanced by the fact that in reality baseline load forecasts can be made using weekly seasonalities combined with special treatment for weekends.

2.3.3. Metrics based on temporal evolution of residuals

In addition to the error metrics defined above, it is also instructive to take a closer look at the residuals themselves. For instance, the autocorrelation function of residuals can be used to establish whether there is any remaining structure in the time series that could have been exploited for a better forecast. More concretely, when the autocorrelation function (ACF) of (forecast) residuals is non-zero for some non-zero lags, we can conclude that there is still some structure in the time series which has not been modelled sufficiently. This can be ascertained visually or using a Ljung Box test on out of sample residuals [39]. Finally, even though the time series under consideration are non-stationary, forecast errors should ideally be stationary over time. When they vary over time, it can indicate that the skill level of the forecaster varies over time as well (based on changes in the forecaster or the forecasted time series). Therefore, we also estimate whether the forecast errors are stationary or not, using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for the load [40] and a visual test for wind and solar using clock plots.

The choice of these error metrics raises an important point: all time series under consideration vary considerably over time, both due to foreseen seasonalities and trends, and unforeseen disturbances such as Covid-19. Consequently, the aforementioned metrics are calculated for two different contexts for the load forecasts: (1) business-as-usual case, which includes predictable disruptions such as holidays etc., and (2) unforeseen disruption caused by Covid-19-induced lockdowns. For renewable energy, the latter two considerations are not relevant, but the forecast error trend over time is considered.

3. Results

We follow the same sequence to evaluate forecast errors for all three time series (load, wind and solar) for all countries in this section. First, we present the distribution of the residuals for each country as a scatter plot and a boxplot (to identify effects such as heteroskedasticity and biased forecasts etc.). Second, we explore the autocorrelation function of the forecast residuals, disaggregated by country. Third, the aggregated error for all countries is visualized as a function of time (from 2017 to 2021). This is followed by a table summarizing the error metrics for all countries discussed earlier. Finally, for load forecasts, we include a sub-section to discuss how Covid-19 lockdowns effected different TSOs' forecast errors. Likewise, for renewable energy forecasts, we include a clock plot to explore the seasonalities and trends of forecast errors.

3.1. Load

3.1.1. Normal operating conditions

Fig. 2 shows load forecast errors in the different countries. Day-ahead demand forecasts appear to be quite accurate in most countries with only small deviations from the line of perfect fit. There are some notable departures from this trend. For instance, based on the scatter plot, the Austrian day-ahead load forecasts seem to be inaccurate when the load is rather low. Additionally, a few outliers with extremely poor forecasts are evident as well, including for the Netherlands and Switzerland. The data for the Netherlands was verified against that shown in ENTSO-E dashboards, but it is unclear why these forecasts are consistently biased for extended periods of time. The German forecast also likewise seems to be biased. The type of errors made for these countries indicate that this is perhaps a data quality rather than a forecast issue, as it is unlikely that a TSO can continue making hugely biased errors for years on end. Practitioners should therefore proceed with caution when using TP for obtaining load (or other) forecasts, especially for these countries.

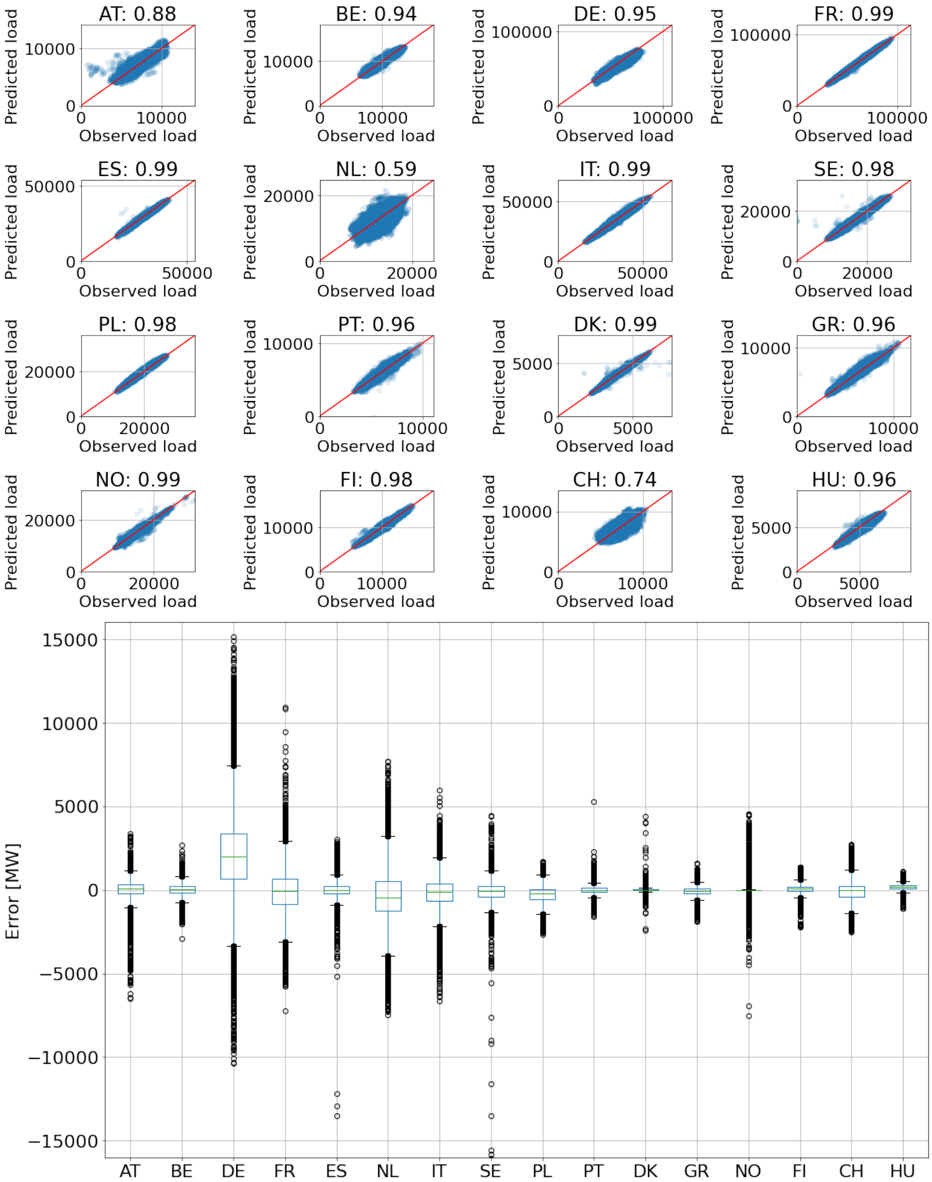


Figure 2: (Top): A scatter plot showing the coefficient of determination for each country, along with the distribution of load forecasts (y-axis) as a function of observed load (x-axis); (Bottom): a boxplot showing the error distribution for the load forecasts, disaggregated by country

Despite the high R^2 and tight fit between forecast and observed demand, Fig. 3 shows the autocorrelation function for the residuals from most countries is non-zero at numerous lags. This provides strong evidence that the forecasts can be further improved, while illustrating two facts. First, that there is considerable serial structure in the residuals. This holds true for all countries considered in the study. Second, there is a considerable periodic element to the errors as well, which differs for different countries. More specifically, we were able to identify four different archetypes based on the autocorrelation function of their residuals:

1. The first archetype, which includes Greece, Belgium, Hungary, and Finland, demonstrates primarily daily seasonality in the forecast errors;
2. The second archetype, which includes the Austrian, Dutch and German regions (given the likely data quality issues, this needs further investigation), demonstrates both daily and weekly autocorrelations;
3. The third archetype, which contains all remaining countries except Spain, demonstrate seasonality in the residuals at multiples of 12 hours;
4. The fourth archtetype, which consists only of Spain for now, shows very limited seasonal autocorrelations (as well as a low serial autocorrelation as well).

Unfortunately, the residuals are not just autocorrelated, they are also non-stationary for many countries as indicated by the KPSS test. This indicates that the error distribution changes over time. A common cause for this includes higher errors during peak hours. The exceptions to this, at $p = 0.05$, include Belgium, France, Spain, Sweden, Portugal, Denmark and Hungary where we fail to reject the null hypothesis.

