Arbitrage energy and wind power forecasting in the presence of diverse actors.

Damien Fay, Soumyendu Sarkar

[1]Hewlett Packard Enterprise

Abstract

In this paper we examine wind power forecasting in a grid setting where consumers and generators of power have differing forecast lead times. We first examine a set of leading time series forecasting models and asses their performance using data from the Irish grid. We then design an algorithm to calculate the arbitrage energy which is the energy available to an actor due to the uncertainty in the energy market. Our simulations suggest that a significant amount of energy might be available to data centers but this is highly dependent on the other actors present in the grid.

1 Introduction

In the past, energy grid operation was relatively straightforward: a forecast of the required demand was made, and then the only actors to be orchestrated were the different generation plants. These were scheduled, in advance, to meet this demand (i.e. unit commitment (Bukhsh 2024)). This was complicated by the fact that generation plants have differing lead times (coal plants require 24 hours notice while a gas plant typically requires 4 hours notice). However, once scheduled the generation was known, deterministic. Modern grid systems are very different (Meena et al. 2024), generation now consists of the old actors plus stochastic generation (wind, solar and to a lesser extent wave energy) while some demand can be shaped using price signals but yet others can be stochastic[1]. Storage technologies add another layer of diversity to the system; for example Lithium-Ion batteries account for 300Mw of storage capacity in the Irish grid alone (6% of typical system demand (ESB 2023)). As an actor, a battery is not simply a sink for excess demand, their profitability depends on the cycle time. Discharging a battery is when an operator generates income and so doing this as often as possible is their aim. In addition, across the entire spectrum of actor types there is an impressive array of emerging technologies (See (ESB 2023, 2024) for a review/evaluation of these technologies) and so the grid is destined to be become a crowded eco-sytem of actors with diverse goals.

[1]For example, a drop in sunshine can appear as a rise in demand to the grid from customers with solar panels. In addition, Data Centers with on-site generation can require grid power for reasons that are not visible to the grid operator.

In this complicated eco-system one might wonder how a time series forecast of generation is relevant to an actor. In this paper we begin to analyze this situation, the key insights are that i) actors have different forecast origins and ii) face cost functions which are asymmetrical. Differing forecast horizons lead to a difference in certainty, specifically the forecast distribution of each actor differs. Asymmetrical cost functions arise typically because commitments to generate power which are not met can be very costly. The combination of asymmetrical cost functions and differing forecast distributions lead to us to develop the Arbitrage Energy (AE) approach in this paper.

This research is in part motivated by the environmental problems with energy generated from fossil fuels. Indeed transitioning to renewable sources of energy has of late become an imperative with much of the world intending to to become carbon neutral by 2040.[2] In parallel, data center growth is projected to double by 2026 (IEA 2024) which will complicate achieving these goals. One grid in particular (Ireland) has already arrived at a situation where generation is dominated by stochastic renewables (wind) while demand includes a large data center component (20% and rising rapidly). As data centers have the capacity to shape their demand the matching of stochastic wind generation with data center demand looks like an ideal combination. However, as mentioned above in a complex grid eco-system the available power depends on the actions of other actors in the system.

Ireland boasts the 3^{rd} highest % of electricity generation from wind in the world (at 33% of the total) (Ember 2024). The temperate climate also means that air temperatures do not vary excessively and so are ideal for (air cooled) data center operations. This and a (relatively) low population density mean that data center operations consumed 5% of total electricity generation in 2015 rising to 21% in 2023 and is projected to rise to 33% by 2026.(G. Głaczyński 2023) This figure is far higher than the world average of 2%, or the 4% consumed in the USA (IEA 2024).

The national grid in Ireland is jointly operated by Eirgrid (south) and SONI (northern Ireland) as a single all Ireland grid system. The market mechanisms are quite complex (see [3]

[2]www.statista.com/chart/26053/countries-with-laws-policy-documents-or-timed-pledges-for-carbon-neutrality/

[3]www.sem-o.com

for details) but for our purposes we can start with the system demand. Any grid must balance the system demand/load and generation almost identically (a small deviation can be tolerated by varying the supply frequency) and excess generation must be either stored or curtailed. Different generating plants have different lead times (and constraints) with coal powered stations requiring up to 24 hours advance notice to ramp up and down, gas requiring in the order of 4 hours to ramp up but these stations can be put on standby in the short term (up to several minutes in advance). Wind and solar generation in contrast will generate whatever is available from the environment and while they can be reliably curtailed their production is stochastic and so must be forecast in advance. There is a preference towards renewable sources of generation and in the event a gas plant has been committed it will be asked to curtail its production in favour of a renewable source. In the case of curtailment (for any generator) they must be compensated for the load they could generate but are being asked not to. In addition, the grid is required to generate excess energy called a spinning reserve which can be switched into the grid to meet unexpected short term spikes in demand.

