The Alan Turing Institute

Data Study Group Final Report: Greenvest Solutions

14 – 25 September 2020

Forecasting wind energy production using satellite data

The Alan Turing Institute

Data Study Group
Final Report

GREENVEST SOLUTIONS

Forecasting Wind Energy Production
Using Satellite Data

doi.org/10.5281/zenodo.4534349

Contents

1 Executive summary

1.1 Challenge overview

Wind resource assessment is an important step when deciding where to build a new wind farm and to design the best plant by considering the type of wind turbine to be used, plant geometry and orientation. A more accurate, faster and cheaper approach to wind resource assessment could be seen as one of the contributing factors to higher the rate of transition towards renewable energy resources. This challenge aims to develop an algorithm to assess wind speed and direction at any location in the UK (and potentially the Earth) by exploiting satellite data and artificial intelligence.

When assessing the wind resources at a certain location to build a new wind farm, old maps of the surrounding area can be used to roughly determine its potential. Then, meteorological masts are placed at the locations of interest and the data it records are collected for at least one year. This procedure is often inaccurate, slow, expensive but it is the standard for the renewable energy industry.

An alternative solution to the problem is to use satellite data and mesoscale models in conjunction with computational fluid dynamics calculations to assess the wind resource at a certain location. However, this approach can suffer from poor accuracy and convergence when the surrounding terrain and meteorological conditions are complex, as well as high computational costs. This challenge investigates the use of machine learning to estimate wind resource from satellite data, to obtain a faster, cheaper and potentially more accurate solution.

The outcome of the challenge will help predict the wind energy production of wind farms and thus offer a valuable product for wind resource assessment that will benefit us all in transitioning towards a clean-energy driven society. This could lead to faster development of wind farms and optimisation/hybridisation of current ones to ultimately rely more and more on less-polluting energy sources.

1.2 Data overview

To tackle this challenge, Greenvest has provided data measured from different satellites that give an overview of an area, its climate and its topology as well as time-series of wind speed and direction at 10 meters above the ground.

The dataset includes time-series of wind speed, direction and other meteorological parameters, as well as topological maps, measured by satellites in the vicinity of 172 meteorological stations across the UK. Wind speed and direction measurements from each ground station are also provided and will serve as ground true measurements.

1.3 Main objectives

This challenge aims to predict wind speed and direction at 10 meters above the ground for a specific location. The key difficulty is to understand the relationship between the topography of the terrain, its roughness, meteorological parameters and wind resources, and to effectively interpolate the satellite data, which is on a 9km rectangular grid, to any given point within the grid.

The key questions to be answered by this challenge are:

- Can we train an algorithm that can predict wind speed and direction at the selected ground station, using satellite data?
- What is the accuracy of the algorithm on an hourly, daily and yearly basis?
- What are the expected errors when using the algorithm to predict the wind resource at a new location (not among the meteorological stations used for training)?
- What are the features with most weight and can we engineer better ones?
- What is the computational cost in terms of time and processing power needed for a new prediction?
- Are there any suggestions to extend the dataset with more ground stations based on considerations about climate, proximity to shore,

etc. to improve the algorithm performance?

The minimal achievement of a successful challenge would be to predict the long term yearly wind resource of any given station with an error of less than 6% Mean Absolute Percentage Error (MAPE). This would lead to a better estimation of wind energy production in the long term compared to existing methods.

The best possible outcome of the challenge would be to be able to predict the wind resource with an accuracy better than 10% MAPE with 1-hour time step. This would allow for a better overall estimation of the wind resources and could be used to optimise the output of a wind farm by deciding on the wind turbines to be used and the desired energy output profile. An accurate hourly prediction of wind resource could also be used in the future for near-real-time monitoring and forecasting of wind energy production.

1.4 Approaches

As an overview of our models, there are two key strategies which form our overall approaches to the problem. The first line of attack involves focusing more on the temporal features and forecast. Examples of these involve LSTM and SARIMA. The second strategy weights heavily on spatial and geographical information, typically trying to upsample the spatial resolution of information by interpolation methods. A representative of this method is Kriging (a.k.a. Gaussian-process interpolation).