Next, we compare the TSO forecasts against a simple, daily-naive baseline model. This leads to the relative MAE metric (rMAE), which mirrors findings from other metrics as well. Norway, Denmark and Spain have very low rMAE values (all three also score very well on the other scale-invariant metrics of R^2 and WAPE). On the other hand, as expected, the (likely) data quality issues mean that the Dutch forecast performs poorly even when compared to the naive baseline. For the other countries, there remains a non-monotonic relationship between the WAPE and rMAE. This provides some evidence for the fact that there may be inherent differences in the forecastability of underlying time series, which effects results in addition to the TSO forecasting skill.

Finally, Fig. 4 visualizes the error observed in all regions between 2017 and 2021, i.e. the absolute forecast errors are simply added for all countries. Overall, the forecast error does not appear to have changed considerably over the past years. While it is interesting that load forecasts have not necessarily gotten more accurate, this must be tempered with the realization that 2020 and 2021 saw large disruptions due to Covid-19, which we turn our attention to in the next section. The overall results for load forecasts are summarized in Table 1.

3.1.2. The effect of Covid-19

This section focuses on the pandemic-related disruptions and their effect on electricity demand forecasts. Fig. 5 shows that the overall electricity demand in 2020 (i.e. the aggregated electricity demand in the sixteen countries under consideration) deviated significantly

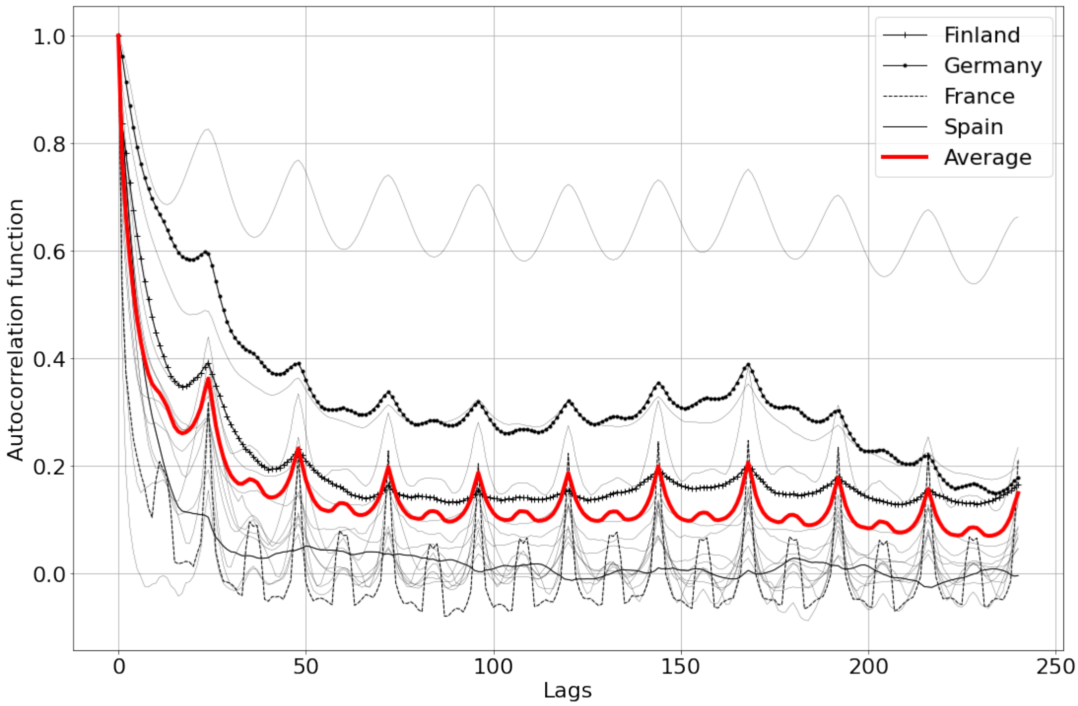


Figure 3: The autocorrelation function of the load forecasts; four different countries are highlighted as representing the archetypes identified in the text, i.e. those showing 1. strong daily autocorrelations; 2. strong daily and weekly autocorrelations; 3. strong sub-daily, daily and weekly autocorrelations; and 4. only a serial autocorrelation

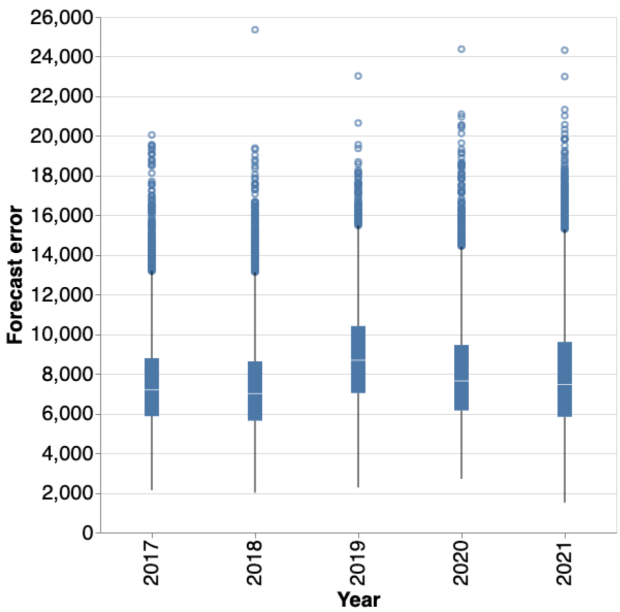


Figure 4: The distribution of load forecasts between 2017 and 2022, aggregated over all 16 countries; Covid-induced forecast errors in 2020 do not seem to lead to a departure from expected forecast accuracy

Country	Symbol	R^2	ME [MW]	MAE [MW]	WAPE [%]	rMAE
Austria	AT	.88	38	352	4.95	.61
Belgium	BE	.94	34	250	2.56	.45
Germany	DE	.95	2082	2436	4.27	.56
France	FR	.99	-78	917	1.71	.29
Spain	ES	.99	-9	307	1.09	.17
The Netherlands	NL	.59	-242	1271	10.14	1.59
Italy	IT	.99	-154	678	2.06	.21
Sweden	SE	.98	-79	388	2.48	.42
Poland	PL	.98	-281	424	2.21	.29
Portugal	PT	.96	-4	157	2.77	.37
Denmark	DK	.99	-1	41	1.08	.14
Greece	GR	.96	-75	178	3.07	.49
Norway	NO	.99	17	107	.70	.14
Finland	FI	.98	59	184	1.93	.39
Switzerland	CH	.74	-68	416	5.84	.92
Hungary	HU	.96	184	201	4.10	.7

Table 1: Summary of load forecast error metrics for different countries

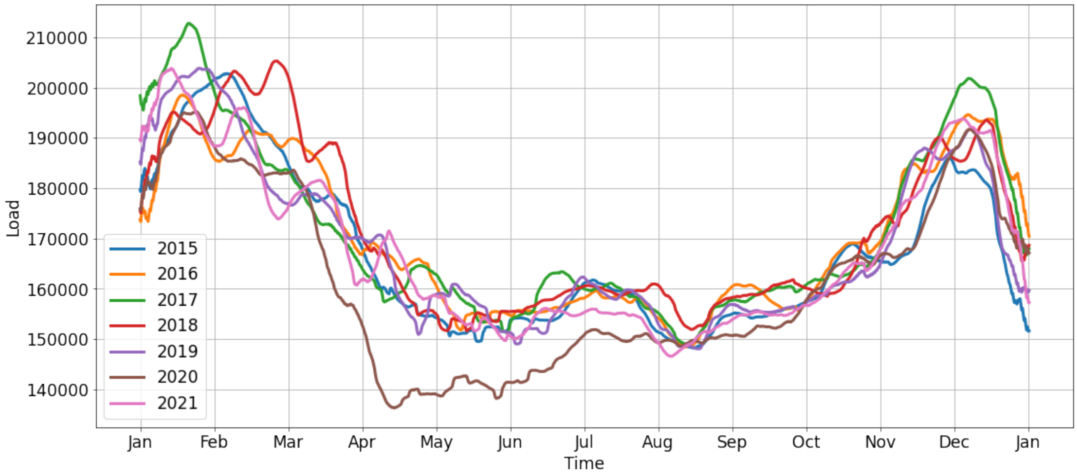


Figure 5: Comparing aggregated load in all sixteen countries for different years; the first lockdown period in spring of 2020 shows a significant departure from its expected value, thereafter no discernible difference is obvious