The stochastic nature of wind generation complicates unit commitment. Figure 1 shows the wind power carried by the grid between April and July in 2024. In addition, the system forecasts are available and these show that on average there is a 474MW error. This is a considerable amount of energy given that system demand in Ireland is 5GW (i.e. approximately 10% of system demand and half the consumption of all data centers). In the event this is an over-prediction of production then standby stations must make up the slack. Conversely, in the event of an under-prediction this power is consumed by asking other players to either reduce their generation or increase their consumption. Both actions are costly but wind power is by its very nature highly unpredictable.

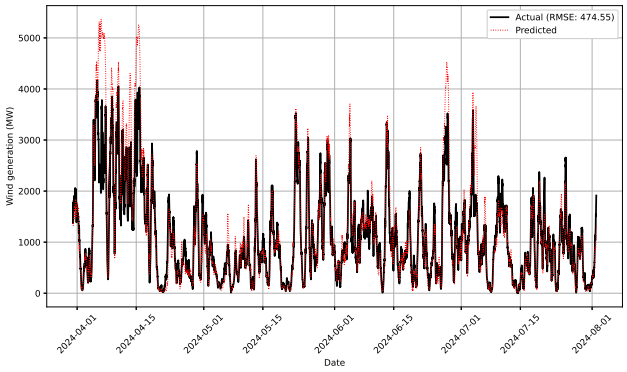

Figure 1: Wind generation forecasts and actuals. Note these figures are the actual wind generated i.e. with curtailment. Collected from **www.smartgriddashboard.com** between April to July 2024.

The data in Figure 1 does not take into account *curtailment*. This is wind generation that is not carried by the grid due a lack of sinks for the energy or a lack of capacity on the network to carry this power. The figures without curtailment are larger (see Section 5) and in this research we use the *available* wind energy provided to us by Eirgrid and so we examine the true stochastics of wind, independent of system infrastructure limitations.

The aim of this paper is to establish a framework for evaluating how an actor might benefit in a grid eco-system. For example it may seem that a data center in Ireland could easily shape its demand to soak up excess wind generation at very short notice. However, if other actors have already booked this power then pursuing this approach has little reward. Conversely, building storage might not be necessary if demand shaping actors are already available. It is not immediately clear how one should approach this situation and so this paper seeks to:

- establish a reasonable model for predicting the wind power generation in Ireland,
- show how this model may be used by different actors to establish the arbitrage energy available to an actor (the energy available to that actor in the presence of others),
- obtain a ballpark estimate for the arbitrage energy that might be available to data centers under different situations,
- highlight issues with energy forecasting models that are important for their use in future grid eco-systems (for example accurate forecast posteriors (see Section 5).

The remaining paper is organized as follows. Section 2 is the related work. Section 3, introduces the approach and gives the background. Section 4 describes the data used, and Section 5 presents the results and simulations. Section 6 concludes.

2 Related work

Wind Power Forecasting (WPF) has attracted much interest in the research community due to its importance in determining the relationship between wind and the energy both from a construction standpoint and that of grid operation on several different horizons. The ultra-short term horizon (15 minutes to several hours ahead) has in particular attracted much attention (Wang et al. 2021). Many statistical approaches have been taken including classical models, Gaussian Processes (Takao et al. 2022), decomposition techniques ((Hou, Wang, and Fan 2024)) and deep learning methods (Liu and Zhou 2024) amongst others. These models usually incorporate a temporal (Liu et al. 2024b) and sometimes a spatial element (Zhang et al. 2024) to account for the fact that wind at proximate locations can be used as a lead indicator.

The forecasting techniques employed in WPF mirror those used in time series forecasting in general. Recurrent neural network methods such as LSTM's (Liu and Zhou 2024; Huo et al. 2024; Liu et al. 2024b; Hou, Wang, and Fan 2024) and GRU's (Fantini et al. 2024) are popular but like all recurrent neural networks suffer from vanishing and exploding gradients. Methods such as temporal convolutional networks seek to bypass the long-short term dependencies by integrating information through convolutional filters. A recent paper added attention to identify salient parts of the convolutional output to the target further focusing the models attention on relevant parts of the time series (Lin, Koprinska, and Rana 2021).