Many models do not fall under the umbrella of these two sharply-defined categories, since they typically combine *both* spatial and temporal aspects of the problem. Examples of these are: the Inverse-Distance-Weighting Models (IDWs), Gaussian-process regression for predictions of the mean year (GPR), XGBoost tree-like models and finally a RNN model called Gated recurrent unit (GRU) conditioned on geographical data.

We have summarised our models in Fig. 1, colour-coded according to their complexity.

Finally, we provide a high-level summary of these 7 models that we have considered:

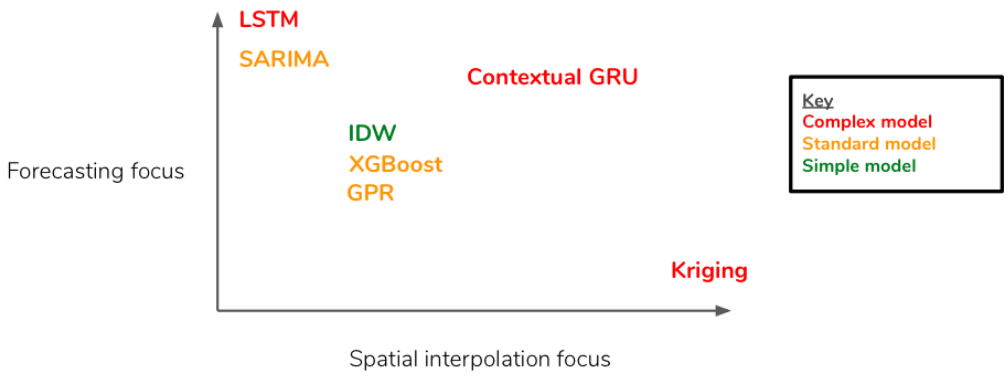


Figure 1: Overview of models

1. **Inverse Distance Weighted Interpolation (IDW)**

   These are models involving only the satellites' wind data and their distances to the respective ground station. Each of the satellite's contribution to the prediction is weighted (inversely) by their distance to the ground station. These simple models provide coarse spatial interpolations which will be used by other methods, both as baselines/benchmarks and as part of the pipeline of a more complicated model.

2. **Seasonal Autoregressive Integrated Moving Average (SARIMA)**

   In this model, the goal is to predict the satellite's wind data based on historical measurements, using traditional Seasonal Autoregressive Integrated Moving Average (SARIMA) models. The forecast satellite wind information can be used as inputs to the IDW model. This then provides a simple forecasting model.

3. **Gaussian Process Regression (GPR) for the mean year)**

   This is a simple approach to predict the monthly ground station data using Gaussian process regression, given the monthly annually-averaged wind and geographical data around the ground station. This approach is able to provide uncertainties for its predictions.

4. **XGBoost**

   By treating each hourly time-series data as independent data point, and incorporating the geographical information near the ground station, this XGBoost model leverages a tree-like ensemble learning model to predict ground wind speed at each station. This is a very interpretable model which is also flexible to incorporate any new set of data or information.

5. **Regression-Kriging**

   The Regression-Kriging method combines the deterministic regression part and stochastic Kriging (Gaussian Process Regression) part, which assumes the target variable is spatially autocorrelated. The model pools knowledge from several satellite-derived wind speeds/directions point locations with several predictors, to estimate the wind speed at a ground location.

6. **Long short-term memory network (LSTM)**

   LSTM is type of recurrent neural network model which has a built-in long and short-term memory capacities. For our model, we input all the time-series data from the satellites near a particular ground station, with the outcome being the forecast of the same data for the next period of time as well as the wind data for ground stations. This is a very powerful neural network model which allows us to make medium-term prediction such as next-month forecast.

7. **Contextual Gated Recurrent Unit (GRU)**

   This neural network model aims to predict the wind speed and direction distribution, based on the distribution of all time-series data. The model is further conditioned on the output of a convolutional neural network (CNN) which extract important features from the spatial context. This is a powerful method incorporating both spatial and temporal features, with the trade-off of being a very computational expensive and complex model.

