Wind power ramp forecasting and characterisation using Machine Learning methods

Russell Sharp

Student number: [redacted]

Supervised by: Hisham Ihshaish
Co-supervised by: Ignacio Deza

Word count: 13,141

Independent project report submitted for the degree of Master of Science in Data Science

University of the West of England

January 2022

Acknowledgements

First and foremost, I wish to thank my mother [redacted], for keeping me fed and watered throughout my time at UWE. My next biggest thanks undoubtedly go to Dr Hisham Ihshaish and Dr Ignacio Deza; two giants who allowed me to stand on their shoulders as I worked on this thesis. Dr Ihshaish's teaching superpowers and Dr Ignacio Deza's guidance with the wavelet transform methodology improved the quality of this research no end. I wish to also thank the remaining lecturers, assistants, and my fellow students of the 2021 UWE MSc Data Science course for sharing their knowledge with me. Thanks also to Dr Alexis Tantet, without whom this adventure never would have started. Finally, thanks also to Dr Cristóbal Gallego Castillo for his kind guidance at the start of this project and excellent back catalogue of work.

Abstract

The forecasting of large and rapid changes in wind power output known as *ramp events* is crucial for the incorporation of large volumes of wind energy into national electricity grids. Large variations in wind power supply must be compensated by ancillary energy sources which can include fossil fuels along with their associated costs and carbon emissions. Improved prediction of wind power will help to improve the competitiveness of wind energy against non-renewable energy sources, enable the global transition to cleaner fuels and facilitate climate change mitigation. This paper investigates the potential application of Machine Learning methodologies to the task of ramp event prediction. Using a non-binary wavelet transform approach to ramp characterisation, the ability of three Machine Learning models to predict ramp events in the Engie La Haute Borne wind farm dataset is compared. Results indicate that Recurrent Neural Network models show the greatest potential for ramp event forecasting.

Table of contents

List of figures

List of tables

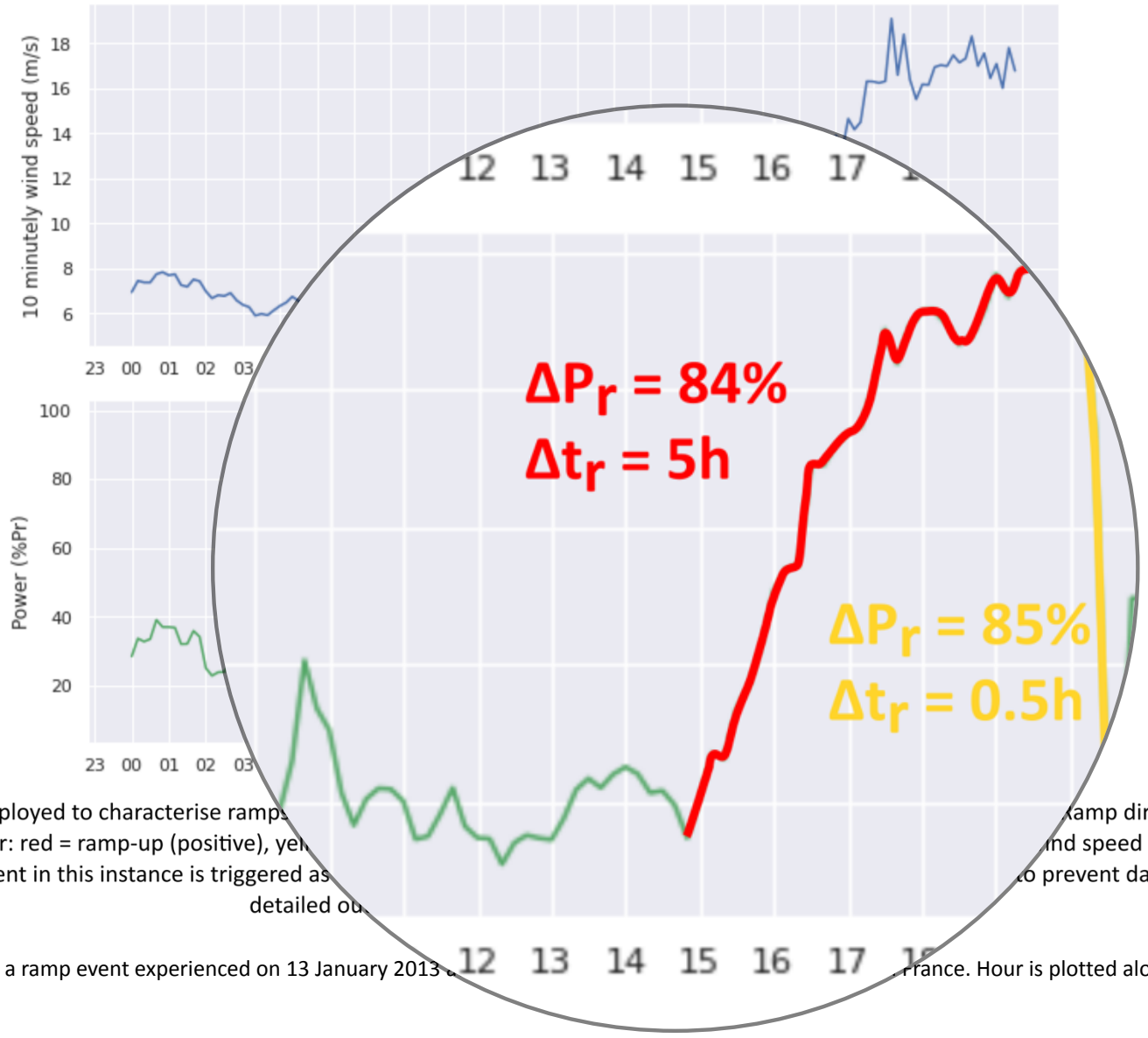
1 Introduction, rationale, and objectives

This chapter introduces and contextualises the ramp event problem that this study seeks to address. It includes an outline of the specific aims and success criteria, a consideration of ethical implications, and information required to access the source code repository.

1.1 Introduction

In recent decades, renewable [illegible] leading renew[illegible]d significant attention due to worldwide concerns [illegible]ns. As part of their commitments to [illegible]nal impact on the environ[illegible]ve committed to supplying cer[illegible]s by the year 2030. Win[illegible]rm dataset from France, a countr[illegible]s relatively low cost [illegible]16). This study f[illegible]has pledged to increase its installed wi[illegible]neet its climat[illegible]of 38.4 GW by 2028 (European Commiss[illegible]om 16 GW in[illegible]

Despit[illegible]by its inheren[illegible]ind power, its potential as an energy[illegible]e wind. Generati[illegible]as *ramp events*. Ran[illegible]to its direct correlation to the p[illegible]ind farm or portfolio of w[illegible]arge increases (positive ramp) o[illegible]udden large ramps in wi[illegible]to as *wind ramps*. Ramp events are manife[illegible]s, generally over a short period of up to a few hours (Cutle[illegible]astillo et al., 2015) (Figure 1).



Parameters employed to characterise ramps … Ramp direction is given by colour: red = ramp-up (positive), yel… …nd speed (1). The ramp-down event in this instance is triggered as … …o prevent damage. A detailed ou…

Figure 1. Example of a ramp event experienced on 13 January 2013 … France. Hour is plotted along the x-axis.

1.2 Rationale and user needs

The variability of wind power reduces its dependability in the eye of electricity Transmission System Operators (TSOs) wishing to incorporate commercial volumes into national power grids. TSOs must continually manage their power networks so that supply meets demand (Cutler et al., 2007). This is commonly done through the scheduling of *ancillary reserves* which consist of supplementary power sources that are flexible enough to adapt to variations in load and supply. This, however, carries additional costs and emissions. Most electricity markets operate around a set of short-term procedures known as Day-Ahead (DA) operations that enable them to [illegible] (RT) energy dispatch operations. Part of the DA operation[illegible] by a given deadline, market participants su[illegible]rovide for the following day to the [illegible]meframe. This is especially true

Wind [illegible] negative ramp in wind power coul[illegible]gy into power net[illegible] From a TSO p[illegible]haintain the supply-demand balance, wh[illegible]onse from a[illegible]amp might r[illegible]ow be kept below planned maxima in co[illegible]e network (C[illegible] increases (k[illegible] As the amount of wind energy in a [illegible]ind power pre[illegible]*tration*), so too does the impor[illegible]

Wind po[illegible]uture time period spanned by [illegible]nto three categories, depending on the [illegible]

Table 1. Categories of wind powe[illegible]ínez-Arellano, 2015).

Category	Horizon
Very short-term	2 – 4/9 hours
Short-term	4/9 – 48/72 hours
Medium-term	72 hours - 7 days

Due to the increasing volume of wind power that is set to be incorporated into electricity markets by TSOs, and due to the DA operational timescales outlined above, short-term forecasts of wind power are currently an active area of research and form the horizon of interest in this and future research planned by the author.

1.3 Current practices

Timeseries (TS) forecasting is a specific area of Machine Learning (ML) in which trained models predict new data, even though actual outcomes may not be known until some future time or date (Brownlee, 2019). Machine Learning and timeseries forecasting techniques have been applied to the problem of wind power ramp event forecasting with some promising results (Ahmadi and Khashei, 2021; Gallego-Castillo et al., 2015; Sim and Yung, 2020; Wang et al., 2019; Yang et al., 2021 and references therein) but the field of study is still in its infancy.

There are two types of method for wind power forecasting: those based on timeseries analysis alone (TS), and those based on a combination of timeseries analysis and Numerical Weather Prediction (NWP) model outputs (TS+NWP). NWP models[1] resolve a set of physical equations in order to estimate the dynamics of the atmosphere and output forecasted values of target variables to a 3D grid, but they do so at significant computational cost. A high-resolution NWP run of 0.5 km for example may take 48 hours to complete with limited resources, effectively rendering it useless for DA operations. *Downscaling* is a procedure that reduces computational cost by transforming NWP outputs from the low-resolution grids of NWP models to higher resolutions at specific physical locations of interest. It is commonly performed by statistical analysis of historic data to establish systematic relationships between NWP forecasts and measured observations (Cutler et al., 2007). Nevertheless, the process of downscaling is prone to error from two primary sources: the NWP estimates themselves, and the methods used to convert wind forecasts to power (Martínez-Arellano, 2015). Attempting to reduce the level of error in the downscaling process provides the initial motivation for this research.

[1] e.g., US National Centre for Atmospheric Research (NCAR) Weather Research and Forecasting (WRF) model, European Centre for Medium-Range Weather Forecasts ERA5 model.

Whilst it is well established that TS+NWP prediction models typically out-perform TS approaches after a 3 to 6-hour time horizon (Giebel et al., 2011; Martinez-Arellano, 2015), the computational demands of NWPs still exert challenging constraints on the deployment of downscaling in any domain-specific application. Driven by this, this paper focusses on one aspect of the downscaling process: the statistical features of historic wind power data. The rest of this study investigates the potential of applying Machine Learning methods to the task of establishing systematic relationships between measured and forecasted observations of wind power. As will be outlined in Section 3.4, the inclusion of NWP outputs can still be used as input to the developed approach and the effect of their inclusion on modelling outcomes can be examined. This will be the aim of further research planned by the author.

1.4 Objectives

The specific aim of this study is to provide a conceptual overview of a family of Machine Learning models that could be ... wind power and associated ramp events. This w... tigate improving the accuracy of short... P outputs.

To achieve ... nalysis (Exploratory Data Analysis, ... her in the following ...

1. Und... Borne open dataset) in order to:
   - ... d farm da...
   - ... statistical characteristics of the data; ... and
   - ... ent ML ... onarity,
   - ... siderations and pre-processing requir...
2. Compare ... enges, and opportuniti... onsiderations are the timeseri... vents in the La Haute Borne dat... mprises:
   a. Autoregressive ... end, seasonality and r...
   b. Autoregressive Integrated M... (ARIMA)

c. Seasonal Autoregressive Integrated Moving Average (SARIMA)
d. Prophet
e. Recurrent Neural Network

3. Test each ML algorithm's ability to fit the La Haute Borne wind farm dataset. This is done by modelling the relationship between historic wind power observations (training set) and wind power predictions (test set) and evaluating power prediction and ramp capture accuracies[2]. This objective essentially forms a precursory step to the inclusion of NWP outputs as planned in future research (see Section 3.4).

1.5 Success criteria

The success of this research will be gauged by three main criteria:

1. The predictive accuracy achieved by the best ML model. State-of-the-art wind power forecasting systems [illegible] total installed capacity over a 36-hour hori[illegible] Whilst final short-term forecasti[illegible]roaching this level of ac[illegible]
2. Th[illegible]3.3). The best p[illegible]
3. [illegible]ast (24-72 [illegible]-perform [illegible]ecessarily [illegible]ho (2015) [illegible]ution took a[illegible]. Given the sin[illegible] respectively accor[illegible]P run time of 6 hours to [illegible]rea. Considering an average work[illegible] hours would enable DA wind power to be [illegible]. Model run times of less than 1.5 hours will therefore be considered a success.

nces of the ML model be considered for future develop As highlighted previously, most utiliti vell established that TS+NWP prediction er a 3 to 6-hour time horizon, planned fu NWP outputs. Using the NCAR WRF mo el run covering Spain at a 30 km y to forecast a 12-hour horiz nd France (505.95

[2] It is important to note that this approach permits only a limited evaluation of predictive performance since only one single future time step is forecast at a time. Short-term forecasting (where a 9-72 hours-ahead sequence of time steps is predicted) will be the subject of further research.

1.6 Ethical, legal, and professional considerations

The dataset used in this study (Section 3.2) is published under Open License version 2.0 published by Etalab, France[3]. As such, there are no ethical, legal, or professional restrictions on its use. This work contains no personal data of any sort.

1.7 Code availability

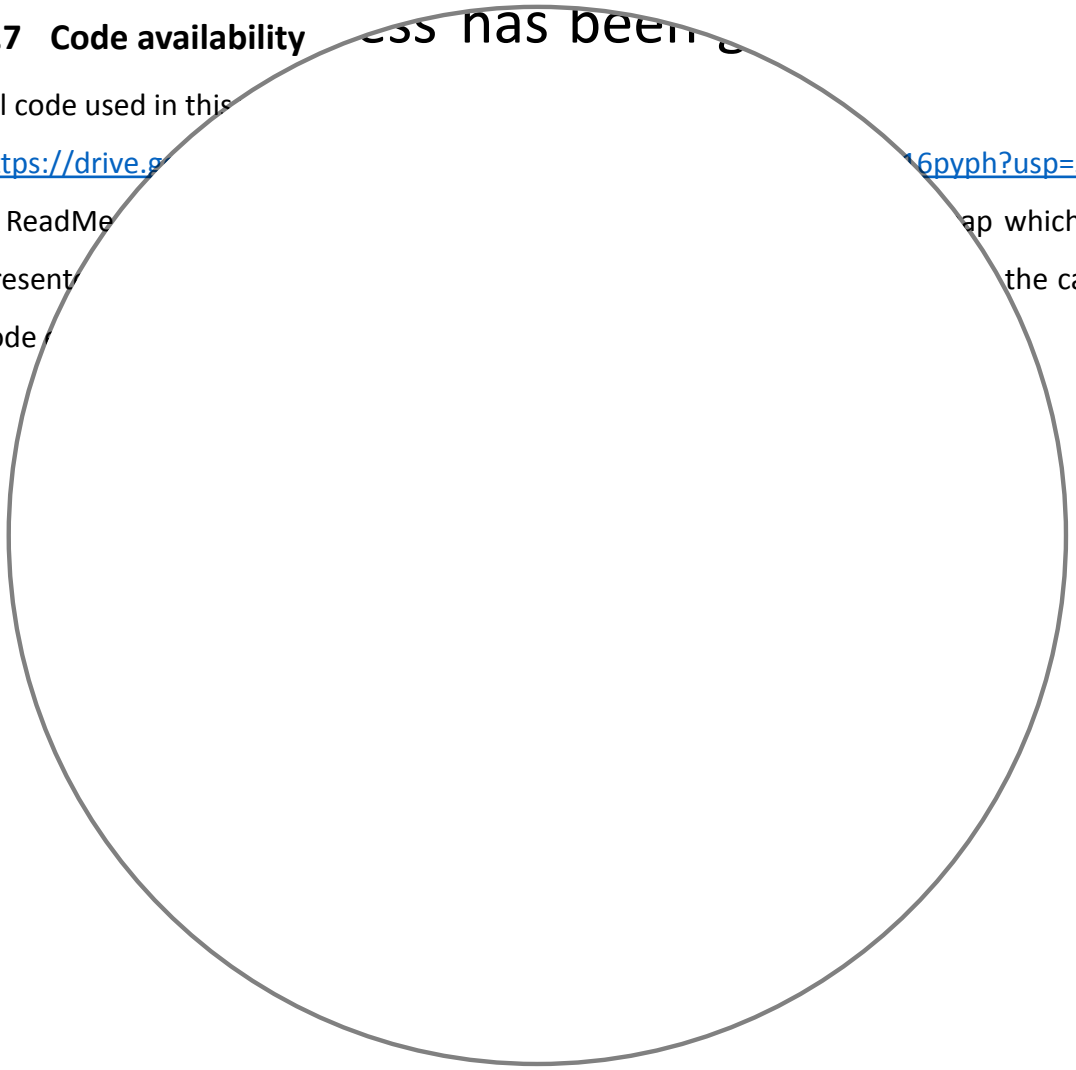
All code used in this

https://drive.g 16pyph?usp=sharing

A ReadMe ap which is also

present the case that

code



[3] https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf

2 Related work

This chapter is divided into three subsections: the first provides a synthesis of the meteorological causes of ramp events, the second summarises the various ramp definitions that exist within the literature, and the third reviews the various methodological and computational approaches used by previous authors in the field.

2.1 Meteorological processes associated with ramp events

An understanding of the meteorological phenomena that cause large ramp events is foundational to any credible ramp forecasting procedure. The methods and data used to successfully forecast ramp events depend on the underlying meteorological or engineering causes and the desired forecast time period (Gallego-Castillo et al., 2015; Zack, 2007). In particular, the spatiotemporal scales of the meteorological processes involved have implications for how such phenomena are interpreted and for the accuracy that can be expected from their interpretation. Figure 2 depicts some of the key atmospheric processes that have been linked to ramp generation in the literature and categorises them by scale. The main processes considered relevant to the current study area and dataset are considered qualitatively in this section and further elaboration is provided in Appendix B. A detailed quantitative analysis is beyond the remit of this study.

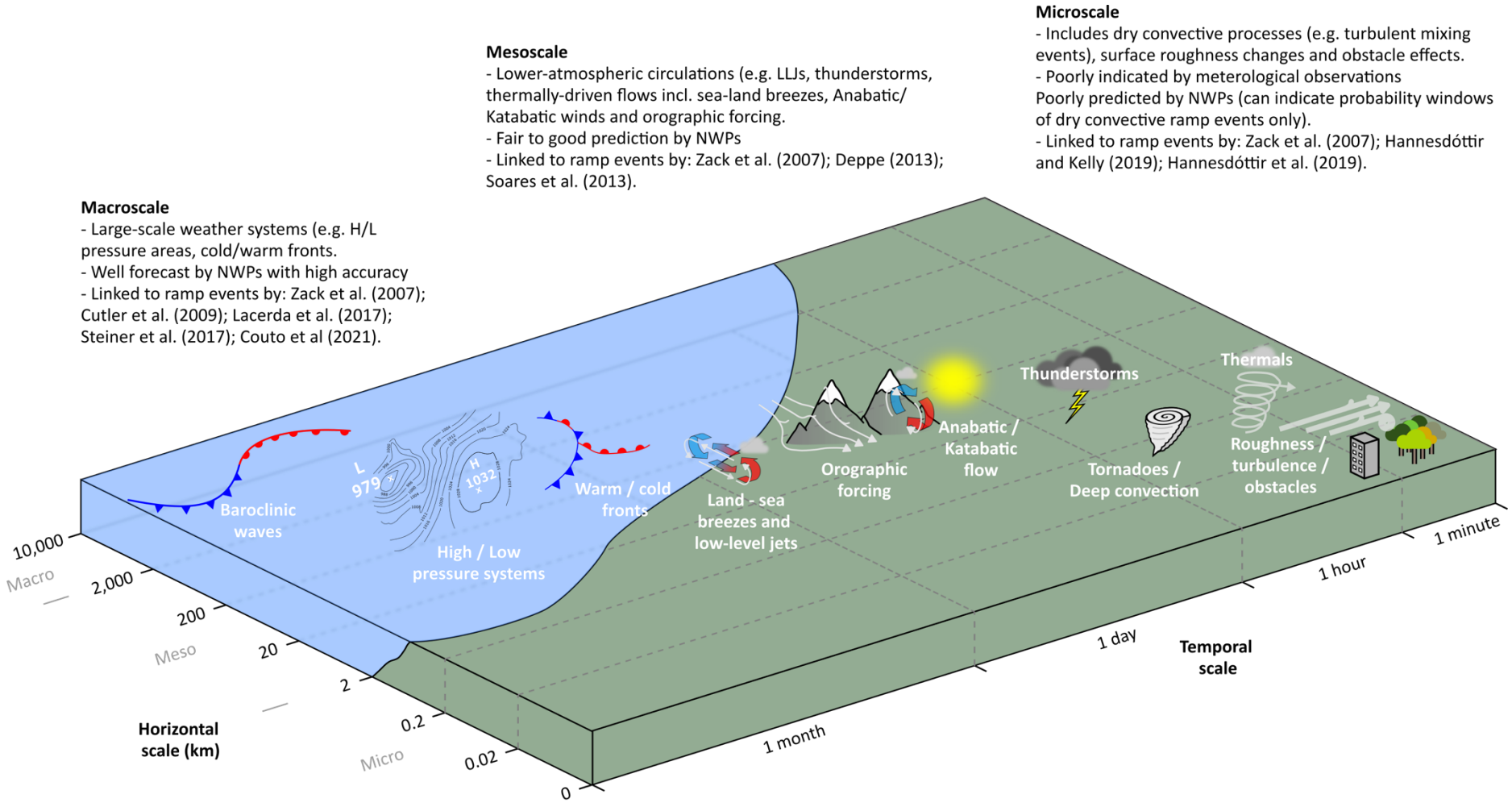


Figure 2. Atmospheric processes known to affect wind speed with characteristic spatiotemporal scales and links to ramp events identified in the ramp literature. Atmospheric scale definitions used in this study are defined on the y-axis. Additional sources: Orlanski (1975); Landberg (2015).

2.1.1 Planetary scale

When studying wind profiles, it is important to consider the planetary scale atmospheric setting of the study area (Landberg, 2015). NE France resides in a mid-latitude setting, predominantly influenced by westerly trade winds (Figure 3).

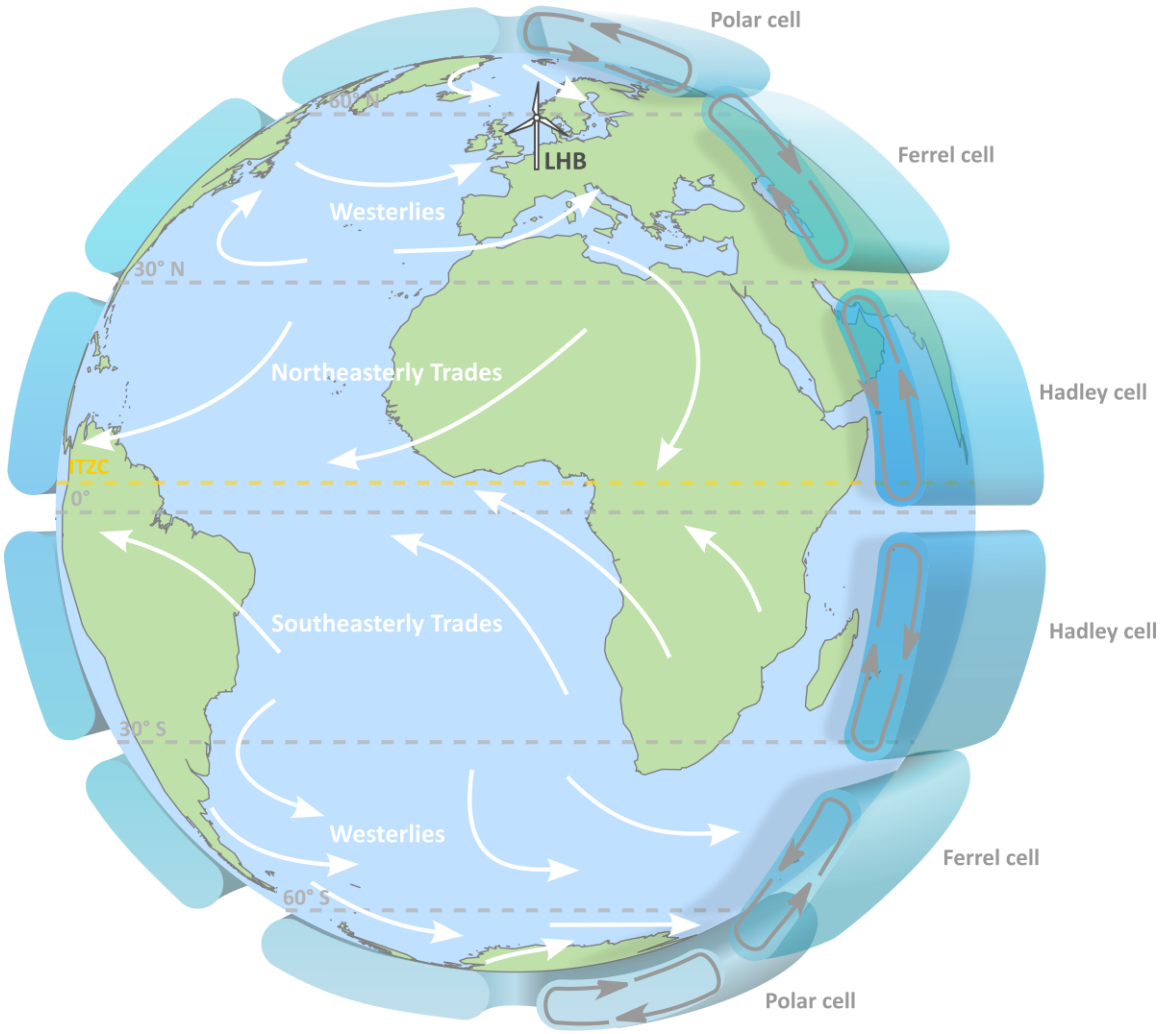
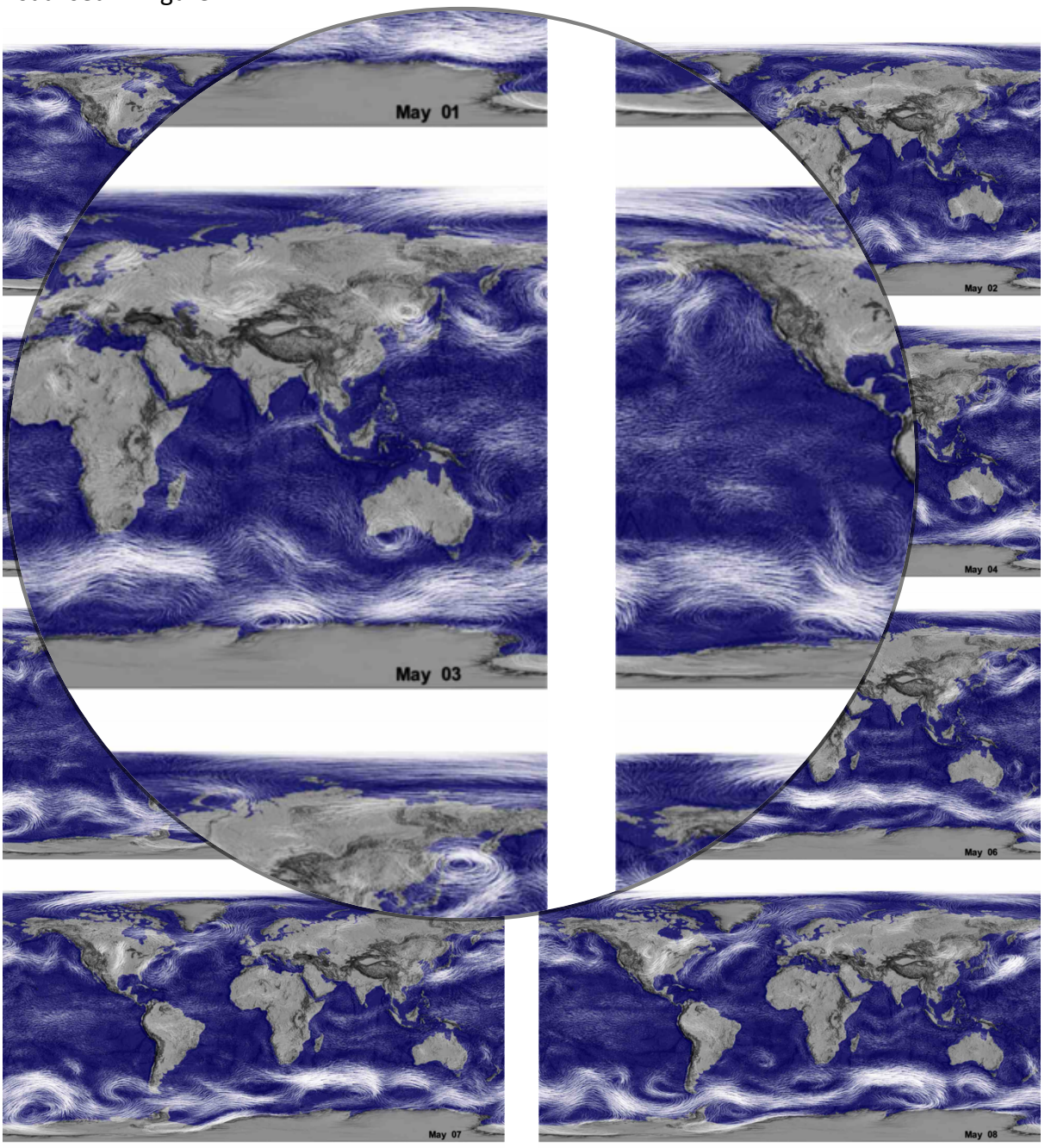


Figure 3. Global circulation model schematic highlighting the influence of westerly trade winds over the LHB site (after Wikipedia 2021a).

2.1.2 Macroscale

Macroscale processes associated with ramp events include large-scale weather systems such as low- and high-pressure areas and cold and warm fronts. Westerly trade winds, generated by solar heating of the atmosphere and the Earth's Coriolis force, are a key driver of such

weather systems. Their development over Western Europe during an eight-day period is visualised in Figure 4.



The image shows just over one week (1 to 8 May 2017) of visualisations from the NASA MERRA reanalysis dataset. White colouring represents winds at 850mb level (between 1,170 and 1,590 m). Increasing opacity reflects higher windspeed. A low-pressure mid-latitude cyclone visible over France and western Europe on 1 May (top left) can be seen to dissipate by 3 May. It is replaced on 4 and 5 May by an area of high pressure before a new lower-pressure weather system develops on 6 and 7 May. Relatively stable conditions return once again on 8 May (bottom right).

Figure 4. Planetary scale atmospheric setting of the study area (NASA Scientific Visualization Studio, 2017).

Seasonality in wind speed most likely related to the passing of warm and cold fronts is observed in the LHB data set (Figure 5).

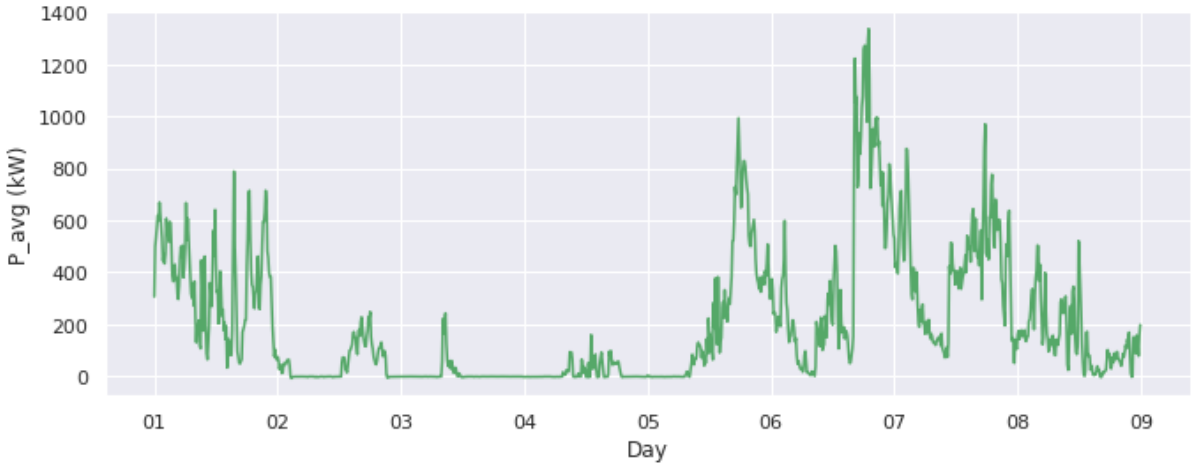


Figure 5. LHB power visualisation, 1 to 8 May 2017, corresponding to the time-period shown in Figure 4. Atmospheric scale events described in Figure 4 are clearly observable in the power data.

2.1.3 Mesoscale

Mesoscale processes include lower atmospheric circulations such as low-level jets (LLJs) and thunderstorms, thermally driven flows such as sea-land breezes and Anabatic/Katabatic winds, and flows linked to topography (*orographic forcing*) such as mountain-valley winds. Sea-land breezes are not expected to affect LHB given its inland location and quantitative evaluation of the influence of LLJs, thunderstorms and thermally driven winds over the LHB site would require additional data sources that are beyond the scope of this study. A quantitative analysis of orographic forcing is also considered out of scope. Qualitatively however, the various topographic features (i.e., combes and forests) that may orographically influence prevailing winds at the LHB site are considered in Figure 6. Worthy of note is the higher elevation of turbine R80711 relative to the others which is likely, at least in part, to explain the difference in power output observed between the turbines (Figure 7). A detailed cross-turbine analysis is beyond the scope of this study. For this reason, unless otherwise stated, all variables are presented as the average of the four turbines, or as percentage of rated power.

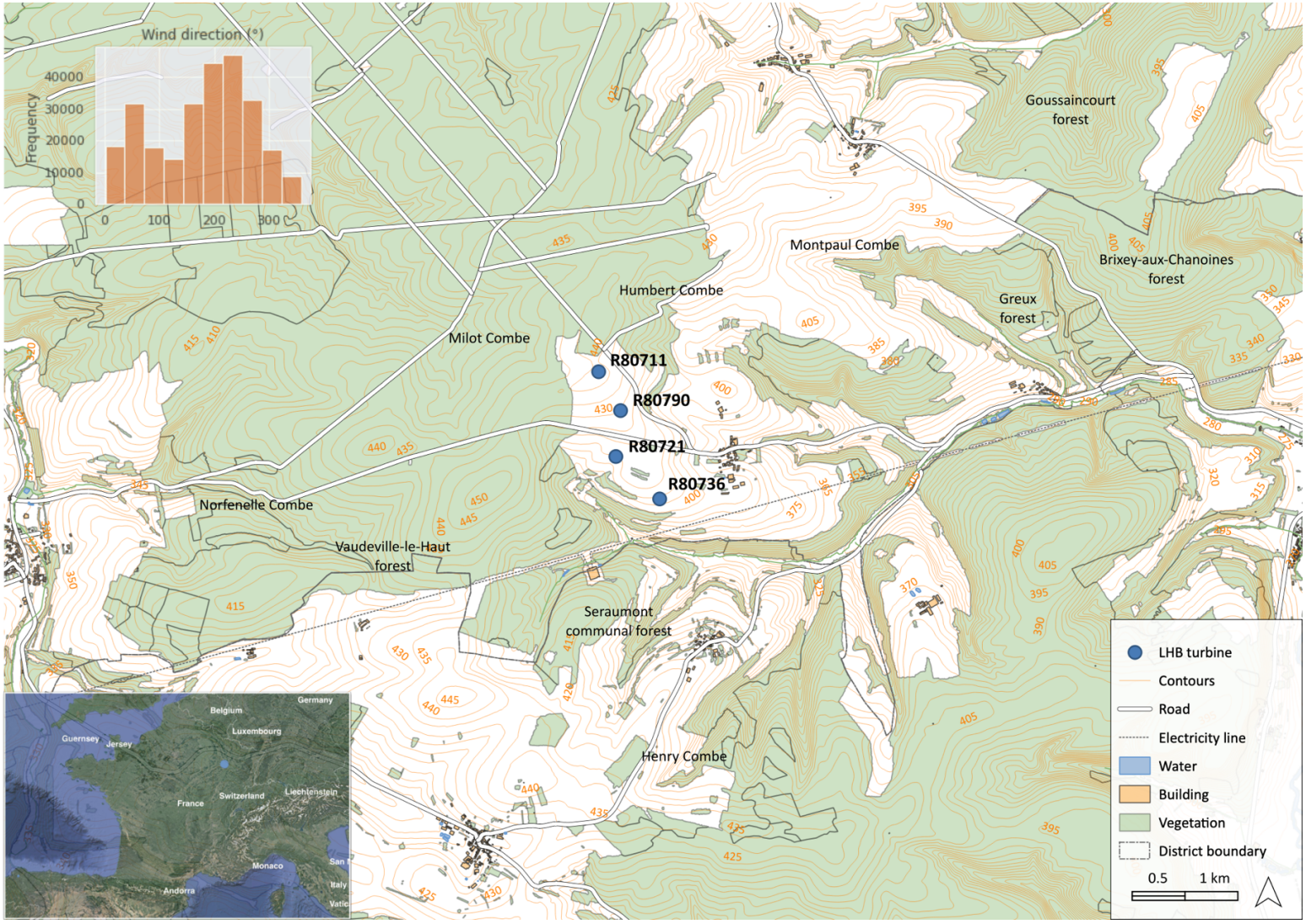

Wind direction (°)
Frequency
40000
30000
20000
10000
0
0
100
200
300
Goussaincourt forest
Montpaul Combe
Brixey-aux-Chanoines forest
Greux forest
Humbert Combe
Milot Combe
R80711
R80790
R80721
R80736
Norfenelle Combe
Vaudeville-le-Haut forest
Seraumont communal forest
Henry Combe
Germany
Belgium
Luxembourg
Guernsey
Jersey
France
Switzerland
Liechtenstein
Monaco
Andorra
LHB turbine
Contours
Road
Electricity line
Water
Building
Vegetation
District boundary
0.5
1 km

Figure 6. Topographic map of the LHB wind farm and prevailing wind direction data. Contour heights are in metres above sea level. LHB resides in gently sloping terrain far from significant mountain ranges (see inset) and predominantly experiences prevailing winds from 50° and 220° (GIS data source: Institut National de l'Information Géographique et Forestière, 2021).