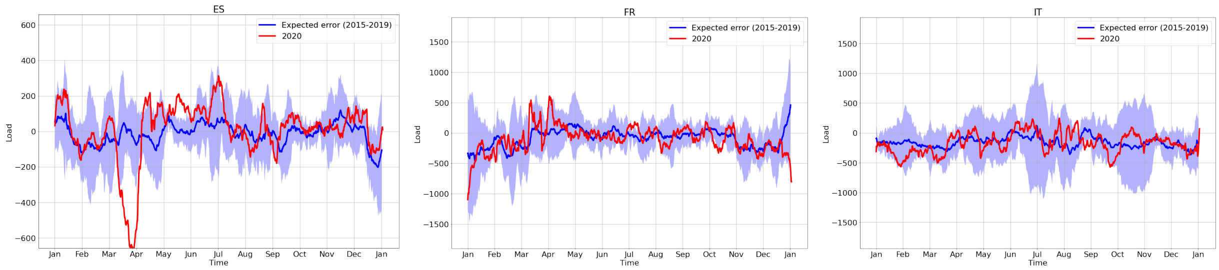


Figure 6: Load forecast errors for three different countries (Spain, France, and Italy): shaded regions represent the expected error values (i.e. based on 2015-2019) compared with the realized forecast errors in 2020

from the norm around mid-March and only reverted to its expected value around July or August. Different countries contributed differently to this deviation, with countries in the South of Europe (such as Spain and Italy) being especially hard hit. However, in terms of forecast accuracy, we found that after an initial adjustment period, most TSOs responded quickly and managed to minimize the error magnitude in a few days to weeks. One notable exception to this rule is Spain. Despite boasting one of the most accurate load forecasts in general, it had one of the highest increases in forecast error during the initial lockdown in 2020. As seen in fig. 6, it was hit particularly hard by lockdowns while other countries such as Italy, France and Belgium also saw comparable lockdown periods without a corresponding sustained spike in forecast error. This is not to say that most other TSOs managed to get through the disruption unscathed, just that they were rather nimble in adjusting to the evolving conditions, which is after all a desirable property in forecasting the future.

3.2. Wind power generation

For wind power generation, only the case of normal operating conditions is considered, as pandemic-induced disruptions could not directly effect generation. As before, we first consider the error distributions, both as a scatter plot and as a boxplot in Fig. 7. From the scatter plot, it is already obvious that the error distribution is much less tight when compared to the load forecast errors. Considerable country-specific differences exist as well. As before, Swiss and Dutch forecast data appears of dubious quality. Several countries, such as Austria, Germany, Germany and Sweden, also show strong deviations from the line of perfect fit, indicating considerable periods of sustained over- or under-predictions. Interestingly enough, Sweden still manages to have a very low overall error rate.

The residuals are next tested using the autocorrelation function of the residuals, where we see a somewhat similar trend as before (Fig. 8). Most countries exhibit strongly (daily) seasonal residuals, which indicate they can be further improved. In this case however, there is no evidence of weekly or 12-hourly seasonality. Unlike in the load case where we needed to carry out the KPSS test, here the errors are obviously non-stationary for most countries due to the nature of wind power generation. This can also be seen in Fig. 9, which shows that the median forecast error aggregated over all 16 countries has grown by a factor of almost two since 2017. This is primarily due to increases in installed capacity. At the same time, the worst case errors have also grown considerably during this period, with the 95th percentile growing from 6 GW to almost 10 GW. Fig. 9 also provides further evidence for the non-stationarity of the forecast errors by visualizing the errors as a clock plot. Here, it is evident that the errors differ based on time of day, season of year, and year of study. For instance, the forecast error aggregated over all countries in the study at 22.00 during winter 2021 is over three times the average error in 2017. This rapid growth in forecast errors is a serious cause for concern, as wind power generation (and consequently forecast error) is expected to rise steeply in the years to come.

Finally, we turn our attention to the relative MAE for the wind forecasts. Here, as before, we see some poor forecasts due to potential data quality issues (e.g. in the Netherlands and Switzerland). On the other hand, most other countries show high levels of skill, when compared with the naive baseline forecast. This is especially true for Sweden and Spain. Most of the remaining countries also perform much better than the naive baseline (often by a factor of three to four times), demonstrating considerable skill at forecasting wind generation in the next 24 hours when compared to the daily persistence model.

3.3. Solar power generation

Finally, we take a look at the solar generation forecasts. Solar power production is, by definition, zero for a considerable period of time, which can cause issues for some evaluation metrics. This was one reason for the choice of WAPE as a relative error metric rather than MAPE. As before, we begin the analysis with the scatter and boxplots, shown in Fig. 10. There is very little data for solar production from the Nordic countries (Norway, Sweden and Finland), and Switzerland and the Netherlands exhibit the same data quality issues as before. In the remaining countries, Austria and Portugal also exhibit poor forecasting performance. For Portugal, this translates to a considerable amount of time where the TSO

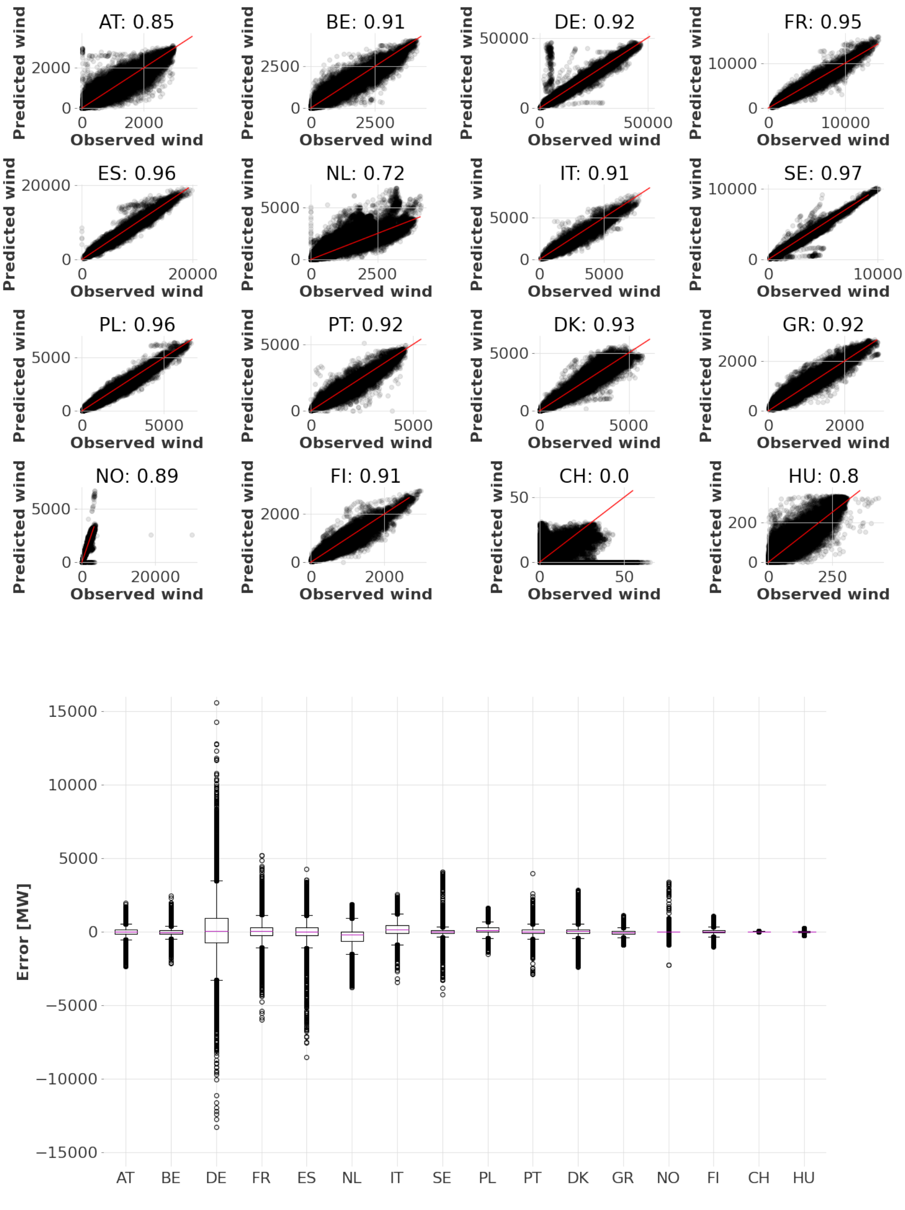


Figure 7: (Top): A scatter plot showing the coefficient of determination for each country, along with the distribution of wind forecasts (y-axis) as a function of observed wind generation (x-axis); (Bottom): a boxplot showing the error distribution for the wind forecasts, disaggregated by country