In a latter section, where we present the methodology and results of our approaches, we will group the IDW, SARIMA and GPR methods under the baseline approaches due to their simplicity and inexpensive

computational time. They provide baseline benchmarks for other more sophisticated models.

1.5 Main conclusions

In this work we demonstrate how the above approaches can be trained on historic satellite data to provide estimates for ground wind speed and wind direction. We show that simple approaches such as IDW, SARIMA, and GPR can provide rapid baseline estimates for both wind speed and direction, and that for a handful of stations these baseline estimates are hard to improve upon. We directly compare these baselines against the alternative approaches, highlighting the performance of the contextual GRU and XGBoost - despite the latter's strong assumptions of intra-time-series independence.

We conclude that whilst the baseline estimates often perform well for estimating wind speed and direction, they fail to generalise well across the full range of stations. In contrast, the XGBoost, LSTM and contextual GRU models exhibit consistent and strong performance for modelling both wind speed and direction across all available stations. We also highlight how the extensibility provided by Regression-Kriging and the GRU would allow both approaches to make full use potential additional stations and features.

Comparing the LSTM and contextual GRU approach, we are also able to show how incorporating both temporal and fixed, spatial information yields the highest and most consistent performance on withheld wind speed and direction data. Moving forward, we recommend that Greenvest consider further development of both the Regression-Kriging and contextual GRU models which can take both temporal and spatial information, using XGBoost as a rapid and capable benchmark for performance and evaluation of feature importance. Recall that we had set a performance benchmark of 6% MAPE at any given station for the challenge. This level of performance could be achieved for several stations using LSTM and the GRU models, but not for all stations.

1.6 Limitations

Each of our methods has different limitations; where some methods perform stronger, others perform poorer, and vice versa. Some of the main limitations are listed below:

1. The IDW, SARIMA and LSTM approaches do not incorporate spatial data (elevation and surface roughness) in their predictions and so potentially miss important predictive information contained in these features.
2. Only the IDW, Regression-Krigging, XGBoost and contextual GRU methods are able to predict the wind speed and direction at an arbitrary location around the ground station, whilst the other methods are restricted to predicting these quantities at the ground station only. However, IDW, Regression-Krigging and XGBoost are unable to forecast wind speed and direction into the future, given current data, whilst the other methods are able to do this.
3. XGBoost and GPR assume that the distribution of wind speed and direction is stationary over time, while in reality we observe that these distributions vary over time.
4. All methods are limited by their training data: some ground stations have poor quality/noisy data, while others only have a short time period of data available. Only the closest 4 satellite points and spatial data bounded by these points was available at each ground stations; our methods may well benefit from the data surrounding a larger area around each ground station, or indeed access to a national spatial mapping. Furthermore combining in other data products, such as CFD simulations or satellite imagery may improve performance.

1.7 Recommendations and future work

Our recommendation is the following two-step process for suggesting best locations to build new wind farms:

1. In the presence of vast addition of new data and information, we recommend first using XGBoost as a type of quick or high throughput model for large number of new stations or new features.

This model is very flexible and inexpensive to run. Not only will this tells us if new data will change the current landscape of energy yield prediction, it will also inform us if new features and information about the ground station are of importance and if these information require more data/resolutions.

2. Once XGBoost shortlists a manageable number of candidate ground stations, we propose using *both* Regression-Kriging and contextual GRU to perform fine-grained and more accurate comparisons across the candidates. The expensive computational cost and time of Regression-Kriging and contextual GRU is ideal for a small set of candidates. Moreover, these models incorporate high resolution spatial information which can be used to distinguish among stations which are otherwise similar in the XGBoost model.

For future work, we briefly summarise a few promising strategies going forward.

The first direction involves optimising the deep neural network models. This could involve finding better-suited deep generative models to capture the wind data distribution. Moreover, in thinking about modelling distribution, there are perhaps much efficient representations of the information (in the distribution) which can be leveraged to build more efficient models in this class.

The second future improvement involves building hybrid models, incorporating all simple but less accurate models, such that these ensemble models could incorporate all information more efficiently. We believe that such ensemble models will amplify the strengths of some of the models as well as being flexible enough to include new data or new models.