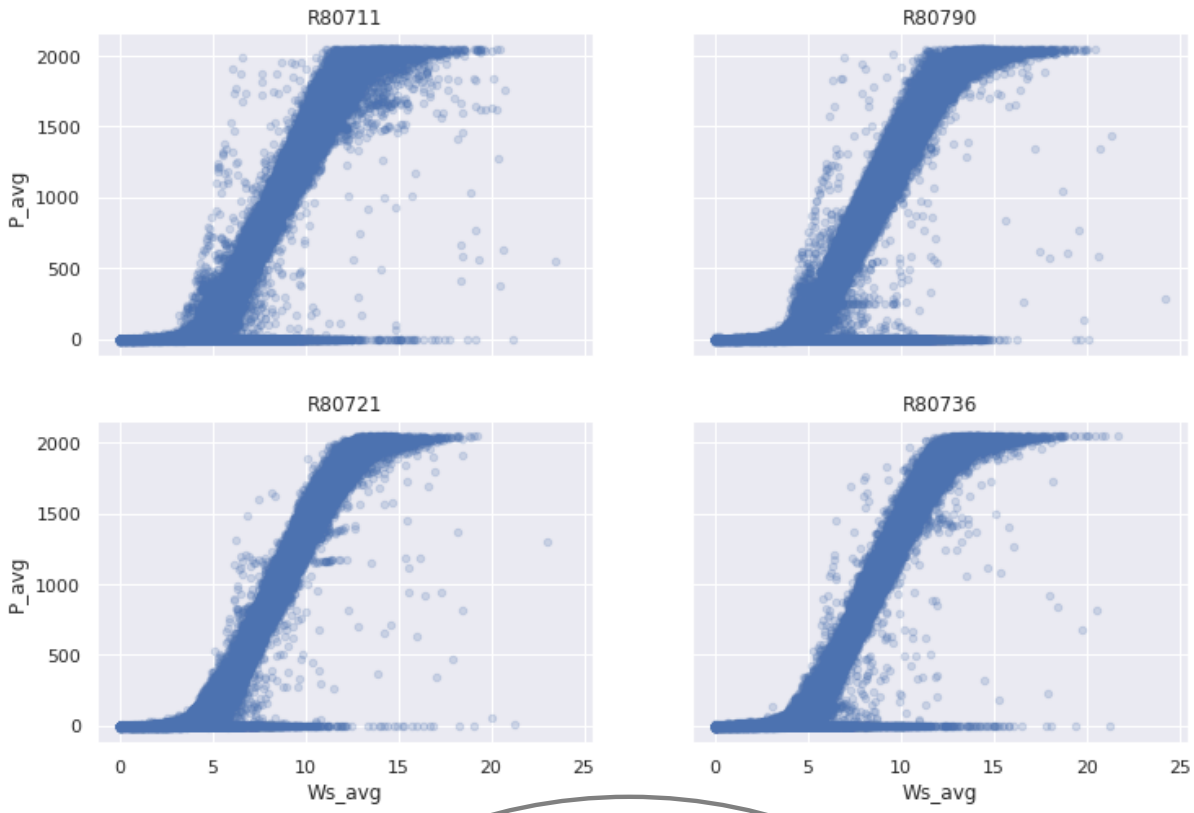


Figure 7. Cross-turbine powe ely to reflect its higher

tive processes related

ce roughness and obstacles. Exc

s generally free from obstructive buil , and

line

ase turbulence. The short and highly v sts,

ral

cesses (minute scale – Figure 2) means t be

5-minute sample rate of the LHB data (s

tantial

p events

e that wind

ically around 25

ed to this point have t to predict accurately

2.1.4 Microsc

The microsc

influence f

utility bu

however

scales of

below th

2.1.5 En

The causes

changes in wi

can also be cause

speed is already at o

m/s for most modern turbi

l factors

as small errors in the wind speed forecast can result in large errors in power output forecasts (Zack, 2007). This highlights the importance of considering the scale of the ramp events and thus how ramp events are defined.

2.2 Ramp event definition

In general terms, a ramp event is a large and rapid variation in wind power output (Cutler et al., 2007; Gallego-Castillo et al., 2015). However, the relative interpretation of 'large' and 'rapid' will differ according to:

- The end use and time horizon of the forecast. For example, a wind farm operator wishing to anticipate market penalties will be interested in different time scales as compared to an energy trader focussed on instantaneous market demand and spot prices (Cutler et al., 2007; Martínez-Arellano, 2015).
- The size of the wind farm/ portfolio. As an example, when defining ramps using an absolute power amplitude threshold value, a higher frequency of ramp events is likely to be observed as the installed capacity increases (Gallego-Castillo et al., 2015).
- The cost function considered. For instance, costs of ancillary reserves and electricity market penalties (Gallego et al., 2013).

Throughout the wind ramp literature, ramp events are identified and characterised by the following features:

Table 2. Ramp characterisation terms and parameters used in the literature (after Gallego-Castillo et al., 2015).

Term	Parameter	Description
Magnitude	ΔP	Variation in power observed.
Duration	Δt	Time period over which a variation takes place.
Ramp rate	$\Delta P/\Delta t$	Variation divided by duration. Indicative of ramp intensity.
Timing	t_0	Time instance of ramp event. Can be start or central time.
Direction	+/-	Increase/ ramp-up (+) or decrease/ ramp-down (-) in power output.

Taking ramp events from wind power timeseries, the parameters of Table 2 can easily be analysed (e.g., Figure. 1). However, ramp forecasting usually entails the reverse; given certain characteristics or criteria, a forecaster must identify ramp events in order to determine underlying causes and create accurate predictive models. This establishes the

need to set such criteria, or in other words: a ramp definition. Ramp event forecasting is a relatively immature research fie[illegible] of a standard formal ramp event definition, the literature [illegible] ding on wind farm size and quantity (if aggre[illegible] d (Gallego-Castillo et al., 2015; Ma[illegible]

2.2.1 Bi[illegible]

Most r[illegible] ction 2.3, Table[illegible] defined thres[illegible] 1). This appr[illegible] ecome high[illegible] set to 50%, [illegible] is that a binary [illegible] different character[illegible] forecaster from explo[illegible] nd continuous explanatory va[illegible] eorological tower measurements (Galle[illegible] iously used binary ramp event definitions and their [illegible] Ultimately, a bi[illegible], despite the clear drawbacks of binary ramp event definitions, many recent works continue to use and refer to them (Section 2.3, Table 5).

sified ramp events using a bi[illegible]

is determine whether a ramp exis[illegible]

gnitude and duration (ΔP_r and Δt_r r[illegible]

two major disadvantages. The first is tha[illegible]

ften arbitrary) threshold values used. F[illegible]

n power output of 49% may not be [illegible]

all ramps as similar, despite [illegible]

Table 3. Binary ramp definitions used in the literature.

* Ramp-up is signified by $P_t < (Pt + \Delta t)$, ramp-down by $P_t > (Pt + \Delta t)$

** Wind power can exhibit high variability over timescales shorter th signal between two t characterisation may become sensitive to noise. To overcome this issue, Bossavy et al. (2010) introduced the idea pre-set threshold.

Reference	Definition	Limitations
Kamath (2010)*	$P_t + \Delta t - P_t$	Based on start and end values of Δt so doesn't account for amps that may occur during e interval.
Kamath (2010)	$\max([P_t, t + \Delta t]) -$	s not characterise rate of ge (ramp rate).
Zheng and Kusiak (2009)*	$P_t +$	itive to threshold value
Bossavy et al. (2010)**	$\lvert P_t^f \rvert$	sitive to threshold value

And:

$P_t^f = mean(P_{t+h} - P_{t+h-}$

A ramp event occurs if the difference be maximum and minimum power output m exceeds a pre-set threshold.

A ramp event occurs if absolute difference and end values of Δt and the size of Δt its than pre-set power ramp rate value, P_{TT}

Where: P_t^f = filtered version of power using k-order differences in power ed of nber of average power mea vents are id

2.2.2 Non-binary approach

To overcome the drawbacks of the binary definition approach, Gallego et al. (2013) applied *wavelet transform analysis* to the characterisation of ramp events. One of the first practical applications of wavelet analysis was in the field of hydrocarbon exploration (Morlet et al., 1982a and 1982b). Since then, it has been successfully applied in a variety of fields including image processing, timeseries forecasting, chemical analysis and tool condition monitoring (Gallego et al., 2013 and references [illegible] frequency decomposition of wavelet analysis has also been [illegible] link abrupt changes in meteorological vari[illegible]esdóttir and Kelly, 2019). Wavelet [illegible] of wind ramps (e.g., Ji et al[illegible]ical analysis (Pichault[illegible]

[illegible] uses wavelike functions known as

Wavel[illegible]n into another representation which [illegible]m the signal [illegible] more succi[illegible]e Section 3.3, Figure 11). The transform[illegible]nal is known[illegible]cularly useful [illegible]nsform (WT). The WT method has bee*filters,* wavelets [illegible]cies (the *passband*)[illegible]eriodic or noisy signals. By effective[illegible]equencies that are not [illegible]ontext of this study, the freque[illegible] signal components within a[illegible]c processes that were linked to ramp[illegible]ver hourly to weekly timescales (Figure 2). Micr[illegible]by the energy sp[illegible] sub-hourly timescales are considered noise and are rem[illegible]velet function wi[illegible]

During wavelet t[illegible]f the wavelet (Figures 8a a[illegible] is calculated at various locati[illegible]ically, this is done throu[illegible]it is done through [illegible]avelets and the wavelet transform methodge WT

[4] For a [illegible]red to Addison

coefficients are obtained where the wavelet matches the shape of the signal well at a specific location and scale. Low WT coefficients are obtained where the wavelet and the signal do not correlate well. WT values can then be plotted as a wavelet transform plot (Figure 8c), providing information on both the location and scale, in time, of different ramp events.

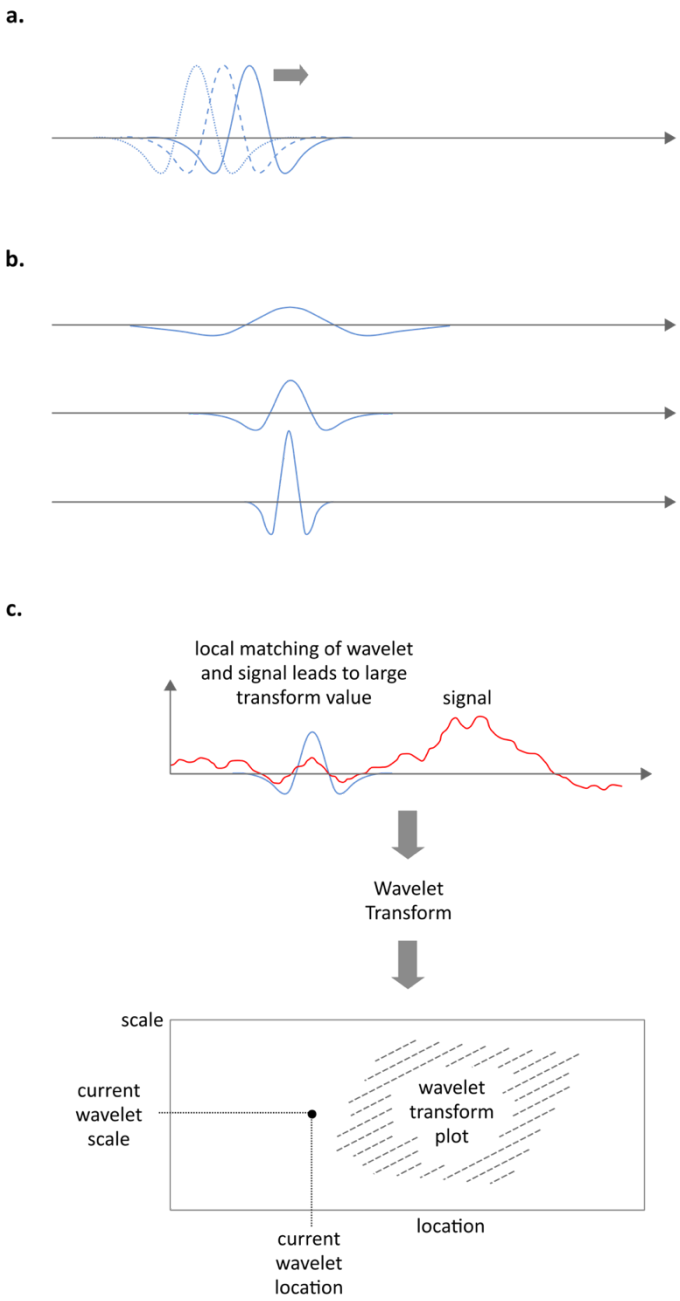


Figure 8. a. Shifting of the mother wavelet. **b.** Dilation of the mother wavelet. **c.** The wavelet transform process and the wavelet transform plot (after Addison, 2017).

The wavelets are shifted and dilated versions of a so-called mother wavelet, ψ^{τ}, and are derived from Equation 1:

$$\psi^{\tau,\lambda}(t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t-\tau}{\lambda}\right) \#(1)$$

where:
τ relates to the shift (expressed on a continuous scale between min and max)
λ relates to the dilation

The WT of a timeseries, $\{y_t\}$, consists of a set of coefficients obtained from Equation 2[5]:

$$W^{\tau,\lambda} = \sum_{-\infty}^{\infty} \psi_t^{\tau,\lambda} \cdot y_t \#(2)$$

where:
$\tau \in \mathbb{Z}$
$\lambda \in \mathbb{Z}^+$
and $\psi_t^{\tau,\lambda}$ relates to the wavelet function employed.

Several wavelets exists a[illegible] ramps, Hannesdótt[illegible]urpose of the study. Gallego et al. (2013) cho[illegible]n about the gradient experienced a[illegible]elet (DOG1) as their choice of a[illegible] events where a large gradi[illegible]udy aiming to characte[illegible]ot discuss the reasons behind this che[illegible]ed the first deriva[illegible]rtunately, howe[illegible]

Wav[illegible] be applied to ramp characterisation [illegible]mental conc[illegible]ecutive time[illegible]t is that a certain large gradient is ma[illegible]ity. This means[illegible]ords, the shape o[illegible]ies. Therefore, a ramp can be said[illegible] event, the wavelet tra[illegible]range of times scales, λ, since e[illegible]event is similar to a smaller [illegible]ated in a time scale

[illegible]different scales. Du[illegible]

[5] Note that the wavelets are define[illegible]cient obtained is opposite to that of the ramp event gradient, thus, the sign of $W_{\tau,\lambda}$ in Equation 2 is inverted so that positive coefficients are obtained for positive gradients.

[6] The increase of $W_{\tau,\lambda}$ with respect to occurs because the scale is contributed as -12 in Equation 1.

$\Delta t = \lambda$. Hence the gradient is high for both short and long-time scales, where 'long' means close to the length of the overall ramp event. If a period is considered where $\{y_t\}$ exhibits many high-frequency fluctuations, similar WT coefficients are observed for short scales but not for long scales, in other words, the gradient is no longer self-similar at longer time scales. Finally, if a period is considered where $\{y_t\}$ does not exhibit high gradients, the WT coefficients $W^{\tau,\lambda}$ would be close to zero for every scale, λ, considered.

This is utilised in the r ramp function, $\{R_t\}$, as the sum of the ales, λ, given by $[\lambda_1, \lambda_N]$:

$$R_t(\lambda_1, \lambda_N) = \sum_{\lambda=\lambda_1}^{\lambda_N} W^{\tau,\lambda} \quad \#(3)$$

related to the sharpness of the ramp e

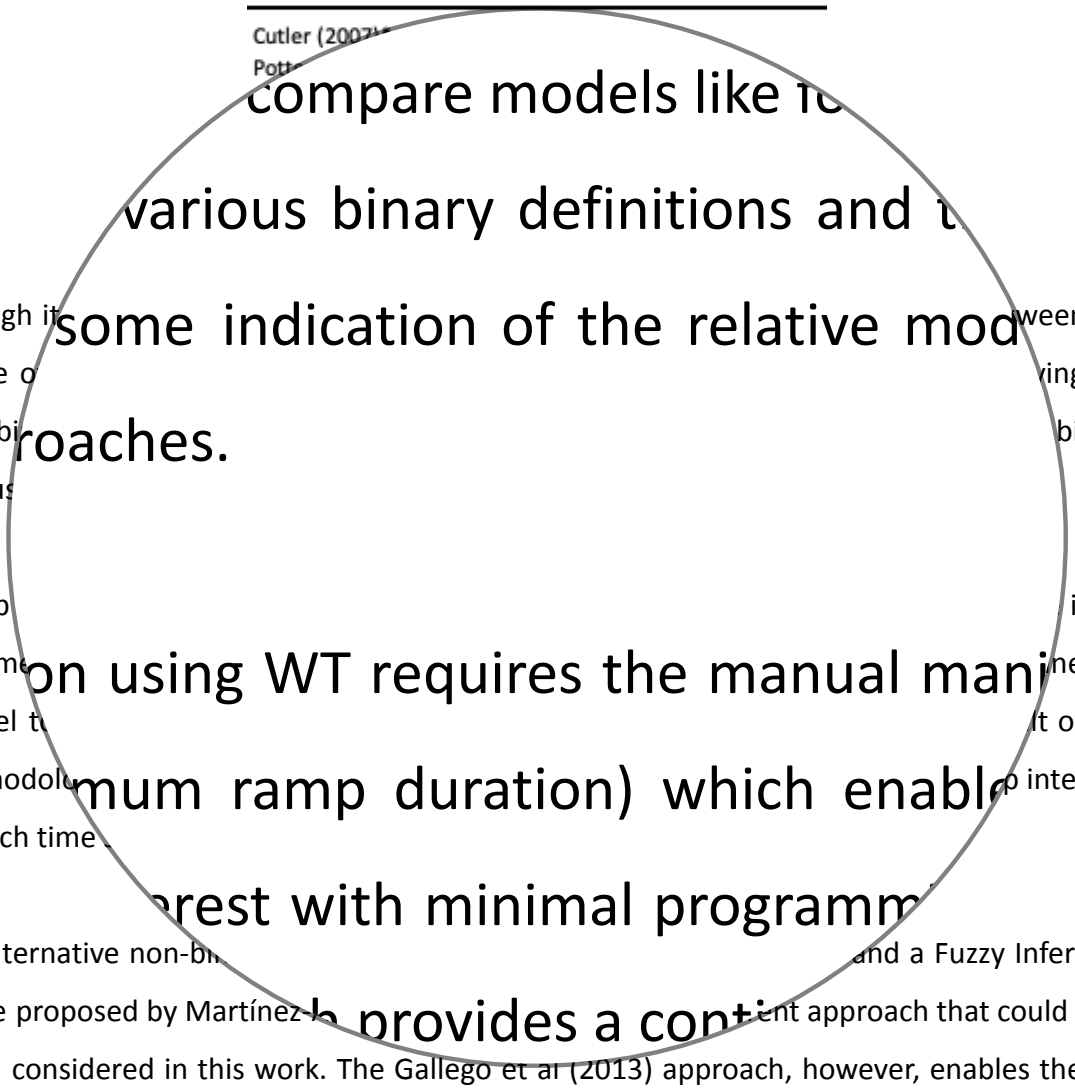
In thi ers at each bution of the gradient evaluated under Time scale $\lambda_1 = 2$ because t can be considered, and the minimu window (in terms Gallego et al. (2013), ap etween 5 and 10. Irrespective re required to define a gradien ighly sensitive to small changes in λ, p transform coefficients across the range of scales in ted, λ_N, requires en the impact of any additional scale λ_{N+1}.

Gallego et al. (2013) showed that the use of a continuous index-based ramp definition led to more reliable ramp characterisations than four binary ramp classifications presented by Cutler (2007), Potter (2009), Greaves (2009) and Bradford (2010) because it is less sensitive to input parameters (Table 4). It is noted however that these works were based on different

datasets. Comparative relative ramps frequencies observed using the same dataset were as follows:

Table 4. Relative frequency of ramp events (%) observed by different ramp definitions using the same dataset. It should be noted that the models were not designed using the same datasets/ site locations. * = Binary, ** = Non-binary. *** = range given by varying λ_N from 5 to 8 (see text for details).

Reference/ methodology	Relative ramp frequency (%)
Cutler (2007)*	
Pott	



Though it ween the range o ying the non-bi binary versus

Ramp input parame ne the model t t of the methodolu intensity at each time

An alternative non-bi and a Fuzzy Inference Score proposed by Martínez-b ent approach that could have been considered in this work. The Gallego et al (2013) approach, however, enables the use of well-established methodologies and packages (e.g., PyWavelets for Python – Lee et al., 2019)[7]. It is for these reasons that the non-binary, WT transform approach to ramp characterisation is used in this study.

[7] For potential drawbacks of the WT approach, the reader is referred to Gallego et al. (2013).

2.3 Approaches

A comprehensive review of wind power ramp forecasting was undertaken by Gallego-Castillo et al. in 2015. Table 5 provides a summary of the relevant work in the field that has been carried out since its publication. Recent statistical analyses of ramp events include Aguilar (2019), Dalton et al. (2019, 2021), DeMarco and Basu (2018), Kelly et al. (2021), Pereyra-Castro et al. (2019), Pereyra-Castro et al. (2020), Pichault et al. (2021). Literature reviews of wind forecasting include Ahmadi and Khashei (2021), Sim and Yung (2020), Wang et al. (2019) and Yang et al. (2021).

Table 5. Main ramp forecasting literature reviewed.

Author	Year	Model type	Definition	Evaluation metrics	Horizons	Outputs	Location
Fernández et al.	2013	TS	Gabor Filtering, BPNN	CM	3 h	Ramp occ.	ESP
Sevlian and Rajagopal	2017	Binary	AR logit models	CM	Varied	Power output	USA
Martínez-Arellano et al.		Binary	Orthogonal Test, SVM	CM, FIS score	100 h	Power output	ESP
Martínez-Arellano		Binary	Reservoir computing, Echo state net	FIS score	12-36 h	Power output	ESP
Ji et al.		Binary	Swinging Door Algorithm	n/a	n/a	Power output	n/a
Heckenbergerova		Binary	RNN, Echo State, Delay line reservoir	Box plots	4 h	Ramp occ. (prob)	USA
Gallego-Castillo		Binary	SVR, WT	CRPS	12 h	Power output (prob)	USA
Zha et al.		Binary	NN, ARMA, WSR, PSO, KDE	MAE, MAPE, RMSE, CC	3.5-5.5 h	Power output	USA
Li et al	TS	Binary	KDE, Back Propagation Neural Network (	FA, RC, RMSE	24 h	Power output	CAN
Taylor		Binary	MSAR	Brier score	1-2 h	Power output (prob)	GRC
Liu et al	+ TS	Binary	RF, RUS, ROS, SMOTE, Laplacian kernel	RC, FA, CSI	0.5 to 24 h	Ramp occ.	CHN
Dorado		Non-binary	Bayesian Network	CM, Sensitivity, Specificity, GMS	6-24 h	Ramp occ.	ESP
Zhang		Binary	Multi-parameter segmentation algorit	POD, CSI, FBIAS	1-6 h	Ramp occ. (prob)	USA
Dorad		Non-binary	Wave Division, Grey Wolf Optimizer	Geometric mean, CCR, AMAE	6 h	Power output	ESP
Dhim		Binary	Extreme Learning Machine, SVR, S	AE, RMSE, MAE, CPU time	0.5-3 h	Power output	ESP, USA, AUS, IND
Zhan		Binary	Multi-task Learning, Deep Neur	MAE, MAPE, MSE, RMSE	16 h	Wind speed	CHN, ESP
Liu e		Binary	Multi-parameter segmentat	RMSE	4 h	Power output	CHN
Ouya		Binary	K-NN	BIAS, MAE, RMSE, CM	72 h	Power output	CHN
Fujin	+ TS	Binary	MSAR	RMSE, Precision, Recall, CSI	0.5-48 h	Ramp occ., power output	JPN
Zhao		Binary	RF, RUS, ROS, SMOTE, Laplacian kernel	CM	30 min	Ramp occ. (prob)	CHN
Lyners	+ TS	Non-binary	Bayesian Network	PDFs	25 h	Ramp occ.	ZAF
Ye et a		Binary	Multi-parameter segmentation algorit	NMAE, NRMSE, ACR	24-72 h	Power output	CHN
Cornejo		Non-binary	Wave Division, Grey Wolf Optimizer	ROC, CM	6 h	Ramp occ.	ESP
Dorado-M		Binary	Extreme Learning Machine, SVR, S	Accuracy, Sensitivity, GMS,	6 h	Ramp occ.	ESP
Lyners et a		Binary	Multi-task Learning, Deep Neur	POD, CSI, FBIAS, SR	n/a	Power output	ZAF
Hirata et al		Binary	Multi-parameter segmentat	MAE, Ignorance score	1-12 h	Power output	Japan
Pichault et al	TS	Non-binary	K-NN	Statistical analysis	1 h	Ramp characterisation	AUS
Couto et al.		Binary	Extreme Learning Machine, SVR, S	Statistical analysis	24 h	Power output	PRT
Dhiman and Deb		Binary	Multi-task Learning, Deep Neur	RMSE	10 min	Power output	UK, NLD, AUS
Zhou et al.		Binary	Multi-parameter segmentat	MAE, MAPE, RMSE, FA, RC	42 h	Power output	BEL, CHN

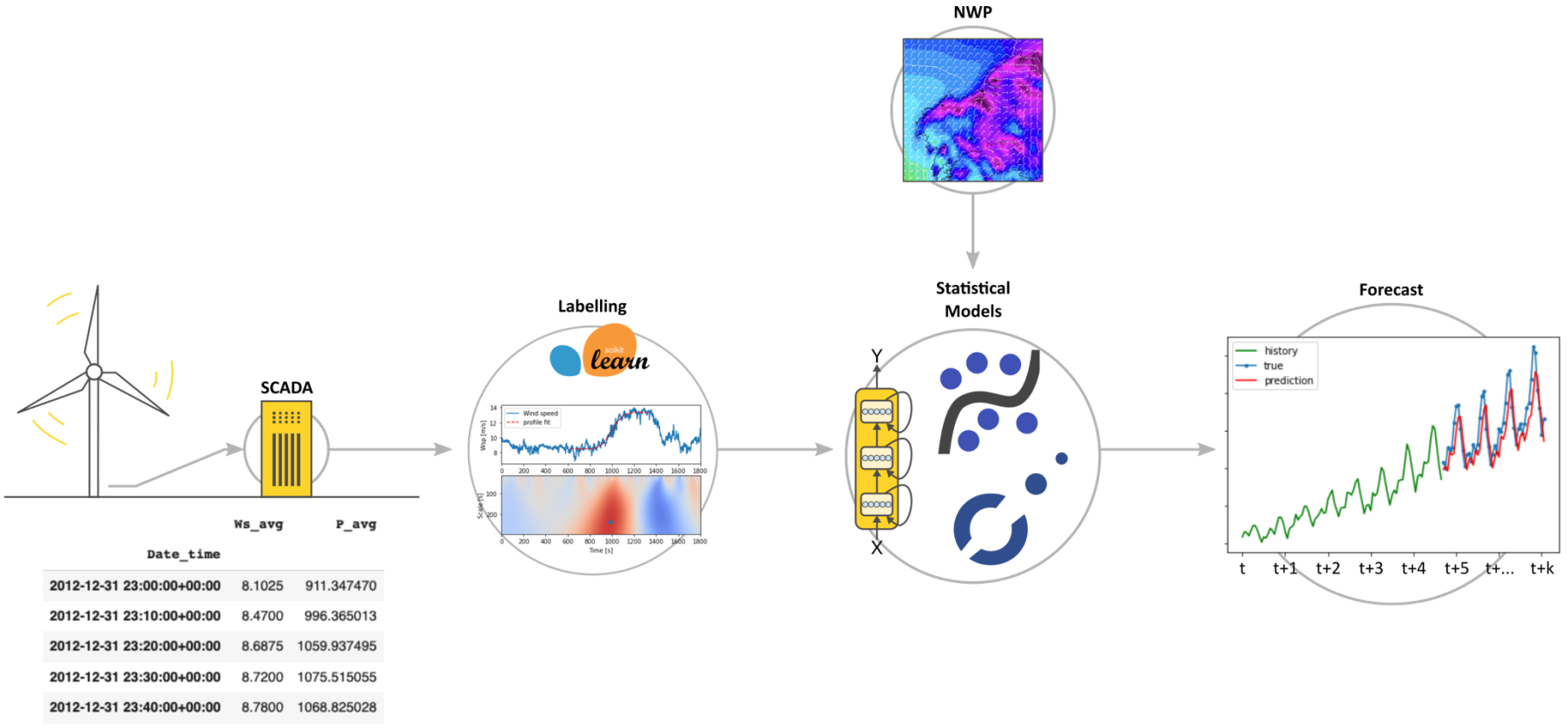
Model abbreviations: Genetic ry; Principal Component Analysis, PCA; Support Vector Machines, SVM; Auto-Regressive, AR; Artificial/Convolutional/Recurrent ; om Forest, RF; Long Short-Term Memory, LSTM; Markov-Switching-Auto-Regression, MSAR
Evaluation Metric abbreviations: Confusi ; solute (Percentage) Error, MA(P)E; Root Mean Squared Error, RMSE; Ramp Capture, RC; Probability Density Function, PDF; Receiver Operating Characteristic; Forecast Accu
Other abbreviations: Time series, TS; Numerical Weather Prediction Model, NWP.

3 Methods

This chapter outlines the methodological approach used in this study. It summarises the dataset, the chosen ramp event definition/ characterisation method, the ML models that are tested, and the metrics used in their evaluation. It also provides a consideration of how outputs from this work are hoped to be included in future research.

3.1 Technical design and approach

The methodology depicted in Figure 9 has been designed for this study.



Date_time	Ws_avg	P_avg
2012-12-31 23:00:00+00:00	8.1025	911.347470
2012-12-31 23:10:00+00:00	8.4700	996.365013
2012-12-31 23:20:00+00:00	8.6875	1059.937495
2012-12-31 23:30:00+00:00	8.7200	1075.515055
2012-12-31 23:40:00+00:00	8.7800	1068.825028

Ramps are identified ('labelled') in the raw data using wavelet transform. The family of ML algorithms is trained on the raw data, then used to forecast power output and associated ramp events. Using this methodology, NWP outputs may be used as an alternative data source (see Section 3.6). Ramp function plot from Hannesdóttir and Kelly (2019). NWP output from Xunta de Galicia (2021).

Figure 9. Flow chart of the methodology employed in this research.

3.2 Dataset

The dataset used in this project is a Supervisory Control and Data Acquisition (SCADA) dataset collected from the La Haute Borne wind farm between December 2012 and January 2018 (Engie 2020). La Haute Borne is an 8,200-kW capacity onshore wind farm located in north-eastern France, centred at 48.45°N, 5.59°W (Figure 6). The farm is owned by the French energy and utilities company, Engie, and comprises four Senvion MM82 wind turbines (Figure 10). SCADA data gen… …rray of sensors within each turbine is recorded in two large csv f… …readings of each turbine's components such a… …id information such as frequency … …al information including wi… …riables used in this stu…

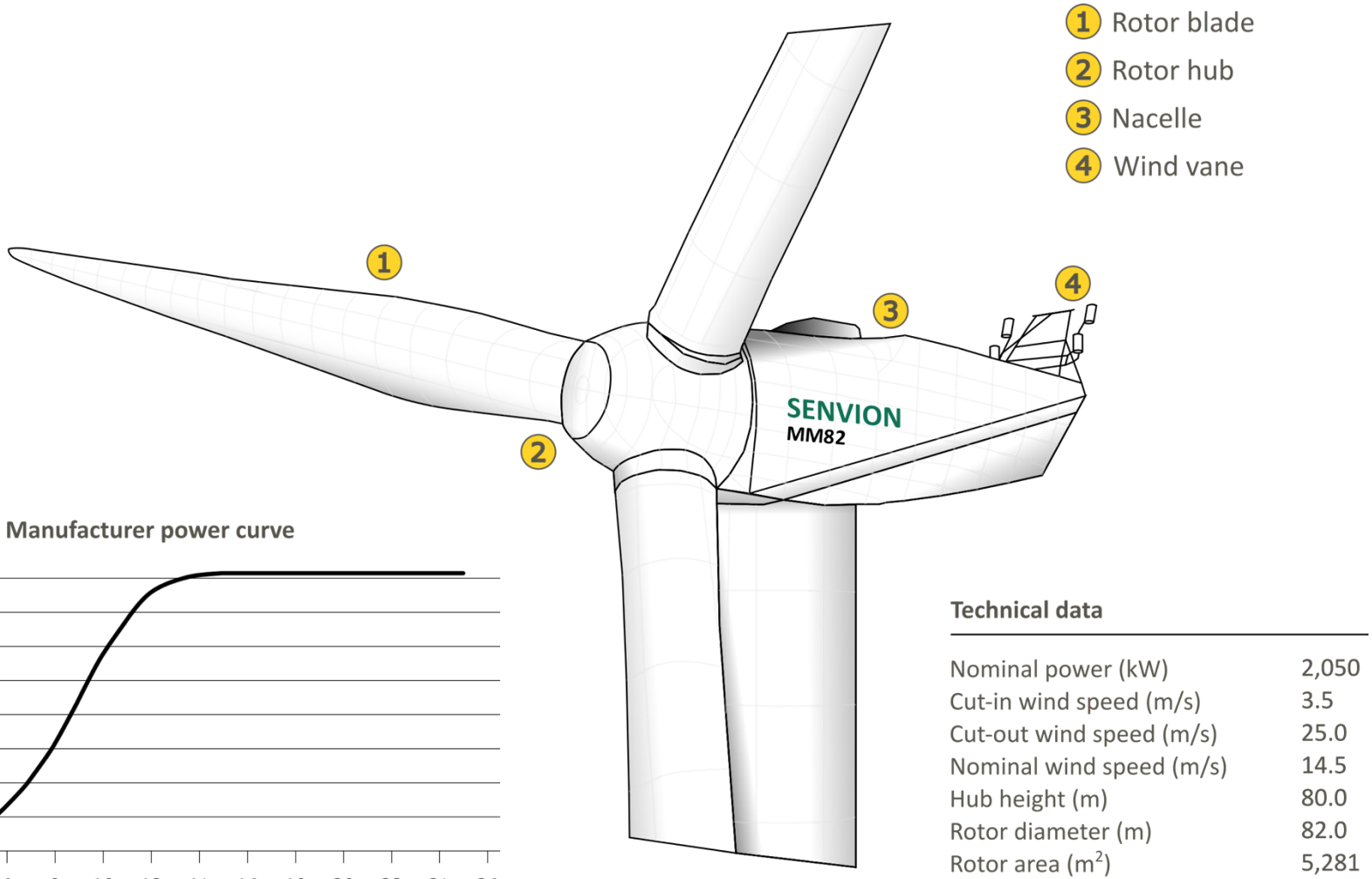
…tion. *Nominal power of LHB wind farm

Table 6. I…

Term	Units	Description
Ws_a…	m/s	10-min avg from two anemometer…
Wa_a…	degrees (°)	10-min avg from the wind vane of …
P_avg…	kW	10-min avg from each turbine
Ot_avg…	°C	10-min avg from the wind vane of…
Calculate…		
P_tot	kW	Sum of P_avg from all turbi…
%P_rated	%	P_tot divided by 8200 k…

The data are re… …W … (00:20) and 14 December 2017 (04:… … observations over a four-year period. Exploratory … … temperature readings with differences of 50°C+ compared to surrounding readings. These readings were deemed erroneous and were replaced using linear interpolation between surrounding readings. EDA also revealed 552 duplicate rows related to daylight saving clock changes. The second instance of these duplicates were removed from the dataset leaving a total of 1,057,416 observations.