Kolmogorov–Arnold Networks (KAN) (Liu et al. 2024a) have recently garnered a lot of attention due to their ability to outperform MLPs at general function approximation. When applied to time series forecasting they have not yet outperformed more established techniques (Vaca-Rubio et al. 2024) but then MLP's never performed well at this task either and so much attention has now focused on combining KAN's with convolutional filters etc. (Han et al. 2024).

Time series modeling has undergone a revolution in the types and variety of models available in recent years. The advent of attention mechanisms and transformers has led to the development of time series models which are considerably simpler than their LLM equivalents but still very powerful in terms of their representation abilities (Ahmed et al. 2023). This hybrid approach of combining different computational units (ex: LSTMs + attention, convolutional + KAN etc) is a common thread in the approaches cited above. One consequence however is that there is no one times series model (or even a compact set of models) that the practitioner can use, rather a time series model needs to be tailored to the data set at hand.

With particular reference to the underlying data generating process (DGP) for wind (Wei et al. 2024) proposes an interesting approach which looks at how the distribution of the wind process changes over the short term and may be used to estimate future values. This is important for the research presented here as we are interested in the inherent uncertainty in wind. Despite the many works looking at WPF few examines the underlying statistics of the wind process itself, (Bigdeli and Sadegh Lafmejani 2016) being one notable exception. (Takao et al. 2022) note the need for uncertainty forecasting for renewables in the UK for grid operation. They then employ a Gaussian process model which gives a posterior estimate of the forecast distribution which can be used to quantify uncertainty. (Liu et al. 2023) note that uncertainty in WPF is an understudied area. In particular, they use a Bayesian approach to estimate the volatility of the wind power forecast. A highly volatile forecast indicates that the actual power generated could be far below the mean forecast, **and this shortfall can be costly for generators**. This is an example of the asymmetrical cost function faced by actors. However, this paper is concerned with the generators while in our study we are concerned with unit commitment, dynamic demand shaping and actor interaction. It is argued in this research that it is the difference between knowledge at different horizons which is important and the serial correlation between errors at different time scales.

(Liu et al. 2023) discuss the idea of aleatoric and epistemic uncertainty loss in their paper which aligns with the concepts of process and measurement noise in state space systems, respectively. The aleatroic noise is due to the inherent noise on the wind process and this can only be estimated statistically. In contrast epistemic noise is caused by modeling error, measurement errors, etc. As our forecast horizon grows both sources of noise increase, the uncertainty in the wind process becomes greater as there is a strong auto-regressive (AR) component in the wind and innovations in the process compound each other (as with all AR processes (Hamilton 1994)). The aim in this paper is to estimate the effect of aleatoric uncertainty with forecast horizon but unfortunately we cannot separate the two sources. This led us to pick an assortment of time series models from those discussed above in order to find a model with a reasonably low epistemic noise.

3 Background

Figure 2 below shows a typical forecasting scenario. A forecast is made at an origin time t using information up to that point $t : t - r$ where r denotes the regressor in classical time series or the sequence length in recursive variants[4]. In addition, depending on the forecasters need the forecast horizon can often exceed 1 step ahead but rather includes a patch/window or size h which may not start until a delay of δ. For example a gas powered station requires four hours to dispatch generation which they will then maintain for say 6 hours. Therefore from their point of view $\delta : \delta + h = 4 : 10$ hours ahead. The process from $0 : 4$ hours ahead is of no interest as they dont have time to react to it in any case. It is interesting to note that as δ increase the mix of actors starts to change. Often the demand shaping actors can react on fast time scale but the generation actors (in particular base generation) requires a longer time frame. In addition to the above there exists lead indicators which can make an estimate of some variable in the future, for example a weather model based wind forecast, but we do not address this type of data in this paper.

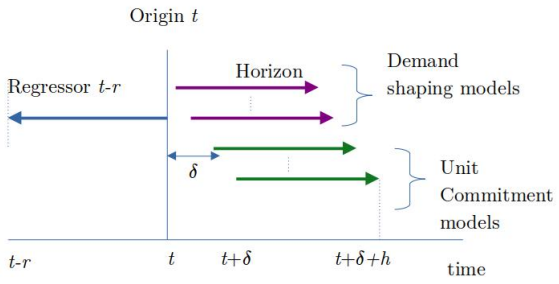

Figure 2: Sketch showing the actors interested in models as a function of their forecast horizons.