A third future improvement includes conditioning more of our models on the location of the ground station, so that they could be used to predict the wind speed and direction at an arbitrary location instead of only being able to predict these quantities at the ground station only.

Finally, in an effort to explain or interpret predictions from the deep learning, we propose using the LIME and SHAP methods (and other sophisticated extensions of them). These methods provide a first glimpse of what features matter most in making wind speed prediction. Our hope

is that this will in turn inform us of how to build accurate but smaller and much efficient models.

It would also be very interesting to investigate how well our approaches perform for different countries.

We believe we have learnt many valuable lessons from the reasonably many models that we have considered. The recommendations and future proposals above distil this knowledge into actionable plans. Our hope is that our work provides a lamp-post for the wind-energy industry to better leverage data in their future decision-making.

2 Data overview

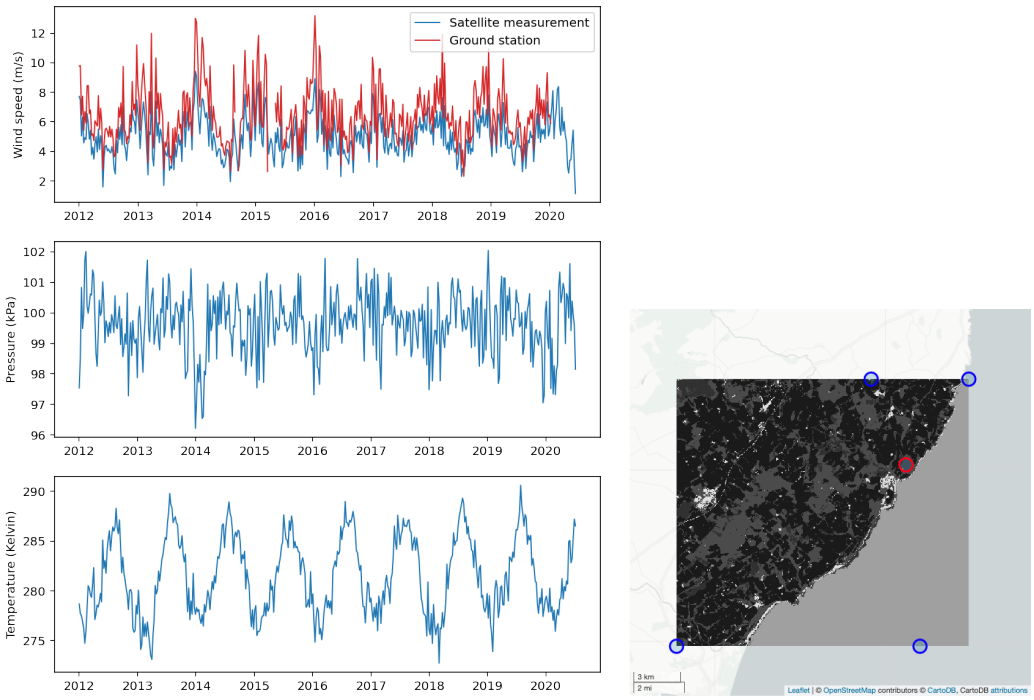
2.1 Dataset description

The dataset comprises satellite measurements in the vicinity of 172 meteorological ground stations as well as measurements from each ground station.

The satellite measurements available at each station include both time-series and spatial data. The time-series satellite data includes measurements of wind speed, direction, atmospheric pressure and temperature at the 4 closest grid points to the ground station, on a sparse 9km grid (Figure 2a). The spatial satellite data includes a high spatial resolution elevation model and surface roughness classification, both on a common grid and bounded by the 4 time-series locations (Figure 2b).

The available ground station measurements include wind speed and direction, and are taken as ground truth measurements.

The satellite time-series data is available at a hourly resolution, and typically over the last 8 years. The ground station measurements also have a hourly resolution, and are mostly available up to 8 years.



(a) (b)

Figure 2: Example ground station (red point/line) with spatial and temporal satellite data plotted. (a) Example temporal data available at each satellite point, plotting weekly mean for upper left satellite point. (b) Spatial data. The 4 available satellite points of time-series measurements (blue markers) and surface roughness classification plotted.