[8] See 'eda.ipynb' in accompanying Colab repository.



Manufacturer power curve

Technical data

Nominal power (kW)	2,050
Cut-in wind speed (m/s)	3.5
Cut-out wind speed (m/s)	25.0
Nominal wind speed (m/s)	14.5
Hub height (m)	80.0
Rotor diameter (m)	82.0
Rotor area (m^2)	5,281

Figure 10. Senvion MM82 characteristics and terminology referred to in this study (after Senvion GmbH, 2018).

3.3 Ramp characterisation

The ramp characterisation method used in this paper focuses on the day-ahead TSO end-use case introduced in Section 1.2. Use of the WT methodology (Section 2.2.2) enables the end user to tune the model to the time scale of interest. WT coefficients (see Equation 2) are obtained using the PyWavelets package. The *ramp function,* R_t, of the power (P_avg) series is then obtained using Equation 3, reproduced here.

on the range

he size of the wind farm. Ther

The range -1 and 1 by defining the relative ramp nd power timeseri re-scale the ra

$$r_t = \frac{R_t}{max\{|R_t|\}} \#(4)$$

To isolat events, the ramp f :

luring ramp up, r_t^u, ramp down

r_t three timeseri (5)

$r_t^0 = 1 - r_t - r_t^u \#(7)$

This provides the non-binary/ continuous index related to ramp intensity at each time step of the series (Figure 11), thus adding the desired outcomes or 'labels' to the datasets that are introduced to the ML algorithms. In this way, the ramp event characterisation process is reframed as a supervised learning ML problem.

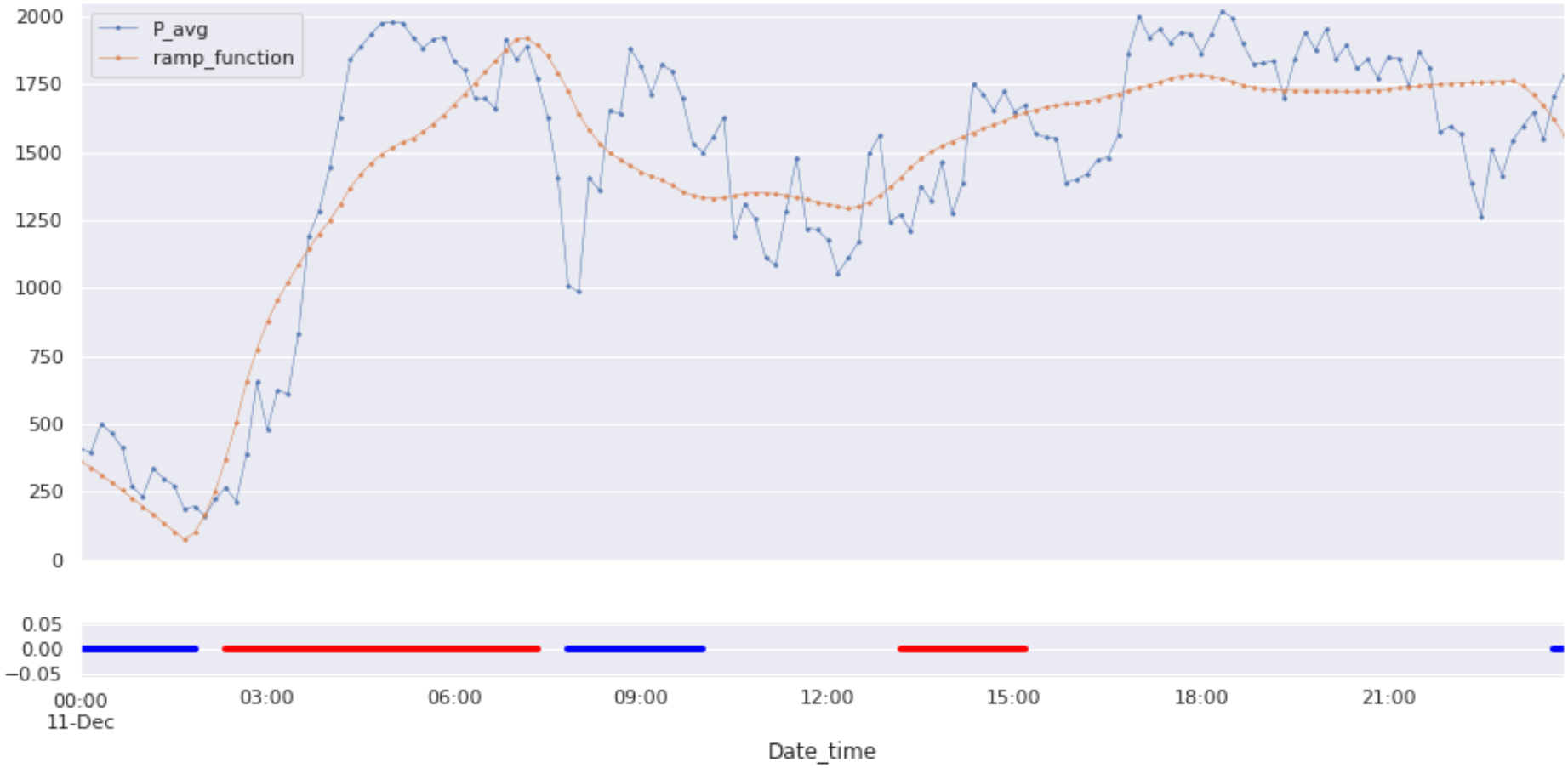


Figure 11. Ramp function plot against wind power. The ramp function is plotted in orange. The subplot displays decomposed ramp performance series: positive ramp, r_t^u (red), negative ramp, r_t^d (blue), and non-ramp, r_t^0 (blank space).

The ramp frequency, f_r, of a timeseries is the percentage of times in which a ramp is observed. It can be calculated using Equation 8:

$$f_r(\%) = 100. \frac{\sum_{t=1}^{N} |r_t|}{N} \#(8)$$

where:
r_t = relative ramp function
t = time
N = number of samp

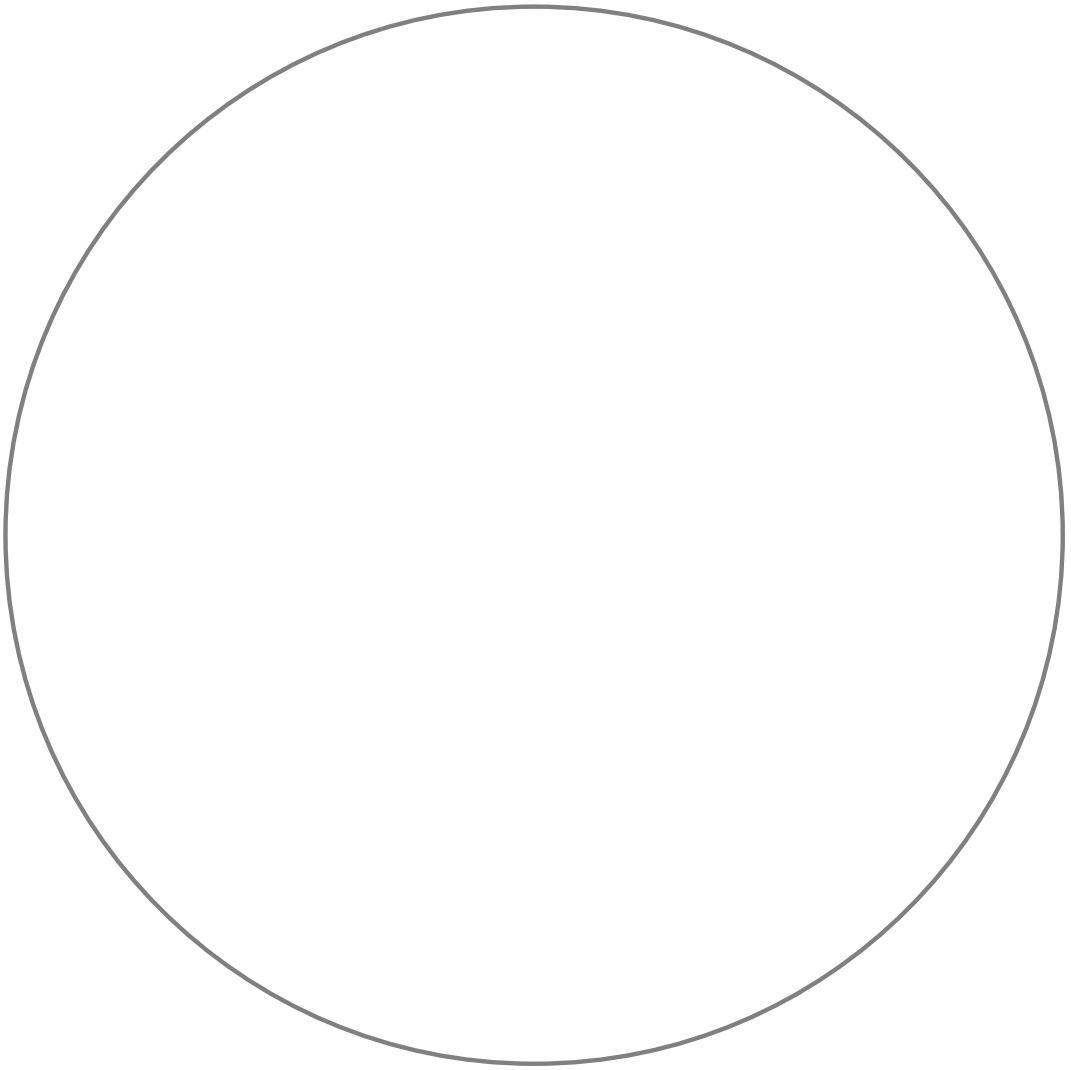
Applying Equatic in which ramps were observe the f_r of the forecaster t by each of the r ISE and MAE) a

Wave

The w in the examina 2021b). As such, udy is Daubechies 2, known as DB2 of several wavelets we ne LHB data most appropri level of noise removal). For the p el comparison), the choice of wavelet is of m sistent across models. For comparative purposes however, ramp r tained by DB2 and the Haar wavelet, as used by Gallego et al. (2013), are presented in Table 7.

Table 7. Comparative ramp frequencies between DB2 and the Haar wavelet using the LHB dataset.

Ramp frequency (%)	db2	Haar
Total ramp freq.	12.2	9.75
Positive ramp freq.	6.18	4.9
Negative ramp freq.	6.03	4.85
Non-ramp freq.	87.8	90.25



3.4 Numerical Weather Prediction model outputs

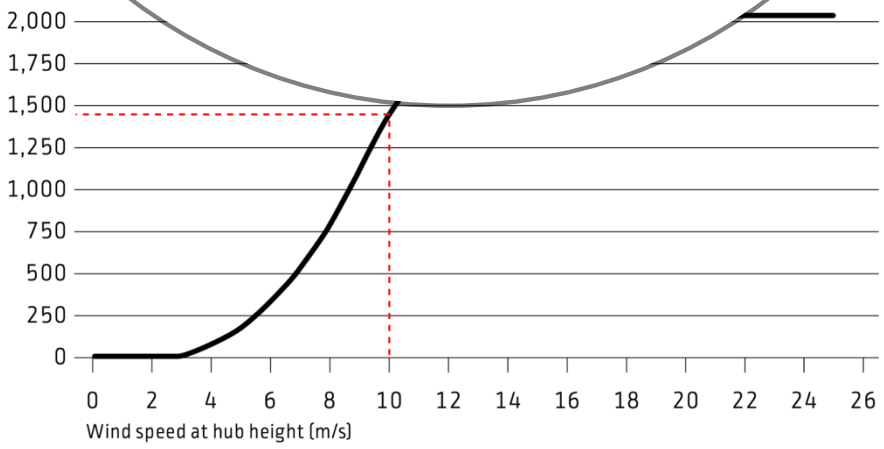
A benefit of the methodological design outlined above is that NWP outputs can readily be incorporated into the forecasting procedure during future research (Figure 9). By dividing the datasets into training and test sets, the steps undertaken as part of Objective 3 (Section 1.4) effectively test each algorithm's ability to model the relationship between historic wind power observations (training set) and forecasted wind power predictions (test set). However, historic wind power observations m[illegible]stituted for NWP predictions of wind speed and the models m[illegible]d from wind speed for[illegible]recasts. As mentioned in Section 1.3, error is [illegible]conversion of wind forecasts to po[illegible]lenezes et al., 2020):[illegible]vantage.

Prediction[illegible]example:

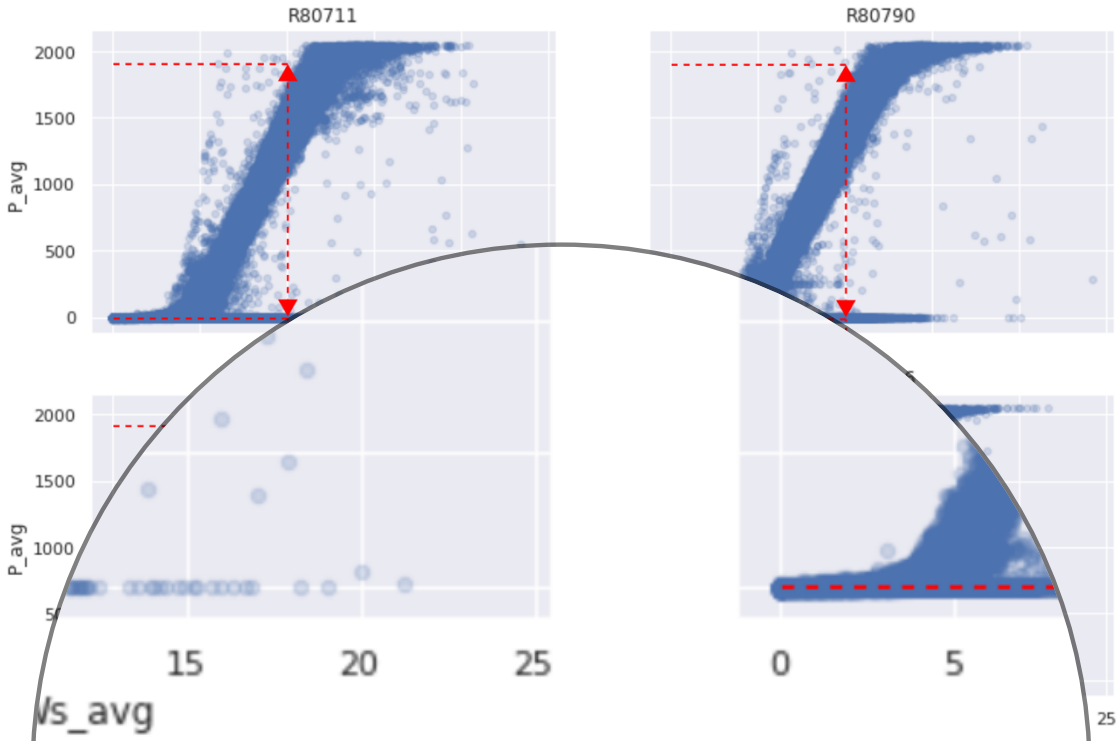
$$P = \frac{1}{2}\rho A v^3 \#(9)$$

1. [illegible]

2. [illegible]



3. Empirically using historic site data (reproduced from Figure 7):



Examp ble 3 is the rea peed of 10m/s, fo ween the minimum a This can be compared wit not model the variability at all. H conversion of wind speed to power with su improve the accuracy of the downscaling process. This is th planned by the author.

ised solutions and arguably do not refl

d at LHB. The issue, however, is t

he actual power output

NWP models output their estimates at hourly intervals. It is with this in mind that the ML models in this study are tested not only using raw 15-minute sampled data, but also with hourly re-sampled datasets.

3.5 ML modelling

Generally, if no information is assumed about a given set of ML problems, all optimisation and search algorithms should perform equally well when their performance is averaged over all possible problems — as established by the so called 'no free lunch theorem'. Put differently, if algorithm A outperforms algorithm B on a given number of problems, there must exist exactly as many other problems where B outperforms A (Brownlee, 2019; Wolpert and Macready, 1995. 1[illegible] free lunch theorem is that, as better understanding a[illegible] the choice and tuning of 'suitable' learn[illegible] the analysis on their perform[illegible]

Since the [illegible] power predict[illegible] ndency. In the [illegible] e, the algori[illegible] asting future [illegible] st be capabl[illegible]

3.5.1 [illegible]

ARIMA mo[illegible] lysing and forecasting t[illegible] model itself (Brownlee, 2019[illegible]

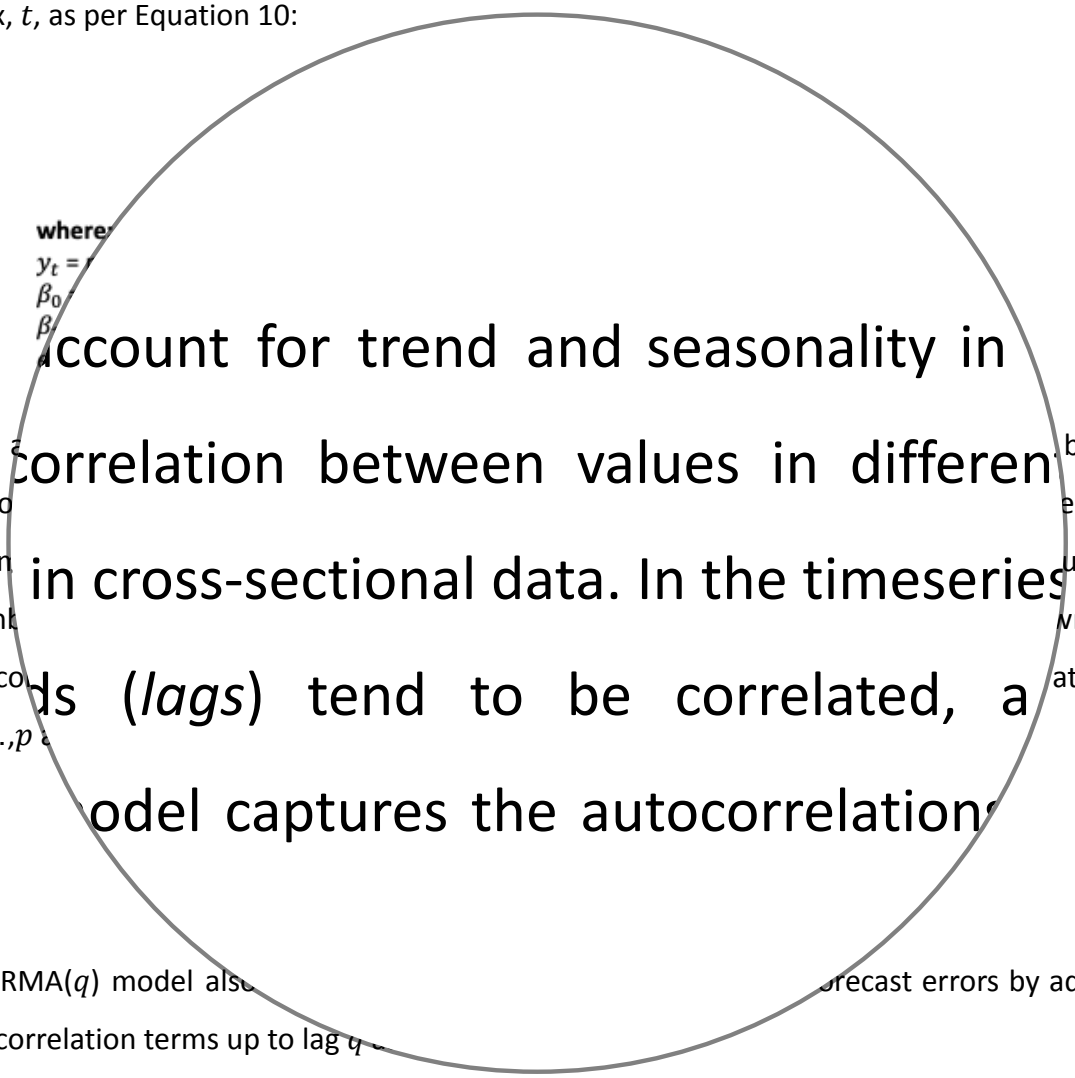
- **AR**: AutoRegressi[illegible] ween an observation and past values of a serie[illegible]
- **I**: Integrated. ARIMA uses differencing (see below) to make the timeseries *stationary*.
- **MA**: Moving Average. ARIMA uses the dependency between an observation and a residual error from a moving average model applied to lagged observations (i.e., a moving average process is a linear combination of past errors).

A standard notation of **ARIMA(p,d,q)** is used to describe the ARIMA model parameters that represent these aspects where:

- **AR** = **p** = The number of lag observations included in the model.

- **I** = **d** = The number of times that the raw observations are differenced.
- **MA** = **q** = The size of the moving average window.

Constructing the three ARIMA components helps to visualise how the model itself works (Shmueli and Lichtendahl, 2016): An AR model is similar to a linear regression model, except that the predictor variables are the past values of the series. A linear model would be fit between *power* and *time* by setting *power* as the output variable, y, and *time* as the *time* index, t, as per Equation 10:



Such a … but it cannot … e this information … ues in neighb… wn as autoco… at lags 1,2…,p …

An ARMA(q) model also … forecast errors by adding autocorrelation terms up to lag q …

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \#(12)$$

Data preparation

In this way, AR and ARMA models are not capable of modelling trend or seasonality. An ARIMA(d) model therefore incorporates a step known as *differencing* which has the effect of

making a timeseries stationary[9] and removing a trend that can negatively affect the regression model. The technique of differencing and the effect of first order differencing of the LHB power (P_avg) series is illustrated by Figure 12.

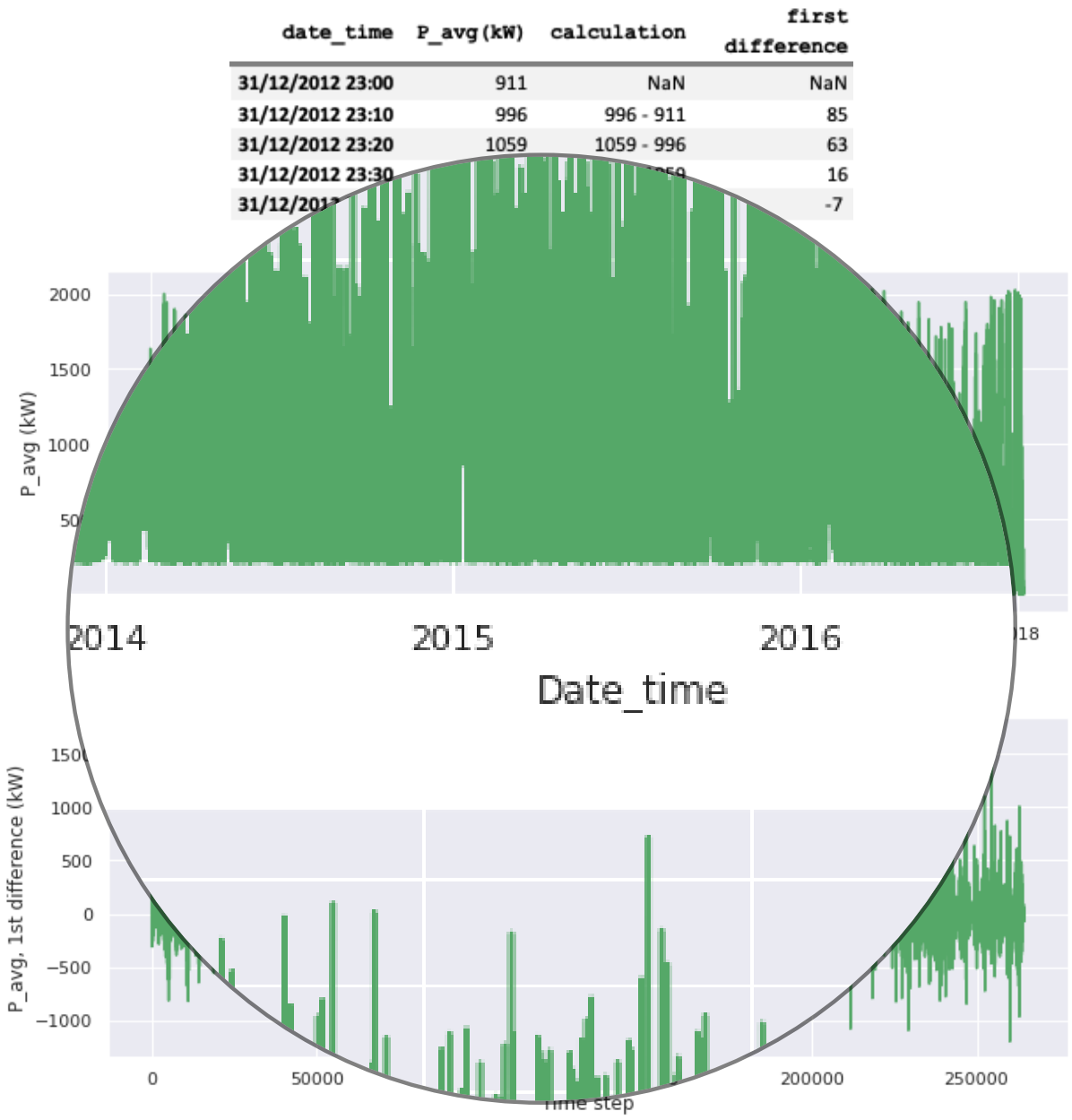


Figure 12. Method for calculating the first difference of a timeseries (top), original P_avg timeseries (middle) and 1st differenced P_avg timeseries (bottom).

[9] See Section 4.1.2.

3.5.2 SARIMA

A seasonal ARIMA (SARIMA) model is formed by including additional seasonal terms in the ARIMA model. The additional seasonal terms are multiplied by the non-seasonal terms of the model. The expression is written:

$$ARIMA(p, q, d)_{\curvearrowright\uparrow non-seasonal\ part\ of\ the\ model} \times (P, Q, D)m_{\curvearrowright\uparrow seasonal\ part} \#(13)$$

where:
P = number of seasonal autore
Q = number of seasonal
D = number of sea
m = seasonal f iod.

A summary of Table 8.

Table 8. Adv 019).

Feature	s to a suite of standard structures in	Disadvanta ARIMA foreca
Parsim	data, and as such provides a simple	errors when th end
Tunab		r the Tuning of ARIM differencing, ar average compo ng nd n a understandin model outco number of

3.5.3 Prophe

Prophet is a timese dle timeseries data that exhibits multiple st d changes, outliers, and holiday effects. It has also been desig ve adjustable parameters that do not require extensive knowledge of the model. It learns a model that decomposes timeseries into three main components: trend, seasonality and holidays (Taylor and Letham, 2017). These components are combined as per Equation 14:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \#(14)$$

where:
$y(t)$ = Forecasted value
$g(t)$ = Trend function which models non-periodic changes in the time series
$s(t)$ = Periodic changes (e.g., weekly, and yearly seasonality)
$h(t)$ = Effect of holidays (which may be irregular)
ϵ_t = Idiosyncratic changes not accommodated by the model, assumed normally distributed

Trend

Prophet models trend ng growth model and a piecewise linear

Prophet w number of users of n growth in natu rtain *carrying capacity*. For example, t iling or becom of the numb s might be the number of people that h ernet. To inc to an increa ved to change analyst (for insta ncy with which the n model by Equation 15[10]. gepoints can be set automaticall f known product launch

[10] The reader is referred to Taylor and Letham, 2017 for a more comprehensive explanation.

$$g(t) = \frac{C(t)}{1+exp(-(k+a(t)^T\delta)_{\hookrightarrow \text{rate of growth adjustment term}}\left(t-\left(m+a(t)^T\gamma\right)\right))} \#(15)$$

where:
$g(t)$ = trend function
$C(t)$ = carrying capacity at time t
k = growth rate
m = offset parameter (required to connect endpoints of projected segments)
$\boldsymbol{a}(t)^T$ = vector containing all previous rate adjustments
$\boldsymbol{\delta}$ = change in rate
γ = number of change points multiplied by change in rate

For timeseries that do [illegible] a linear, constant rate of growth that is [illegible] model is given by Equation 16:

[illegible] phet is a type of additive regression

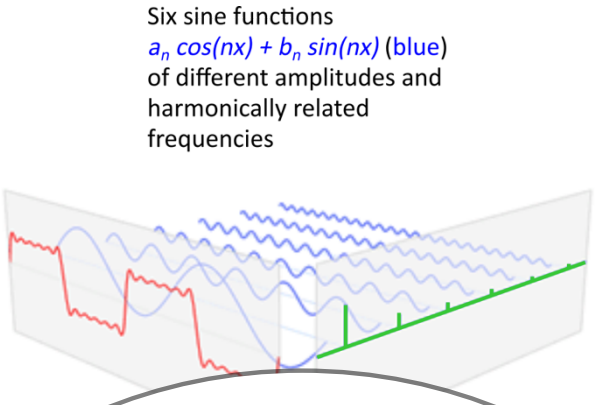
With t[illegible] with the potential to have several linea[illegible] as the indepe[illegible] ctions of tim[illegible] this way, Prophet effectively reframes [illegible] asting as a [illegible] that explici[illegible]e, which is inherently different to mo[illegible]7).

Seasona[illegible] to-regressive structure in the data (Ta[illegible]

Prophet es[illegible] composed of a weighte[illegible] cally in Figure 13[11]. As such, Pro[illegible] ry periodic signal that best fits the data.

[illegible] Fourier series

[11] For further explanation of Fourier series, the reader is referred to Taylor and Letham (2017) and Wikipedia (2021b) and references therein.



Fourier series
function
a sum
six
transform *s(f)*
ts the
nd

Data prmodel must always be a Pandas DataFr

The in ns, *ds*

and mp) column must be of the Pandas DD or

YYYY-M cture,

Prophe and column *y* must be numeric. Becau

A summa odelling univariate timeseries data. ble 9. The

holiday com is produced

continuously.

The data does, hov es and outliers (see

Section 4.1) making Prophe disadvantages of Pro g of wind power.

Table 9. Advantages and disadvantages of the Prophet model (after Taylor and Letham, 2017; Skander, 2020).

Feature	Advantage	tage
Additive model	Prophet can readily de New components c seasonality is id Additive mod	m
Fourier seasonality model	Prophet ac different	eries to model seasonality increases the risk of unc
Trend growth model		Proph continue to change at the same frequency and en projected forward, this can produce wide smooth e best curve to the time series. By using arbitrary growth it can reduce to standard logistic or linear approach zero.
Tunable parameters	Pro	sts to prove that parameter adjustment leads erpretable parameters. Only ar
Robustness to missing data	Inte Meas	to mor
Assumption of no causal relationship between past and future values		odated by the model (i.e., noise in the signal) are ed.
Simulated Historical Forecasts (SHFs)*	Economise Can be used relative to bas outliers, or where	ularly spaced. cy measurements. e, the more correlated their estimates of error are which nation and slows down forecast evaluation. casts to be computed which becomes computationally expensive

* To generate historical forecast errors, Prophet uses a procedure called simu recasts at various cutoff points in the history, chosen such that the horizons lie within the history and the total error can be evaluated.

3.5.4 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a type of Artificial Neural Network (ANN) capable of predicting future values of a timeseries. Artificial Neural Networks are ML models inspired by biological neural networks. Biological neural networks consist of millions of neurons organised such that each neuron is connected to thousands of others via synapses. Neurons produce electrical impulses that cause synapses to release chemical signals known as neurotransmitters. When a connected neuron receives a sufficient volume of neurotransmitter, it releases its own electrical impulses, propagating the signal along the network. The more positive feedback they provide to the overall system, the stronger and more established the synaptic connections become.

By way of imitation, artificial neurons consist of one or more binary inputs (activated/ not activated) and a binary output. This output forms the input/ connection to subsequent neurons. Each neuron activates its output when a predetermined number of inputs are active, propagating the signal across the network. Error in the output of each neuron is measured and the neuron's *weighting* within the system is adjusted accordingly, mimicking the feedback mechanism of biological neural networks (Gerón, 2019). Neurons can be grouped together in layers, and layers can be stacked to form deep neural networks where the emphasis of the model is on learning successive layers of increasingly meaningful representations of the data (Chollet, 2021). Recurrent Neural Networks are very similar to the ANNs described above except that their neurons also have a backwards-facing output (Figure 14). This enables them to handle sequential data very well.

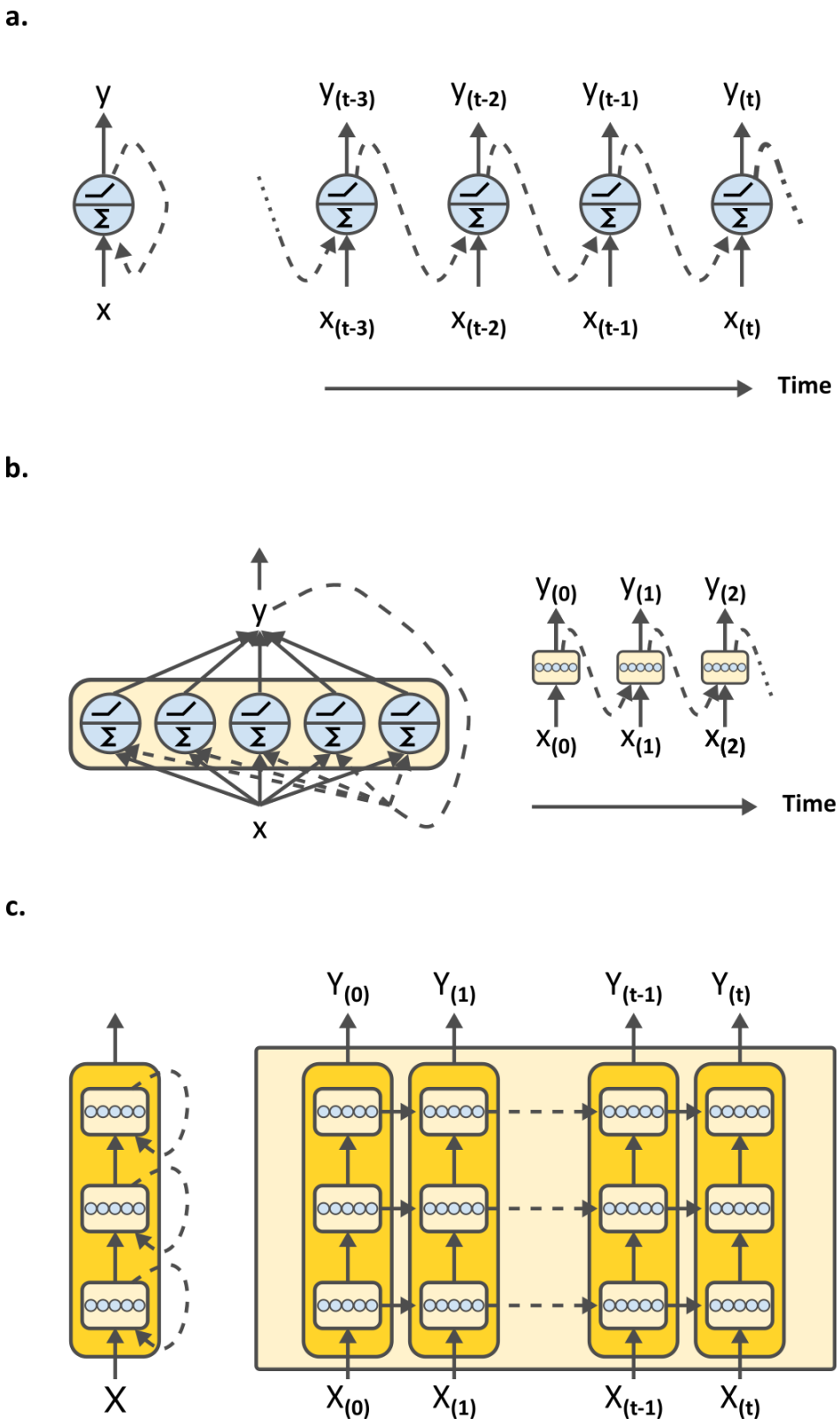


Figure 14. RNN structure schematics (left), unfolded in time (right). **a.** A recurrent neuron. **b.** A layer of recurrent neurons. **c.** A deep RNN. Backwards-facing outputs are highlighted by dashed lines (after Gerón, 2019).

As a person reads a sentence, he or she processes it word by word while keeping a memory of what came before. This results in a fluid representation of the passage of text. In this way, human neural systems process information incrementally while maintaining an internal model of what is being processed. This model is built from past information and is constantly updated as new information is taken in. The backwards-facing outputs of a RNN's neurons enable it to do the same, albeit in a simplified version: sequences are processed by iteration through the sequence elements while an internal *state* is maintained that contains information regarding what has gone before (Chollet, 2021). This internal state can be thought of as a type of memory. Figure 15 provides a schematic explanation of how it is achieved and maintained in a LSTM cell (see below).