Specifically, we assume N actors, each characterized by their forecast horizon δ_i, confidence level $1 - \alpha_i$, posterior predictive mean $\hat{y}_i$, and posterior predictive standard deviation $\hat{\sigma}_i$, consume sections of the wind power generated. $\hat{y}_i$ and $\hat{\sigma}_i$ are estimated by the WPF model used by the actor. The first actor to commit is the one with highest forecast horizon. They commit to consuming $(1 - \alpha_i)$ percentile of their posterior predictive distribution of the forecast generation as the cost function they face is asymmetrical. Any remaining resource is updated after each actor's consumption and it is critical that **each actor is aware (or can estimate) the consumption of those in the chain before them.** Under these conditions the amount consumed by each actor can be calculated using Algorithm 1 below.

[4]Recursive models rarely recurse over the entire data set but rather go back r steps and assume the model state has passed any

Algorithm 1: Recursive Resource Allocation Algorithm

Input:

- y: Total available energy (true)
- $\mathcal{A} = \{(\delta_i, 1-\alpha_i, \hat{y}_i, \hat{\sigma}_i)\}_{i=1}^{N}$: Set of N actors s.t. $\delta_{i+1} < \delta_i$

Output:

- c_i: Resource consumed by each actor
- y_{ae_i}: Remaining energy after each actor's consumption

Initialize remaining energy $y_{ae_0} \leftarrow y$.

for $i = 1$ to N **do**

Compute estimated consumption:

$$y_c(i) = \Phi^{-1}(1-\alpha_i; \hat{y}_i, \hat{\sigma}_i)$$

Update remaining energy:

$$y_{ae_i} = y_{ae_{i-1}} - y_c(i)$$

Ensure remaining energy >0:

$$y_{ae_i} = \max(0, y_{ae_i})$$

Compute actual AE consumed:

$$c_i = y_{ae_{i-1}} - y_{ae_i}$$

end for

4 Data

The data used in this work comes from several different sources and covers the period of 01/01/24 to 30/06/24. Eirgrid provided the system demand data which contains the available wind generation (the target variable here) and the actual wind generation. The Eirgrid dashboard provides the generated wind and grid system wind generation forecasts (used in Figure 1).[5]. Actual weather readings for 18 weather stations are obtained from Met Eireann.[6]

In this research the models specifically have available 3 weather readings from 18 stations, $\{x^i, u^i, v^i\}$ where x is the wind speed, u, v are the cos and sin of the wind direction, and i denotes the station. In addition, the available wind power generation is denoted $y(t)$. As weather systems can last up to 12 hours (48 steps) and the aim of the model is to predict power from an evolving system we take a sequence length of 48 steps denoted $\{x^i_{t-48:t}, u^i_{t-48:t}, v^i_{t-48:t}, y_{t-48:t}\}$ where t denotes the forecast origin. In addition, the target is a frame as described in Figure 2: $y_{t+\delta:t+\delta+h}$ and so each model seeks to estimate $f : \mathbb{R}^{48\times 55} \rightarrow \mathbb{R}^{16}$. The input and targets are normalized prior to training using z-score normalization.

The original test/training set split was not considered appropriate (it had the test set entirely in summer when the wind is generally lower and better behaved) Therefore we took January and added this to the test set. The data is quarter hourly resulting in 11k training set points and 5k test set points (see Table 1 below)

Table 1: Data Set Overview

Set	Training	Test
Range	2024-01-31 to 2024-05-31	2024-01-01 to 2024-01-31 ∪ 2024-05-31 to 2024-06-30
N#	11649 (66%)	5825 (33%)

5 Results

In order to estimate aleatroic uncertainty in wind power we examined the performance of eight time series models that are widely used in the literature. A reasonably accurate model is required and in particular the RMSE obtained should be close to the optimal. The resulting models show prediction accuracies that are in line with those in Figure 1 which is encouraging. [7]

For each model type we train a model for a given value of δ (Figure 2) which makes a forecast from $t+\delta$ to $t+\delta+h$ where h is a window of length 16 (4 hours). δ is varied from 0 to 24 resulting in 24 models.

Table 2 shows the models that are used in this study (the relevant citations are included in the table). We chose a variety of models from a standard LSTM and GRU to the more recent PATe. In addition, we examined the KAN networks proposed in (Han et al. 2024). The number of parameters for each model can be seen to vary considerably with the TCM containing only 7.8K parameters compared to 3.2M for the PATe model. Training times are not reported but the KAN and PATe models took considerably longer and were trained on a separate machine with a GPU.