Country	Symbol	R^2	ME [MW]	MAE [MW]	wMAPE [%]	rMAE
Austria	AT	.8	-14	201	24.6	0.3
Belgium	BE	.87	-32	184	19	.29
Germany	DE	.88	-77	1343	10.1	.2
France	FR	.92	36	414	11.5	.24
Spain	ES	.9	10	437	7.3	.19
The Netherlands	NL	.69	-367	492	49.4	.85
Italy	IT	.22	195	404	18.4	.36
Sweden	SE	.92	-8	151	5.9	.14
Poland	PL	.92	119	208	12.3	.22
Portugal	PT	.77	17	201	14.1	.25
Denmark	DK	.87	38	220	12.2	.21
Greece	GR	.86	-55	128	17.5	.33
Norway	NO	.88	-4.5	66	8.7	.21
Finland	FI	.81	11	120	16.6	.3
Switzerland	CH	0	5.6	8.9	83	1.28
Hungary	HU	.73	-5.6	24.4	31.7	.37

Table 2: Summary of wind forecast error metrics for different countries

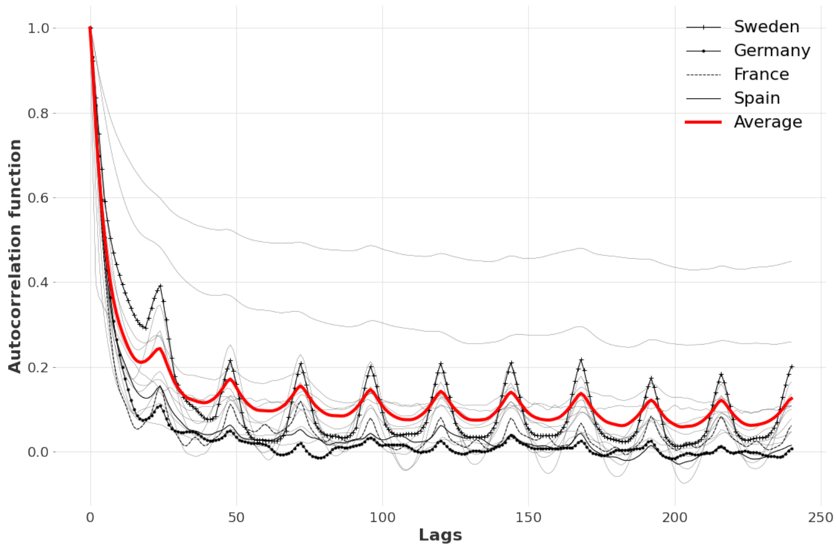


Figure 8: Autocorrelation function of the wind forecast residuals, with the average highlighted; the ACF shows non-zero values at many lags indicating both serial and seasonal autocorrelations; barring countries with probable data quality issues, most countries exhibit a largely similar behaviour with daily peaks in the ACF

forecast solar generation when there was in fact none and vice versa. Austria, on the other hand, shows a curious break in the data where the forecasts continue to grow between 2017 and 2021. The actual generation data, on the other hand, does not grow and actually falls below the previous levels in 2020. We can only hypothesise that this is yet another data quality issue. The remaining countries (Belgium, Germany, France, Spain, Italy, Poland, Denmark, Greece and Hungary) exhibit fair performance based on the scatter plot. However, even in this subset, the boxplot shows that Italy has a number of very high positive errors (reaching almost 10 GW), which could potentially destabilize the grid.

The autocorrelation plot (Fig. 11) shows a largely similar pattern as before for wind forecasts, with significant serial and daily correlations for every country. The ACF in general goes down much slower than for wind generation and load, indicating potentially greater serial autocorrelations and consequent scope for improvement. However, this is naturally influenced by the intermittent nature of solar production as well. Furthermore, unlike for wind, some of the countries again show a sub-daily significant autocorrelation function, which could potentially be related to the morning and evening ramps.

As with wind and unlike the load series, the solar forecast errors (aggregated over all countries) have grown tremendously since 2017, with the median value growing by over five times (Fig, 12). Likewise, the worst case error has grown by over two times and is roughly of the same order of magnitude as wind generation forecast errors at this point. At the same time, there is an understandable element to the forecast errors which are consistently higher

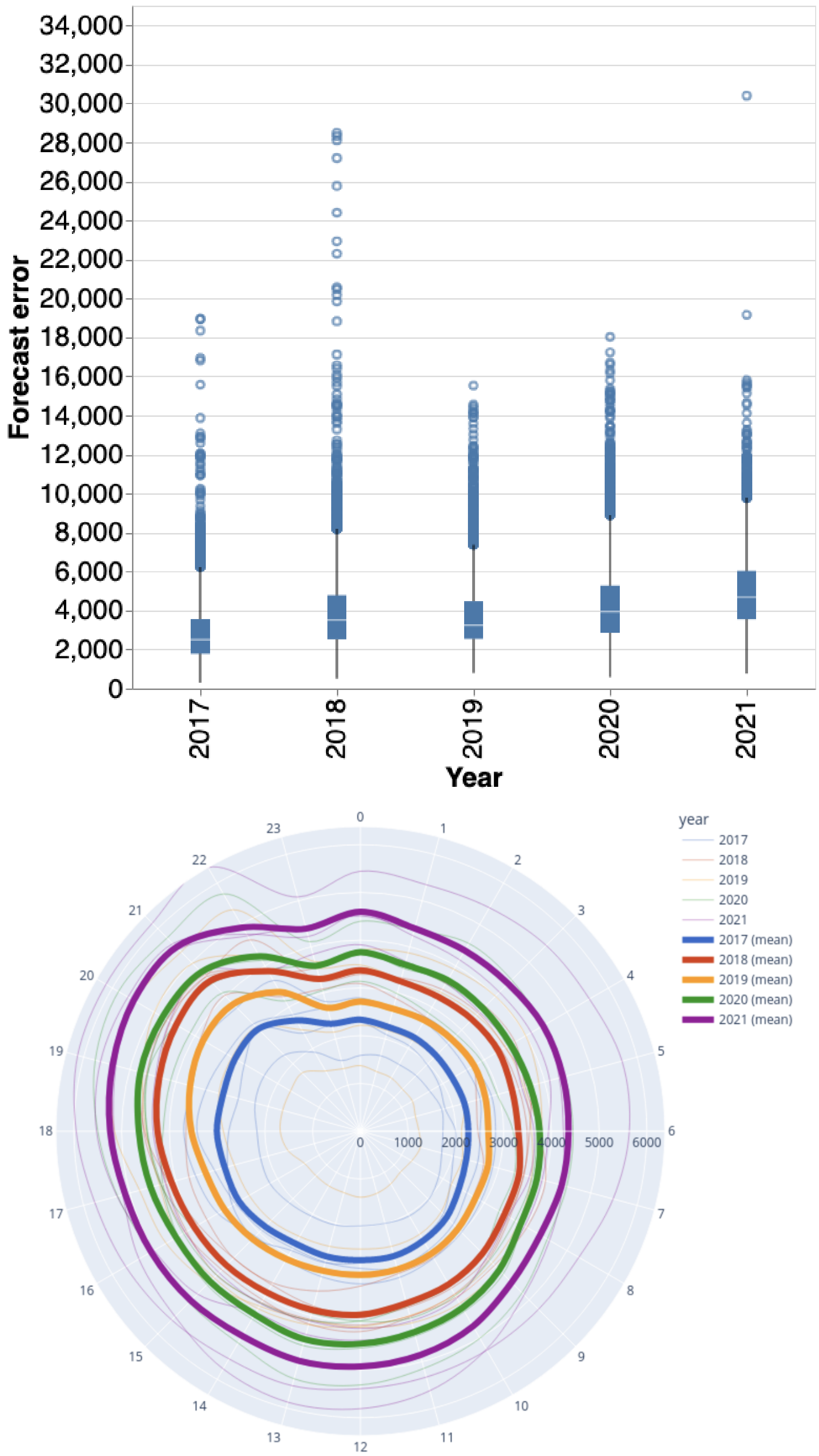


Figure 9: Wind forecast errors aggregated over all countries by (top) year, and (bottom) visualized as a function of time of day and season of year; in addition to visualizing the rising error, the plot also showcases how diurnal and seasonal variations can be quite important