Because of this design, RNNs are often able to perform better than other ML algorithms on large, complex problems. Their form means that they are inherently data-hungry models, able to make use of very large datasets. This makes them a good fit for the LHB dataset.

Disadvantages

Despite the advantages outline above, RNNs can suffer from two main issues:

1. Unstable gradients[12] and
2. Very limited short-term memory

Unstable gradients can be managed using a technique known as *dropout* (see Appendix D) and limited RNN short-term memory can be augmented using *Long Short-Term Memory (LSTM)* or *GRU* cells (Figure 15).

The use of LSTM cells results in better performance, faster convergence during training, and better detection of long-term dependencies in the data. Due to the transformations performed on the data as it passes through the RNN, some information is lost at each time step. As this effect accumulates, it can become critical. In principle, LSTM cells reduce this effect by enabling the network to learn what information to save, and what to discard over the longer term (Figure 15).

[12] Gradient Descent is introduced in Appendix D (esp. Figure 38). See Gerón (2019) for further explanation.

For the reasons outlined above and in Appendix D, these techniques are incorporated into the RNN models used in this study. Where appropriate, the specific LSTM-RNN configuration is henceforth referred to as LSTM-RNN.

Data preparation

Another drawback of RNNs is that they require a significant amount of data preparation in order to function. An explanation of the steps undertaken in this study, along with a discussion of some further drawbacks of the RNN is presented in Appendix D.

A summary of the advantages and disadvantages of RNNs are summarised in Table 10.

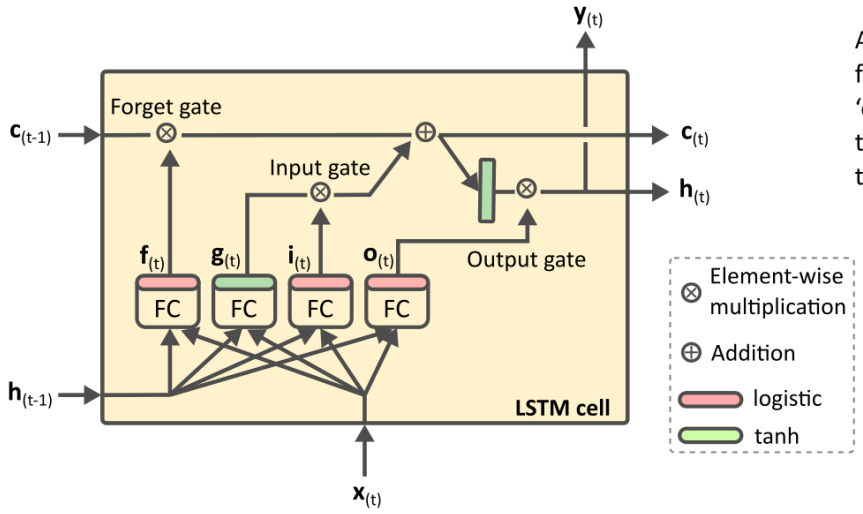
Table 10. Advantages and disadvantages of the RNN model (After: Gerón, 2019; Chollet, 2021).

Feature	Advantage	Disadvantage
Backwards-facing neurons	Enables the maintenance of an internal *state* or model of information that has already been processed. Enables the (iterative) processing of large Enables the capture of highly complex patterns within the data	Iterative exposure to the sequence of data results in long fitting times.
Layering and stacking of neurons (Deep Neural Networks)	Facilitates the learning of successive layers of increasingly meaningful representations of the data	Increases fitting time
Data preparation	Increases training (convergence) times	Requires advanced statistical and programming expertise and additional overall model construction time.
LSTM cell	Results in better performance, faster convergence during training, and better detection of long-term dependencies in the data	Requires additional expertise and knowledge
Use of stochastic gradient descent to minimise loss function	Increases fitting speed and enables training on large training sets	Stochasitc nature means that once a minimum is reached, the algorithm continues searching for a solution. Final parameter values are therefore good but not necessarily optimal.

As the long-term state $\mathbf{c}_{(t-1)}$ proceeds through the cell, it passes through the 'forget gate', disregarding some information and later adding new information through an operation that adds data selected by the 'input gate'. $\mathbf{c}_{(t)}$ then passes out of the cell without further transformation.

The 'input gate', controlled by $\mathbf{i}_{(t)}$, controls which parts of $\mathbf{g}_{(t)}$ should be added to the long term state

An LSTM cell is similar to a normal cell except that its incoming signal or 'state' is divided into two vectors. These can be thought of as a short-term state, $\mathbf{h}_{(t)}$, and a long-term state, $\mathbf{c}_{(t)}$.



A copy of $\mathbf{c}_{(t)}$ passes through a tanh function and is later filtered by the 'output' gate. This produces the short-term state $\mathbf{h}_{(t)}$ which is the output of the cell at this time step, $\mathbf{y}_{(t)}$.

The 'forget gate', controlled by $\mathbf{f}_{(t)}$, controls which parts of the long-term state are discarded

The principal layer, $\mathbf{g}_{(t)}$, analyses the current input, $\mathbf{x}_{(t)}$, and the previous short-term state, $\mathbf{h}_{(t-1)}$

New information is added to the system as the current input vector, $\mathbf{x}_{(t)}$, and the previous short-term state, $\mathbf{h}_{(t-1)}$, pass through four different, connected layers that serve different functions

The 'output gate', controlled by $\mathbf{o}_{(t)}$, controls which parts of the long-term state should be read and generate the output for this time step, for both $\mathbf{h}_{(t)}$ and $\mathbf{y}_{(t)}$.

Figure 15. LSTM configuration (after Gerón, 2019).

3.6 Model evaluation

Many methods have been applied to assess model performance during ramp event characterisation (Table 5), depending on the output being predicted and whether the problem is one of classification or regression. Since the aim of this study is to predict wind power output and associated ramp events, the primary problem is one of regression. A typical error evaluation measure for regression problems is Root Mean Square Error (RMSE). RMSE gives an idea of the percenta… lgorithm makes in its predictions (zero indicating perfect fit). An … s in which there are many outliers is Mean Ab… as is the case with ramp events (s… erally preferred (Gerón, 201… ses.

3.7 Ur…

NWP … rface pressure, precipitation etc. Given …, wind directi… … with powe… ind speed is considered the dominan… vents. Howe… … s and changi… udies have shown (Section 2.1/ Appe… utors. Martíne… …her the tendencie… es also act as major (perhaps, …erisation. The same a… … d to reduce error in the p… mmended that future research …ese ideas, the predictive ability o… … with a multivariate LSTM-RNN using physica… ould improve ramp eve… temperature, and power from the LHB dataset.

…iate data

…ray of meteorological variables[13] i…

[13] e.g., https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview

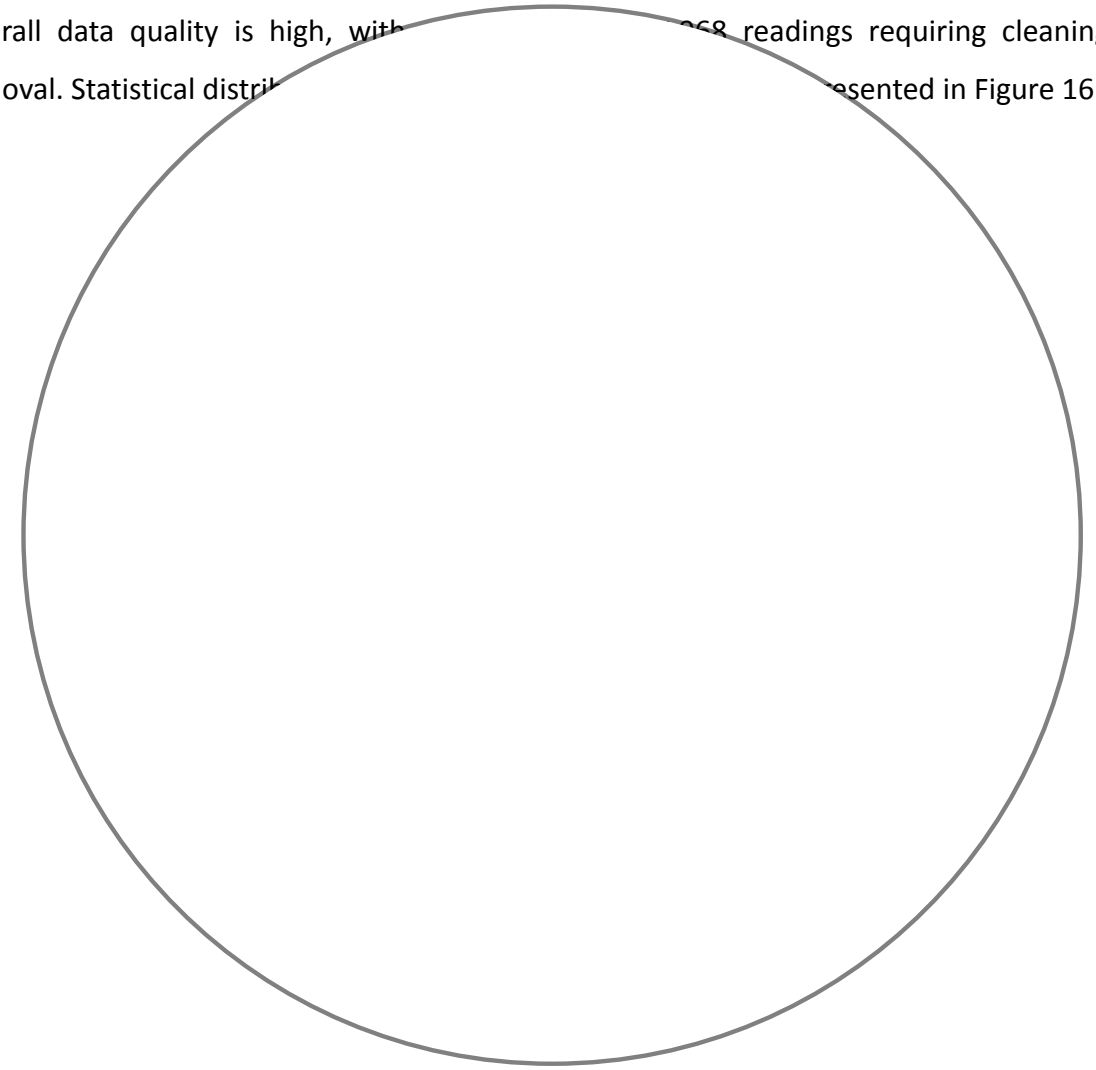
4 Results

This chapter is organised into three sections: 1) results of the EDA 2) a comparison of the high-level modelling results and 3) key outputs from each model.

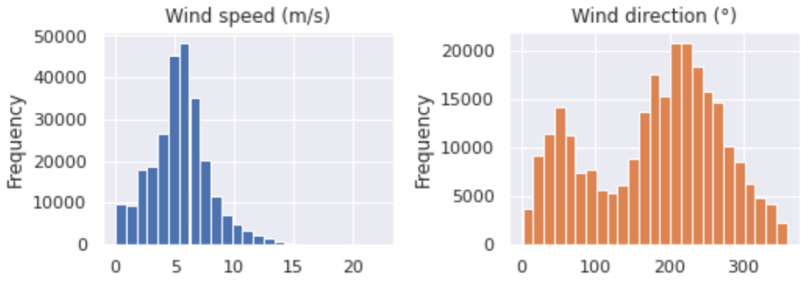
4.1 EDA results

4.1.1 Data quality and statistical distributions

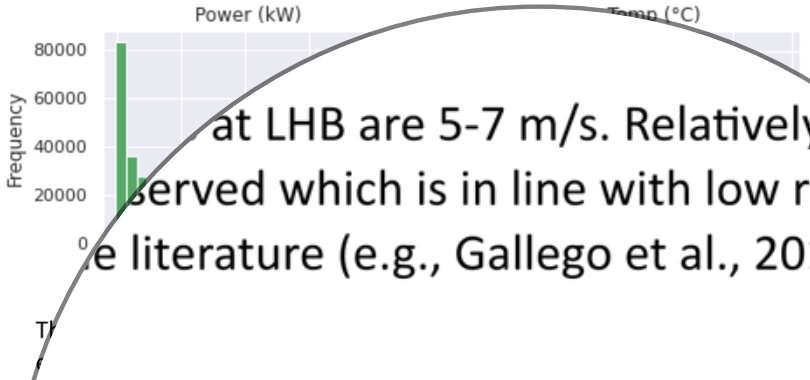
Overall data quality is high, with [illegible]68 readings requiring cleaning or removal. Statistical distri[illegible]esented in Figure 16.



a.



Wind direction is variable but two prevalent directions are apparent (~50° and ~220°) from the bimodal distribution.



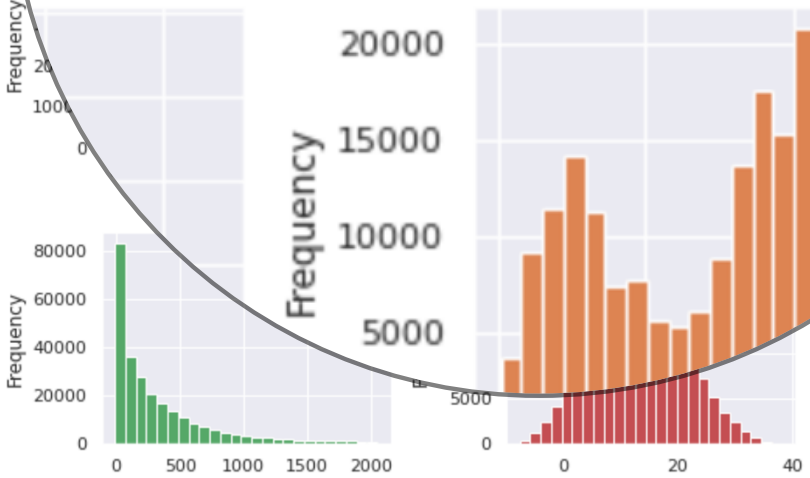
The uncleaned tetmperature plot shows the anomalous readings of in summertime which ely skew the data.

at LHB are 5-7 m/s. Relatively fe
served which is in line with low ramp
e literature (e.g., Gallego et al., 2013).

b.

(m/s)

Wind direction (°)



eaned temperature ot shows an approximately symmetric distribution centred around 10°C which is in line with expectations.

The power histogram shows a Lorenz curve-like distribution which indicates that average power output from each turbine is generally low (most frequently less than ~150 kW) and rarely reaches the maximum capacity of 2050 kW. This is reflective of the wind speed data and is in line with expectations.

Figure 16. Statistical distributions of the selected LHB variables **a.** raw **b.** cleaned.

4.1.2 Stationarity

When modelling timeseries data, it is assumed that the summary statistics of the observations are consistent over time. A timeseries is considered stationary when this is the case, and non-stationary otherwise. The assumption of stationarity in a timeseries can easily be violated by the presence of trend, seasonality, and other time-dependent structures (Brownlee, 2019). Reviewing pl 17) shows obvious trends in all features: higher wind sp ong short-term cyclicity in wind direction es that all features are non-statio

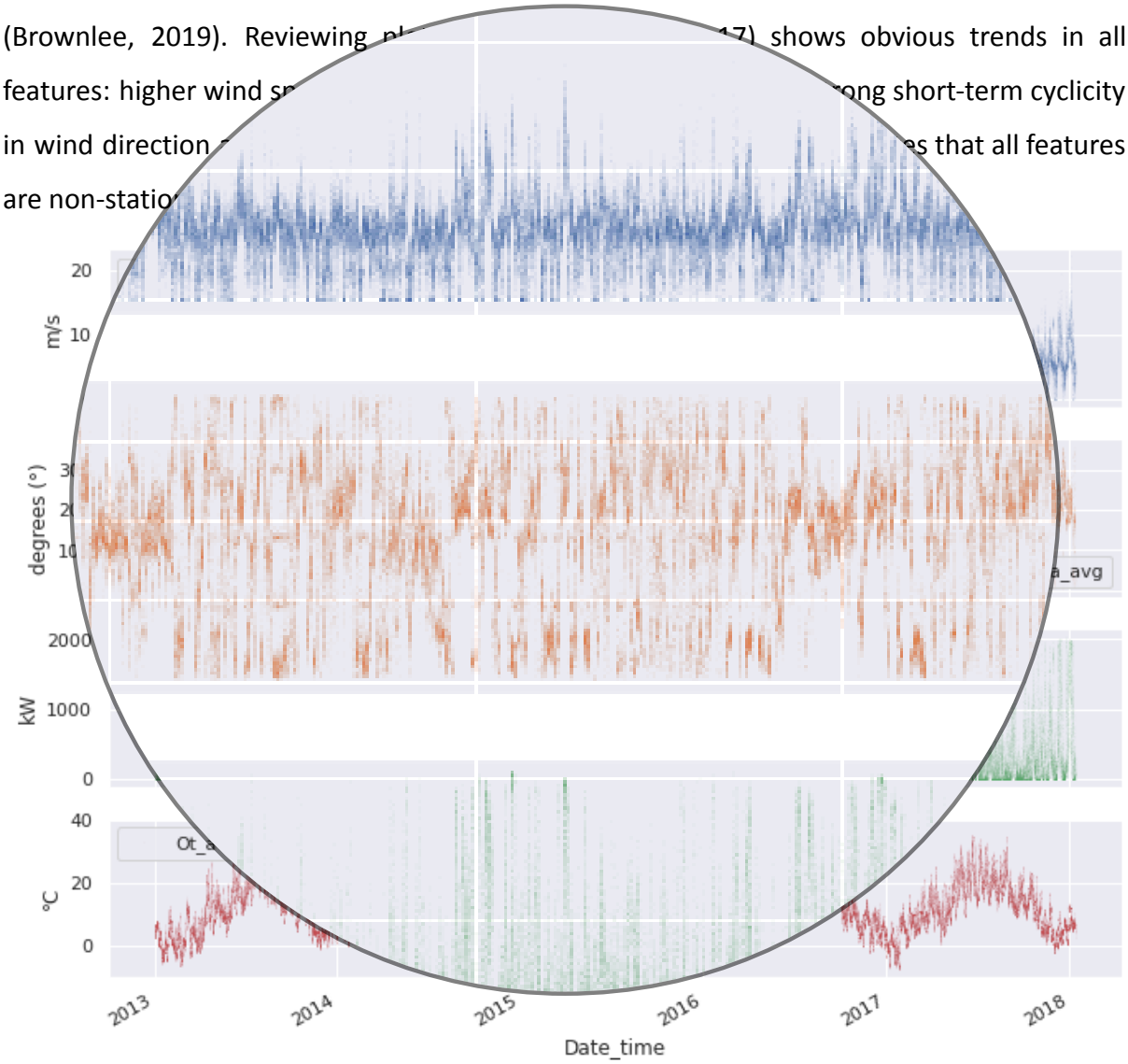


Figure 17. LHB physical features averaged at wind farm level. Ws = wind speed, Wa = wind direction, P = active power, Ot = outside temperature. Outliers are clearly visible in the Power series.

The Augmented Dickey-Fuller Test (ADF) is a test that can be used to provide statistical evidence of stationarity. The null hypothesis, H_0, of the test is that the series is stationary.

H_0, the data is non-stationary

The alternativ terpreted

using the re the data is stationary

- p-valu
- p-val n both the wind speed and the pow

The ADF ned

p-values that they are in fact stationary. This resu r in

Section 5

4.1.3 Au

Autocorrela vations

in a timeserie tial Autocorrelation between

the series and lag re 18.

strength and type of rel

Date_time	t	t-	
31/12/2012 23:00	0	Nan	
31/12/2012 23:10	1	0	Nan
31/12/2012 23:20	2	1	0
31/12/2012 23:30	3	2	1
31/12/2012 23:40	4	3	2
31/12/2012 23:50	5	4	3
01/01/2012 00:00	6	5	4
01/01/2012 00:10	7	6	5
01/01/2012 00:20	8	7	6
01/01/2012 00:30	9	8	7

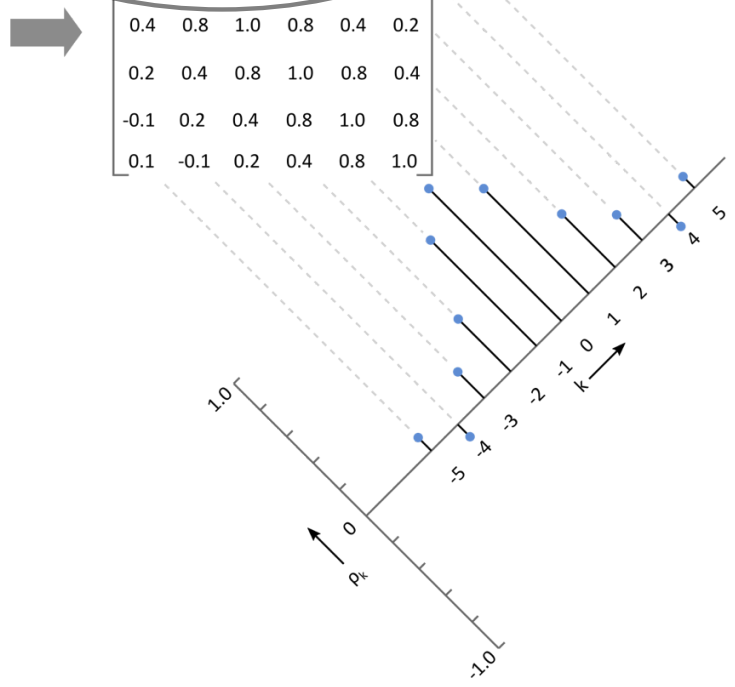


Figure 18. Calculation of autocorrelation (schematic). After Box et al. (2015).

A timeseries is considered autoregressive when previous observations are predictive of later observations. Autoregressive data shows a gradual decrease in the autocorrelation plot.

Partial autocorrelation quantifies the relationship between an observation in a timeseries with observations at prior time steps after the effect of intervening observations is removed. It can be used to determine the order of an AR(p) model.

The autocorrelation and partial correlation plots of the 10-minutely and resampled hourly power series are presented in Figures 19 and 20.

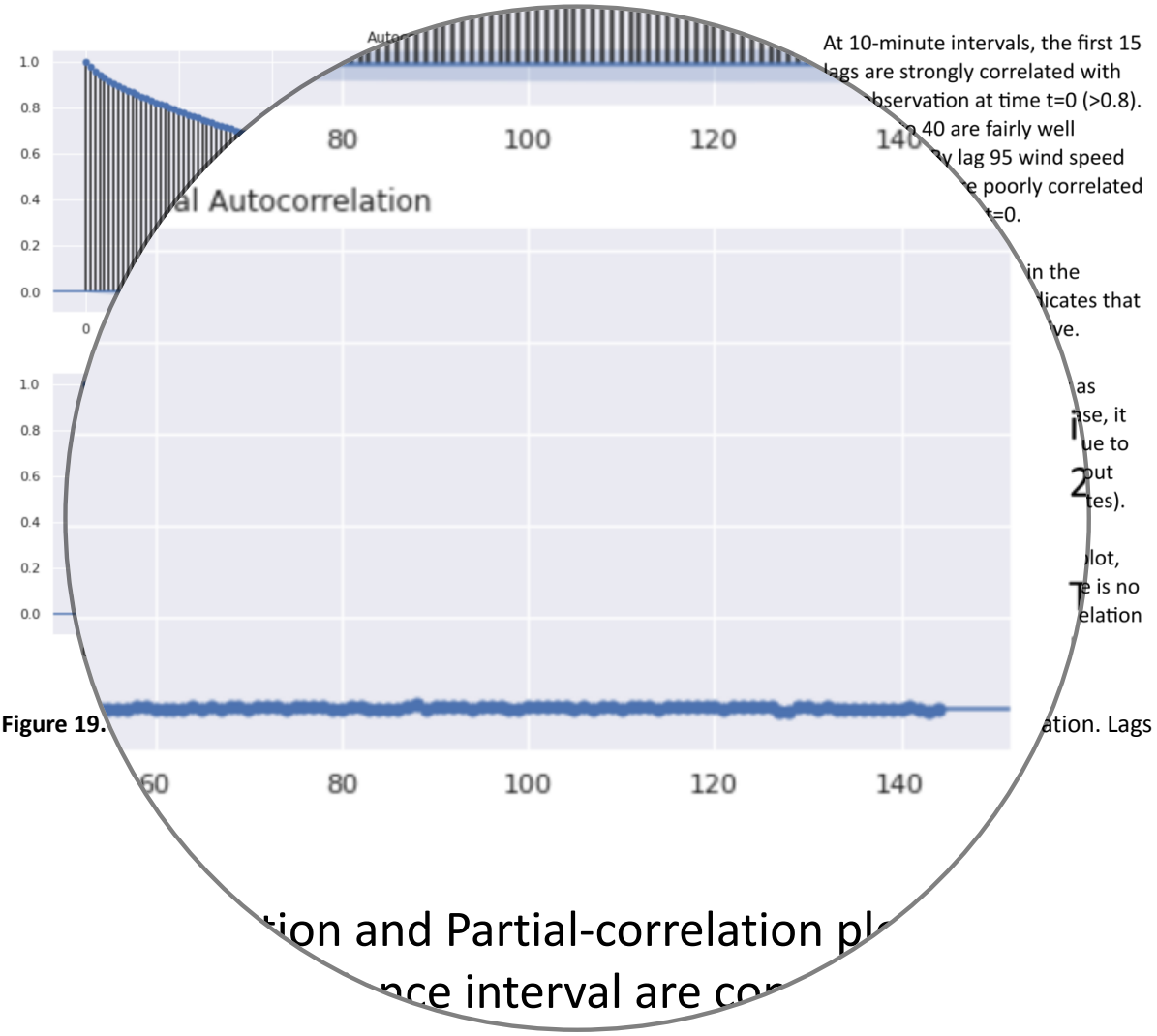
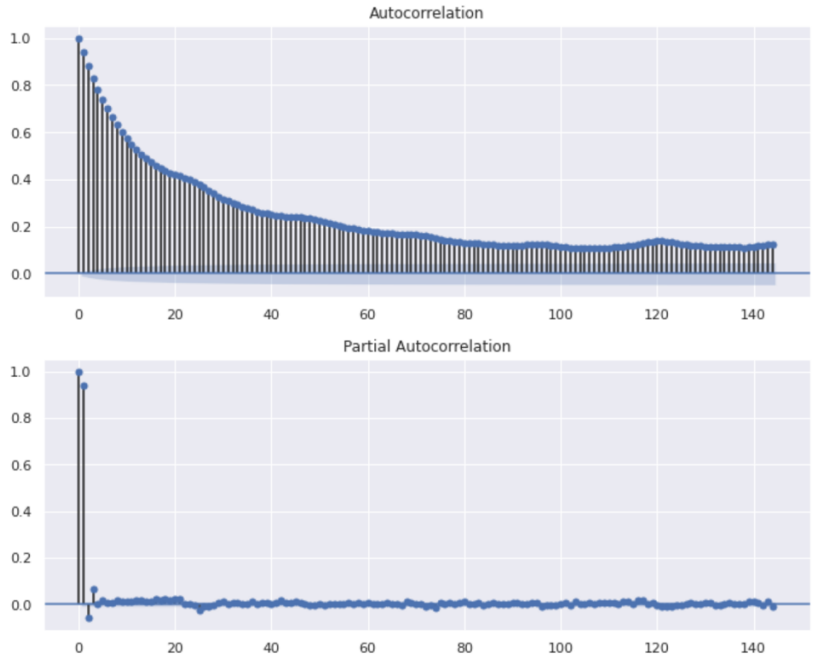


Figure 19. ation. Lags



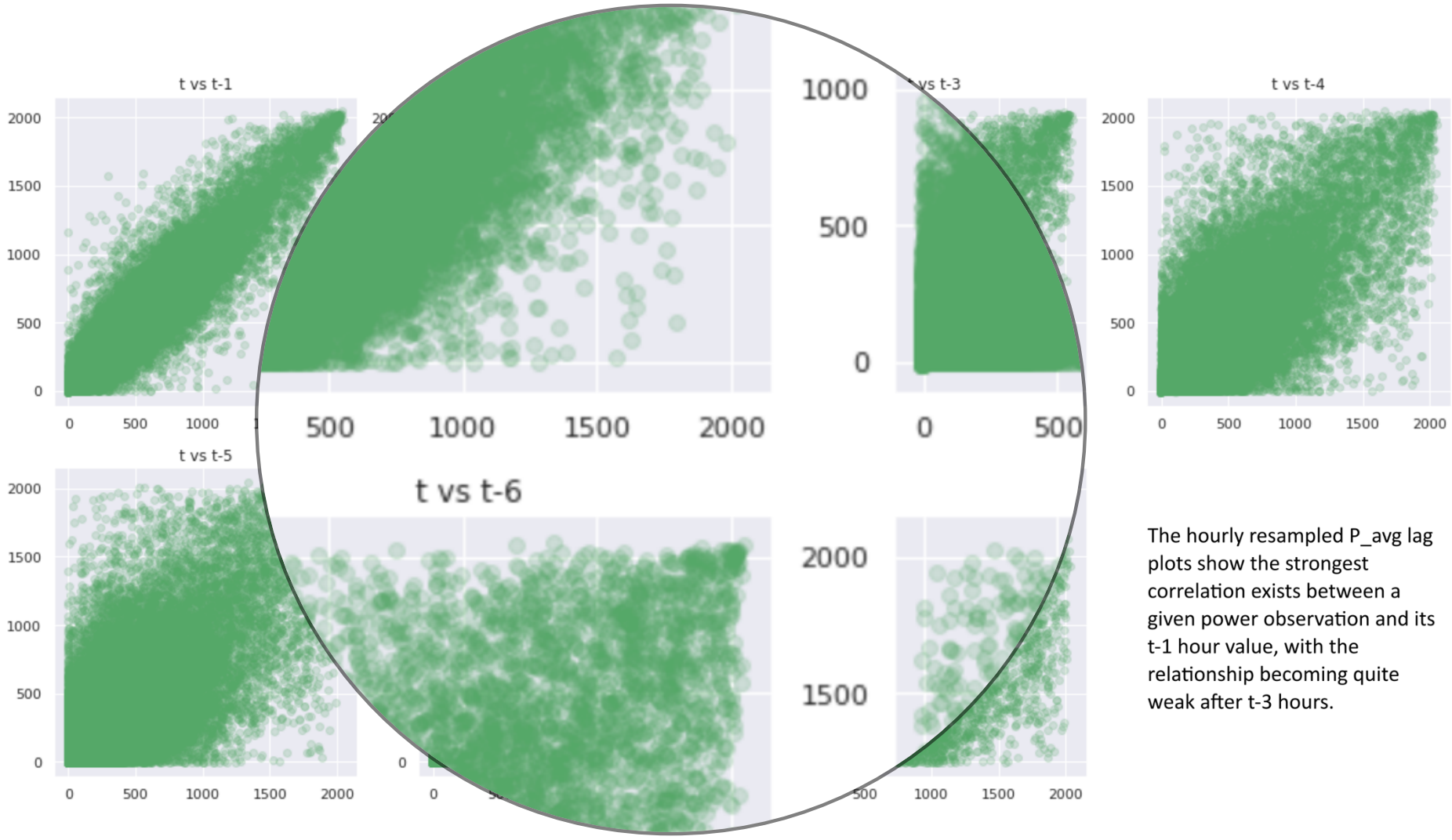
The hourly power series is less autoregressive – shown by the sharper decline in autocorrelation.

The autocorrelation plot also reveals a diurnal seasonality - indicated by peaks in the correlation coefficient every 24 hours. This indicates that the primary atmospheric processes that affect wind speed over the LHB site occur over daily cycles.

The partial autocorrelation plot, however, again suggests that there is no statistically meaningful correlation for lag values beyond lag 1

Figure 20. 1-hour P_avg Autocorrelation and Partial-correlation plots (x-axis = lag, y-axis = correlation).

Autocorrelation is also demonstrated in the lag plots of Figure 21. Lag plots are scatterplots of observations against their lagged values.



The hourly resampled P_avg lag plots show the strongest correlation exists between a given power observation and its t-1 hour value, with the relationship becoming quite weak after t-3 hours.

Figure 21. P_avg lag plots.