Table 2: Model Information with Citations and N_p Values

Model	Paper	N_p
TCN	(Lin, Koprinska, and Rana 2021)	7.8K
PATe	(Nie et al. 2023)	3.2M
GRU	(Yamak, Yujian, and Gadosey 2019)	49.4K
LSTM	(Yamak, Yujian, and Gadosey 2019)	65.6K
LSTMA	(Zhang et al. 2019)	65.6K
LFT	(Zhou et al. 2021)	154K
TST	(Ahmed et al. 2023)	208K
KAN-MOK	(Han et al. 2024)	109K

Figure 3 compares the performance of the models by RMSE as a function of the forecast horizon h (test set). These results show that the TCN and PATe model outperform the others significantly and consistently. This is interesting as these are the simplest and most most complex models respectively. The KAN model did not perform well and it structure may need adjusting to perform better on this data set. Table 3 gives further metrics. It shows that the LSTM, GRU

transient by time t.

[5]https://www.smartgriddashboard.com/

[6]https://www.met.ie/climate/available-data/historical-data

[7]Caveat: the system forecasts are generated at midnight for the entire day ahead. They also take into account curtailment and transmission restrictions and so do not forecast the available wind power as we do. Therefore they are not directly comparable but one would expect the RMSE to be in the same ballpark.

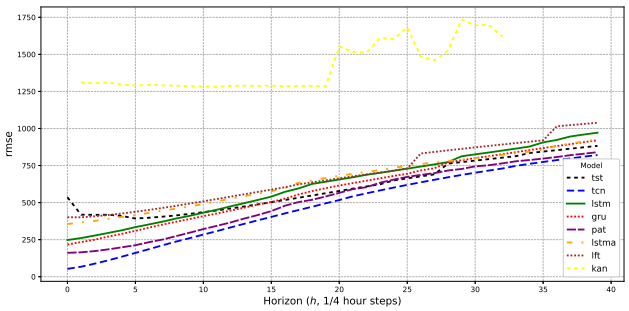

Figure 3: RMSE (test set) for each of the models as a function of h

and LSMTA achieve far higher accuracy on the training set than in the test set, indicating that they failed to generalize. Therefore the TCN is chosen as our forecasting model for the rest of the paper.

Table 3: Performance Metrics across Different Horizons

Model	**h**	**RMSE**		**MAE**	
		Trn	**Test**	**Trn**	**Test**
TCN	0	**50**	**161**	**41**	**41**
	4	139	**198**	103	**96**
	10	300	**321**	221	**200**
	16	458	**479**	339	**297**
PATe	0	140	355	110	110
	4	183	403	141	138
	10	324	490	242	221
	16	488	598	376	335
GRU	0	64	247	49	147
	4	**58**	313	**44**	203
	10	**57**	432	**43**	293
	16	**62**	572	**47**	388
LSTM	0	73	217	56	166
	4	64	289	49	222
	10	60	409	45	316
	16	66	527	50	428
LSTMA	0	80	401	61	256
	4	64	426	48	295
	10	65	509	49	363
	16	72	603	54	446
LFT	0	340	536	262	297
	4	385	410	291	319
	10	508	436	378	378
	16	625	517	472	451
TST	0	555	53	437	425
	4	323	135	250	308
	10	363	284	278	326
	16	475	427	352	382
KAN-MOK	0	-	-	-	-
	4	1105	1296	894	1063
	10	1093	1284	883	1053
	16	978	1284	771	1062

Figure 4 shows the RSME in the training and test sets (as a function of h) for the TCN model. As each model predicts a window of size 16 and there are 24 models, this plot shows 24 line segments (or length 16) laid on top of each other (leading to the slightly ragged appearance of the lines). The first thing to note is that the RMSE appears to be essentially independent of δ. For example the RMSE for a 14 step ahead forecast is present in several models. For the model with $\delta = 10, h = 4$ the RMSE is 395 which is very close to that from the model with $delta = 4, h = 10$ at 402. This is not surprising, although the models differ in the domain of their prediction windows, the 14 step forecasts are still based on the same information (its 14 steps from the origin). The second point to note is the similarity in the training and test set RMSE's. This is an encouraging sign the models have generalized well. In addition, we see that the forecast error rises with h as expected (the further from an event we are in time the harder it becomes to predict).