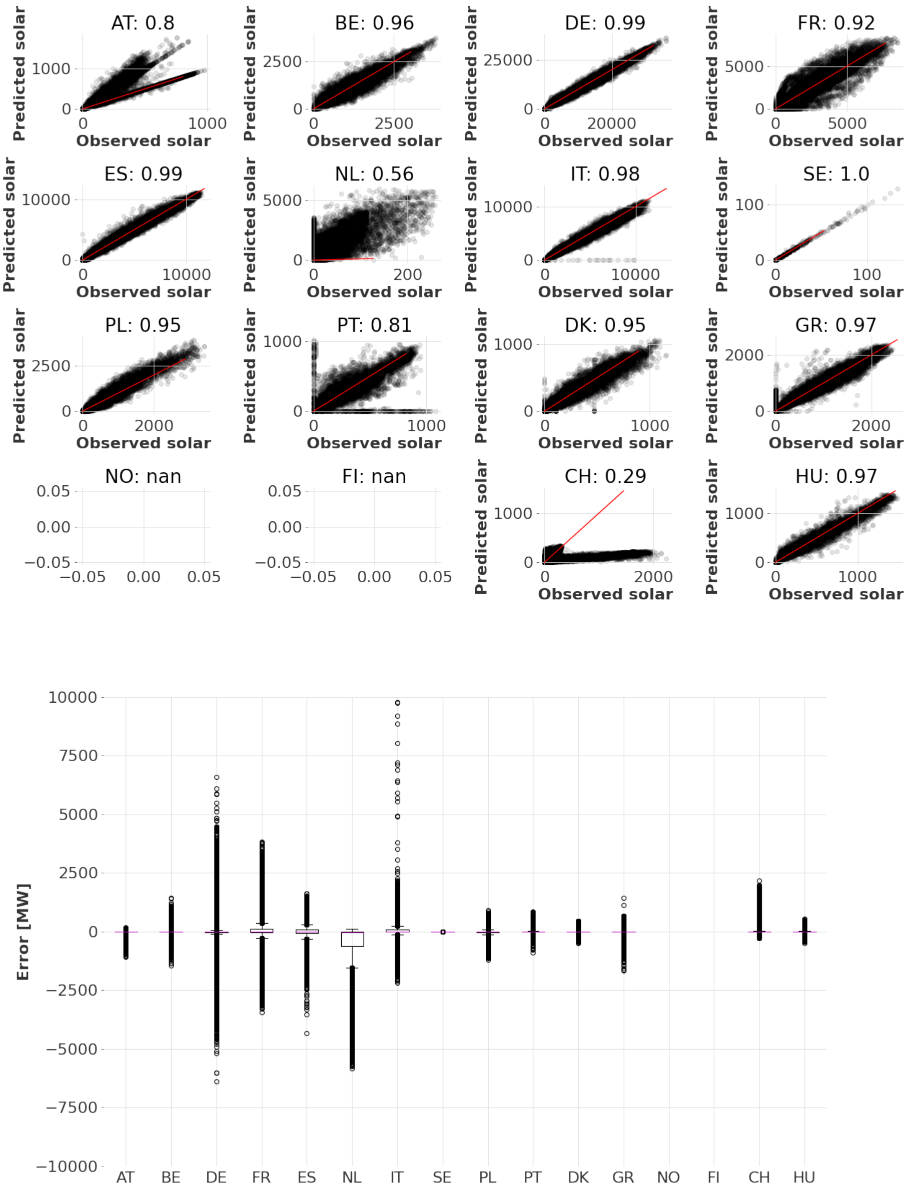


Figure 10: (Top) Scatter plot visualizing solar forecasts x-axis against observations y-axis by country, (bottom) box plot visualizing solar forecast error on an individual country basis

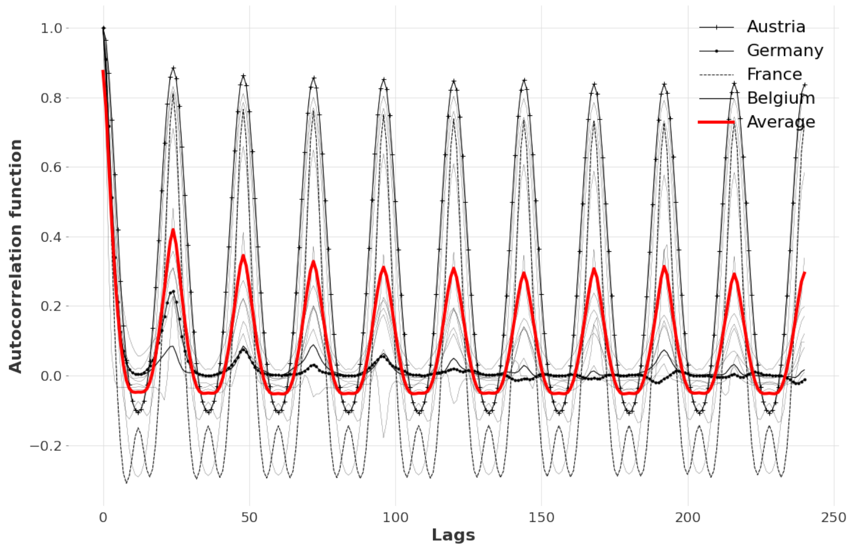


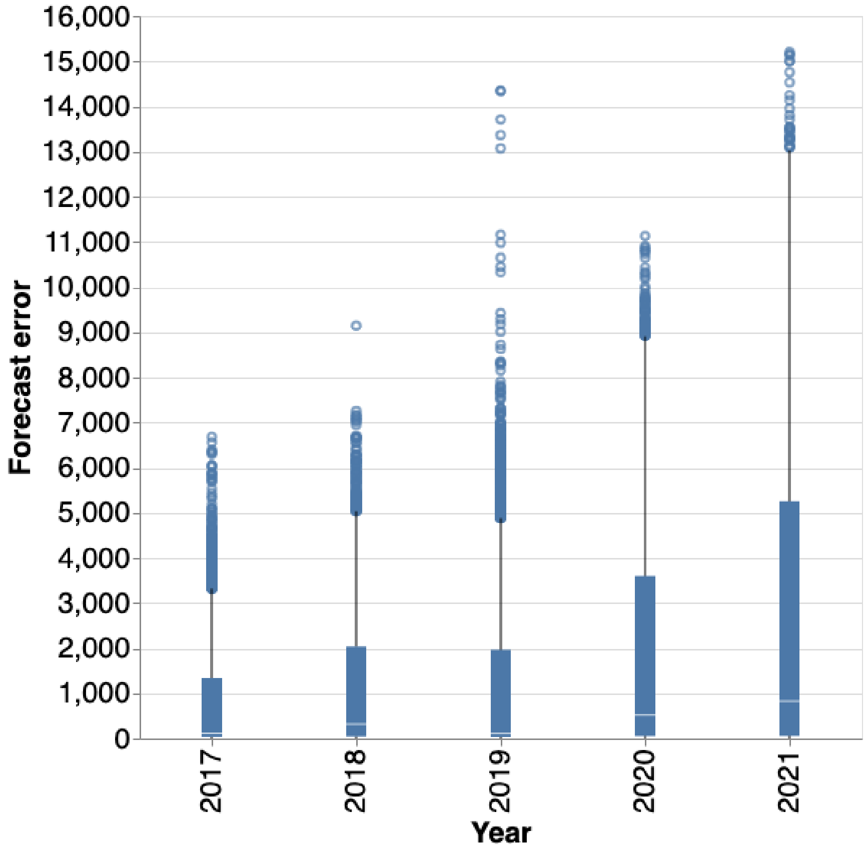
Figure 11: Autocorrelation function of solar forecast errors with the average highlighted; the ACF shows non-zero values at many lags indicating both serial and seasonal autocorrelations; most countries demonstrate a very strong diurnal seasonality in the forecast errors, with some also showing a sub-daily seasonality

during the day when the sun is actually shining. The diurnal, seasonal and annual variations of the forecast error can also be seen in the clock plot (Fig. 12). However, this strongly non-stationary nature of forecast error means that the grid operator now has to contend with dramatically different levels of uncertainty and forecast errors at different times of the day and year.