4.1.4 Level, Trend, Seasonality and Noise

Timeseries decomposition is an abstraction of the raw series into components of level (the average value in the series), trend (the increase or decrease in the series), seasonality (the repeating short-term cycles in the series) and noise (the random variation in the series). Statsmodels offers a seasonal decomposition method that first estimates the trend by applying a convolution filter to the data. The trend is then removed from the series and the average of this de-trended series for [illegible] returned seasonal component. This is a naive decomposition [illegible]he winter months. Wh[illegible] he investigated as part of future research (Se[illegible] itions of the power series (P_avg p[illegible] shown in Figure 22. With m[illegible] es in power output a[illegible] plexity of the seasonality. The residua[illegible] ever, the trend b[illegible] ethod is unable[illegible] at noise in the series is random. S[illegible] ematic relatic[illegible] onally displa[illegible]

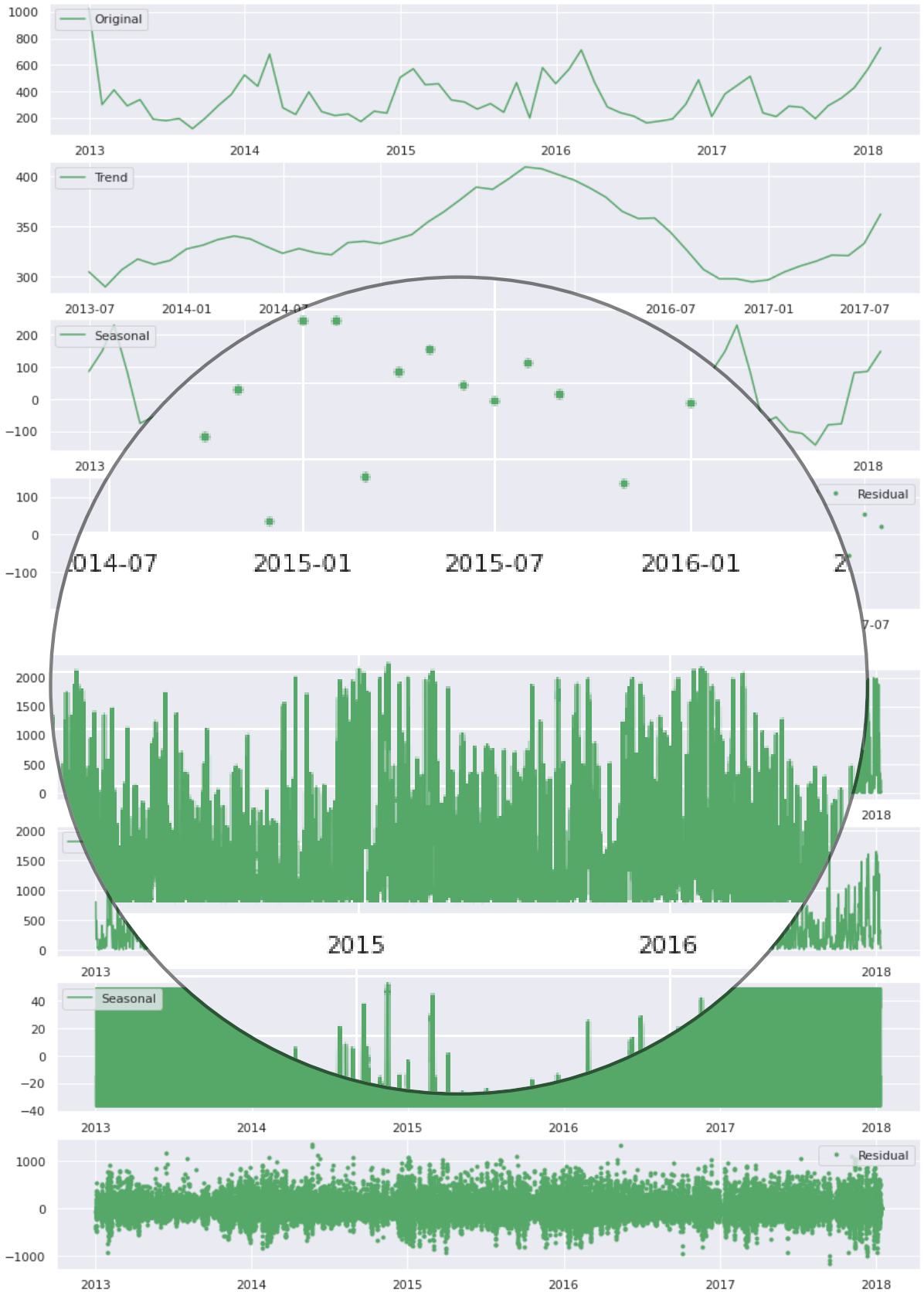
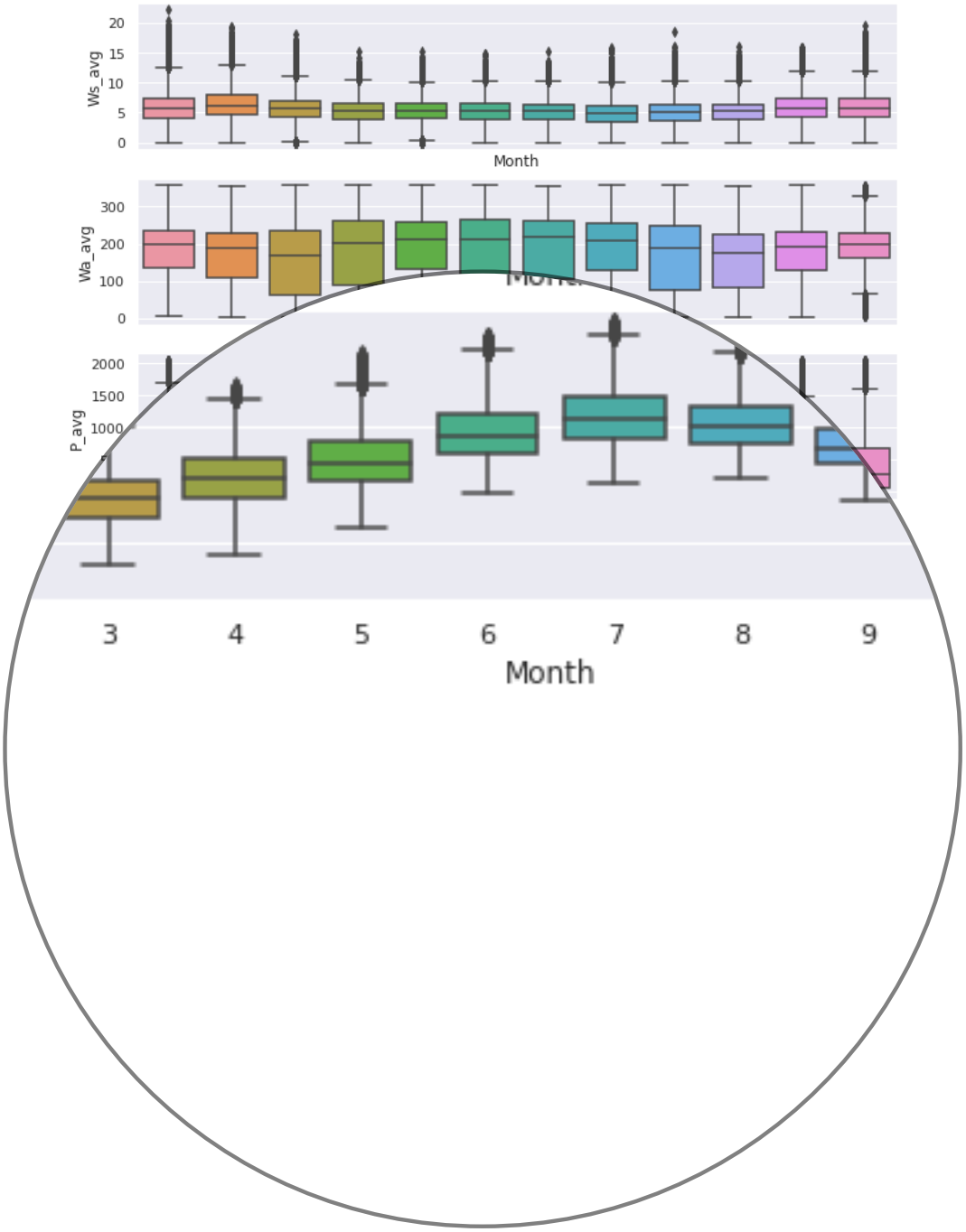
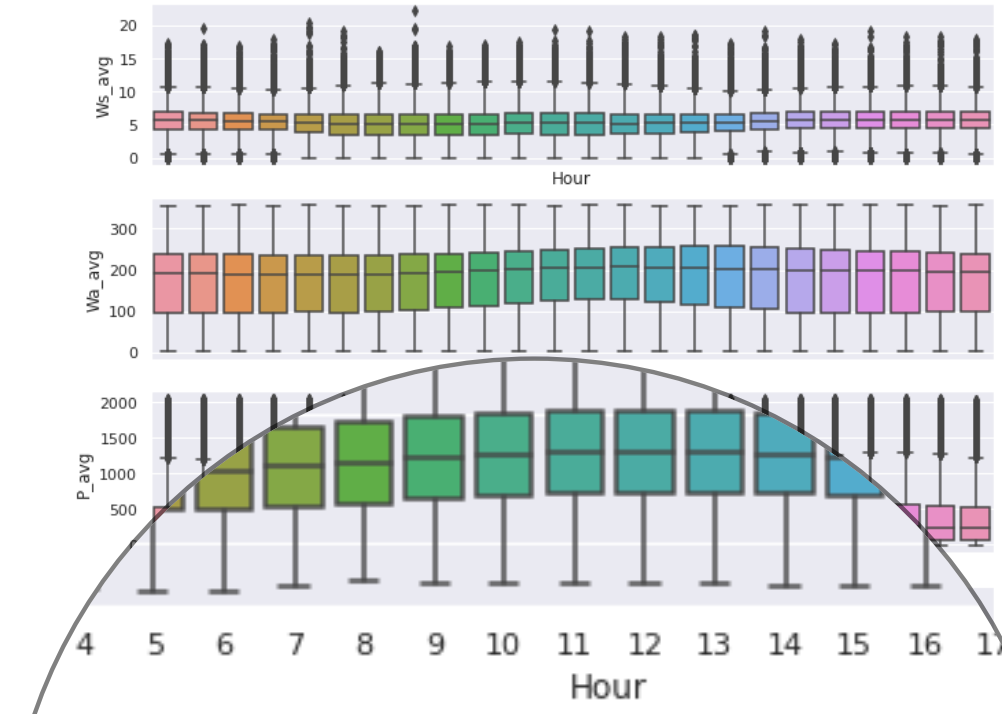


Figure 22. Monthly (top) and hourly (bottom) seasonal decomposition plots of the P_avg timeseries components Level (Original), Trend, Seasonality and Noise (Residual).


Ws_avg
Month
Wa_avg
P_avg
3
4
5
6
7
8
9
Month



Figur grouped by month and hour. Seasonal differenc re 22
are ag o visible are slight increases in wind speed and as during
ght-time hours and diurnal changes in wind direc

4.2 Re

Untuned r study are
presented in son between
models. Model t p frequencies of
the datasets are give s in ramp frequencies
between the hourly and 10- performanc tant to note that valid model
comparisons are not possible between the different datasets. For the same reason, multivariate LSTM-RNNs results are presented separately in Section 4.3.

Predictive accuracy

Using 10-minutely data, the most accurate model is the LSTM-RNN with a test MAE of 27.34 kW. The LSTM-RNN is followed by ARIMA and ARMA with test MAEs of 49.42 and 50.06 kW respectively. Prophet is the least accurate model with a test MAE of 336.16 kW. Using hourly

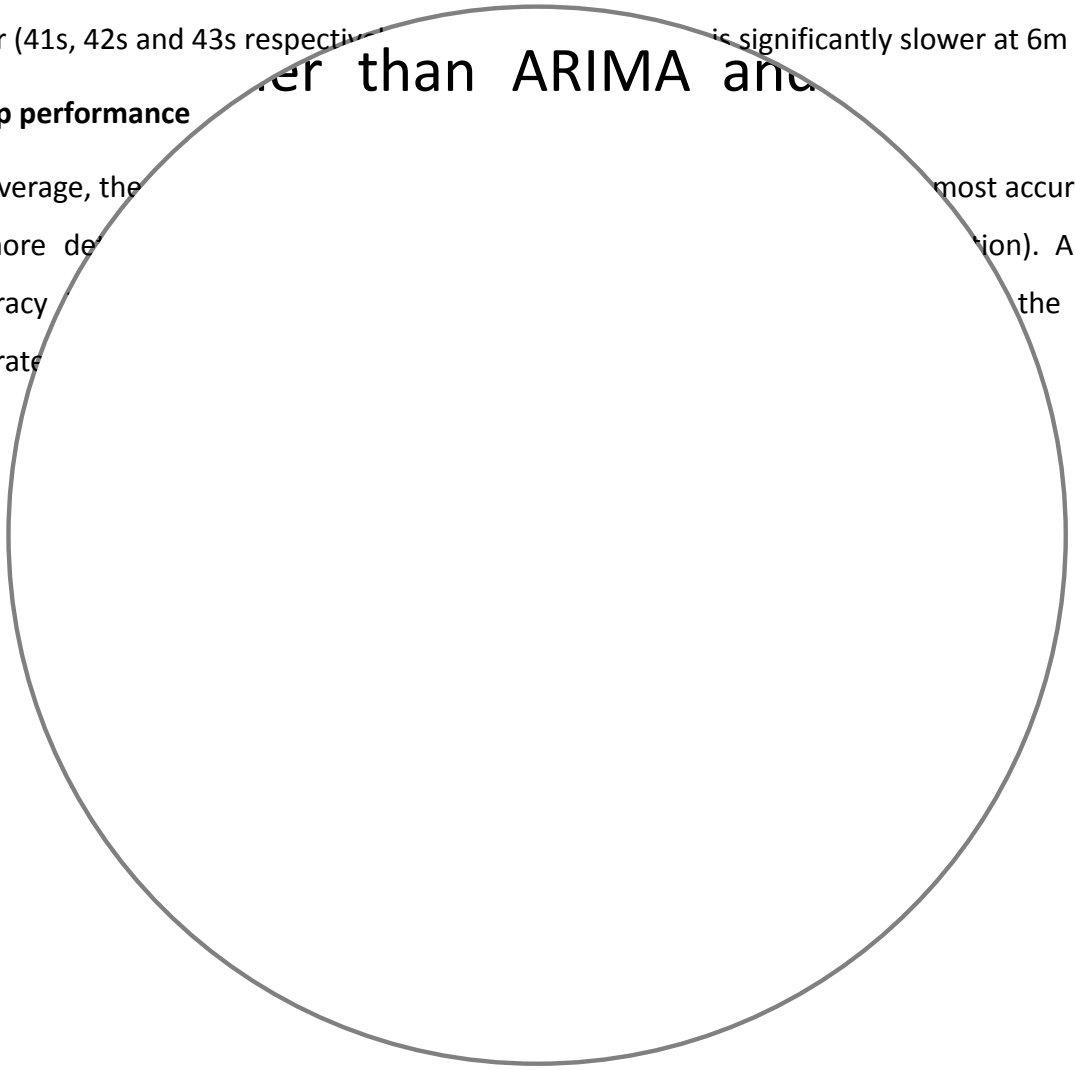
data, ARMA slightly outperforms ARIMA (test MAEs of 93.1 to 93.26 kW respectively). Otherwise, the ranking remains unchanged.

Computation time[14]

ARIMA is the fastest model to fit the 10 minutely data (1m 35s), followed by ARMA (4m 46s) and Prophet (8m 2s). By a significant margin, the LSTM-RNN is the slowest fitting model (29m 34s). ARMA, ARIMA and Prophet all fit the hourly data within one second of each other (41s, 42s and 43s respectiv is significantly slower at 6m 56s.

Ramp performance

On average, the most accurately (a more de ion). ARMA accuracy the least accurate



[14] Computation times are provided in this section for comparative purposes only. See Appendix C.

Table

	Lags	Fit time (mm:ss)	Forecast time (mm:ss)	Train RMSE		Positive ramp acc. (RMSE)	Positive ramp acc. (MAE)	Negative ramp acc. (RMSE)	Negative ramp acc. (MAE)	Non-ramp acc. (RMSE)	Non-ramp acc. (MAE)
nivariate	3	01:35	00:07	79.84	87	123.68	83.85	129.51	86.88	79.65	43.62
nivariate	3	04:46	00:13	79.69	87	124.39	84.58	127.74	85.04	79.62	44.45
nivariate	1	29:34	00:03	-	45	85.6	61.06	57.28	36.28	39.41	23.86
ultivariate	1	44:02	00:03	-	4	81.16	58.09	57.6	38.41	38.41	21.79
						757.64	595.44	540.79	388.9	411.63	310.72
A						258.57	183.22	189.31	141.91	122.75	76.36
AR variate		08:02	07:28	381.53		258.99	183.47	189.19	141.47	122.81	76.17
LST					2	22.07	20.8	21.94	20.82	16.61	15
LSTM					95	51.73	45.91	47.72	42.74	47.6	42.65
Prophet					19.66	719.97	564.23	550.11	394.93	386.21	280.17
iate	3	00:42	00:03	133.16							
	3	00:41	00:04	133							
	1	06:56	00:01								
		07:05	00:0								

the datasets used to calculate ramp performance.

Ramp frequency (%)	Entire, cleaned LHB dataset	10-min test set	Hourly test set
Total ramp freq.	12.2	13.9	19.45
Positive ramp freq.	6.18	7.07	9.81
Negative ramp freq.	6.03	6.83	9.47
Non-ramp freq.	87.8	86.1	80.55

4.3 Model outputs

SARIMA was found in this study to quickly run into memory issues when fitting hourly univariate data datasets of more than one week in size (i.e., 168 observations)[15]. Since sufficient seasonality cannot reliably be captured in a one-week sample of the LHB dataset, SARIMA is not considered further in this section.

4.3.1 10-minute data

ARMA

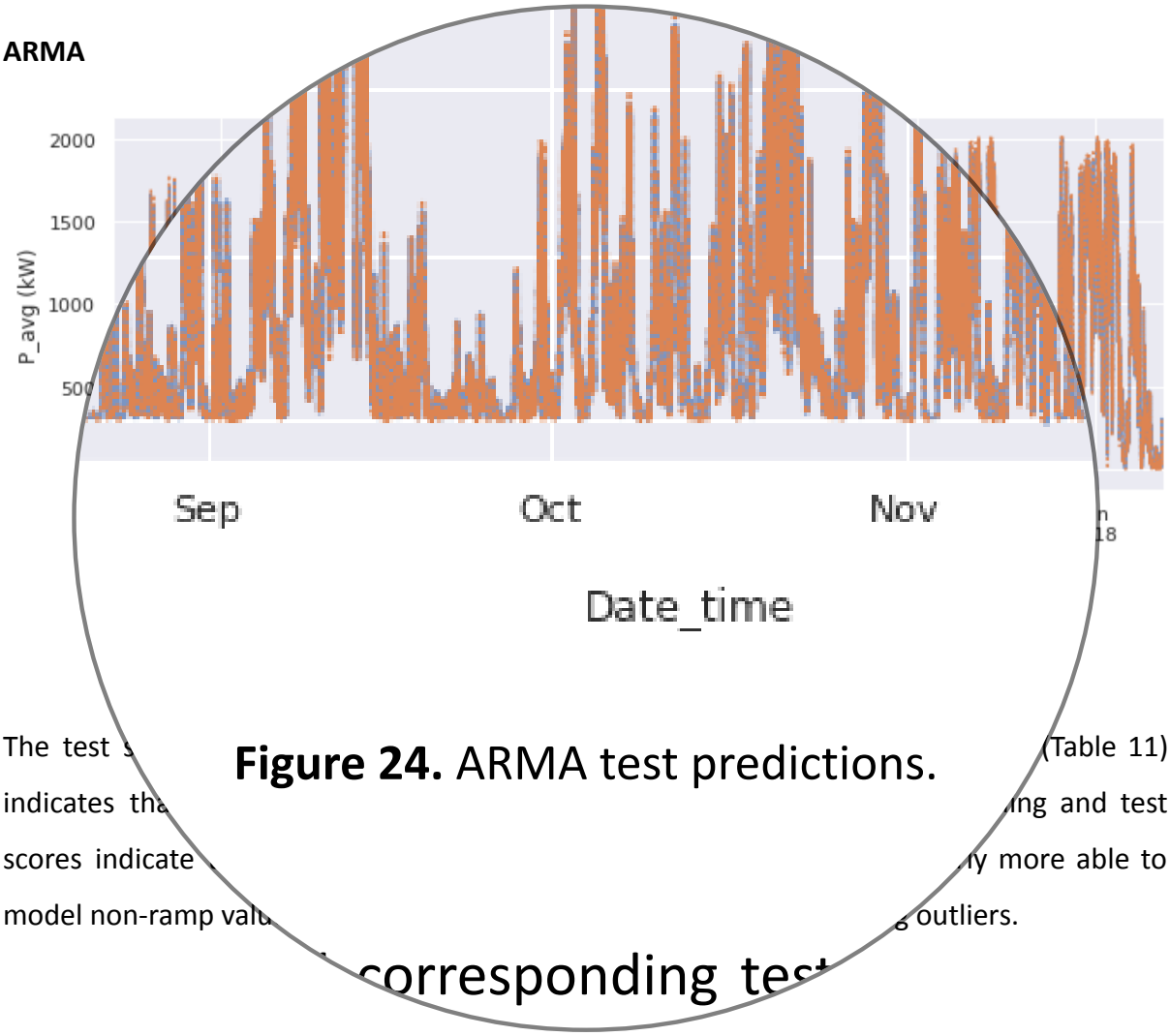


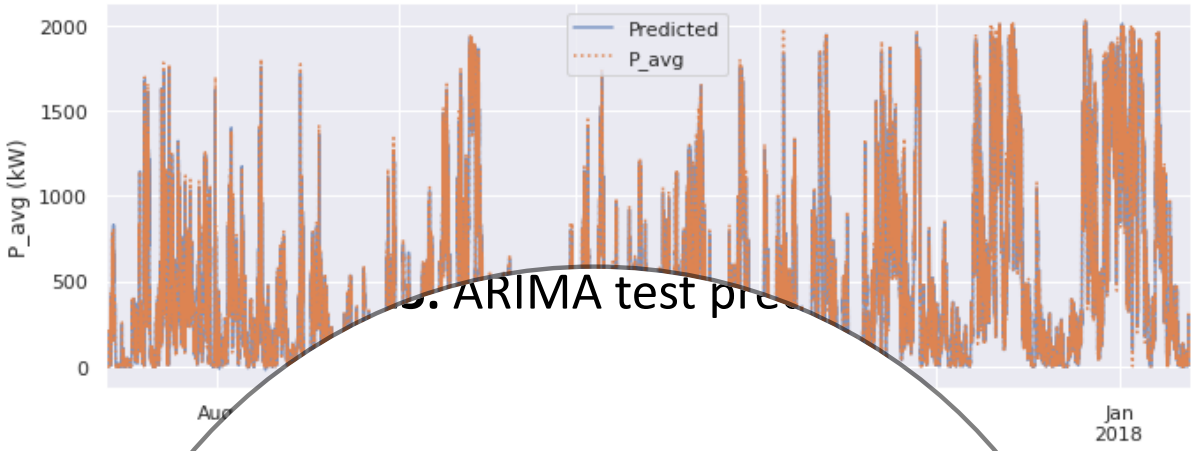
Figure 24. ARMA test predictions.

The test s (Table 11) indicates tha ng and test scores indicate y more able to model non-ramp valu outliers.

corresponding test

[15] See Appendix E.

ARIMA



5. ARIMA test pre

A and ARIMA are very close (e.g.,

tive similarity of the ARIMA training ar

verfitting the data. Interestingly, when co

predict positive ramps and non-ramps

The acc d 49.42 kW re e that the m RIMA is slig redict negati

Prophet

In all Prophet plots, power predictions are represented by an opaque blue line. Black dots represent actual power values. Prophet's default 80% uncertainty interval is shown by the light blue area. Power output in kW (P_avg) is plotted on the y axes.

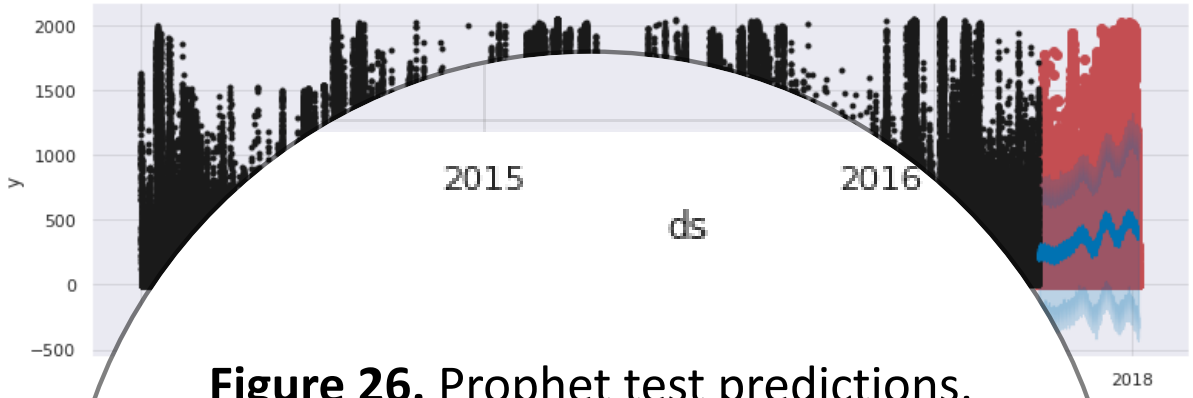


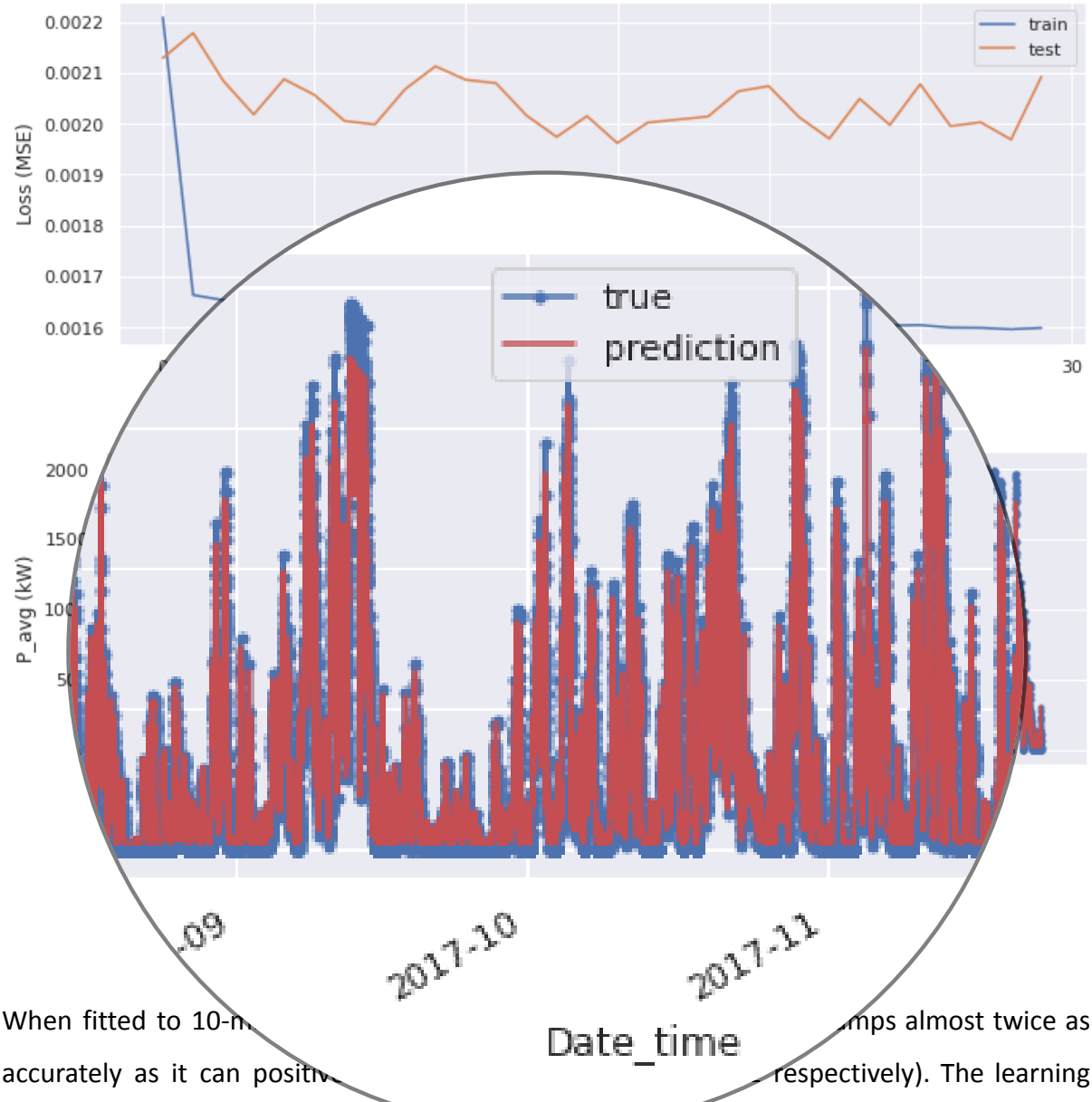
Figure 26. Prophet test predictions.

Prophet is able to capture the trend an
ty interval, however, widens quickly i
his is mirrored by Prophet's poor ra
than the other models (Tab
order of magni

Fitted -scale
seaso utliers
(i.e., la which
are on a able to
predict no erforming
model, the LS

Recurrent Neural Network

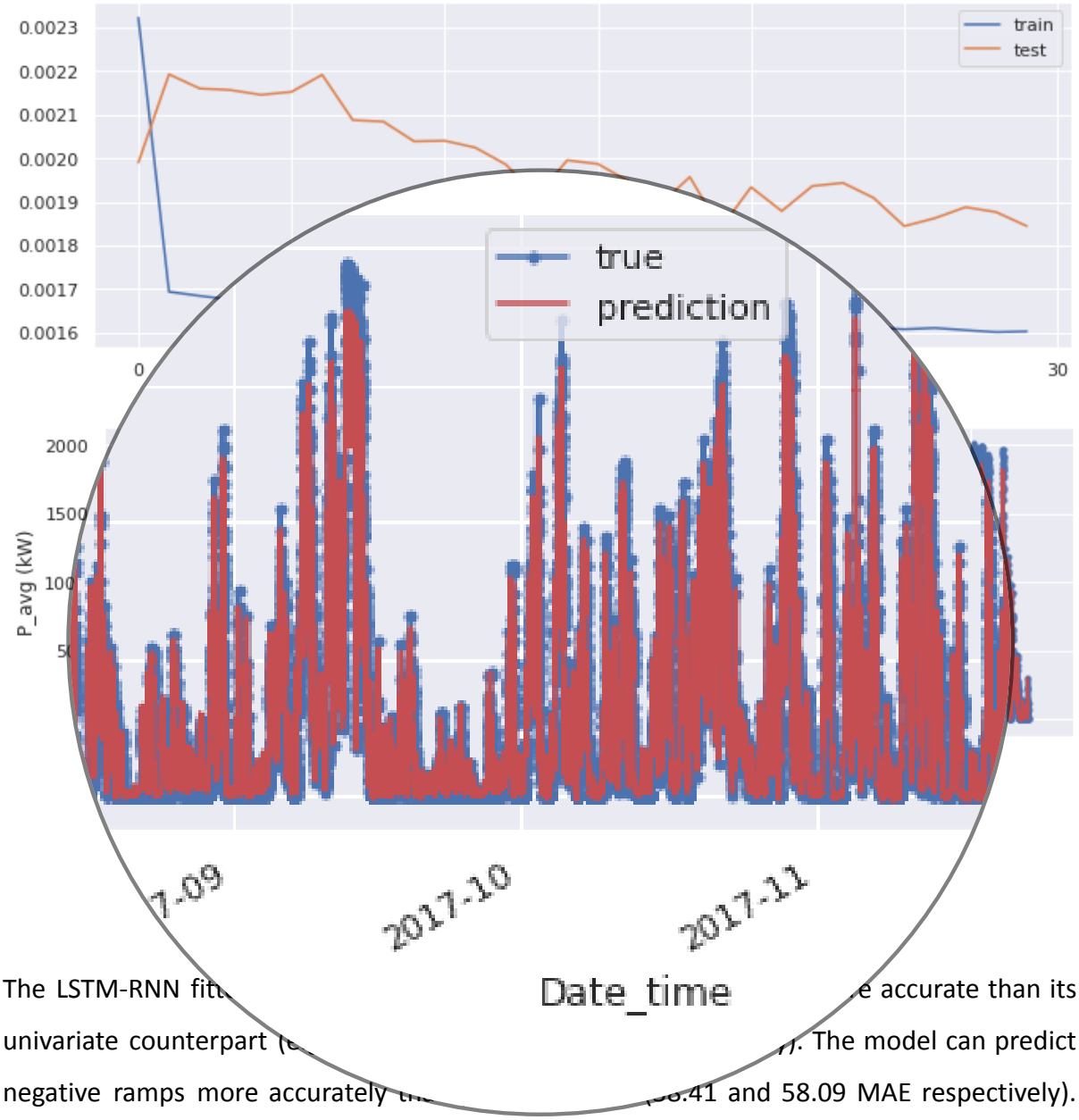
Univariate data



When fitted to 10-n… …mps almost twice as accurately as it can positiv… …respectively). The learning curves, however, indicate that the model is overfitting (training performance is much better than test) and is sub-optimally adjusted (see Section 5.2.1).

Recurrent Neural Network

Multivariate data



The LSTM-RNN fitt e accurate than its univariate counterpart (e). The model can predict negative ramps more accurately th (58.41 and 58.09 MAE respectively). The test predictions show that the model often misses extreme high readings. The learning curves, again, indicate that the model is overfitting and is sub-optimally adjusted.

4.3.2 Hourly data

ARMA

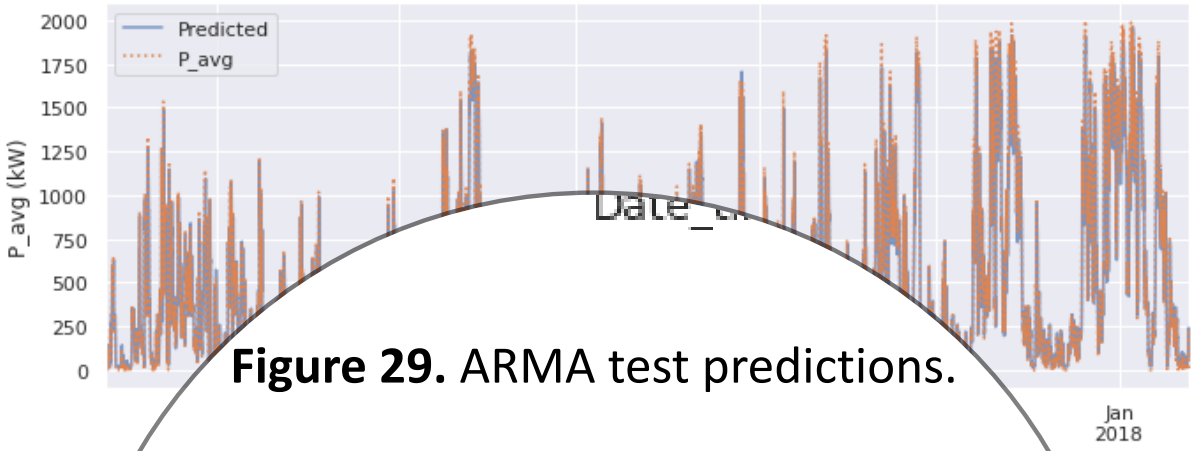


Figure 29. ARMA test predictions.

and test scores are similar, the test sco

While

licates

that t

r. ARMA is again much more able to moc

11).

ARIMA

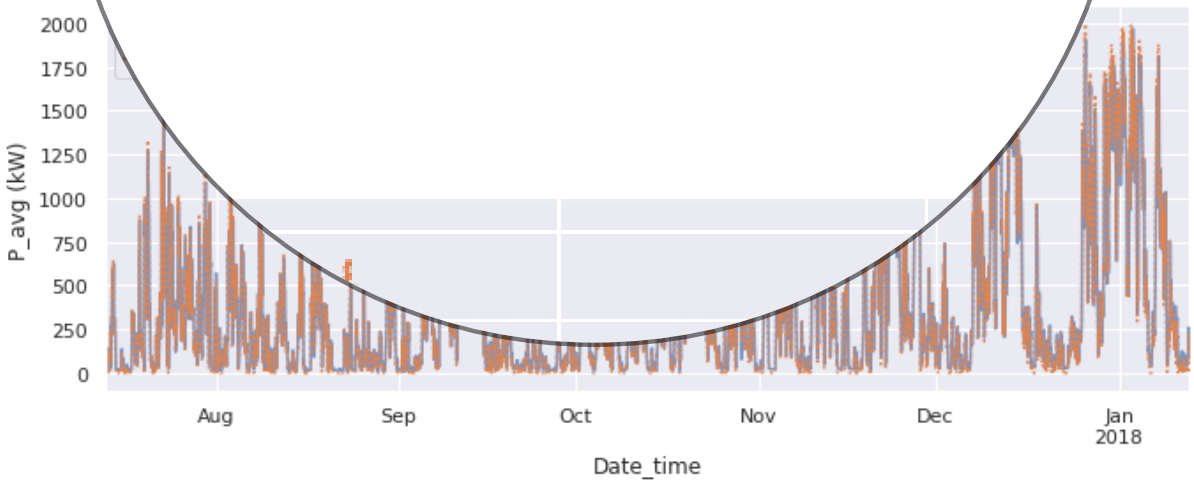


Figure 30. ARIMA test predictions.

The ARIMA test score of 149.1 RMSE indicates that the model is indeed very poor. The similarity of the training and test scores (Table 11) indicate that the model is not overfitting

the data. ARIMA is again more able than ARMA to predict positive ramp events, but less able to predict non-ramps and negative ramps (Table 11).

Prophet

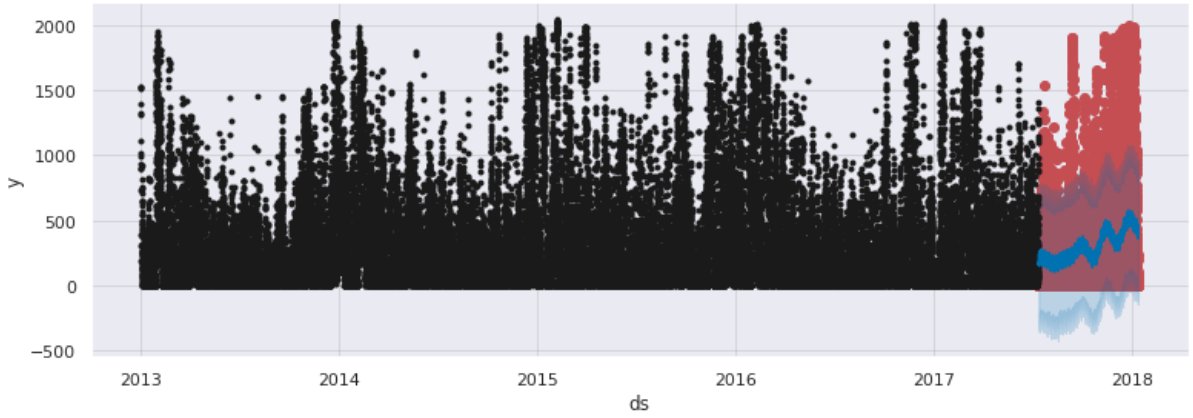
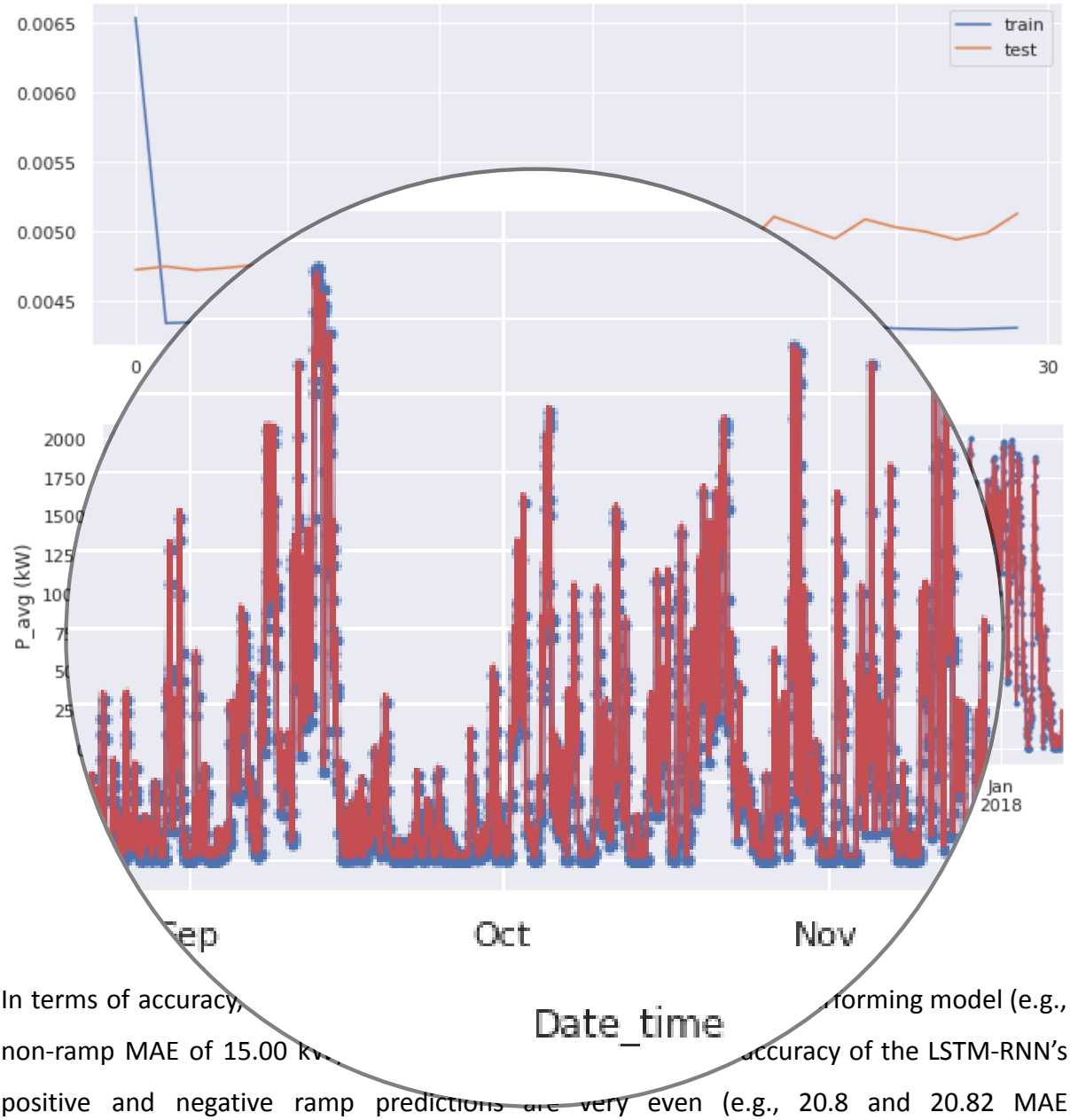


Figure 31. Prophet test set predictions.