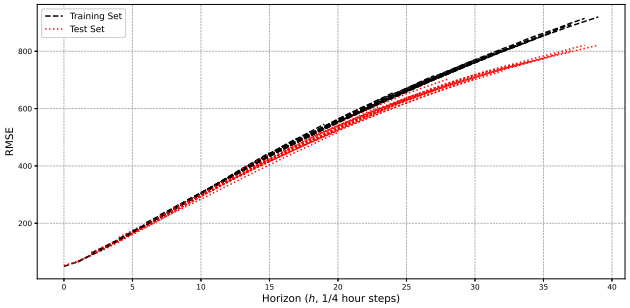

Figure 4: RMSE (test and training sets) as a function of h. Note the graphs are composed of multiple 16 step ahead forecasts.

Figure 5 shows and example of a 16 step forecast made by 3 different actors in the test set. The actors have different forecast origins starting at $\delta = 4, 16$, and 24 respectively. The larger δ the greater the expected error and this is the case here.

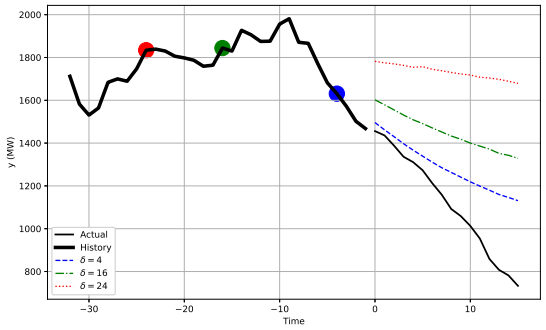

Figure 5: Forecast made by 3 actors. The circles show the forecast origin from which each actor makes their forecast.

The $1 - \alpha$ percentile is estimated under the assumption that the forecast posterior follows a Gaussian distribution

with $e(t) \sim \mathcal{N}\left(\mu = \hat{y}(t+\delta), \sigma = \hat{\sigma}(\delta)\right)$, where $\hat{y}(t+\delta)$ is the forecast at time $t+\delta$ and the forecast standard error is assumed to be a function of δ only and is estimated using the training set RMSE (Figure 4).

Figure 6 below examines how accurate the assumption that $\sigma(t,\delta) = \hat{\sigma}(\delta)$ by setting $\alpha = 0.2$ and measuring the actual percentage of residuals that lie above the corresponding level. We would expect that 20% of the actuals will lie below this percentile but figure 6 shows that the percentile discrepancy can vary from 35% to 15%. We will come back to the importance of minimizing this discrepancy in the conclusion.

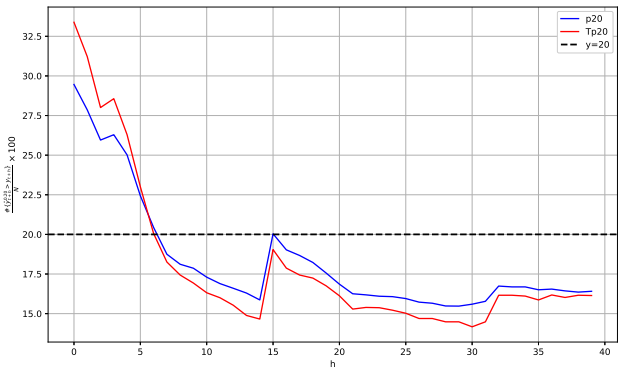

Figure 6: Percentile discrepancy as a function of δ.

The parameters described in Table 4 mimic the situation in which a gas generator and a data center operate with 4 hours and 1 hour horizons respectively. The forecasts produced by the TCN model allows us to simulate the AE that is allocated to each actor.

Table 4: Model Configuration

Model	**Actors**	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\delta_1}$	$\boldsymbol{\delta_2}$
TCN	2	0.2	0.2	16	4

Figure 7 shows the AE that is assigned by 1 for a period of 75 hours (300 steps) (this is a zoom in of the full result shown in Figure 8). As can be seen Actor 1 has the option to consume most of the generated energy (67%) while actor 2 has AE at 20% of the total. Even though actor 2 is only 1 hour (4 steps) from the origin there is still 13% of the energy remaining unused. The picture is similar when we zoom out to look at the entire test set in Figure 8 where 64%, 23% and 13% of the generation is consumed by actor 1 and 2. It is interesting to note that the AE available to actor 2 is bursty; the energy arriving in blocks. Next we examine the behavior of the AE over the range of δ.

Figure 9 shows the percentage of available energy consumed by actor 1 as a function of δ_1 and δ_2. The figure shows that this is invariant with δ_2 which is expected as actor 1's decision are not based on the actions of actor 2. Figure 10 in contrast shows the consumption level for actor 2. This is highly influenced by the actions of actor 1 and show that under some circumstances actor 1 can cannibalize the AE for actor 2.