Finally, the MAE of the solar forecasts can be compared against the daily naive baseline to estimate the rMAE metric. In this case, the result is meaningless for several countries due to missing or (likely) incorrect data. For the remaining countries, the rMAE is generally higher than what was witnessed for wind generation and is, on average, only about twice as good as the baseline forecast. Sweden shows deceptively good performance but is excluded from the rankings because of its very low overall reported generation.

4. Discussion

This paper fills a significant gap in existing literature on the evaluation of publicly available forecasts of load and renewable generation across sixteen different countries. Owing to the diversity in energy demand and generation in these countries which are spread across Europe, the results can be expected to generalize well to other regions in the world as well. In this section, we cover some of the most important implications of these results.



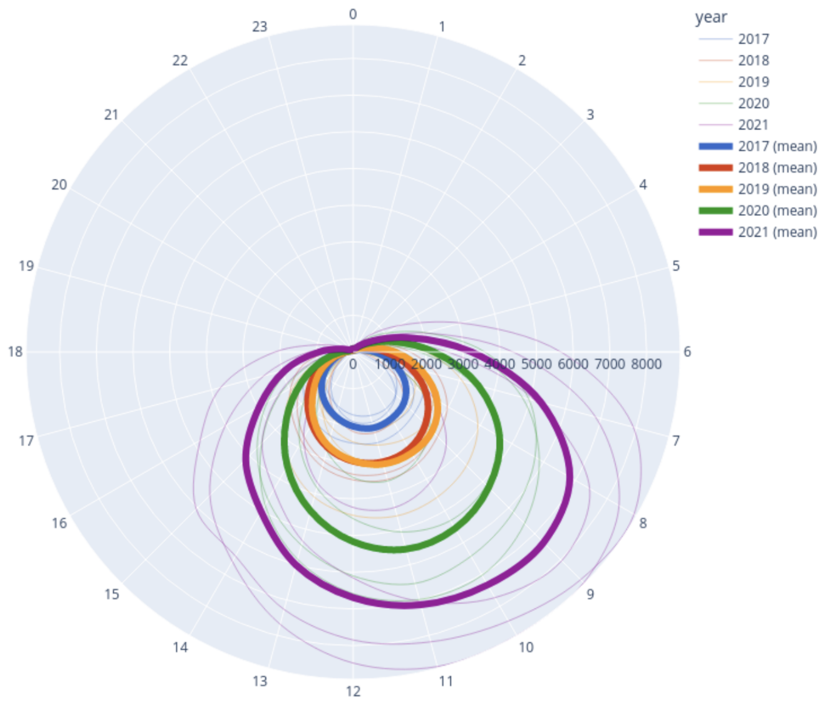


Figure 12: (Top) Evolution of solar forecast error aggregated across all countries from 2017 to 2021, (bottom) clock plot to visualize solar forecast error by time of day and year; in addition to visualizing the rapidly rising error, the plot also showcases how diurnal and seasonal variations will become increasingly important in grid operation

Country	Symbol	R^2	ME [MW]	MAE [MW]	WAPE [%]	rMAE
Austria	AT	.7	-42	43.6	34.3	1.61
Belgium	BE	.88	2.8	65.1	15.3	.45
Germany	DE	.89	.6	332	7.0	.31
France	FR	.82	-12.3	277	21.6	1.2
Spain	ES	.93	-22	164	8.4	.45
The Netherlands	NL	.56	-487	488	4361	116
Italy	IT	.44	79.3	212	9.6	.59
Sweden	SE	.82	0	.22	3.5	.05
Poland	PL	.96	-46	92.9	22.8	.91
Portugal	PT	.69	-.91	32.2	24	1.17
Denmark	DK	.87	1.9	20	16.5	.48
Greece	GR	.83	-7.7	40.2	9.3	.49
Norway	NO	NA	NA	NA	NA	NA
Finland	FI	NA	NA	NA	NA	NA
Switzerland	CH	0.27	97	109.8	78.3	2.04
Hungary	HU	.91	4.5	26.5	12.4	.44

Table 3: Summary of solar forecast error metrics for different countries

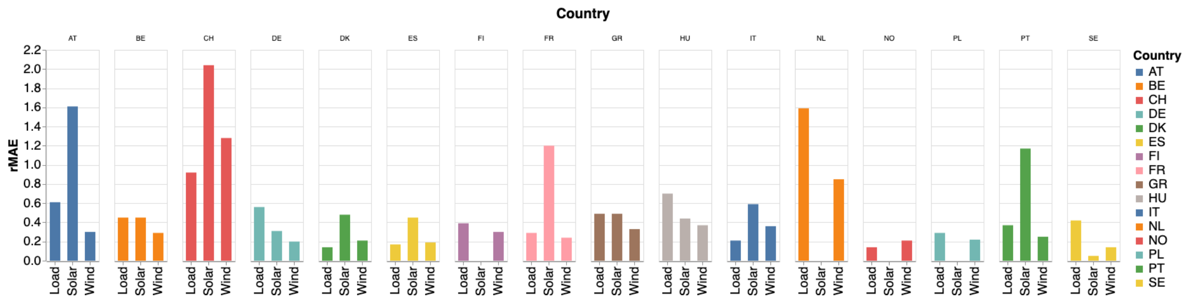


Figure 13: Relative MAE (rMAE) for different countries for load, solar and wind generation; a value of one would indicate a forecast that is no more skilled than the naive daily persistence model, while a value below one indicates a model which outperforms the baseline

4.1. Cross-country effects

As we observed earlier, there are obvious differences between the forecast accuracy of different countries, both in load and renewable generation forecasts. For instance, with seemingly comparable average electricity usage, the forecast error in France is less than half that in Germany. Likewise, a similar trend is visible for both wind and solar generation: we see a three-fold increase in forecast error between Poland and Greece, even though the average electricity generation through solar is roughly the same. This latter is of course partially explained by the different installed capacities in different countries and local conditions. However, at the end of the day, the installed capacity is far less important from an operational perspective than the energy and power actually generated and available to the grid.

The errors in the three time series under consideration are largely decorrelated or show only weak positive or negative correlations depending on the geographic zone. For instance, wind generation forecast errors are weakly negatively correlated for the Iberian peninsula (Spain and Portugal). This however does not hold in most other locations. Consequently, as a whole, it is fair to assume that the forecast errors different TSOs make tend to be largely decorrelated from one another. This is of course only a general guideline and does not necessarily hold in every instance. Practitioners and TSOs should therefore keep this in mind when planning for potential worst case outcomes. As an example, the 95th percentile of errors (load, wind and solar), assuming positive correlation, amounts to over 38 GW in 2021, up from around 22 GW in 2017. Most of the growth has come from renewable energy source forecasts.

4.2. Learning curves

Data quality and access issues notwithstanding, it is obvious that the TP is a valuable resource for data on forecasts of solar and wind generation as well as the expected load on the grid. However, there are considerable differences among the skill level of these various forecasts as shown in Fig. 13. The rMAE benchmarks all TSO forecasts against naive baselines. Overall, wind forecasts are the most accurate, followed by load and finally solar. However, this conclusion is somewhat biased by the fact that countries for which solar generation data is available is still rather low.