2.2 Data quality

Generally, we ascertain the quality of the data from the following two grand perspectives:[1]

1. **Missing data and completeness**

   The amount and the reason for data missing from the dataset are of paramount importance to the task of modelling. If there is a

[1]Other dimensions of data quality include timeliness, relevance, uniqueness and validity. Since this set of data has been carefully curated for the purpose of this project, these criteria are either irrelevant or have been conscientiously satisfied.

significant amount of missing data for a particular ground station or time-span, we might choose to discard such a subset of data. This not only will affect the performance of the models during training or model-building, it will hamper future generalisations of the models to unseen data. Moreover, as much as one can, one should obtain and understand the reasons behind significant missing data. For example, if missing data for a particular ground station is due to sensor malfunctioning, this also casts a suspicious light on the rest of the available data from this particular ground station. In addition, if we further knew that sensor malfunctioning is due to the remote nature of the station (probably due to infrequent maintenance), this should encourage us to scrutinise in greater details data coming from ground stations which are more remote.

2. **Available data**

   We focus our analysis of data quality on the following 2 criteria:

   - **Consistency**

     Do we adhere to the same standard and protocols of data collection? For the same quantity of measurement, are they all recorded in the same units and precision. In particular, are wind-direction angles all measured with the same convention (i.e. measured east from true-north)? Moreover, are they all with the same precision?

   - **Accuracy and reliability**

     How do we know that the readings are accurate? Are there human manual-entry error? Are there machine outage/power outage that would affect accuracy ? Without a comprehensive conversation and feedback from the data source, this is a tricky issue to tackle.

   Next, we will proceed to analyse and discuss the quality of our dataset in terms of the categories above. In each of them, as we encounter issues of data quality in said category, we will elaborate on how we interpret the issues and our approaches to deal with them.

2.2.1 Completeness

As an overview of the data, there are 164 stations with a maximum of 74497 hourly time-point data, while 7 stations have 75958 rows, and 1 station has 83383 data points. Since the first case is the typical case, we shall focus on that in this discussion. The 74497 number of points correspond to hourly data from 2012/01/01 to 2020/07/01, i.e., around 8.5 years of data. This data comes mainly from the satellite geographical data, although the satellite wind data are one-month shy of this range.

For the ground station, the data range is not very consistent. Typically, it starts somewhere in 2012 or 2014, and provide hourly data up for about a period of 8 years. The minimum time-span of data is around 4 years.[2]

To discuss the data quality in terms of completeness, this seems to depend on what we set out to accomplish. If we follow the maximum time-span of data, which is the satellite geographical data, then around the order of 10-20 ground stations will have a range of 20%-50% of missing data. Any model-building should be mindful of such stations, if they would like to perform analysis of the full 8.5 years. Alternatively, for certain analyses (such as supervised machine learning models), we need to have data both for the predictor (for e.g. the ground wind speed) and the features (for e.g. the satellite data). In this case, focusing on timespan where *both* types of data are available is the sensible way to proceed. By definition, such an approach will encounter no missing data. However, one should be mindful of the subtleties: Firstly, due to the non-uniform start dates of the stations, making inter-station comparison is not straight-forward due to year-to-year trends. Secondly, even within the ground station data, there might be hourly periods where significant data are missing. This happens to several stations, where there are about 15%-50% of missing data. Whenever appropriate, we will discard such stations due to sharp discontinuities in the time-series data caused by long periods of missing data.

The satellite measurements, though not always consistent with the time range, contain no missing data. This is due to the availability of more than one independent satellite measurements.

[2]Finer details: 151 stations have data starting in 2012, while 13 stations start at 2014. One station starts in 2016, 2 start in 2015 and 5 start in 2017.

Regarding data availability, we observed the distribution of data availability (in terms of hours) available for each station considering the two types of data: satellite and ground station data. The result in Figure 3 shows that all 172 stations have more than 70000 hours of satellite data, and the majority of stations have around 70000 hours of ground data. However, there are a few stations with exceptionally low number of hours available, mostly in the range of 20000 and 30000 hours.