When fitted to hourly data, Prophet is very conservative in its predictions of power output, not venturing very far from the mean (Figure 31). The width of the uncertainty interval is reduced compared to the model fit to 10-minute data, a fact reflected by the model's improved non-ramp accuracies. No outliers are captured by the model - reflective of the model's poor ramp capture statistics (Table 11).

Recurrent Neural Network

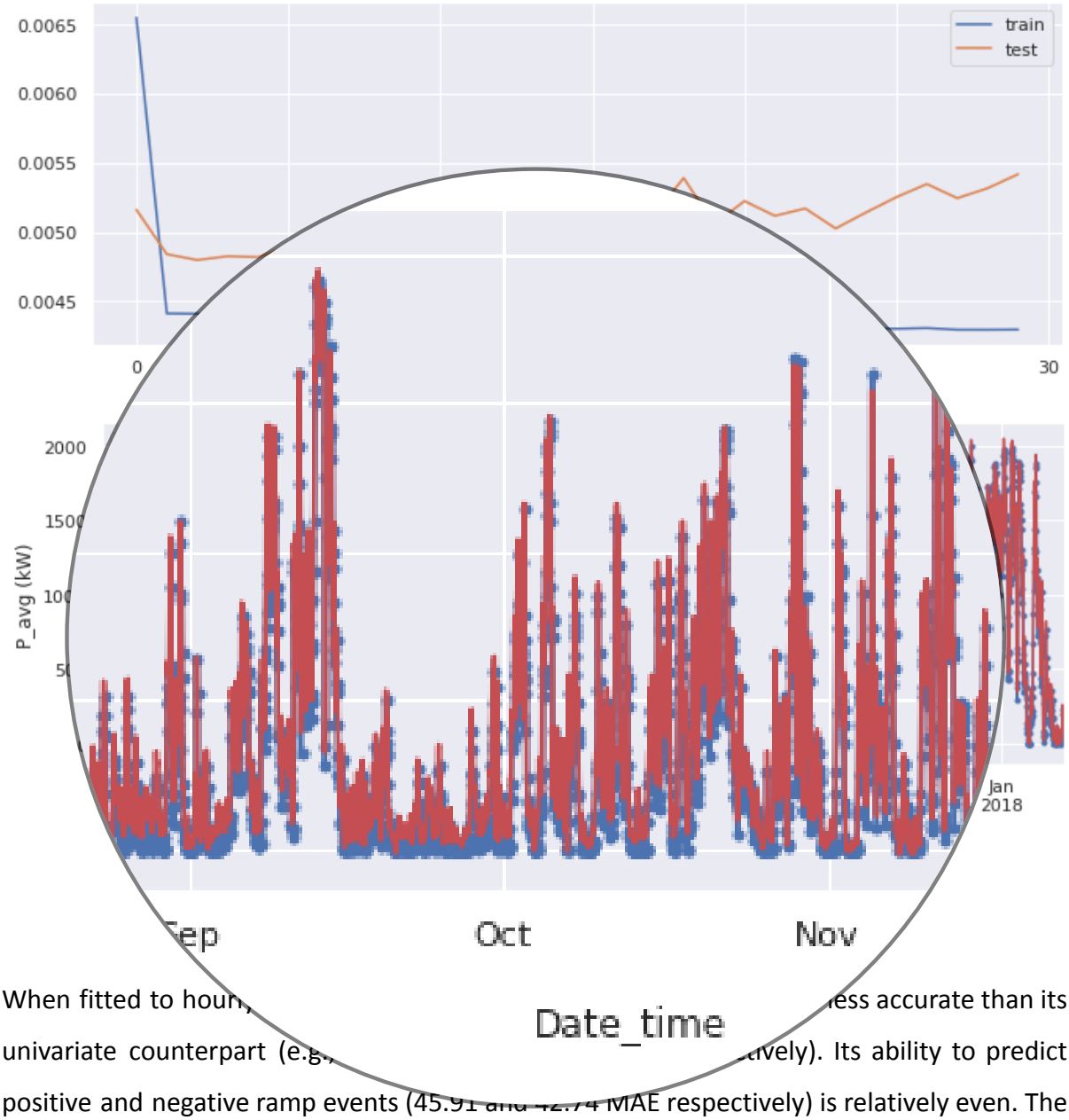
Univariate data



In terms of accuracy, forming model (e.g., non-ramp MAE of 15.00 kW accuracy of the LSTM-RNN's positive and negative ramp predictions are very even (e.g., 20.8 and 20.82 MAE respectively). The LSTM-RNN appears to capture outliers very well, however, the learning curves once again indicate overfitting and sub optimal adjustment.

Recurrent Neural Network

Multivariate data



When fitted to hour ess accurate than its univariate counterpart (e.g. tively). Its ability to predict positive and negative ramp events (45.91 and 42.74 MAE respectively) is relatively even. The test predictions show that the LSTM-RNN misses some of the extreme low values and the learning curves, again, indicate overfitting and sub optimal adjustment.

5 Evaluation and discussion

This section provides a critical and reflective review of the findings, learning, and quality of the research produced by this study, and how this thesis contributes to the initial research problem.

5.1 Findings and learning

5.1.1 EDA

The results obtained f… at the wind speed and power timeseries … t. To understand this, it is imp… d on Equation 17:

$$\gamma y_{t-1} + \delta \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_t$$

…he time trend
… autoregressive process
…pot test

The in… ionarity dependen… wer series over the per… icate a stark upwards trend … efficient, β, essentially make… the entire five years of data return… much more gradual trend, or indeed, no trend … ich it is observed. Fo… tion of relying on statistical tests alone and stresses the benefit of data visualisation.

5.1.2 ARMA and ARIMA

ARMA and ARIMA are discussed together due to the similarity of their results (Table 11). This similarity can be explained by the aforementioned apparent stationarity as measured by the ADF test: ARIMA is able to model trend by differencing the data to in order to make it stationary. However, as explained above, over the entire period of the dataset, the trend of the power series exerts only a negligible effect on the stationarity of the data and therefore

on the ARMA model. Thus, ARIMA is only able to offer slight improvements over ARMA. A discussion of SARIMA is presented in Appendix E.

5.1.3 Prophet

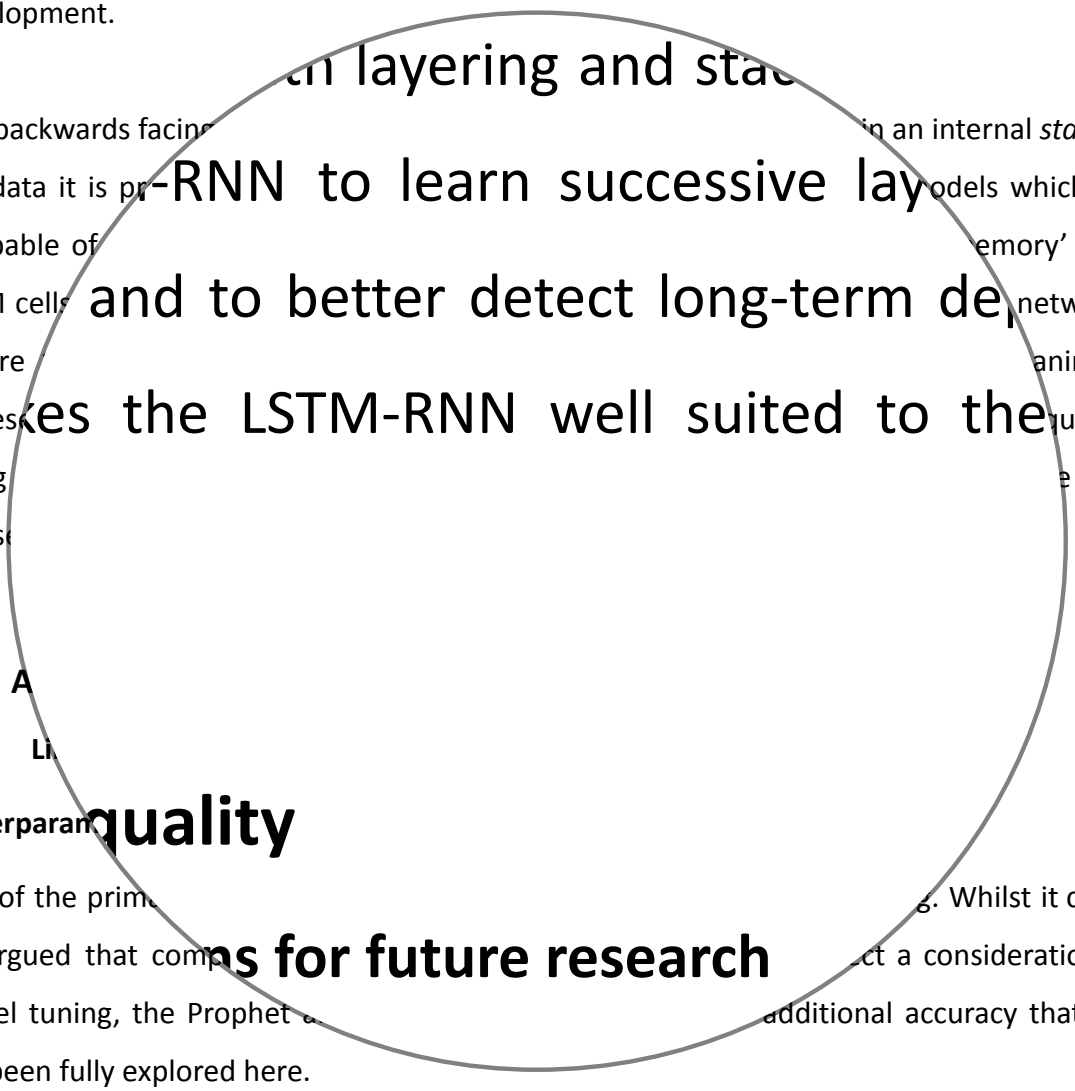
Although Prophet is advertised as 'typically able to handle outliers well' (Facebook Open Source, 2017), its inability to handle outliers in the LHB dataset was demonstrated in the poor ramp performance statistics of Table 11 and Figure 26, where outliers resulted in a widening of the model's uncert[illegible] the design of the model and the typical datasets upon [illegible] designed for use by Facebook analysts [illegible] oducts as outlined in Section 3.5 [illegible] em with trend changes. W[illegible] y assuming that the [illegible] s seen in the pas[illegible] are not expect[illegible] or and Letha[illegible]

Proph[illegible] of the numbe[illegible] cebook Open So[illegible] nt of the 'expected' [illegible] d adjusting NWP outputs [illegible] uce additional error since neith[illegible] remain constant. Finally, since outliers a[illegible] dicting ramp events, the option of removal is not appr[illegible] d necessit[illegible] the least-suited model to the characterisation of ramps in the LHB dataset.

5.1.4 LSTM-RNN

In terms of accuracy, the LSTM-RNN is the best performing model (Table 11). It has been shown to predict wind power and all ramp event types most accurately. However, when using default settings, overfitting was evident in every use case. None of the other algorithms are capable of modelling the multivariate nature of the LHB data which the

design of the RNN, as outlined in Section 3.5.4, enables it to do. Modelling of multivariate data improved accuracy using 10-minute data, but not when using hourly data. This has implications for the planned use of (hourly) NWP outputs and will require analysis prior to further model development. Possible reasons (and lines of future investigation) may include the resampling and imputation methods used, or the differences in ramp frequencies between the datasets. Despite this, the RNN shows the greatest potential for further development.



The backwards facing in an internal *state* of the data it is pr odels which are incapable of emory' with LSTM cells networks (Figure aningful repres quence being e LHB datas

5.2 A

5.2.1 Li

Hyperparam

One of the prim g. Whilst it could be argued that com ct a consideration of model tuning, the Prophet a additional accuracy that has not been fully explored here.

One of the tuneable parameters that Prophet offers relates to the number of changepoints used when forecasting. As discussed in Section 5.1.3 however, this is unlikely be of benefit when predicting future changes in wind speed or power output. A second tuneable parameter that Prophet offers is the number of terms used in the derivation of the Fourier sum (Section 3.5.3, Figure 13) to determine how quickly the seasonality can change.

Prophet's default values are based on data with yearly and weekly seasonality. Increasing this parameter allows for fitting seasonal patterns that change more quickly but with an increased risk of overfitting (Taylor and Letham, 2017). This is indeed reflected in the results of cursory experimentation with seasonality tuning where negligible gains in accuracy are made at the expense of increased disparities between training and testing accuracies (Table 13).

Table 13. Results of Prophet seasonality tuning

Model	Data sample rate	Data selection	Seasonality Tuning	Negative ramp acc. (RMSE)	Negative ramp acc. (MAE)	Non-ramp acc. (RMSE)	Non-ramp acc. (MAE)
Prophet	Hourly	Univari				422.49	299.8
Prophet	Hourly	Uni				86.21	280.17
Prophet	Hourly	U				.82	320.31

ch larger amount of tuning fle

be considered one of their primary

Neural MA and

Prophe and the outputs of the models are hig many

hyper . Even

in a s ustable parameters include the numbe er of

neuro ic etc.

The m ivation function for each layer, the we te (see

Appendi training.

These stra arameter in a neural network, howe is study, a

power sche on with the

learning rate was rategies to establish a suitab

n as *learning sch*

[16] For further detail, the reader is referred to Gerón (2019), Chapter 10.

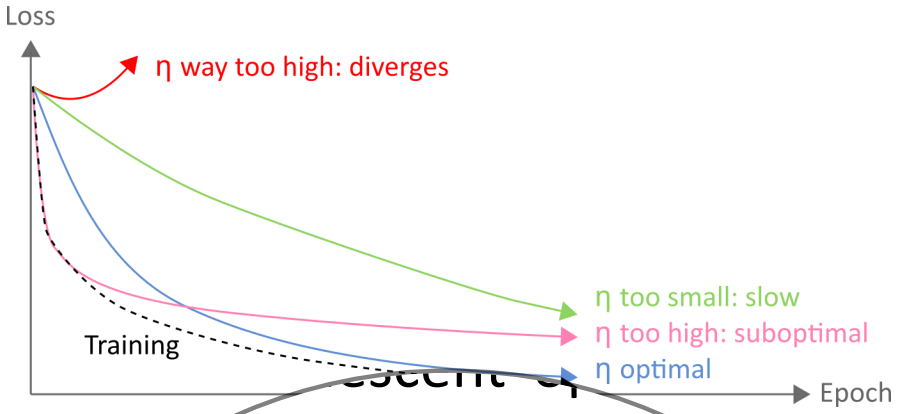


Figure 34 [illegible] (2019).

Modern deep [illegible] dy employ a backpropag[illegible] therefore stochasti[illegible] al model should [illegible] a given datase[illegible] pping functi[illegible] nction minim[illegible] nce is measu[illegible] ariate LSTM-R[illegible] ntuned and tune[illegible]

Table 14. Resu[illegible]

Model	Data sample rate	Dat[illegible] selecti[illegible]			Non-ramp acc. (RMSE)	Non-ramp acc. (MAE)
RNN	Hourly	Univariate	0.01		38.03	27.62	20.39
RNN	Hourly	Univariate	0.04		13.18	11.94	10.15

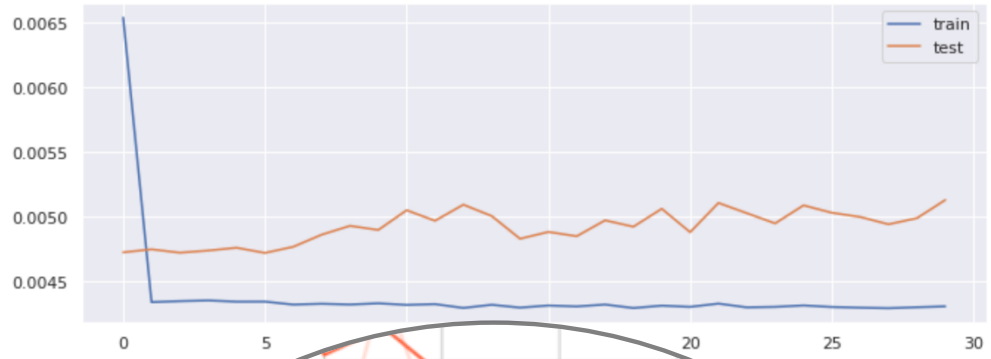
The results show the LSTM-RNN's sensitivity to the learning rate parameter and reflect the need for rigorous systematic tuning experimentation if accuracy limits are to be tested. They also demonstrate the potential margin for model improvement left unexplored by this study. Although the demonstrated increase in accuracy is significant, tuning experiments are temporally and computationally demanding and are therefore beyond the scope of the

[illegible] be averaged over various [illegible] one that is the best performing [illegible] ent with learning rate tunings, the ke[illegible] pirically determine the best and most st[illegible] resources by stopping training once [illegible] the learning rate experiments condu[illegible] a are shown in Figure 35. The acc[illegible] in Table 14.

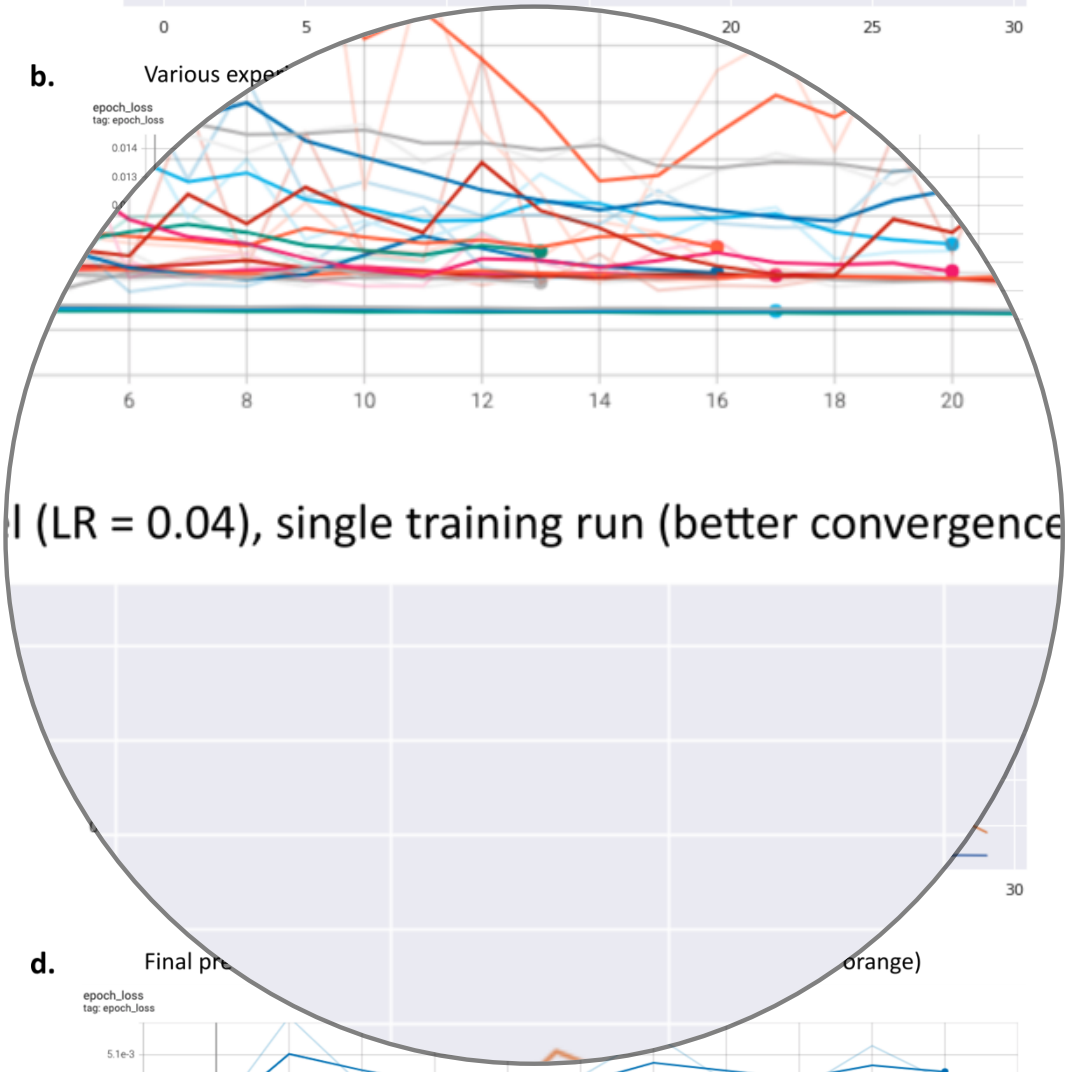
[17] Introduced in Appendix D.

current study. The author would hope to include such experimentation, as well as model generalisation testing (i.e., using different datasets) in future research.

a. Untuned (default LR = 0.001), single training run



b. Various expe



d. Final pr orange)

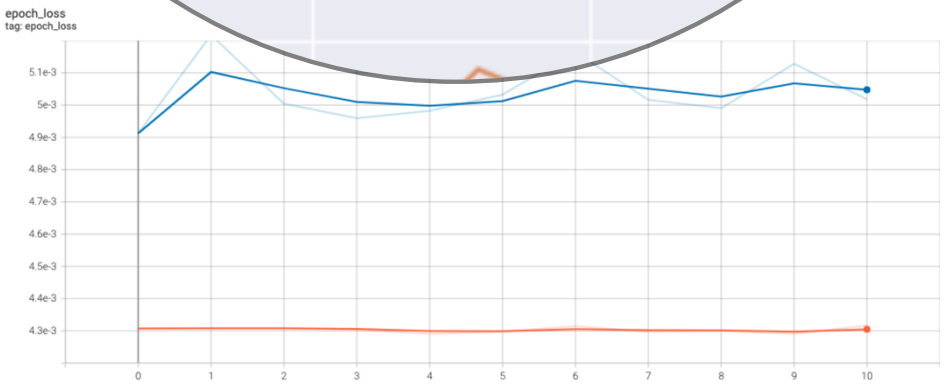


Figure 35. LSTM-RNN learning curves for hourly, univariate LSTM-RNN. LR = Learning Rate.

Advanced partitioning and roll forward validation

Throughout this study, the predictive accuracy of each model has been evaluated by partitioning the data into fixed-size training and tests. This approach permits only a limited evaluation of predictive performance since only one single future time step is forecast at a time. An alternative approach would be to use a roll-forward validation period which involves creating multiple train-test [illegible] partitioning one period at a time. This methodology would [illegible] scenario in which ramp forecasts were refr[illegible]

NWP outputs

The need [illegible] breakdown progressively with each [illegible]st clearly demonst[illegible]wer and its lagg[illegible]ch that inclusion of NWP outputs can [illegible]chnical design[illegible]ted in subse[illegible] respect, an additional feature that c[illegible]prove trainin[illegible]ty is a fundar[illegible]d to a more deployable model could be [illegible]nd the inclusio[illegible] MAE in wind reso[illegible]f Equation 9, used to convert win[illegible]

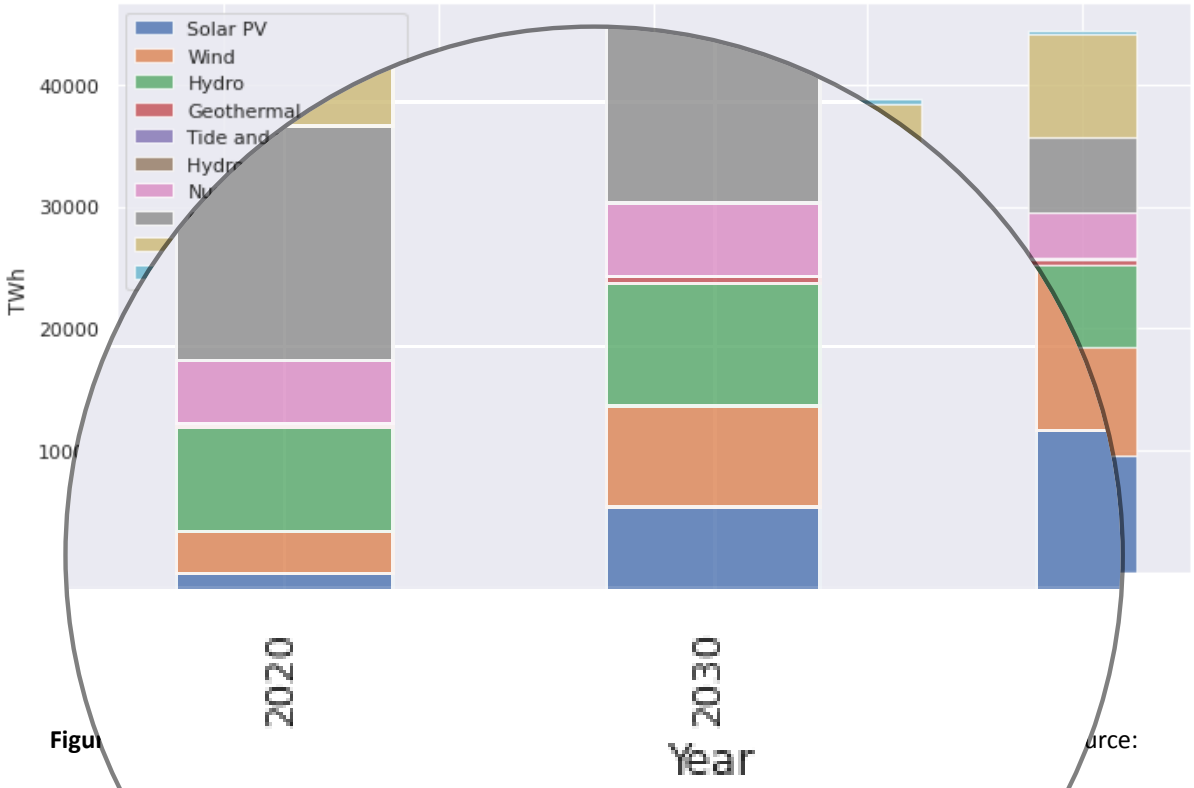
Reflection

[illegible]ates of air density has recently [illegible]

On reflection of [illegible]as spent running pre-processing experi[illegible]d Nielsen, 2019). [illegible]or ultimately, wholly necessary (e.g., running se[illegible]riments to gain only small improvements in accuracy). A more effective approach might have been to spend more time investigating additional multivariate models or feature engineering. Interesting results could have been obtained by expanding the family of models to include Genetic Algorithms as applied by Martínez-Arellano (2015), Vector ARIMA, or RNN variations such as GRU cells for example. The application of ensembles, an approach recently shown to increase predictive performance (e.g., Liu et al., 2019; Dhiman et al., 2020), could also have been investigated.

5.3 Research contributions

This thesis has contributed new knowledge to the field of ramp event forecasting which is currently an area of active research due its importance in the growing integration of wind energy into electricity markets worldwide (Figure 36).



Figu urce:

As the wind i ng installed in large farms where er to the grid. This will especially be the ca as part of its Renewable Energy Directive, the European Co investments equating to installed capacity of between 240 and 450 GW of offshore wind power will be needed by 2050 to keep global temperature rises below 1.5°C (European Commission, 2021). As wind farm sizes and capacities increase, the effects of sudden changes in wind speed on power output and grid stability become even more significant. This is especially true during ramp events.

While conventional power plants have a reliable production capacity per hour, wind farms are inextricably linked to the natural variability of the wind. Therefore, it is not possible to

treat wind power as a conventional energy source. If wind power is to be integrated into the electricity grid as per any other source, it must be considered during the operational procedures of the electricity market, which involves making decisions based on future demand and production. Until such time as new energy storage technologies or alternative innovations are developed, this requires a wind power forecast.

One such alternative innovation currently being developed by turbine manufacturing giant, Siemens Gamesa (SG), is that of *green hydrogen*: At its Brande Hydrogen pilot test site, SG has coupled an existing onshore wind turbine with an electrolyser stack to investigate the production of hydrogen from the electrolysis of water. The company aims to use commercial volumes of hydrogen to decarbonise refining processes and ammonia production. SG is also using the Brande Hydrogen site to explore whether integrating new battery technology can contribute to grid stability and help address issues around wind variability. Batteries can store energy that allows electrolysers to run for longer and produce more hydrogen, or they can distribute energy to the grid to help ease supply bottlenecks (Siemens Gamesa Renewable Energy, 2021). Until such alternatives become mainstream, however, management of energy supply through improved forecasting remains a cost-effective solution to the problem of wind power integration.

Using current practices, wind farm costs arising from imbalances between contracted and produced energy are directly proportional to forecast errors (Girard et al., 2013). This relationship was quantified in the case of a Dutch wind farm where improved forecasting was shown to reduce annual (regulatory) costs by 39% (Pinson et al., 2017). Currently, state-of-the-art wind power forecasting systems can achieve a RMSE of 10-15% of total installed capacity over a 36-hour horizon (Martínez-Arellano, 2015; Giebel, G. et al., 2011), thus representing a target for continued development. It is evident then, that improvements in wind power forecasting capability such as those targeted by this work have and will continue to drive down costs of wind power. Cost reductions will play an essential role in improving the competitiveness of wind energy against non-renewable energy sources, enabling a global transition to cleaner fuels, and facilitating climate change mitigation (Figure 36).

Motivated by this, the work presented in this thesis provides precursory information that will inform planned further research aiming to improve ramp event forecasting accuracy. This information takes the form of:

5.3.1 Statistical analysis of a typical wind farm dataset

A major contribution of this research was the thorough statistical analysis of the LHB dataset. Most published work on ML ramp event forecasting presents model performance results with little, if any, discussion of the links between the ML methods and the datasets used. When applying any ML model to a given task, it is important to match as many of the characteristics of the problem as possible to the ML model (Brownlee, 2019). The results showed that the LHB dataset exhibits:

- An autoregressive character that can be modelled over the very short-term
- Multiple, complex trends and periodicities (seasonalities) that are challenging to model
- Many outlying values and signal noise that adversely affects the modelling of trends

These results can also be used to inform future research that focuses on similar tasks using similar datasets.

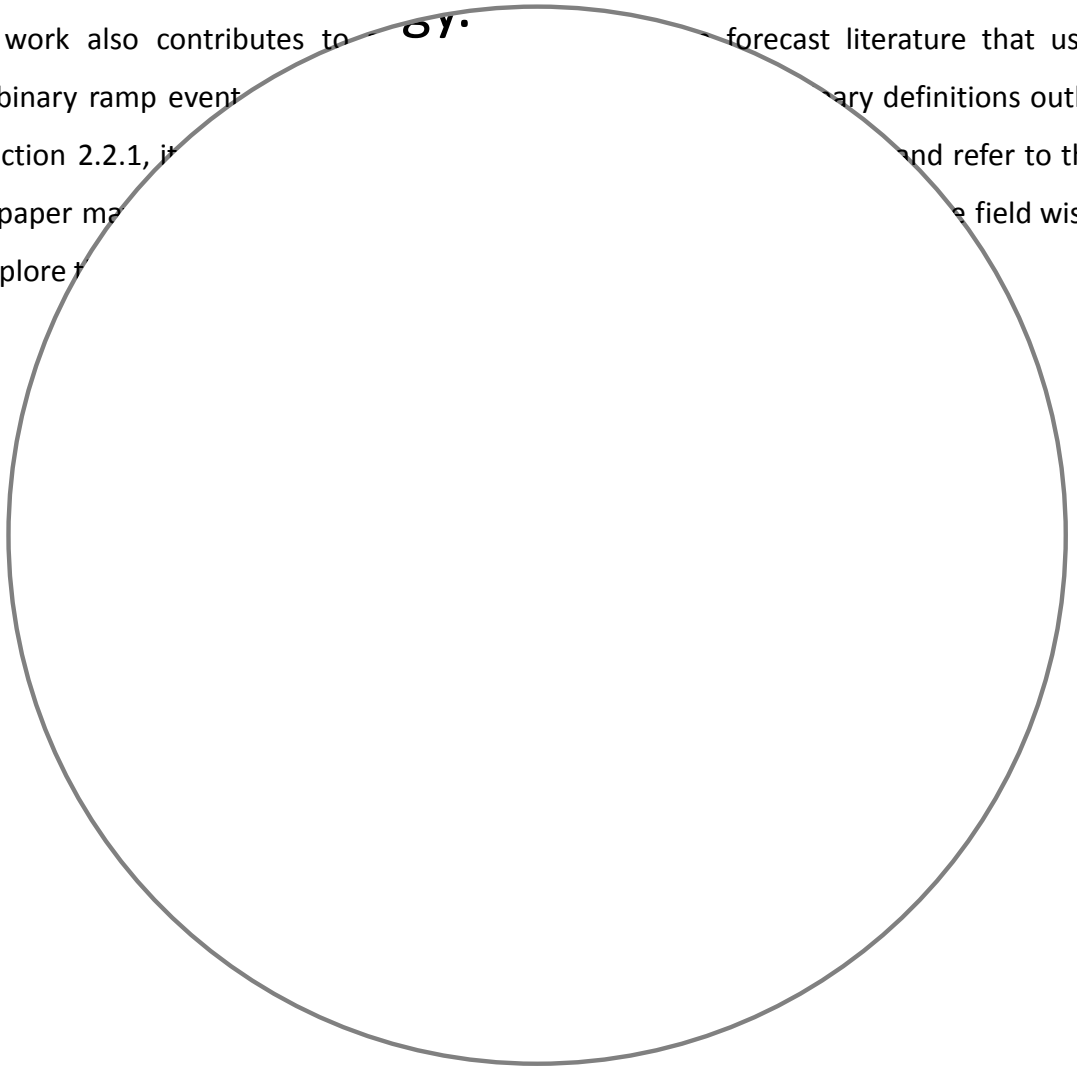
5.3.2 Model comparisons

A second contribution of this work was the consideration of the strengths and weaknesses of the family of ML models and their applicability to the modelling of a typical wind farm dataset. All model run times were below the 1.5 hour target time set out in Section 1.5 and are therefore considered successful in terms of computation time. The results show the potential for LSTM-RNNs to outperform all other models on the task of ramp event prediction. Prophet was shown to be the model least capable of predicting ramp events. The way in which Prophet models trend and seasonality (Section 3.5.3) coupled with the complex trends and periodicities in the LHB data (e.g., Figure 22, Section 4.1.4) was proposed as the reason for this. ARMA and ARIMA were shown to be the next best performing models. The autoregressive methodology that ARMA/ARIMA uses to model time-series data (Section 3.5.1), combined with the very short-term autocorrelative nature of the LHB data (Section 4.1.3) was put forward to explain this. The LSTM-RNN was the model most capable of predicting ramp events. The LSTM-RNN's ability to maintain an internal state or model of the information being processed, coupled with the complexities of

the LHB dataset was the suggested explanation for this. In addition, some interesting nuances of the models (e.g., more accurate modelling of negative ramps than positive ramps) have been revealed that could have implications for further ML-based ramp modelling. The results of this study provide scientific grounds on which model selection in further research can be based.

5.3.3 Non-binary ramp characterisation methodology

This work also contributes to [illegible] forecast literature that uses a non-binary ramp event[illegible]ary definitions outlined in Section 2.2.1, it[illegible]nd refer to them. This paper ma[illegible]e field wishing to explore t[illegible]



6 References

Abbasi, S.A., Tabassum-Abbasi and Abbasi, T. (2016) Impact of wind-energy generation on climate: A rising spectre. *Renewable and Sustainable Energy Reviews*. [online]. 59, pp.1591–1598.