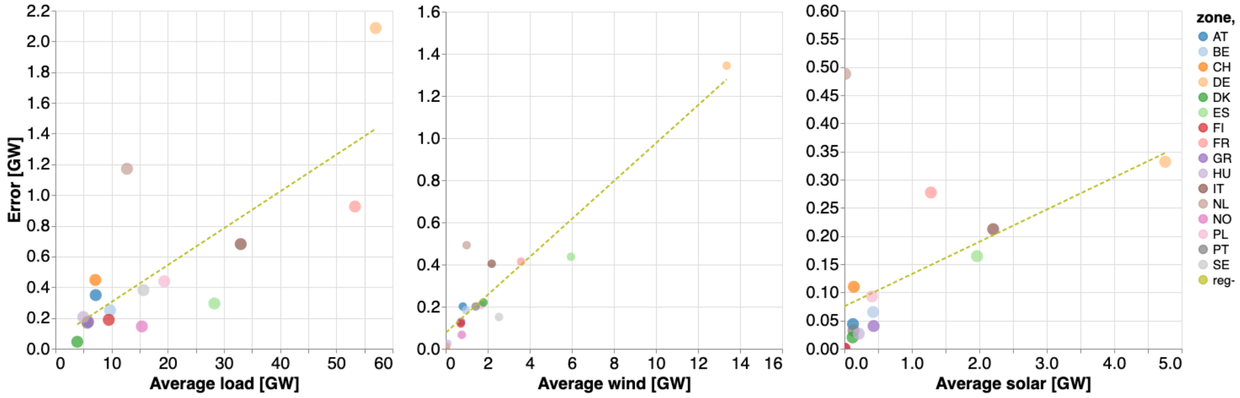


Figure 14: Forecast error as a function of increasing (left) load, (center) wind, and (right) solar; the y-axis represents the absolute error, which is relevant for system planning especially to prevent worst-case scenarios

In 2021, even though the median wind power forecast error was considerably higher than solar, taken together the two amount to around 5.5 GW, assuming decorrelated errors. This has grown roughly two-fold since 2017, and is now of roughly the same order of magnitude as the load forecast error, which has a median value of 7.5 GW. In fact, as Fig. 14 shows, there is an obvious trend of increasing error with increasing energy demand and generation for wind and solar. This trend is most pronounced for wind, where for each additional GWh/h generated, the forecast error rises by roughly 100 MWh/h. For solar, this amounts to 82 MWh/h and for load to only 25 MWh/h.

4.3. The road ahead

As the European Union pursues its ambitious climate and energy security goals, this error will take increasing centre stage in grid operation and planning. According to the latest climate goals, the EU aims to have renewables make up 45% of its overall energy mix by 2030, up from 40% under previous plans. This means doubling installed solar generation capacity by 2025, and expanding it by over 3.5 times by 2030. Similar growth targets are foreseen for wind power generation as well. This will lead to a worst case scenario with forecast errors almost an order of magnitude higher than what they are at present. Naturally, this will require large scale prescriptive analysis to understand and mitigate the effect of these forecast errors [41]. At the same time, these developments also necessitate both more accurate forecasts and the activation of distributed energy flexibility [42] for grids to continue to operate in a stable and secure manner.

The autocorrelation plots in our analysis have identified that, without exception, TSO forecasts can still be improved. There are several suggestions to achieve this:

1. Based on [33], a number of algorithmic remedies are possible, including increasing model order, feature transformations (such as smoothing, normalisation, differencing or log transformations) and use of techniques such as meta- and multi-task learning which are increasingly relevant for forecasting contexts as well [14, 43].

2. The TSOs should embrace greater transparency regarding their methodologies, and actively engage with the research community. An improved understanding of how these forecasts are made, possibly in conjunction with hackathons and competitions, will inevitably lead to breakthroughs helping improve their performance.

4.4. Limitations of the study

At this point, it is also important to discuss some important limitations of this study. First, the dataset we analysed was not perfect and there were significant amounts of missing data. Likewise, results for a number of TSOs represent not the actual forecast accuracy, but arguably underlying data quality issues. This holds true for especially the Netherlands and Switzerland. We also only consider the day-ahead point forecasts available on TP. However, many TSOs do provide other time horizons and interval forecasts on their own open data portals. A follow-up study could analyse these, along with carrying out a deeper analysis of the drivers behind large forecast errors.

Even for countries with no data quality concerns, it is unclear what fraction of forecast skill can be attributed to (improvements in) forecasting models or whether errors have simply evolved as a function of changing electricity (supply) mix. Likewise, it is difficult to ascertain whether the forecast model used by any TSO in 2021 is the same as the one used by that TSO in 2017 (or if several countries are using the same or similar models). It is also unclear how cross-border couplings have influenced the electricity demand and generation forecasts or whether neighbouring TSOs consider each others' forecasts, when making their own. We hypothesise that this consideration - if not already implemented - could further improve performance, but it must also account for the elevated risk of correlated errors. Additionally, what is included in the load and generation data that is being forecast varies as well; for instance, there are often important differences in the jurisdiction of TSOs in European countries and no single, unified framework to demarcate TSO-DSO (distribution system operator) responsibilities exists. Greater transparency on part of the system operators can considerably alleviate this.

5. Conclusion

In this paper, we have analysed openly available load and renewable generation forecasts from ENTSO-E TP, leading to several interesting insights. Rather surprisingly, the foremost insight was driven by what we expect to be data quality issues. Several forecasts (or, in one case, the actual generation series) provide garbled values which bear little resemblance to reality. The issue is arguably most obvious with data from the Dutch and Swiss TSOs, but other countries, including Germany and Austria, exhibit dubious quality data as well. Coupled with the frequent time-outs in retrieving data, the first conclusion is therefore that, despite its obvious utility, TP's real-world usability is limited and the data cannot necessarily be relied upon. ENTSO-E must consequently take steps to rectify this situation - potentially incorporating automated error detection techniques to avoid providing corrupted or incorrect data to downstream users.

Secondly, we discovered that the forecast error grows almost linearly in all three cases with increasing demand or generation using renewable energy sources. In fact, the combined forecast error due to solar and wind has roughly doubled during just the last five years, and is now of roughly the same order of magnitude as the load forecasts. As EU ramps up its renewable generation, this rapidly growing forecast error poses grave concerns about system stability. We also found that the TSO forecast skill measured using rMAE (i.e. compared to a naive baseline) is arguably the highest for wind forecasts, followed by load and finally solar. At the same time, while the forecast errors are not strongly positively correlated with one another, all of them exhibit very strong autocorrelation. This leads to the second and most important conclusion of the paper that, even when the data is not garbled, TSO forecasts should not be blindly relied upon. In fact, as our analysis shows, it is possible to improve on them by exploiting the structure still remaining in the residuals. This can be done by increasing model complexity, feature engineering as well as using more infomrative features in the first place etc.

Finally, in this paper, we have focused exclusively on day-ahead forecasts. The TP provides these in the form of point forecasts. There are also several other lower resolution forecasts, which include month ahead and year ahead forecasts. However, some TSOs (such as Elia the Belgian TSO), already provide interval forecasts for all three time series at multiple time horizons. Such interval forecasts, when properly calibrated, can then be used by downstream stakeholders and market agents to plan in risk-aware contexts. For instance, high uncertainty around demand forecast can force generators to bid conservatively. This leads to the final conclusion that the TP should introduce an additional requirement to the reporting TSOs to provide more frequently updated interval forecasts rather than focusing solely on day-ahead point forecasts.

In an increasingly coupled European electricity network, having observed and forecast energy demand and generation in a single, openly available repository can create tremendous value, besides facilitating several downstream users ranging from aggregators to balance responsible parties. At the moment however, the data and the forecasts made available through the TP are not always at a quality level that can be relied upon for downstream services. In this, the current work builds further on highlighting limitations of the TP, but also presents a way forward to help realize the ongoing energy transition.

Acknowledgment

Hussain Kazmi acknowledges support from Research Foundation – Flanders (FWO), Belgium (research fellowship 1262921N).

References

[1] K. C. Hoffman, D. O. Wood, Energy system modeling and forecasting, Annual review of energy 1 (1) (1976) 423–453.

[2] J. Á. González Ordiano, S. Waczowicz, V. Hagenmeyer, R. Mikut, Energy forecasting tools and services, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (2) (2018) e1235.