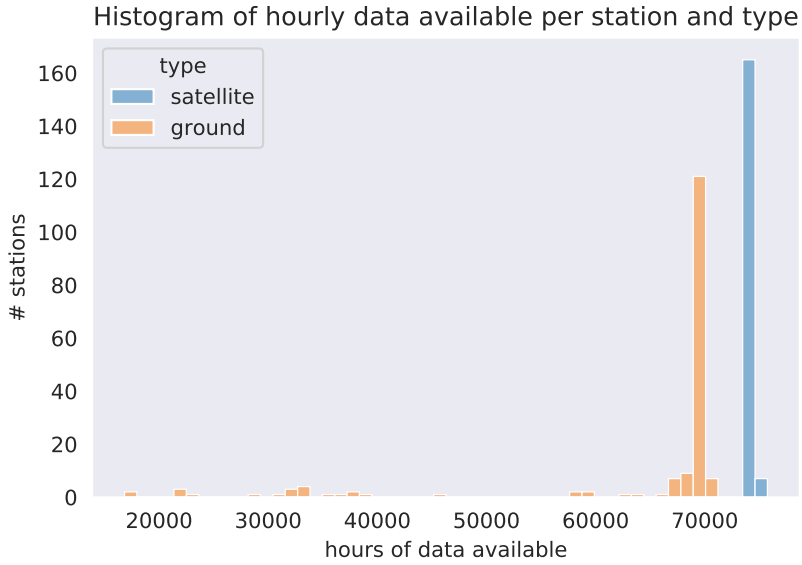


Figure 3: Data availability per data type.

2.2.2 Consistency

There are two key issues which we have discovered regarding the consistencies of the data:

1. **Misalignment of wind directions**

   We observed that the satellite wind direction data are persistently 180 degrees to the one of the ground stations. This has enabled us to further communicate to the data provider and confirm the misalignment.

2. **Granularity**

   The wind-speed data provided by the ground stations are originally measured in knots. In converting to meter/second (m/s), this has introduced discreteness and discontinuity in the data. In details, the speed measurements (in m/s unit) are discrete with step-size of about 0.514444 m/s.

   Regarding the satellite speed measurements, the raw data seems to be precised up to 6 decimal points. Further discussion with the data provider confirms that indeed that is an artefact of intermediate data processing that yields artificially high precision. A reasonable estimate of the instrumental precision of the satellite speed measurement should be in the order of 0.1 m/s.

   The granularity of both ground station and satellite measurements might have implications on the modelling and performance metric. We have not done perform any comprehensive analysis to study this issue, though we acknowledge it will be an interesting future direction to pursue.

2.2.3 Accuracy and Reliability

In the absence of comprehensive first-hand experience with the measuring devices as well as independent accuracy verification of the data, it is not clear how to proceed to verify the accuracy the data. However, in our exploration of the data, we have found one important accuracy issue, which has significant effect to the MAPE score.

First, we have studied how often the mean satellite speed (average wind speed across four satellites computed every hour) dips below one knot (i.e., 0.514444 m/s). This is an indication of how frequent we expect ground-station wind speed to register a 0 value. In all stations, this happens below 2% of the time.

For the ground station wind speed measurements, however, we found that many stations have recorded zero speed (since its converted from knot unit) measurements more than 5% time. In fact, several stations have 20-50% data being zero speed. By looking at some of the time-series of such zero speed data (see Fig. 4 for an illustration of how it looks like) and with

further confirmations from data provider, we conclude that ground wind speed measurements of zero are mostly errors.