Addison, P.S. (2017) *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance.* CRC press.

Aguilar, T.A. (2019) *Detecting the long-term frequency of large-scale wind power ramp events observed in ERCOT's aggregated wind power time-series data.* Thesis [online]. Available from: https://ttu-ir.tdl.org/handle/2346/85506 [Accessed 12 October 2021].

Ahmadi, M. and Khashei, M. (2021) Current status of hybrid structures in wind forecasting. *Engineering Applications of Artificial Intelligence* [online]. 99, Pergamon, p.104-133.

Akansu, A.N., Serdijn, W.A. and Selesnick, I.W. (2010) Emerging applications of wavelets: A review. *Physical Communication* [online]. 3 (1), pp.1–18.

American Meteorological Society (2020) *Glossary of Meteorology* 2020 [online]. Available from: https://glossary.ametsoc.org/wiki/Welcome [Accessed 27 September 2021].

Amjady, N. and Abedinia, O. (2017) Short Term Wind Power Prediction Based on Improved Kriging Interpolation, Empirical Mode Decomposition, and Closed-Loop Forecasting Engine. *Sustainability*. [online]. 9 (11), Multidisciplinary Digital Publishing Institute, p.2104.

Amolins, K., Zhang, Y. and Dare, P. (2007) Wavelet based image fusion techniques — An introduction, review and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*. [online]. 62 (4), pp.249–263.

Aoki, I. et al. (2014) Development and Operational Status of Wind Power Forecasting System. *Electrical Engineering in Japan*. [online]. 189 (4), pp.22–29.

Bessa, R.J. et al. (2011) 'Good' or 'bad' wind power forecasts: a relative concept. *Wind Energy*. [online]. 14 (5), pp.625–636.

Bianco, L. et al. (2016) A Wind Energy Ramp Tool and Metric for Measuring the Skill of Numerical Weather Prediction Models. *Weather and Forecasting*. [online]. 31 (4), American Meteorological Society, pp.1137–1156.

Bossavy, A., Girard, R. and Kariniotakis, G. (2015) An edge model for the evaluation of wind power ramps characterization approaches. *Wind Energy*. 18 (7), Wiley Online Library, pp.1169–1184.

Bossavy, A., Girard, R. and Kariniotakis, G. (2013) Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*. [online]. 16 (1), pp.51–63.

Bossavy, A., Girard, R. and Kariniotakis, G. (2012) Forecasting Uncertainty Related to Ramps of Wind Power Production. p.10.

Box, G.E. et al. (2015) *Time series analysis: forecasting and control* John Wiley & Sons.

Bradford, K.T., Carpenter, R. and Shaw, B. (2010) *Forecasting southern plains wind ramp events using the WRF model at 3-km.* AMS Student Conference Citeseer.

Brownlee, J. (2018) *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python.* Machine Learning Mastery.

Brownlee, J. (2021) No Free Lunch Theorem for Machine Learning. *Machine Learning Mastery*. 16 February 2021 [online]. Available from: https://machinelearningmastery.com/no-free-lunch-theorem-for-machine-learning/ [Accessed 23 November 2021].

Cali, U. and Sharma, V. (2019) Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. *International Journal of Smart Grid and Clean Energy*. [online]. pp.103–110.

Canny, J. (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (6), pp.679–698.

Cheneka, B.R., Watson, S.J. and Basu, S. (2020) A simple methodology to detect and quantify wind power ramps. *Wind Energy Science*. [online]. 5 (4), Copernicus GmbH, pp.1731–1741.

Chollet, F. (2021) *Deep learning with Python.* Simon and Schuster.

Cornejo-Bueno, L. et al. (2017) Wind power ramp events prediction with hybrid machine learning regression techniques and reanalysis data. *Energies*. 10 (11), Multidisciplinary Digital Publishing Institute, p.1784.

Cornejo-Bueno, L. et al. (2020) Wind power ramp event detection with a hybrid neuro-evolutionary approach. *Neural Computing and Applications*. [online]. 32 (2), pp.391–402.

Correia, J.M. et al. (2017) The influence of the main large-scale circulation patterns on wind power production in Portugal. *Renewable Energy*. [online]. 102, pp.214–223.

Couto, A., Costa, P. and Simões, T. (2021) Identification of extreme wind events using a weather type classification. *Energies*. 14 (13), Multidisciplinary Digital Publishing Institute, p.3944.

Cui, M. et al. (2015) *An optimized swinging door algorithm for wind power ramp event detection.* In: *2015 IEEE Power Energy Society General Meeting*. [online]. 2015 IEEE Power Energy Society General Meeting 1–5.

Cui, Y. et al. (2021) Algorithm for identifying wind power ramp events via novel improved dynamic swinging door. *Renewable Energy* [online]. 171, pp.542–556.

Cutler, N. et al. (2007) Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy* [online]. 10 (5), pp.453–470.

Cutler, N.J. et al. (2009) Characterizing future large, rapid changes in aggregated wind power using numerical weather prediction spatial fields. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*. 12 (6), Wiley Online Library, pp.542–555.

Dhiman, H.S. and Deb, D. (2021) Machine intelligent and deep learning techniques for large training data in short-term wind speed and ramp event forecasting. *International Transactions on Electrical Energy Systems* [online]. 31 (9), e12818.

Dhiman, H.S., Deb, D. and Balas, V.E. (2020) *Supervised machine learning in wind forecasting and ramp event prediction.* Academic Press.

Dhiman, H.S., Deb, D., and Guerrero, J.M. (2019) Hybrid machine intelligent SVR variants for wind forecasting and ramp events. *Renewable and Sustainable Energy Reviews* [online]. 108, pp.369–379.

Domingues, M.O., Mendes, O. and da Costa, A.M. (2005) 'On wavelet techniques in atmospheric sciences' *Advances in Space Research*. *Fundamentals of Space Environment Science* [online]. 35 (5), pp.831–842.

Dorado-Moreno, M. et al. (2017) Robust estimation of wind power ramp events with reservoir computing. *Renewable Energy* [online]. 111, pp.428–437.

Dorado-Moreno, M. et al. (2018) Predicción ordinal de rampas de viento usando Echo State Networks de complejidad reducida. *Conferencia de la Asociacion Espanola para la Inteligencia Artificial*. XVIII, p.7.

Dorado-Moreno, M. et al. (2020) Multi-task learning for the prediction of wind power ramp events with deep neural networks. *Neural Networks*. 123, Elsevier, pp.401–411.

Engie (2020) *Engie OPEN data* [online]. Available from: https://opendata-renewables.engie.com/ [Accessed 18 September 2021].

European Centre for Medium-Range Weather Forecasts (2021) *ECMWF Advancing global NWP through international collaboration ECMWF* [online]. Available from: https://www.ecmwf.int/ [Accessed 18 September 2021].

European Commission (2019) *National energy and climate plans (NECPs) Energy* [online]. Available from: https://ec.europa.eu/energy/topics/energy-strategy/national-energy-climate-plans_en [Accessed 12 September 2021].

European Commission (2021) *Renewable energy directive Energy* [online]. Available from: https://ec.europa.eu/energy/topics/renewable-energy/directive-targets-and-rules/renewable-energy-directive_en [Accessed 13 September 2021].

Facebook Open Source (2017) *Prophet, Forecasting at scale* [online]. Available from: http://facebook.github.io/prophet/ [Accessed 12 December 2021].

Fernández, Á. et al. (2013) Diffusion Methods for Wind Power Ramp Detection. In: Rojas, I., Joya, G. and Gabestany, J. (eds.) *Advances in Computational Intelligence*. *Lecture Notes in Computer Science* [online]. Berlin, Heidelberg, Springer, 106–113.

Ferreira, C. et al. (2010) *A survey on wind power ramp forecasting. Technical report* [online]. Argonne National Laboratory. Available from: https://ceeesa.es.anl.gov/pubs/69166.pdf [Accessed 30 September 2021].

Floors, R. and Nielsen, M. (2019) Estimating air density using observations and re-analysis outputs for wind energy purposes. *Energies*. 12 (11), Multidisciplinary Digital Publishing Institute, p.2038.

Florita, A., Hodge, B.-M. and Orwig, K. (2013) *Identifying Wind and Solar Ramping Events.* In: *2013 IEEE Green Technologies Conference* [online]. IEEE, 147–152.

Froude, L.S.R., Bengtsson, L. and Hodges, K.I. (2007) The Prediction of Extratropical Storm Tracks by the ECMWF and NCEP Ensemble Prediction Systems. *Monthly Weather Review* [online]. 135 (7), American Meteorological Society, pp.2545–2567.

Fujimoto, Y., Takahashi, Y. and Hayashi, Y. (2019) Alerting to Rare Large-Scale Ramp Events in Wind Power Generation. *IEEE Transactions on Sustainable Energy* [online]. IEEE Transactions on Sustainable Energy 10 (1), pp.55–65.

Gaberšek, S. and Durran, D.R. (2004) Gap Flows through Idealized Topography. Part I: Forcing by Large-Scale Winds in the Nonrotating Limit. *Journal of the Atmospheric Sciences* [online]. 61 (23), pp.2846–2862.

Gallego-Castillo, C.J. (2013) *Statistical models for short-term wind power ramp forecasting.* PhD [online]. E.T.S.I. Aeronáuticos (UPM). Available from: https://oa.upm.es/21912/ [Accessed 23 September 2021].

Gallego-Castillo, C. et al. (2013) A wavelet-based approach for large wind power ramp characterisation. *Wind Energy*. 16 (2), pp.257–278.

Gallego-Castillo, C. et al. (2016) *Wind power probabilistic forecast in the Reproducing Kernel Hilbert Space.* In: *2016 Power Systems Computation Conference (PSCC)* [online]. 2016 Power Systems Computation Conference (PSCC) 1–7.

Gallego-Castillo, C., Cuerva-Tejero, A. and Lopez-Garcia, O. (2015) A review on the recent history of wind power ramp forecasting. *Renewable and Sustainable Energy Reviews* [online]. 52, pp.1148–1157.

Gensler, A. (2019) Wind Power Ensemble Forecasting [online]. Kassel University Press. Available from: https://www.uni-kassel.de/ub/index.php?id=39129&h=9783737606363 [Accessed 14 October 2021].

Géron, A. (2019) *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media.

Giebel, G. et al. (2011) The state-of-the-art in short-term prediction of wind power. A literature overview [online]. Report number NEI-DK-5521. ANEMOS.plus and SafeWind. Available from: https://www.osti.gov/etdeweb/biblio/1011554 [Accessed 1 January 2022].

Girard, R., Laquaine, K. and Kariniotakis, G. (2013) Assessment of wind power predictability as a decision factor in the investment phase of wind farms. *Applied Energy* [online]. 101, p.609.

González-Sopeña, J.M., Pakrashi, V. and Ghosh, B. (2021) An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renewable and Sustainable Energy Reviews* [online]. 138, p.110515.

Google (2021) *Google Colaboratory Frequently Asked Questions* [online]. Available from: https://research.google.com/colaboratory/faq.html [Accessed 30 December 2021].

Greaves, B. et al. (2009) Temporal Forecast Uncertainty for Ramp Events. *Wind Engineering* [online]. 33 (4), SAGE Publications, pp.309–319.

Grossmann, A. (1988) 'Wavelet Transforms and Edge Detection'. In: Albeverio, S. et al. (eds.) *Stochastic Processes in Physics and Engineering*. Mathematics and Its Applications [online]. Dordrecht, Springer Netherlands, 149–157. Available from: https://doi.org/10.1007/978-94-009-2893-0_7 [Accessed 29 December 2021].

Gualtieri, G. (2021) Reliability of ERA5 Reanalysis Data for Wind Resource Assessment: A Comparison against Tall Towers. *Energies* [online]. 14 (14). Multidisciplinary Digital Publishing Institute. p.4169.

Gutiérrez, A., Porrini, C. and Fovell, R.G. (2020) Combination of wind gust models in convective events. *Journal of Wind Engineering and Industrial Aerodynamics* [online]. 199, p.104118.

Hannesdóttir, Á. and Kelly, M. (2019) Detection and characterization of extreme wind speed ramps. *Wind Energy Science* [online]. 4 (3), Copernicus GmbH, pp.385–396.

Hannesdóttir, Á., Kelly, M. and Dimitrov, N. (2019) Extreme wind fluctuations: joint statistics, extreme turbulence, and impact on wind turbine loads. *Wind Energy Science* [online]. 4 (2), Copernicus GmbH, pp.325–342.

Heckenbergerova, J., Musilek, P. and Janata, M. (2016) *Sensitivity analysis of PCA method for wind ramp event detection.* 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC) IEEE, 1–4.

Hersbach, H. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* [online]. 146 (730), pp.1999–2049.

Hirata, Y. et al. (2021) Forecasting wind power ramps with prediction coordinates. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 31 (10), AIP Publishing LLC, p.103105.

Hong, T., Pinson, P. and Fan, S. (2014) Global Energy Forecasting Competition 2012. *International Journal of Forecasting* [online]. 30 (2), pp.357–363.

Institut National De L'Information Géographique et Forestière (2021) *République Française | Géoservices*2021 [online]. Available from: https://geoservices.ign.fr/ [Accessed 4 December 2021].

Intel (2021) *What Is a GPU? Graphics Processing Units Defined Intel*.2021 [online]. Available from: https://www.intel.com/content/www/uk/en/products/docs/processors/what-is-a-gpu.html [Accessed 10 December 2021].

International Energy Agency (2021), *World Energy Outlook 2021*, IEA. Licence: Creative Commons Attribution CC BY-NC-SA 3.0 IGO.

Ji, F., Cai, X. and Zhang, J. (2015) Wind power prediction interval estimation method using wavelet-transform neuro-fuzzy network. *Journal of Intelligent & Fuzzy Systems* [online]. 29 (6), IOS Press, pp.2439–2445.

Kariniotakis, G. et al. (2004) What performance can be expected by short-term wind power prediction models depending on site characteristics. In *CD-ROM proceedings of the European wind energy conference EWEC*.

Khorramdel, B. et al. (2020) A Generic Convex Model for a Chance-Constrained Look-Ahead Economic Dispatch Problem Incorporating an Efficient Wind Power Distribution Modeling. *IEEE Transactions on Power Systems* [online]. IEEE Transactions on Power Systems 35 (2), pp.873–886.

Krusche, N. and De Oliveira, A.P. (2004) Characterization of Coherent structures in the Atmospheric Surface Layer. *Boundary-Layer Meteorology* [online]. 110 (2), pp.191–211.

Lacerda, M., Couto, A. and Estanqueiro, A. (2017) Wind Power Ramps Driven by Windstorms and Cyclones. *Energies* [online]. 10 (10), Multidisciplinary Digital Publishing Institute, p.1475.

Landberg, L. (2015) *Meteorology for wind energy: an introduction* John Wiley & Sons.

Lee, G.R. et al. (2019) PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software* [online]. 4 (36), p.1237.

Leung, A.K., Chau, F. and Gao, J. (1998) A review on applications of wavelet transform techniques in chemical analysis: 1989–1997. *Chemometrics and Intelligent Laboratory Systems* [online]. 43 (1), pp.165–184.

Li, Y., Musilek, P. and Lozowski, E. (2017) Improving the prediction of wind power ramps using texture extraction techniques applied to atmospheric pressure fields. *International Journal of Data Science and Analytics*. 4 (4), Springer, pp.237–250.

Liu, H. et al. (2019a) Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. *Renewable energy*. 143, Elsevier, pp.842–854.

Liu, Y. et al. (2019b) Quantitative method for evaluating detailed volatility of wind power at multiple temporal-spatial scales. *Global Energy Interconnection* [online]. 2 (4), pp.318–327.

Liu, Z. et al. (2018) Novel forecasting model based on improved wavelet transform, informative feature selection, and hybrid support vector machine on wind power forecasting. *Journal of Ambient Intelligence and Humanized Computing* [online]. 9 (6), pp.1919–1931.

López, E. et al. (2020) *Comparison of Recurrent Neural Networks for Wind Power Forecasting.* In: Figueroa Mora, K.M. et al. (eds.) *Pattern Recognition*. Cham, Springer International Publishing, 25–34.

Lyners, D., Vermeulen, H.J. and Groch, M. (2020) *A Multi-Parameter Algorithm for Wind Power Ramp Detection.* In: *2020 2nd International Conference on Smart Power Internet Energy Systems (SPIES)* [online]. 150–155.

Lyners, D., Vermeulen, H. and Groch, M. (2021) Wind power ramp event detection using a multi-parameter segmentation algorithm. *Energy Reports* [online]. 7, pp.5536–5548.

Ma, H. et al. (2019) An Early Warning Method Based on Improved Affine Arithmetic for Wind Power Ramp Events in Power System. *IOP Conference Series: Earth and Environmental Science*. [online]. 384, p.012199.

Madsen, H. et al. (2005) Standardizing the performance evaluation of short-term wind power prediction models. *Wind engineering*. 29 (6), SAGE Publications Sage UK: London, England, pp.475–489.

Mallat, S. and Hwang, W.L. (1992) Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory* [online]. IEEE Transactions on Information Theory 38 (2), pp.617–643.

Martín Martínez, S. (2013) Análisis y regulación de fluctuaciones de potencia en parques eólicos [online]. Sergio Martín Martínez. Available from: https://repositorio.upct.es/handle/10317/3451 [Accessed 11 June 2021].

Martínez-Arellano, G. et al. (2014) Characterisation of large changes in wind power for the day-ahead market using a fuzzy logic approach. *KI-Künstliche Intelligenz*. 28 (4), Springer, pp.239–253.

Martínez-Arellano, G. (2015) *Forecasting wind power for the day-ahead market using numerical weather prediction models and computational intelligence techniques*. PhD [online]. Nottingham Trent University. Available from: http://irep.ntu.ac.uk/id/eprint/322/ [Accessed 24 June 2021].

Martínez-Arellano, G. and Nolle, L. (2013) *Genetic Programming for Wind Power Forecasting and Ramp Detection.* In: Bramer, M. and Petridis, M. (eds.) *Research and Development in Intelligent Systems XXX* [online]. Cham, Springer International Publishing, 403–417.

Menezes, D. et al. (2020) Wind Farm and Resource Datasets: A Comprehensive Survey and Overview. *Energies* [online]. 13 (18), Multidisciplinary Digital Publishing Institute, p.4702.

Météo-France (2021) *Météo-France - Prévisions gratuites à 15 jours sur la France et sur le monde*2021 [online]. Available from: https://meteofrance.com/ [Accessed 18 September 2021].

Milidiú, R.L., Machado, R.J. and Renteria, R.P. (1999) Time-series forecasting through wavelets transformation and a mixture of expert models. *Neurocomputing*. 28 (1–3), Elsevier, pp.145–156.

Miller, L.M. and Keith, D.W. (2018) Climatic Impacts of Wind Power. *Joule* [online]. 2 (12), pp.2618–2632.

Mishra, S., Leinakse, M. and Palu, I. (2017) 'Wind power variation identification using ramping behavior analysis' *Energy Procedia*. *Power and Energy Systems Engineering* [online]. 141, pp.565–571.

Monteiro, C. et al. (2009) *Wind power forecasting: State-of-the-art 2009* [online]. Argonne National Lab (ANL), Argonne, IL (United States). Available from: https://publications.anl.gov/anlpubs/2009/11/65614.pdf.

Morlet, J. et al. (1982a) Wave propagation and sampling theory—Part I: Complex signal and scattering in multi-layered media. *GEOPHYSICS* [online]. 47 (2), Society of Exploration Geophysicists, pp.203–221.

Morlet, J. et al. (1982b) Wave propagation and sampling theory—Part II: Sampling theory and complex waves. *GEOPHYSICS* [online]. 47 (2), Society of Exploration Geophysicists, pp.222–236.

NASA (2017) *Scientific Visualization Studio: Global Surface- and Upper-Level Winds*29 August 2017 [online]. Available from: https://svs.gsfc.nasa.gov/4571 [Accessed 7 December 2021].

Nielsen, T. et al. (2006) *Short-term wind power forecasting using advanced statistical methods.* The European Wind Energy Conference, EWEC 2006 9-pages.

Olauson, J. (2018) ERA5: The new champion of wind power modelling? *Renewable Energy* [online]. 126, pp.322–331.

Ortega-Vazquez, M.A. and Kirschen, D.S. (2010) Assessing the impact of wind power generation on operating costs. *IEEE Transactions on Smart Grid*. 1 (3), IEEE, pp.295–301.

Ouyang, T. et al. (2019) Prediction of wind power ramp events based on residual correction. *Renewable Energy* [online]. 136, pp.781–792.

Pedregosa, F. et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12 (85), pp.2825–2830.

Pereyra-Castro, K. et al. (2020) Wind and Wind Power Ramp Variability over Northern Mexico. *Atmosphere* [online]. 11 (12), Multidisciplinary Digital Publishing Institute, p.1281.

Pereyra-Castro, K., Caetano Neto, E.D.S. and Martinez-Alvarado, O. (2019) Wind Power Ramp Characteristics over Northern Mexico. AGU Fall Meeting Abstracts 2019, pp.GC53F-1177.

Pichault, M. et al. (2021) Characterisation of intra-hourly wind power ramps at the wind farm scale and associated processes. *Wind Energy Science*. 6 (1), Copernicus GmbH, pp.131–147.

Pinson, P., Chevallier, C. and Kariniotakis, G.N. (2007) Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*. 22 (3), IEEE, pp.1148–1156.

Potter, C.W., Grimit, E. and Nijssen, B. (2009) *Potential benefits of a dedicated probabilistic rapid ramp event forecast tool.* In: *2009 IEEE/PES Power Systems Conference and Exposition* [online]. 2009 IEEE/PES Power Systems Conference and Exposition 1–5.

Rose, S. and Apt, J. (2016) Quantifying sources of uncertainty in reanalysis derived wind speed. *Renewable Energy* [online]. 94, pp.157–165.

Ruiz, C., Alaíz, C.M. and Dorronsoro, J.R. (2020) Multitask Support Vector Regression for Solar and Wind Energy Prediction. *Energies* [online]. 13 (23), Multidisciplinary Digital Publishing Institute, p.6308.

Seabold, S. and Perktold, J. (2010) *Statsmodels: Econometric and Statistical Modeling with Python.* In: *Proceedings of the 9th Python in Science Conference* [online]. Python in Science Conference Austin, Texas, 92–96. Available from: https://conference.scipy.org/proceedings/scipy2010/seabold.html [Accessed 23 December 2021].

Senvion GmbH (2018) *The MM82. Optimized for high wind locations.* 4 October 2021.

Sevlian, R. and Rajagopal, R. (2013) Detection and Statistics of Wind Power Ramps. *IEEE Transactions on Power Systems* [online]. IEEE Transactions on Power Systems 28 (4), pp.3610–3620.

Sherry, M. and Rival, D. (2015) Meteorological phenomena associated with wind-power ramps downwind of mountainous terrain. *Journal of Renewable and Sustainable Energy*. 7 (3), AIP Publishing LLC, p.033101.

Shi, Z., Liang, H. and Dinavahi, V. (2018) Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. *IEEE Transactions on Sustainable Energy* [online]. IEEE Transactions on Sustainable Energy 9 (3), pp.1177–1187.

Shmueli, G. and Lichtendahl Jr, K.C. (2016) *Practical time series forecasting with R: A hands-on guide.* Axelrod Schnall Publishers.

Shumway, R.H. and Stoffer, D.S. (2000) *Time series analysis and its applications* 3. Springer.

Sim, M.K. and Jung, J.-Y. (2020) *A Short Review on Predictions for Wind Power Generation – Its Limitation and Future Directions* [online]. ICIC International. Available from: https://doi.org/10.24507/icicelb.11.10.995 [Accessed 12 October 2021].

Siemens Gamesa Renewable Energy (2021) White Paper, Unlocking the Green Hydrogen Revolution [online]. Available from:

https://www.siemensgamesa.com/en-int/-/media/whitepaper-unlocking-green-hydrogen-revolution.pdf [Accessed 2 January 2022].

Skander, H. (2020) *Inference in Time Series: Prophet vs. ARIMA Stack Exchange* [online]. Available from: https://stats.stackexchange.com/q/472622 [Accessed 15 November 2021].

Smith, T., G. (2017) *pmdarima: ARIMA estimators for Python* [online]. Available from: https://alkaline-ml.com/pmdarima/ [Accessed 16 November 2021].

Soares, P.M.M. et al. (2014) Climatology of the Iberia coastal low-level wind jet: weather research forecasting model high-resolution results. *Tellus A: Dynamic Meteorology and Oceanography* [online]. 66 (1), p.22377.

Sotavento Galicia SA (2020) *Sotavento – Parque Eólico Experimental* [online]. Available from: http://www.sotaventogalicia.com/ [Accessed 18 September 2021].

Steiner, A. et al. (2017) Critical weather situations for renewable energies – Part A: Cyclone detection for wind power. *Renewable Energy* [online]. 101, pp.41–50.

Sweeney, C. et al. (2020a) The future of forecasting for renewable energy. *Wiley Interdisciplinary Reviews: Energy and Environment*. 9 (2), Wiley Online Library, p.365.

Sweeney, C. et al. (2020b) The future of forecasting for renewable energy. *WIREs Energy and Environment* [online]. 9 (2), p.e365.

Szilagyi, J. et al. (1999) An objective method for determining principal time scales of coherent eddy structures using orthonormal wavelets. *Advances in Water Resources* [online]. 22 (6), pp.561–566.

Taylor, S.J. and Letham, B. (2017) *Forecasting at scale* [online]. PeerJ Preprints. Available from: https://peerj.com/preprints/3190 [Accessed 11 November 2021].

VanderPlas, J. (2016) *Python data science handbook: Essential tools for working with data* O'Reilly Media, Inc.

Veron, D.E. et al. (2018) Modeling the electrical grid impact of wind ramp-up forecasting error offshore in the Mid-Atlantic region. *Journal of Renewable and Sustainable Energy* [online]. 10 (1), American Institute of Physics, p.013308.

Wang, H. et al. (2019) A review of deep learning for renewable energy forecasting. *Energy Conversion and Management* [online]. 198, p.111799.

Wang, L. et al. (2021) Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. *Renewable Energy* [online]. 164, pp.642–655.

Wang, Q. et al. (2016) The value of improved wind power forecasting: Grid flexibility quantification, ramp capability analysis, and impacts of electricity market operation timescales. *Applied Energy* [online]. 184, pp.696–713.

Watanabe, T. (2004) Large-Eddy Simulation of Coherent Turbulence Structures Associated with Scalar Ramps Over Plant Canopies. *Boundary-Layer Meteorology* [online]. 112 (2), pp.307–341.

Wikipedia (2021a) *Atmospheric circulation Wikipedia* [online]. Available from: https://en.wikipedia.org/w/index.php?title=Atmospheric_circulation&oldid=1061580476 [Accessed 23 December 2021].

Wikipedia (2021b) *Daubechies wavelet Wikipedia*. [online]. Available from: https://en.wikipedia.org/w/index.php?title=Daubechies_wavelet&oldid=1057591524 [Accessed 31 December 2021].

Wikipedia (2021c) *Fourier series Wikipedia* [online]. Available from: https://en.wikipedia.org/w/index.php?title=Fourier_series&oldid=1061243963 [Accessed 28 December 2021].

Winder, S. (2002) *Analog and digital filter design.* Elsevier.

Wolpert, D.H. and Macready, W.G. (1995) *No free lunch theorems for search.* Technical Report SFI-TR-95-02-010, Santa Fe Institute.

World Bank (2021) *World Bank Open Data* [online]. Available from: https://data.worldbank.org/ [Accessed 30 November 2021].

Würth, I. et al. (2018) Forecasting wind ramps: can long-range lidar increase accuracy? *Journal of Physics: Conference Series* [online]. 1102, p.012013.

Xunta de Galicia (2021) *Xunta de Galicia | Consellería de medio ambiente, territorio e vivenda, MeteoGalicia* [online]. Available from: https://www.meteogalicia.gal/web/inicio.action?request_locale=es [Accessed 18 September 2021].

Yang, B. et al. (2021) State-of-the-art one-stop handbook on wind forecasting technologies: An overview of classifications, methodologies, and analysis. *Journal of Cleaner Production* [online]. 283, p.124628.

Yao, S.J. et al. (2000) Wavelet transform and neural networks for short-term electrical load forecasting. *Energy Conversion and Management* [online]. 41 (18), pp.1975–1988.

Ye, L. et al. (2020) *Combined Approach for Short-Term Wind Power Forecasting Considering Meteorological Fluctuation and Feature Extraction.* In: *2020 IEEE/IAS Industrial and Commercial Power System Asia (I CPS Asia)* [online]. 2020 IEEE/IAS Industrial and Commercial Power System Asia (I CPS Asia) 1334–1343.

Yu, C. et al. (2018) A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. *Energy Conversion and Management*. 178, Elsevier, pp.137–145.

Zack, J.W. (2007) *Optimization of wind power production forecast performance during critical periods for grid management.* Proceedings of the European Wind Energy Conference EWEC, Milano (IT) 8.

Zha, X. et al. (2016) Selection of time window for wind power ramp prediction based on risk model. *Energy Conversion and Management*. 126, Elsevier, pp.748–758.

Zhang, B.L. and Dong, Z.Y. (2001) An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research* [online]. 59 (2), pp.121–129.

Zhang, Y. et al. (2020a) A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic. *Energy Conversion and Management* [online]. 203, p.112254.

Zhang, Y. et al. (2020b) Wind Speed Interval Prediction Based on Lorenz Disturbance Distribution. *IEEE Transactions on Sustainable Energy* [online]. IEEE Transactions on Sustainable Energy 11 (2), pp.807–816.

Zhang, Y. et al. (2018) Wind Speed Prediction of IPSO-BP Neural Network Based on Lorenz Disturbance. *IEEE Access* [online]. IEEE Access 6, pp.53168–53179.

Zhang, Y. et al. (2019) Wind Speed Prediction Research Considering Wind Speed Ramp and Residual Distribution. *IEEE Access* [online]. IEEE Access 7, pp.131873–131887.

Zhao, Y. et al. (2020) Bayesian Network Based Imprecise Probability Estimation Method for Wind Power Ramp Events. *Journal of Modern Power Systems and Clean Energy* [online]. Journal of Modern Power Systems and Clean Energy, pp.1–10.

Zheng, H. and Kusiak, A. (2009) Prediction of Wind Farm Power Ramp Rates: A Data-Mining Approach. *Journal of Solar Energy Engineering* [online]. 131 (3). Available from: https://doi.org/10.1115/1.3142727 [Accessed 18 June 2021].

Zhou, B. et al. (2021) Short-term prediction of wind power and its ramp events based on semi-supervised generative adversarial network. *International Journal of Electrical Power & Energy Systems* [online]. 125, p.106411.

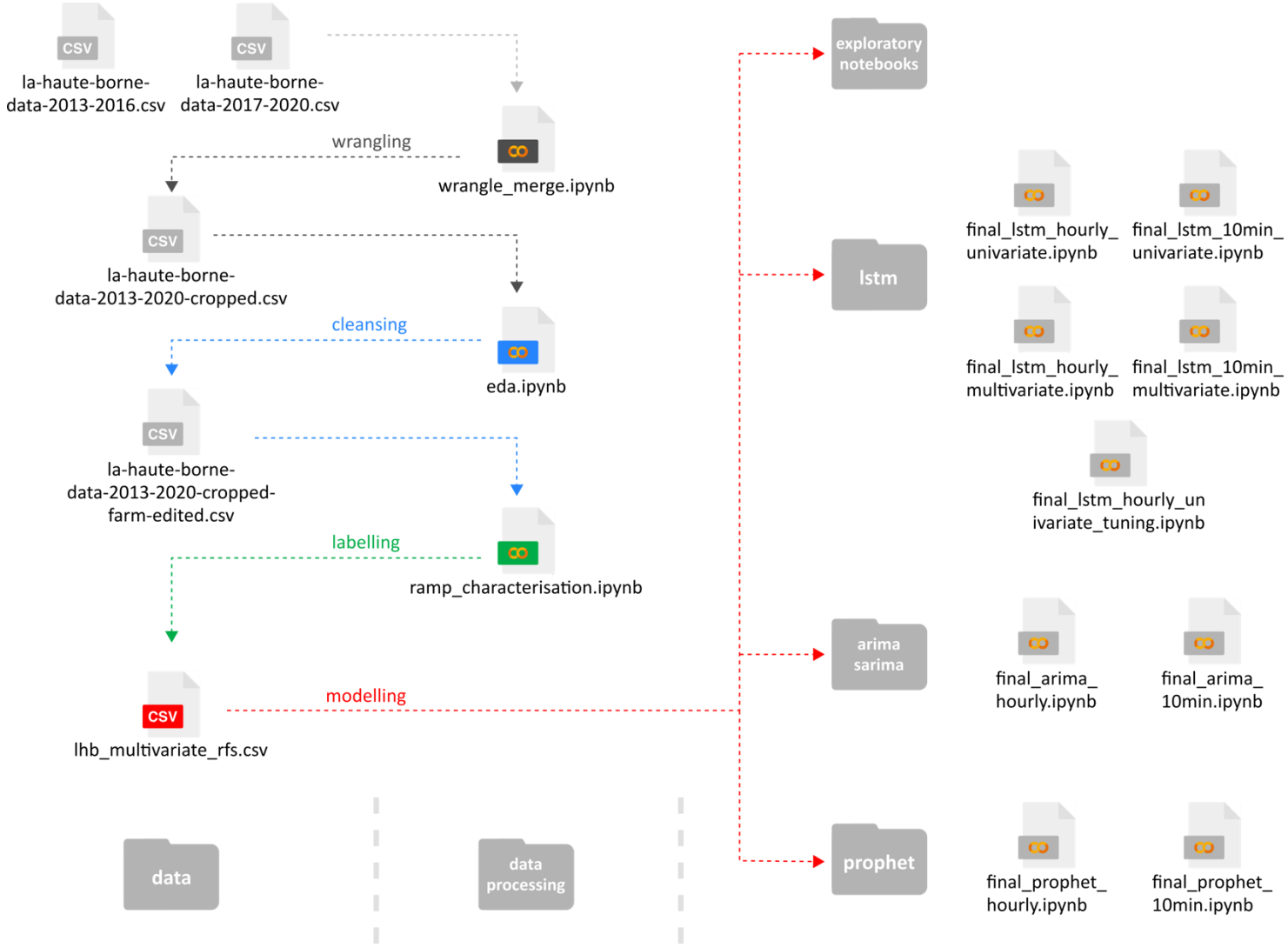
Zhu, K., Wong, Y.S. and Hong, G.S. (2009) Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results. *International Journal of Machine Tools and Manufacture* [online]. 49 (7), pp.537–553.

Zucatelli, P.J. et al. (2021) An investigation on deep learning and wavelet transform to nowcast wind power and wind power ramp: A case study in Brazil and Uruguay. *Energy* [online]. 230, p.120842.

7 Appendices

This section provides a map of the associated code repository, computational specifications for reproducibility purposes, records of supervisor meetings, and supplemental material that attempts to briefly elaborate on some of the topics touched on by this thesis.

7.1 Appendix A – Code file map



7.2 Appendix B – Meteorological processes associated with ramp events

7.2.1 Planetary scale

The planetary scale atmospheric setting of the study area is visualised in Figures 3 and .

7.2.2 Macroscale

Macroscale processes include large-scale weather systems such as low- and high-pressure areas and cold and warm fronts. Ramp events are typically caused by low-level high wind speed features associated with the passage of the fronts (Zack, 2007). Figure 4 shows how these features develop and dissipate over the study areas throughout the course of one week. The long-life cycles, the large horizontal scales and the strong controlling influence of these large-scale weather systems means that they are generally well forecasted by NWPs and the accuracy of their prediction is fairly high. Their development and dissipation are important over DA timescales (Zack, 2007). Cutler et al. (2009) found that the majority of ramp events experienced by three wind farms in Australia were associated with cold fronts, low pressure systems and troughs. Other studies from Germany (Steiner et al., 2017) and Portugal (Couto et al., 2021; Lacerda et al., 2017) have shown strong correlations between power system critical ramp events and fronts and troughs associated with cyclones. Couto et al. (2021) highlight that high wind power variability is to be expected under cyclonic weather regimes whereas ramp events associated with low levels of power generation are more common under anticyclonic conditions.

7.2.3 Mesoscale

Mesoscale processes include lower atmospheric circulations such as low-level jets and thunderstorms, thermally driven flows such as sea-land breezes and Anabatic/Katabatic winds, and flows linked to topography (*orographic forcing*) such as mountain-valley winds.

Low level jets are high wind speed features driven by sharp pressure-temperature gradients. They have an extension of a few hundred kilometres, typically occur at 300-400m altitudes, and heavily influence weather patterns in the lower levels of the atmosphere (Orlanski, 1975). Deppe (2013) showed that low level jets were associated with 29-32% of ramps identified in their Pomeroy wind farm (Iowa, USA) case study. A seasonal wind known as *La Nortada* is an example of a coastal low-level jet that affects the NW coast of Spain. Using the

WRF ERA-Interim reanalysis dataset, Soares et al. (2013) showed that *La Nortada* was active on almost 70% of summer days between 1989 and 2007 and was responsible for north-north easterly winds with mean and maximum speeds of 15 and 25 m/s respectively.