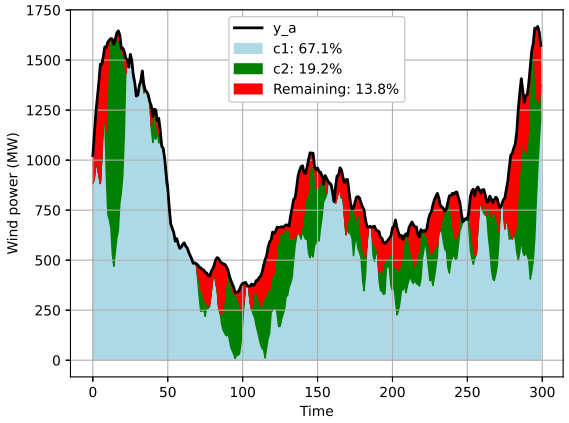

Figure 7: AE which actor 1 and actor 2 obtain during the simulation based on parameters in Table 4 (the full range is shown in Figure 8)

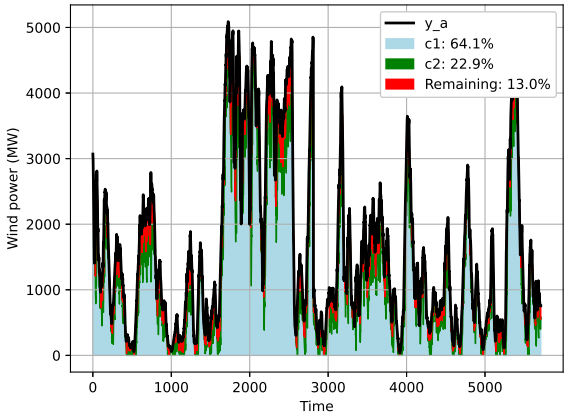

Figure 8: AE which actor 1 and actor 2 obtain during the simulation based on parameters in Table 4

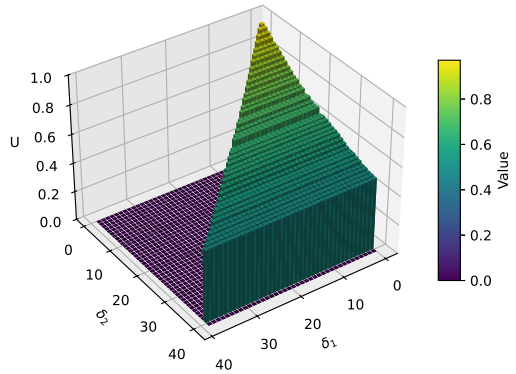

Figure 9: % of energy consumed by actor 1 as a function of δ_1 and δ_2, $\alpha = 0.2$

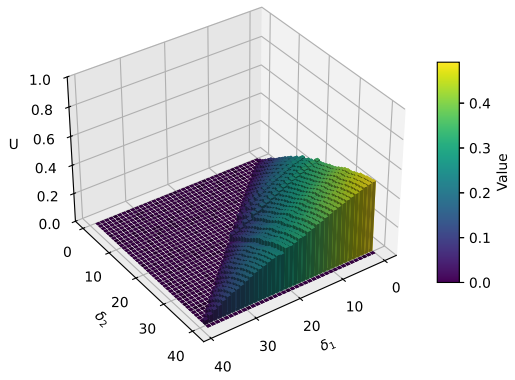

Figure 10: % of energy consumed by actor 2 as a function of δ_1 and δ_2, $\alpha = 0.2$

6 Conclusion

There are several interesting outcomes in this experiment. The first is that the forecasting situation faced by an actor is not as simple as just predicting the mean of the target. Indeed accurate estimates of the whole predictive distribution (or at least a percentile estimate) is necessary to optimize the amount of AE available. The experiments also demonstrated that a simple fixed estimate for the predictive variance based on a Gaussian assumption can lead to a large percentile discrepancy. As this is directly related to the energy an actor consumes there is a need for forecasting models to produce accurate predictive distribution in particular. Also as seen in Figure 10 an eco-system can be in danger of cannibalizing actors if not designed carefully. That said the simulations in Table 4 indicates that a data center could take advantage of approximately 20% of wind power generation and this certainly warrants further investigation.

References

Ahmed, S.; Nielsen, I. E.; Tripathi, A.; Siddiqui, S.; Ramachandran, R. P.; and Rasool, G. 2023. Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, 42(12): 7433–7466.