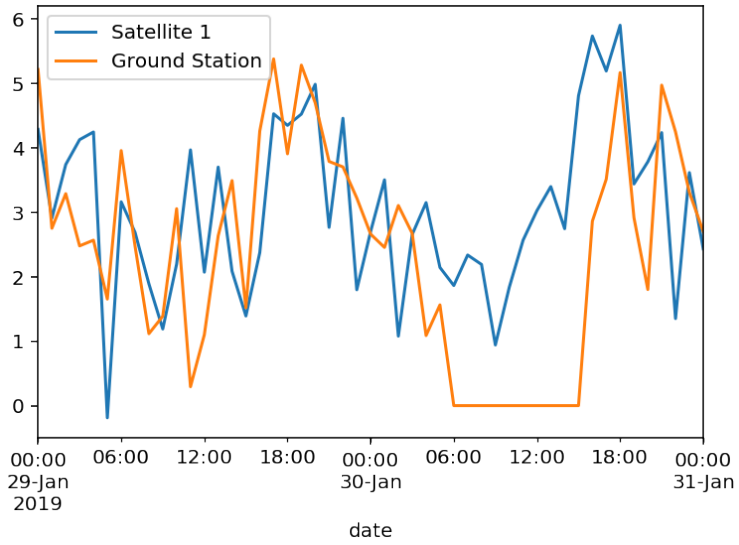


Figure 4: An illustration of the zero speed data. This is a plot of simulated data for wind speed (m/s) vs time (hour). We can clearly see that the satellite reading track the data from the ground station except a period of about 10 hours where the ground station reading is completely zero. We discard this zero speed period, since it probably is due to instrumental/human errors.

For our purpose, even a single zero-speed wind data in MAPE score will skew the whole score to arbitrary large value. Therefore, we choose to discard any zero-speed data. This not only will avoid errors from the ground-station, this will avoid small wind-speed (which is about 2% of the satellite data) to dominate the average MAPE. We believe this choice will help better represent the physical wind speed as well as providing a better representation of how stable and high the wind-energy-yield.

3 Data exploration & analysis

This section describes different venues of data exploration and analysis, each subsection focus on a particular aspect of the dataset.

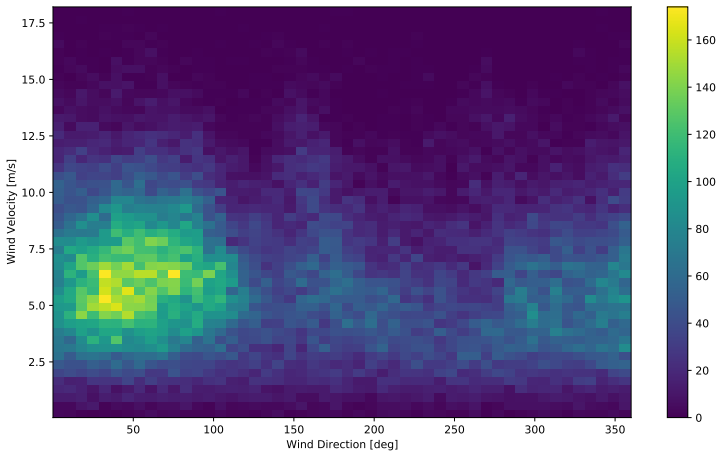
3.1 Decomposing wind speed and direction

Joint distributions of wind speed and direction allows us to intuit broad wind behaviours. For example, Figure 5a shows a heatmap of the wind speed and direction as measured over multiple years at a randomly selected ground station. From this alone, we can identify prevailing wind conditions at a given site. This joint distribution also highlights one potential issue for estimating wind speed and direction, namely that degrees can be a poor feature choice owing to its discontinuity between 359 and 0 degrees. Clearly, a wind direction of 5 degrees is close to a wind direction of 355 degrees (they're both just 5 degrees from north), but in the given form models may struggle to learn this relationship and accurately predict future changes in wind direction. Additionally, it may be argued that the direction of the wind speed does not matter when the wind is not blowing.