Thunderstorms typically have life cycles ranging from one to twelve hours and can have horizontal scales on the order of tens of kilometres. Effective prediction of ramp events induced by thunderstorms relies upon the skill of NWP models which are poor predictors of the timing and location of individual thunderstorms. Meteorological observations typically provide little useful information regarding this type of ramp event but they can act as supplemental or confirmatory data sources. Whilst NWP models cannot be used to predict individual thunderstorm-induced ramp events, they can be used to indicate when the probability of large ramp events is above or below average over a period of several hours (Orlanski, 1975; Zack 2007).

Thermally driven flows are divided here into Anabatic and Katabatic flows. They also include sea-land breezes however, these are not expected to influence the LHB site given its inland location. Anabatic flow can occur when air is heated relative to its surroundings. For example, at a mountain side exposed to the sun, air near the warm surface will initially rise and then begin to cool and descend as its proximity to the surface (the heat source) decreases. This can lead to a weak circulation and cloud formation above the flow (Landberg, 2015). Katabatic flows are formed, often at night, as heat loss from the surface cools the air within its proximity causing it to become denser and to start to descend. This process can be accelerated in downhill settings leading to strong and destructive wind speeds. Although thermally driven windspeeds can reach damaging levels (e.g., 25 m/s reached by the Mistral wind in southern France), their effect on the wind resource is quite limited (Landberg, 2015).

Orographic forcing describes how atmospheric flow near the Earth's surface follows topography. Mountains/hills and valleys affect both wind speed and direction. As air flows over a hill, compression of the air resulting from the displacement imposed by the hill itself, causes the wind speed to increase (hence wind turbines are normally situated on hill tops rather than in valleys). The direction of the wind is also deflected around the hill (Landberg,

2015). La Haute Borne resides in gently sloping terrain. The site predominantly experiences prevailing winds from 50° and 220°. The various topographic features that may orographically influence winds blowing from these directions are labelled in Figure 6, Section 2.

The horizontal scale of mesoscale processes ranges from a few to a few hundred kilometres and their lifespans range from a few hours up to a day (Landberg, 2015; Orlanski, 1975; Zack 2007). The smaller spatiotemporal scales involved make it more difficult to distinguish these processes and to monitor their evolution. The forecasting of the development and dissipation of these systems is more important than monitoring of their movements. NWP models can provide fair to good prediction of the winds associated with mesoscale circulations over DA or even two day-ahead timeframes (Zack, 2007).

7.2.4 Microscale

In contrast to moist convective processes, dry convective processes occur over shorter (and more variable) spatiotemporal scales. Dry-convective ramp events typically occur in two ways. Firstly, small changes in atmospheric stability or increases in vertical wind shear can cause the sudden turbulent mixing of layers of different wind speeds at turbine rotor levels. Many explosive ramp-up events are associated with this process as wind aloft at higher speed is brought down to rotor level. Secondly, sudden near-surface cooling can drastically increase atmospheric stability at rotor level and decrease the amount of vertical turbulent mixing. This can lead to significant and sudden reductions in wind speed and associated power output over a relatively short time frame. The key to predicting dry convective ramp events lies in the ability to anticipate the underlying changes wind shear and thermodynamic stability. Meteorological towers again provide little useful information about the vertical mixing of the surface layer and NWP models have little skill in predicting specific turbulent mixing events, especially over the DA timeframe. However, once again NWPs can be used to indicate time windows in which the probabilities of dry convective ramp events are higher or lower than average (Zack, 2007).

The ramp events described to this point have all been driven by substantial changes in wind speed and hence significant meteorological events. However, ramp events can also be

caused by very small changes in wind speed under the circumstance that wind speed is already at or near the turbine cut-out (overspeed) threshold, which is typically around 25 m/s for most modern turbines. This type of ramp event is very difficult to predict accurately as small errors in the wind speed forecast can result in large errors in power output forecasts. The ability to predict such critical changes in wind speed varies depending on the meteorological cause: as has been discussed, large events can often be anticipated by NWPs with high levels of accuracy, however, local microscale circulations, dry turbulent mixing or thunderstorms are difficult to forecast due to their rapid evolution (Zack, 2007).

Surface roughness is discussed here due to the significant influence that it exerts over vertical wind profiles (and the modelling thereof) at local scales. Although surface roughness has not commonly been identified as a direct cause of wind ramps, Würth et al. (2018) recently suggested that the effects of local roughness dominated over those of larger scale phenomena during ramp events. The same authors, however, point out that their analysis is based on an 'unrealistic' model. Rough surfaces (e.g., forests) generate frictional drag that increases vertical wind shear and reduces the momentum of turbulent flow (American Meteorological Society, 2020). Forests likely to affect the LHB turbines are also labelled in Figure 6, Section 2.

Obstacles are the final microscale atmospheric influence discussed here. They typically consist of buildings or *shelter belts* (e.g., a row of trees). Their effect is to reduce wind speed and increase turbulence from ground level to 100m height. LHB is free from obstructive buildings, however rows of trees are likely to exert an influence. Obstacles are not typically accounted for by NWP models (Landberg, 2015).

7.3 Appendix C – Google Colab CPU specifications

Computation times are provided in this section for comparative purposes only. All experiments in this study were run using a standard Google Colab account with no hardware acceleration (i.e., standard runtime). In order to be able to offer computational resources for free, Colab adjusts usage limits and hardware availability on the fly. Resources available in Colab therefore vary over time (Google, 2021). Colab CPU information as output on 30 December 2021 was as follows:

processor: 0
vendor_id: GenuineIntel
cpu family: 6
model: 79
model name: Intel(R) Xeon(R) CPU @ 2.20GHz
stepping: 0
microcode: 0x1
cpu MHz: 2200.158
cache size: 56320 KB
physical id: 0
siblings: 2
core id: 0
cpu cores: 1
apicid: 0
initial apicid: 0
fpu: yes
fpu_exception: yes
cpuid level: 13
wp: yes
flags: fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush mmx fxsr sse sse2 ss ht syscall nx pdpe1gb rdtscp lm constant_tsc rep_good nopl xtopology nonstop_tsc cpuid tsc_known_freq pni pclmulqdq ssse3 fma cx16 pcid sse4_1 sse4_2 x2apic movbe popcnt aes xsave avx f16c rdrand hypervisor lahf_lm abm 3dnowprefetch invpcid_single ssbd ibrs ibpb stibp fsgsbase tsc_adjust bmi1 hle avx2 smep bmi2 erms invpcid rtm rdseed adx smap xsaveopt arat md_clear arch_capabilities
bugs: cpu_meltdown spectre_v1 spectre_v2 spec_store_bypass l1tf mds swapgs taa
bogomips: 4400.31
clflush size: 64
cache_alignment: 64
address sizes: 46 bits physical, 48 bits virtual
power management: n/a

7.4 Appendix D - RNN disadvantages

Unstable gradients

RNNs can suffer from unstable gradients during training. One solution to deal with this is a popular regularisation technique known as *Dropout*, (Figure 37). Dropout has been shown to improve model accuracy by 1-2%. At each step, each neuron has a probability, p of 'being left out' temporarily (i.e., being ignored completely during a given training step) yet still being included in subsequent time steps. The hyperparameter p, known as the dropout rate, is generally configured at between 20-30% for RNNs. After training, neurons are no longer left out. The principal behind training RNNs with dropout is that neurons are trained not to rely on others, making them more able individually and less sensitive to small changes in inputs. The result is a more stable network that is able to generalise better (Gerón 2019).

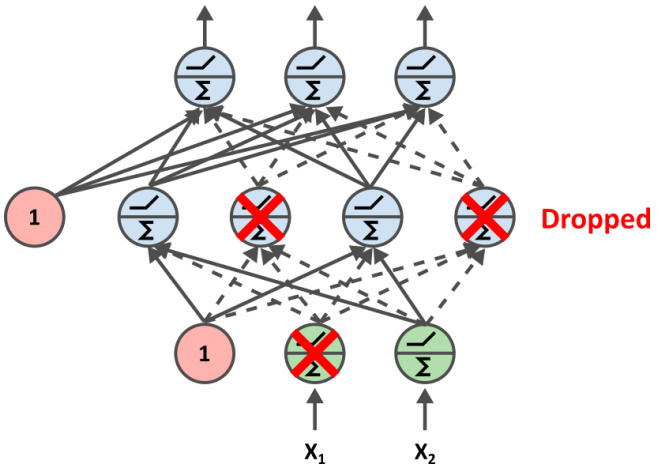


Figure 37. Concept of dropout. With each iteration, a random subset of neurons is left out (dropped). These neurons generate a result of zero in this time step as shown by dashed arrows (modified after Gerón, 2019).

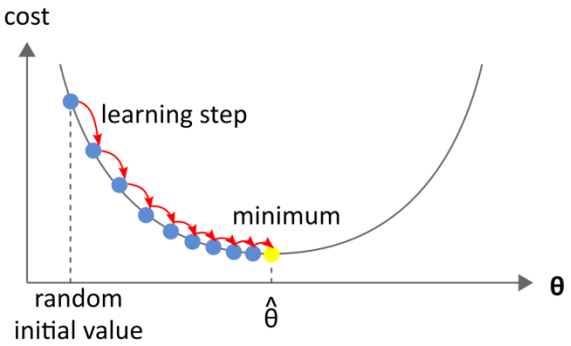
Data preparation – Scaling

One of the primary drawbacks found during the implementation of the LSTM-RNN used in this study was the complexity of the data preparation process. To illustrate, scaling was performed using the MinMaxScaler transformer from Scikit-learn. This transformer was selected so that the effects of outliers (ramp events in this context) were preserved. However, Scikit-learn offers eight different types of transformer classes and even dedicates an entire page to a comparison of the behaviours of each, specifically when applied to a

dataset containing marginal outliers[18]. An understanding of such subject matter requires advanced statistical knowledge which an average TSO analyst may not possess. MinMaxScaler subtracts the minimum value and divides by the maximum minus the minimum (Pedregosa et al, 2011). Scaling is necessary because RNNs often expect input values between 0 and 1. The reason for this is that they employ an algorithm known as 'backpropagation' to minimise the chosen loss function (e.g., Mean Squared Error). At a high level, the backpropagation algorithm is a stochastic *gradient descent* (Figure 38d) with an efficient technique to calculate the gradients automatically: in two steps through the network (one forwards, one backwards), the algorithm can calculate the error of the network. In other words, it can gauge how much to adjust the weights of each neuron connection and each bias term to reduce the error. Once the gradients are calculated, the algorithm calculates the gradient descent, and the process is repeated until the network converges to a solution (Figure 38a,b). If the characteristics being used to calculate the gradients have the same scales, descent along the gradient is most direct (Figure 38c left), and therefore most rapid. If the scales are different, the descent can take longer to converge to the minimum point (Figure 38c, right). Gradient descent itself is an optimisation algorithm capable of iteratively adjusting parameters to minimise the loss function. It is summarised schematically in Figure 38 (Gerón 2019).

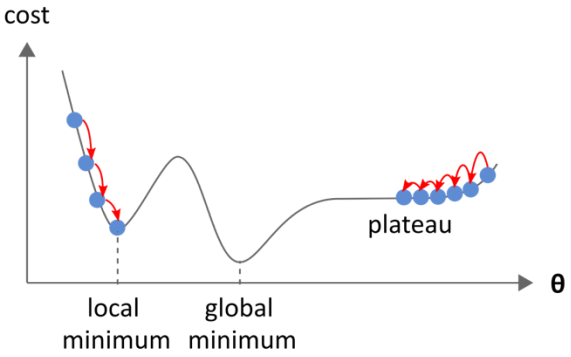
18 https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#sphx-glr-auto-examples-preprocessing-plot-all-scaling-py

a.



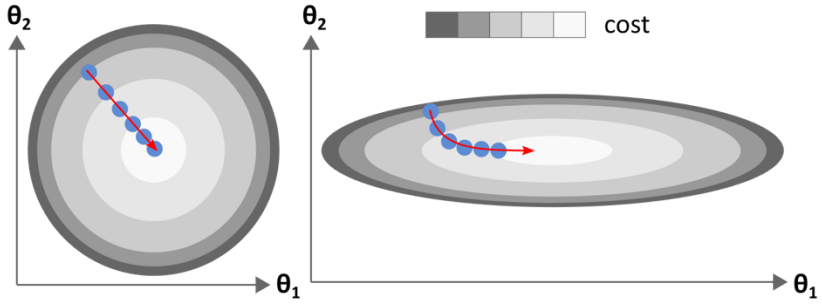
Gradient descent measures the local gradient of the error function with respect to a vector of the parameters, θ, then 'steps' in the direction of the descending gradient until it reaches a minimum. Step size is determined by the *learning rate* hyperparameter and refers to the amount that the network weights are updated during training.

b.



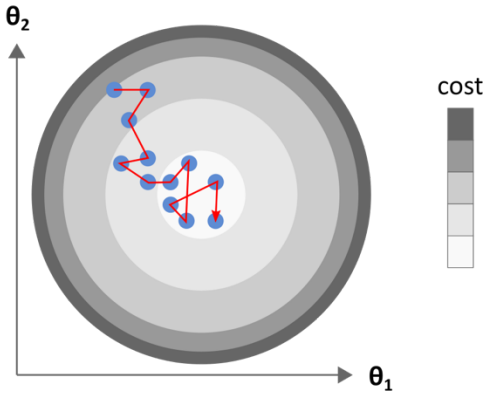
If the learning rate is too small, the algorithm can take too long to reach the minimum or may become stuck in a local minimum or plateau. If it is too high, the algorithm could step past the minimum, causing it to diverge and not reach a solution.

c.



If the characteristics used to calculate the gradients have the same scales, descent along the gradient is most direct, and therefore most rapid. If the scales are different, the descent can take longer to converge to the minimum point.

d.



Stochastic gradient descent selects a random instance in the training set at every step and calculates the gradient based on only that instance. This makes the algorithm much faster and makes it possible to train on huge training sets.

Due to its stochastic nature however, once a minimum is reached the algorithm continues to move around. Final parameter values are therefore good but not optimal.

Figure 38. Gradient descent optimization algorithm schematic. Modified after Gerón (2019).

Data preparation – Reframing a time series as a supervised learning problem

Before a Recurrent Neural Network can be used for time series forecasting, the time series forecasting problem must be reframed as a supervised learning problem. A time series is a sequence of numbers ordered by a time index. A supervised problem is composed of an input pattern, X, and an output pattern, y, such that an algorithm can predict y from X (Figure 39).

Date_time	
31/12/2012 23:00	0
31/12/2012 23:10	1
31/12/2012 23:20	2
31/12/2012 23:30	3
31/12/2012 23:40	4
31/12/2012 23:50	5
01/01/2012 00:00	6
01/01/2012 00:10	7
01/01/2012 00:20	8
01/01/2012 00:30	9

X	y
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9

Figure 39. Time series (left) vs supervised learning pattern (right).

By shifting a time series forwards or backwards, the time series can be reframed as a supervised learning problem (Figure 40). In time series forecasting, past columns (*t-n*) can be taken as input (X) and used to predict output values (y) of present and future columns (*t*, *t+n* respectively).

Date_time	*t*
31/12/2012 23:00	0
31/12/2012 23:10	1
31/12/2012 23:20	2
31/12/2012 23:30	3
31/12/2012 23:40	4
31/12/2012 23:50	5
01/01/2012 00:00	6
01/01/2012 00:10	7
01/01/2012 00:20	8
01/01/2012 00:30	9

Date_time	*t-1*	*t*	*t+1*
31/12/2012 23:00	Nan	0	1
31/12/2012 23:10	0	1	2
31/12/2012 23:20	1	2	3
31/12/2012 23:30	2	3	4
31/12/2012 23:40	3	4	5
31/12/2012 23:50	4	5	6
01/01/2012 00:00	5	6	7
01/01/2012 00:10	6	7	8
01/01/2012 00:20	7	8	9
01/01/2012 00:30	8	9	NaN

Figure 40. Converting a time series to a supervised learning problem.

Rows containing NaN values are discarded as they cannot be interpreted by the algorithm. This concept can be expanded to consider multivariate time series as in Figure 41.

Date_time	var1(t-1)	var1(t-1)	var1(t)	var1(t)	var1(t+1)	var1(t+1)
31/12/2012 23:10	0	100	1	101	2	102
31/12/2012 23:20	1	101	2	102	3	103
31/12/2012 23:30	2	102	3	103	4	104
31/12/2012 23:40	3	103	4	104	5	105
31/12/2012 23:50	4	104	5	105	6	106
01/01/2012 00:00	5	105	6	106	7	107
01/01/2012 00:10	6	106	7	107	8	108
01/01/2012 00:20	7	107	8	108	9	109

Figure 41. Multivariate supervised learning problem

By repeating this sequence, with shifts of 2 or more-time steps, long input sequences (X) can be generated and used to forecast individual output values (y) or sequences of output values (Y) (Brownlee, 2019).

Pre-processing of the data was implemented in this study by defining a function to achieve the 'reframing' described above. An alternative, simpler approach to this step might have been to use the Pandas DataFrame.shift function, but again, this requires Python programming knowledge that the average analyst may not have. This example of data pre-processing warrants a discussion of another of the RNN's major limitations: the fact that RNNs act as black boxes.

Black box models

RNNs act as black box models in the sense that, while they are able to approximate a function, they are unable to reveal any information about the form of the function being approximated. As an example, the multivariate LSTM-RNN used in this study predicts wind power, P, as a function of a matrix input features (wind speed, wind direction, power itself, and temperature), here denoted x. The LSTM-RNN is therefore approximating the function, f, such that:

$$f(x) = P \ \#(18)$$

The issue is that the LSTM-RNN provides no insight into the relationship between its weightings (see Section 3.5.4) and the form of f. The counter argument to using black box models is that it is valid to do so when the model or the explanation of the underlying processes do not matter. However, the perils of this paradigm were highlighted during the

experimental stages of this study, where pre-processing the data using more than one time step was found to be including the current power observation in the matrix of inputs. In other words, the observation being predicted was already contained in the training data. This led to misleadingly accurate results which would not necessarily be instantly noticeable to a non-specialist TSO analyst.

7.5 Appendix E – Auto-ARIMA and SARIMA

Auto-ARIMA

Auto-ARIMA is an algorithm that automatically seeks the most optimal parameters for an ARIMA model. It does this by conducting a series of differencing tests (Kwiatkowski–Phillips–Schmidt–Shin, Augmented Dickey-Fuller or Phillips–Perron) to determine the order of differencing, d and then fitting models within defined ranges of p and q. If the 'seasonal' optional is enabled, Auto-ARIMA can also identify the optimal P and Q hyperparameters after conducting the Canova-Hansen unit root test to determine the optimal order of seasonal differencing, D. In order to find the best model, Auto-ARIMA optimises for a given information criterion (the default is AIC) and returns the ARIMA configuration that minimizes the value (Smith, 2007).

SARIMA

A seasonal ARIMA (SARIMA) model is formed by including additional seasonal terms in the ARIMA model. SARIMA was found in this study to quickly run into memory issues when fitting hourly univariate data datasets of more than one week in size (i.e., 168 observations). The reason for this is the way in which Auto-ARIMA optimises for model parameters: Auto-ARIMA must first estimate model performance for each configuration of the non-seasonal parameters. It must then estimate the performance for each of those non-seasonal configurations, multiplied by each seasonal lag. Thus, the number of calculations required grows rapidly for data with long periods of seasonality and even more so for granular data with high sample rates between seasonal peaks. This is the case for the 10-minutely LHB data, where the sample rate is every 10 minutes, and the seasonality is approximately every 144 lags (Figure 42).

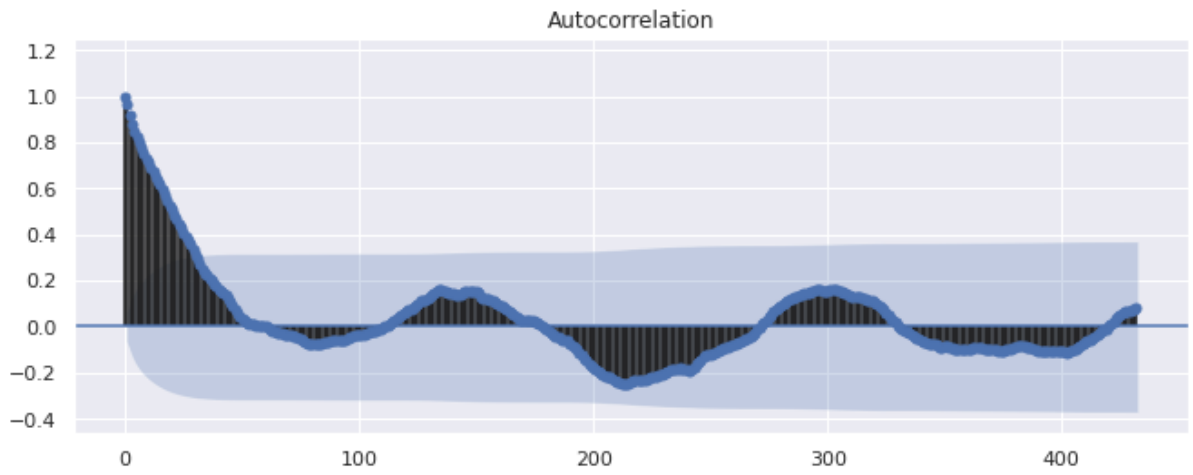


Figure 42. Autocorrelation plot for the raw LHB data showing seasonal peaks at approximately every 144 lags, equivalent to 24 hours.

To confront this problem, Google Colab Graphics Processing Unit (GPU) runtime was employed to accelerate processing power. GPUs were originally designed to accelerate the rendering of 3D graphics. They are capable of processing multiple lots of data simultaneously and as such can deliver incredible acceleration in workloads that can take advantage of their highly parallel nature (Intel, 2021). Despite this, Colab GPU runtime still ran into memory issues on a one-week sub-selection of the LHB dataset. The primary drawback of this is that sufficient seasonality cannot reliably be captured in a one-week sample of a four-year dataset resampled at hourly intervals. A secondary drawback of using Auto-ARIMA is its dependence on AIC optimisation. AIC optimisation returns the ARIMA configuration that minimizes the AIC value and in doing so, emphasizes the model's fit to the data, rather than the model's predictive performance.

7.6 Appendix F - Supervisory Meetings

Date	01/06/2021
Meeting agenda	Topic focus
Attendees	Hisham Ihshaish (HS), Alexis Tantet (AT), Russell Sharp (RS)

	Notes	**Actions**
1	**Meteorological models (AT):** • Balance to be kept in mind btw. theoretical and operational. • What is the added value? • Three hierarchies of climate models: 1. Global weather models (reanalysis): 1.1. Historical - static assumptions of climate change 1.2. Projections - take into account climate change. Must not be constrained by observations. Key is to use climate models 1. Global Atlas models (e.g. ECMWF-ERA5). Key to using these will be to downscale from 50kms to few kms with a refined mesh and to take account variables such as convection. 2. Models that include impact of topography (e.g. scale 5x5km scale, proximity to mountains etc.	
2	**Scale (AT):** • Important to define scale. • Regional assessments of wind potential include fluid dynamics modelling and therefore become complex and computationally expensive. • Large scale models may also mask localised effects. • Downscaling will require reanalysis combined with local observations (model bias corrections). • Climate projections add model uncertainty. • Scale needs to be appropriate so that ramps are visible	

	Notes	Actions

3	**Ramp modelling (AT):**
	• Reanalysis won't be sufficient. Observations from meteo stations will be needed (Anemometers, lidars, wind masts).
	• Key = height at which measurements are made.
		• **RS**:
		• Gather historical statistics on:
		  • Average number
		  • Ramp duration
		  • Local variation
		  • Variations between ramps.
		  • Turbine security:
		  • Shutdowns
		  • Costs
		  • Turbine damage
		  • Effect on market (whether or not to sell, availability of energy)
		  • Low frequency variability
		• Investigate:
		  • Current state of the art
		  • Problems with ramp prediction
		  • Modelling of topography

4	**Datasets (AT):** ERA5 ECAD COREDEX (Euro-, Med-) SAFRAN (reanalysis) COSMO/ COSMOS	• **HI:** follow up on contact with Meteocat.

Date	03/11/2021
Meeting agenda	Project progress and next steps
Attendees	Hisham Ihshaish (HS), Ignacio Deza (ID), Russell Sharp (RS)

	Notes	Actions

1	**Review of progress made (RS):**
	• RS asked about methodology to correct temperature readings. ID/HI suggested:
	  • Take long term average readings or
	  • Interpolate
	• ID suggested plotting lag plots using trend data. HI suggested using statsmodels autocorrelation plots. Remove daily a/c plots already done as these skip the ramps
	• ID/HI suggested making time the index in the autocorrelation plots – see example in data_preparation.ipynb
	• HI suggested removing random walk section (data_preparation.ipynb) as AUTOARIMA will handle this
	• ID suggested performing autcorrelation plots without the seasonality but keeping noise
	• In the differencing plots, ID suggested that the peaks are the ramps once noise is removed using a filter.
	• HI reminded RS that SARIMA/ AUTOARIMA handles differencing automatically
	• ID suggested that removing seasonality and slightly cleaning the noise (e.g., using moving average) will reveal ramps. If ramp duration is 6 hours, look at 7 hours and reduce with low pass filter. HI suggested using SARIMA if not de-noising the signal, if de-noising the signal, use ARMA
		RS
		• Amend temperature corrections
		
		• Plot lag plots using trend data
		• Remove daily a/c plots
		
		• Recode a/c plots from example/ tidy up with/without noise
		• Remove random walk section
		
		• Search for Python low pass filter methods (probably statsmodels)

2	**Wavelet Transform/ Ramp Function (HI/ID):** • RS explained understanding of wavelet transform/ ramp function. HI suggested assessing that at next stage. • ID agreed that it makes sense to use to denoise but it is an advanced topic	**RS** • Park for now

	Notes	Actions
3	**ML next steps (HI/ ID):** • HI suggested that Prophet will be better at picking ramp magnitudes and pointed out that it magnifies seasonality. • HI suggested, if results from these steps are good, the bulk of the work is done. If results from these steps are not good, investigate Wavelet Transform • ID suggested calculating cross-correlation of wind turbines. If >0.8 -> combine, if <0.8 analyse separately. • ID suggested investigating Power output by Turbine: Plots of power output per m/s of wind will give turbine efficiencies. Plot of wind speed vs power should give an almost perfect line. Shift by 10 minutes if distribution is noisy.	**RS**: • Establish baseline model performance (AutoArima SARIMA) • Compare SARIMA, Prophet, LSTM • Look into Anomaly Detection with Random Forests • Stabilized point prediction (applied in LSTM) • Cross turbine analysis • Finish LHB contour map
4	**Next time:** • Wavelet Transform/ Ramp Function • NWP attribute selection • How to handle NWP outputs programatically	RS to email once results are done

Date	18/11/2021
Meeting agenda	Project progress and next steps
Attendees	Hisham Ihshaish (HS), Ignacio Deza (ID), Russell Sharp (RS), Ciaran Haines (CH), Ryan Fellows (RF)

	Notes	**Actions**
1	**Review of progress made (RS):** • RS asked about Augmented Dickey Fuller test results and stationarity of datasets. HI/ID answered that the data shouldn't be stationary. • RS asked about low pass filter. ID said to increase frequency. • ID suggested to derive a statistical model* from wind/power relationship (sigmoidal or linear regression) to convert between the two. • IS suggested to look at MAE between predicted and actual values for each model. • KH asked "Do you have any ground truths? Do you have any ramps labelled?". RS responded "Not yet. This is the purpose of the ramp function". • KH/ ID suggested plotting current wind power against 24hr lagged wind speed (exploratory analysis). • ID suggested using RF regression to predict wind from wind then to transform to power using model* • RS asked ID "What will the ramps look like? Will I apply the ramp function to get the ramps, then delete all the non-ramps and be left with a series of timestamps which will be the timings of the ramps?" ID answered "Yes"	**RS** • Look back at AFD test • RS to read more on filter choice/ redo • Derive statistical model from wind/power relationship • Look at MAE between predicted and actual values for each model. • Proceed with ramp function to define ramps. • RS to send SARIMA to Hisham

Date	25/11/2021
Meeting agenda	Ramp characterisation (labelling)
Attendees	Hisham Ihshaish (HS), Ignacio Deza (ID), Russell Sharp (RS)

	Notes	Actions
1	**Publishable work (RS/ HI):** • RS/ HI reviewed how publishable RS' thesis is. HI recommended *prioritising* 2 days re-working NEURIPS proposal to include most recent approaches and methodologies (Also to include data description). • HI pointed out that there is no requirement to have all code in a gitlab repository but recommended writing a table of contents for all notebooks. HI stated that evaluation and reflection is the most important part. HI said that code should be introduced with some comments but not to go overboard on this. RS showed an example which HI validated.	**RS** • Rework NEURIPS proposal • TOC for notebooks
2	**Filtering/ Ramp characterisation process (RS, HI, ID)** • RS explained difficulties encountered with filtering/ ramp characterisation process. ID explained that there is no need to filter before performing the wavelet transform. ID explained that the Hannesdóttir approach is likely adapted to the field of investigation and not necessary here. • RS, HI, ID discussed wavelet transform and different time scales involved. ID and HI recommended that RS choose 3 scales and compare LSTMs on each.	• Skip filtering process. Just perform wavelet transform.
3	**Next steps** • RS communicated aims to finish result by end of November, write up by 20 December. HI indicated that this was ambitious. ID suggested tackling remaining tasks step by step, completing each one before moving on: labelling, LSTM, Prophet, ARIMA… New wind predictions.	• Labelling, LSTM, Prophet, ARIMA… New wind predictions.

Date	29/11/2021
Meeting agenda	Thesis guidance (NWP outputs)
Attendees	Hisham Ihshaish (HS), Russell Sharp (RS)

	Notes	**Actions**
1	**Thesis guidance:** • RS asked for HI's advice regarding the inclusion of NWP outputs into the thesis. RS was concerned that there was insufficient time remaining to extract information from the ECMWF ERA5 dataset with the associated manipulation of NetCDF files etc. HI advised that this would require at least one month's work from an inexperienced user. HI and RS agreed that this should be included in the further research discussion section.	• RS to include inclusion of NWP outputs in discussion section of thesis.

2	**Proposal for publishing archive**
	• RS commented that his first draft of the proposal was nearly done and that he would amend it in light of this discussion.
	• HI said that he was expecting the following sections for this document:
	  • Motivation
	  • Literature review
	  • Structural approach
	  • Perspective
		• RS to amend proposal to re-frame inclusion of NWP outputs.

3	**Next steps** • RS said that he would focus on writing the main report and tuning of the models to capitalise on early feedback from HI. HI agreed and said to send sections over once done. • HI suggested that Learning rate will be key to tuning the LSTM as this controls how loss is minimised.	• RS to focus on writing main report sections. • RS to look into LSTM learning rate.

Date	02/12/2021
Meeting agenda	Ramp classification
Attendees	Hisham Ihshaish (HI), Russell Sharp (RS), Ignacio Deza (ID)

	Notes	Actions
1	**SARIMA results** • RS asked what options there were to deal with SARIMA memory issues. HI suggested using Colab GPUs (runtime > change runtime selection) or to investigate SARIMA parallel processing. HI confirmed that SARIMA will not be able to capture much of the seasonality when sub-selecting only one week.	• RS to try GPUs/ investigate SARIMA parallel processing.

2	**Wavelet Transform (WT)**
	• ID explained the difference between Fourier Transform (FT) and WT (e.g. FT is infinite, WT is finite in time and space.
	• ID explained that:
	  • The number of coefficients depends on the length of time series.
	  • Coefficients [0] = fundamental frequencies (basically the mean).
	  • Each array of coefficients corresponds to the frequency band to be filtered. Deleting them removes anything that looks like noise. Tuning is by trial and error. Deleting the coefficients that correspond to the 1-hour cycles for example, removes the data that we are not interested in.
	  • Derivatives of the coefficients give the direction and magnitude of the signal.
	  • WRT plots:
	  • 'new data' = WT
	  • 'Derivatives' = ramp index
	  • 'Threshold' is effectively equivalent to λ in Gallego et al (2013).
	• RS asked what scores to expect. ID said that the standard deviation of the noise = the accuracy limit of the model.
		• RS to tune and feed labelled time series to LSTM as a multivariate series (refer to Brownlee TS forecasting with LSTM)

	Notes	Actions

3	**Additional**
	• ID recommended investigating
	  • 'keras show model graph' to visualise RNN
	  • Joblib – to save running notebooks each time
	  • Lorentzian distribution (indicative of a complex system)
		• RS to investigate:
		  • 'keras show model graph'
		  • joblib
		  • Lorentzian distribution re: power series.

Date	16/12/2021
Meeting agenda	Ramp classification
Attendees	Hisham Ihshaish (HI), Russell Sharp (RS), Ignacio Deza (ID), Ryan Fellows (RF), Ciaran Haines (CH), Peter Mayhew (PM)

	Notes	Actions
1	**Round table discussion** • Each member presented progress and issues. • HI told PM that it was important to show how objectives are derived from aims.	• RS to tighten up link between aims and objectives.
2	**RS Progress/ issues** • RS discussed the issue that duplicated data was producing better results than cleaned. HI said this could have been giving your model more flexibility. ID likened it to spending all your life around the Midlands accent, then listening to someone speak English with a Spanish accent; you wouldn't have had any noise in your training data. ID suggested there may be an issue in RS' code. • RS asked whether or not to include both sets of results. ID said, 'If you cannot find a reason why, include both and explain what they show. If you can, just include the clean results'. • RS asked whether or not to include SARIMA if it only uses a sub-selection of 3 days. HI said 'not in results, but discuss it'. • RS asked about how long to spend on model tuning, and if marks will be lost for not including extensive tuning. ID and HI said that you won't lose marks as long as you acknowledge the tunings that would have been possible. ID recommended searching for Keras Tuner, Keras Family, Github. • HI reminded RS that he was not aiming to create the best model but to write a scientific paper evaluating a family of models, including a discussion of and justification for any shortcomings. • HI asked RS to send an elaborated version of his previous proposal to include the approach, an explanation of the ramp characterisation process, and expectations/ preliminary results. Deliverable: Better specified approach with prelim results prior to handing in the thesis (for publication on archive).	• RS to: • Look at cleansing/LSTM code. • Explain different results with cleaned vs dirty data and include one or the other • Remove SARIMA from results • Acknowledge tuning parameters in discussion. • Discuss each model shortcomings and reasons why. • Elaborate proposal

	Notes	Actions
	• HI asked RS to share colab repository and report so far. • ID suggested feeding 'dirty data' to the trained models instead of the test set. The test set makes it too easy for the models (esp. LSTM)	• Send colab and report to HI and ID • Feed dirty data to models
3	**Additional** • Ignacio away from 17/12 – 01/01 • Hisham away 27/12 – 07/01	

Date	28/12/2021
Meeting agenda	HI review of written report
Attendees	Hisham Ihshaish (HI), Russell Sharp (RS)

	Notes	Actions
1	**Motivation** • HI recommended making the motivation for the project more emphatic – link to projections of wind power use, industry dynamics etc. HI said RS may wish to link to management of energy supply, complexity/ chaotic nature of the data, topology, mesoscale considerations etc. Emphasise that this is an exploration of the application of ML to the problems of VRE.	• RS to emphasise motivation.
2	**Wavelet Transform** • HI stated that the WT section was light on the benefits of WT. • HI suggested that RS include that it maintains relevant signal by removing noise, enabling models to capture most significant part of the data.	• RS to include benefits and applications of WT.
3	**Experiments/ results** • HI stated that arguments to support the fact that the 10min LSTM was superior are missing. • Add explicit motivation for investigating multivariate LSTM (to include mention of Vector ARIMA in further research section) – e.g. so that we can arrive at power predictions from other variables. • HI noted that learning rate is terrible – LSTM doesn't converge. HI also noted that since they are stochastic, LSTMs need to be run at least x10. Sklearn or Keras should have a package for this, if not, RS to ask HI for code. This can also be discussed as additional computational complexity. • HI reminded RS that this is an application of ML rather than empirical ML science – for which standardized computational architecture would be required.	• RS to remove fabricated variables from LSTM and re-run x10. Provide arguments for superiority. • RS to add motivation for multivariate LSTM • RS to add additional computational complexity to discussion.
4	**Forecasting**	• RS to refer to further research for 72-hour forecasts.

	Notes	Actions
	• RS asked about whether to include 72-hour forecasts as these are essentially the purpose of further research. HI agreed and said that they can be referred to.	
5	**NeurIPS proposal** • HI said that he will reshape the proposal and upload to Archive.	• HI to upload proposal
6	**Lit review** • HI said that this is sufficient, but that RS may wish to add "X was motivated by Y holes in the knowledge as identified in the literature"	• RS to add additional motivations and references to lit review section