Bigdeli, N.; and Sadegh Lafmejani, H. 2016. Dynamic characterization and predictability analysis of wind speed and wind power time series in Spain wind farm. *Journal of AI and Data Mining*, 4(1): 103–116.

Bukhsh, W. 2024. The significance of time constraints in unit commitment problems. *IEEE Access*.

Ember. 2024. Global Electricity Review 2024.

ESB. 2023. Emerging Technology Insights 2023.

ESB. 2024. Emerging Technology Insights 2024.

Fantini, D.; Silva, R.; Siqueira, M.; Pinto, M.; Guimarães, M.; and Junior, A. B. 2024. Wind speed short-term prediction using recurrent neural network GRU model and stationary wavelet transform GRU hybrid model. *Energy Conversion and Management*, 308: 118333.

G. Głaczyński, G. B. 2023. Data Centres Metered Electricity Consumption 2023. *CSO statistical publication*.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press.

Han, X.; Zhang, X.; Wu, Y.; Zhang, Z.; and Wu, Z. 2024. KAN4TSF: Are KAN and KAN-based models Effective for Time Series Forecasting? *arXiv preprint arXiv:2408.11306*.

Hou, G.; Wang, J.; and Fan, Y. 2024. Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction. *Energy*, 286: 129640.

Huo, J.; Xu, J.; Chang, C.; Li, C.; Qi, C.; and Li, Y. 2024. Ultra-short-term wind power prediction model based on fixed scale dual mode decomposition and deep learning networks. *Engineering Applications of Artificial Intelligence*, 133: 108501.

IEA. 2024. Electricity 2024, Analysis and forecast to 2026. *IEA*.

Lin, Y.; Koprinska, I.; and Rana, M. 2021. Temporal Convolutional Attention Neural Networks for Time Series Forecasting. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Liu, L.; Liu, J.; Ye, Y.; Liu, H.; Chen, K.; Li, D.; Dong, X.; and Sun, M. 2023. Ultra-short-term wind power forecasting based on deep Bayesian model with uncertainty. *Renewable Energy*, 205: 598–607.

Liu, X.; and Zhou, J. 2024. Short-term wind power forecasting based on multivariate/multi-step LSTM with temporal feature attention mechanism. *Applied Soft Computing*, 150: 111050.

Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024a. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*.

Liu, Z.-H.; Wang, C.-T.; Wei, H.-L.; Zeng, B.; Li, M.; and Song, X.-P. 2024b. A wavelet-LSTM model for short-term wind power forecasting using wind farm SCADA data. *Expert Systems with Applications*, 247: 123237.

Meena, S. B.; Patil, P. R.; Kandharkar, S.; Hemalatha, N.; Khade, A.; Dixit, K. K.; and Chinthamu, N. 2024. The Evolution Of Smart Grid Technologies: Integrating Renewable Energy Sources, Energy Storage, And Demand Response Systems For Efficient Energy Distribution. *Nanotechnology Perceptions*, 1098–1109.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Takao, S.; Nassimiha, S.; Dudfield, P.; Kelly, J.; and Deisenroth, M. 2022. Short-term Prediction and Filtering of Solar Power Using State-Space Gaussian Processes. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*.

Vaca-Rubio, C. J.; Blanco, L.; Pereira, R.; and Caus, M. 2024. Kolmogorov-Arnold Networks (KANs) for Time Series Analysis. arXiv:2405.08790.

Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; and Liu, Q. 2021. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, 304: 117766.

Wei, H.; Chen, Y.; Yu, M.; Ban, G.; Xiong, Z.; Su, J.; Zhuo, Y.; and Hu, J. 2024. Alleviating distribution shift and mining hidden temporal variations for ultra-short-term wind power forecasting. *Energy*, 290: 130077.

Yamak, P. T.; Yujian, L.; and Gadosey, P. K. 2019. A comparison between arima, lstm, and gru for time series forecasting. In *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence*, 49–55.

Zhang, J.; Li, H.; Cheng, P.; and Yan, J. 2024. Interpretable Wind Power Short-Term Power Prediction Model Using Deep Graph Attention Network. *Energies*, 17(2).

Zhang, X.; Liang, X.; Zhiyuli, A.; Zhang, S.; Xu, R.; and Wu, B. 2019. At-lstm: An attention-based lstm model for financial time series prediction. In *IOP Conference Series: Materials Science and Engineering*, volume 569, 052037. IOP Publishing.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.