[3] D.-A. Ciupăgeanu, G. Lăzăroiu, L. Barelli, Wind energy integration: Variability analysis and power system impact assessment, Energy 185 (2019) 1183–1196.
[4] P. J. Baruah, N. Eyre, M. Qadrdan, M. Chaudry, S. Blainey, J. W. Hall, N. Jenkins, M. Tran, Energy system impacts from heat and transport electrification, Proceedings of the Institution of Civil Engineers-Energy 167 (3) (2014) 139–151.
[5] J. Lowitzsch, C. E. Hoicka, F. J. van Tulder, Renewable energy communities under the 2019 european clean energy package–governance model for the energy clusters of the future?, Renewable and Sustainable Energy Reviews 122 (2020) 109489.
[6] K. O. Adu-Kankam, L. M. Camarinha-Matos, Towards collaborative virtual power plants: Trends and convergence, Sustainable Energy, Grids and Networks 16 (2018) 217–230.
[7] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, H. Zareipour, Energy forecasting: A review and outlook, IEEE Open Access Journal of Power and Energy 7 (2020) 376–388.
[8] E. Worrell, S. Ramesohl, G. Boyd, Advances in energy forecasting models based on engineering economics, Annu. Rev. Environ. Resour. 29 (2004) 345–381.
[9] J. Zhao, Z.-H. Guo, Z.-Y. Su, Z.-Y. Zhao, X. Xiao, F. Liu, An improved multi-step forecasting model based on wrf ensembles and creative fuzzy systems for wind speed, Applied Energy 162 (2016) 808–826.
[10] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Conversion and Management 198 (2019) 111799.
[11] Y. Hao, C. Tian, A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting, Applied energy 238 (2019) 368–383.
[12] J. Hu, J. Heng, J. Tang, M. Guo, Research and application of a hybrid model based on meta learning strategy for wind power deterministic and probabilistic forecasting, Energy Conversion and Management 173 (2018) 197–209.
[13] C. Faloutsos, J. Gasthaus, T. Januschowski, Y. Wang, Forecasting big time series: old and new, Proceedings of the VLDB Endowment 11 (12) (2018) 2102–2105.
[14] T. Peirelinck, H. Kazmi, B. V. Mbuwir, C. Hermans, F. Spiessens, J. Suykens, G. Deconinck, Transfer learning in demand response: A review of algorithms for data-efficient modelling and control, Energy and AI 7 (2022) 100126.
[15] Z. Tao, J. A. Moncada, K. Poncelet, E. Delarue, Review and analysis of investment decision making algorithms in long-term agent-based electric power system simulation models, Renewable and Sustainable Energy Reviews 136 (2021) 110405.
[16] M. Chawla, M. G. Pollitt, Global trends in electricity transmission system operation: where does the future lie?, The Electricity Journal 26 (5) (2013) 65–71.
[17] A. M. Ersdal, D. Fabozzi, L. Imsland, N. F. Thornhill, Model predictive control for power system frequency control taking into account imbalance uncertainty, IFAC Proceedings Volumes 47 (3) (2014) 981–986.
[18] Z. Xu, J. Ostergaard, M. Togeby, Demand as frequency controlled reserve, IEEE Transactions on power systems 26 (3) (2010) 1062–1071.
[19] S. Goodarzi, H. N. Perera, D. Bunn, The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices, Energy Policy 134 (2019) 110827.
[20] C. Kath, F. Ziel, The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts, Energy Economics 76 (2018) 411–423.
[21] A. Dunbar, F. Tagliaferri, I. M. Viola, G. Harrison, The impact of electricity price forecast accuracy on the optimality of storage revenue, in: 3rd Renewable Power Generation Conference (RPG 2014), IET, 2014, pp. 1–6.
[22] L. Exizidis, J. Kazempour, P. Pinson, Z. De Grève, F. Vallée, Impact of public aggregate wind forecasts on electricity market outcomes, IEEE Transactions on Sustainable Energy 8 (4) (2017) 1394–1405.
[23] M. Farrokhabadi, J. Browell, Y. Wang, S. Makonin, W. Su, H. Zareipour, Day-ahead electricity demand forecasting competition: Post-covid paradigm, IEEE Open Access Journal of Power and Energy (2022).
[24] S. R. Beckman, The sources of forecast errors: experimental evidence, Journal of Economic Behavior & Organization 19 (2) (1992) 237–244.

[25] S. B. Taieb, R. J. Hyndman, et al., Recursive and direct multi-step forecasting: the best of both worlds, Vol. 19, Citeseer, 2012.

[26] S. Schelter, F. Biessmann, T. Januschowski, D. Salinas, S. Seufert, G. Szarvas, On challenges in machine learning model management (2018).

[27] J. W. Busby, K. Baker, M. D. Bazilian, A. Q. Gilbert, E. Grubert, V. Rai, J. D. Rhodes, S. Shidore, C. A. Smith, M. E. Webber, Cascading risks: Understanding the 2021 winter blackout in texas, Energy Research & Social Science 77 (2021) 102106.

[28] H. Holttinen, J. Kiviluoma, N. Helisto, T. Levy, N. Menemenlis, L. Jun, N. Cutululis, M. Koivisto, K. Das, A. Orths, et al., Design and operation of energy systems with large amounts of variable generation: Final summary report, iea wind tcp task 25, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2021).

[29] H. Jeon, B. Hartman, H. Cutler, R. Hill, Y. Hu, T. Lu, M. Shields, D. D. Turner, Estimating the economic impacts of improved wind speed forecasts in the united states electricity sector, Journal of Renewable and Sustainable Energy 14 (3) (2022) 036101.

[30] R. A. Rajagukguk, R. A. Ramadhan, H.-J. Lee, A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power, Energies 13 (24) (2020) 6623.

[31] M. Rahman, M. Shakeri, F. Khatun, S. K. Tiong, A. A. Alkahtani, N. A. Samsudin, N. Amin, J. Pasupuleti, M. K. Hasan, et al., A comprehensive study and performance analysis of deep neural network-based approaches in wind time-series forecasting, Journal of Reliable Intelligent Environments (2022) 1–18.

[32] X. Zhao, S. Wang, T. Li, Review of evaluation criteria and main methods of wind power forecasting, Energy Procedia 12 (2011) 761–769.

[33] H. Hewamalage, K. Ackermann, C. Bergmeir, Forecast evaluation for data scientists: Common pitfalls and best practices, arXiv preprint arXiv:2203.10716 (2022).

[34] ENTSO-E, Transparency platform, URL https://transparency. entsoe. eu (2019).

[35] H. Kazmi, Í. Munné-Collado, F. Mehmood, T. A. Syed, J. Driesen, Towards data-driven energy communities: A review of open-source datasets, models and tools, Renewable and Sustainable Energy Reviews 148 (2021) 111290.

[36] ENTSO-E Transparency Platform, Central collection and publication of electricity generation, transportation and consumption data and information for the pan-european market (2019).

[37] M. Radi, G. Taylor, F. Oliveira, J. Köhlke, M. Uslar, Prospective expansion of the entso-e transparency platform to include tso–dso interaction and wider market participation, in: CIRED 2020 Berlin Workshop (CIRED 2020), Vol. 2020, IET, 2020, pp. 819–822.

[38] L. Hirth, J. Mühlenpfordt, M. Bulkeley, The entso-e transparency platform–a review of europe's most ambitious electricity data platform, Applied energy 225 (2018) 1054–1067.

[39] G. E. Box, D. A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, Journal of the American statistical Association 65 (332) (1970) 1509–1526.

[40] Y. Shin, P. Schmidt, The kpss stationarity test as a unit root test, Economics Letters 38 (4) (1992) 387–392.

[41] C. A. Sima, M. O. Popescu, C. L. Popescu, M. Alexandru, G. Lazaroiu, Increasing ress share using generation and transmission expansion planning-stochastic approach, in: 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), IEEE, 2019, pp. 1–6.

[42] B.-M. Hodge, C. B. Martinez-Anido, Q. Wang, E. Chartan, A. Florita, J. Kiviluoma, The combined value of wind and solar power forecasting improvements and electricity storage, Applied Energy 214 (2018) 1–15.

[43] M. Matijaš, J. A. Suykens, S. Krajcar, Load forecasting using a multivariate meta-learning system, Expert systems with applications 40 (11) (2013) 4427–4437.