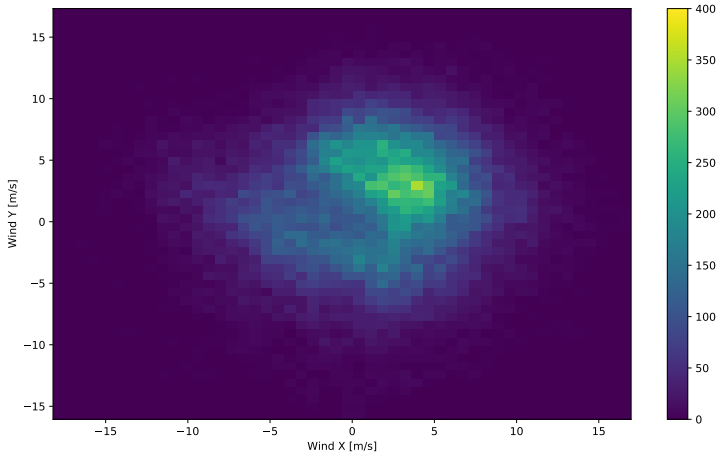
As a solution, it is often suggested to decompose the wind speed and direction measurements to form a wind x component and a wind y component. As shown in Figure 5b decomposing wind speed and direction in this manner eliminates the wrap-around issue associated with using wind direction in degrees directly, as well as engineering features that are easier for machine learning models to interpret. We explore the effects of these feature transformations on model performance in Section 4.5.1.

3.2 Fourier transforms and periodicity

Periodic behaviour is a strongly conserved characteristic of meteorological data. Clear day/night cycles, which we might expect to affect measurements of temperature for instance, we also expect seasonal patterns in precipitation and wind measurements that repeat at yearly intervals. To determine the most important frequencies for the periods contained in our data set, we applied fast Fourier Transforms (FFT) to key time-series variables; namely wind speed and temperature. The FFT transforms the time domain signal into a representation in the frequency domain, which we can examine to test for seasonal effects and to check for systematic errors in the time-series measurements. Frequency (log score) is plotted along the x-axis against the input



(a) Joint distribution of original wind direction and wind speed for station A.



(b) Joint distribution of decomposed wind x and y components for station A.

Figure 5: Heatmaps visualising the impact of, and motivation for, decomposition of wind speed and direction features to independent x and y components.

time-series variable (in the original units). The resulting frequency plots, here shown for a randomly sampled station, highlight the daily and yearly cycles captured in the time-series measurements for temperature [Figure 6]. We can also see how these periods are conserved between satellite estimations of wind speed and their corresponding ground truth station measurements [Figure 7, 8].

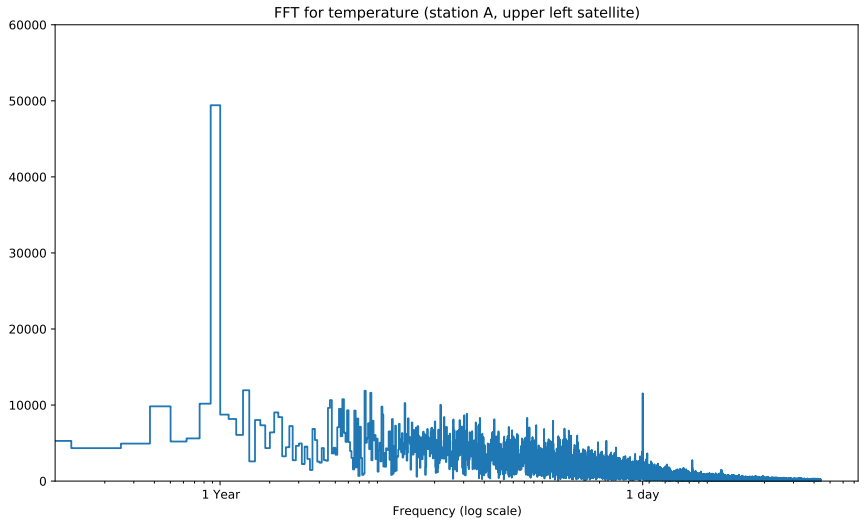


Figure 6: FFT of satellite wind speed time-series for the upper left satellite associated with station A shows daily and yearly periodicity.

3.3 Distribution of features over all stations

To have a comprehensive understanding of the data, namely the minimum, average and maximum values for each feature we observed the distributions of all satellite features (wind speed, wind direction, pressure and temperature) and ground station wind speed and direction. For simplicity we consider the value of each satellite feature as the average among all satellites, which is justified by the high correlation among satellites. These histograms consider hourly time steps where all features are available, i.e., if ground station speeds/directions are missing

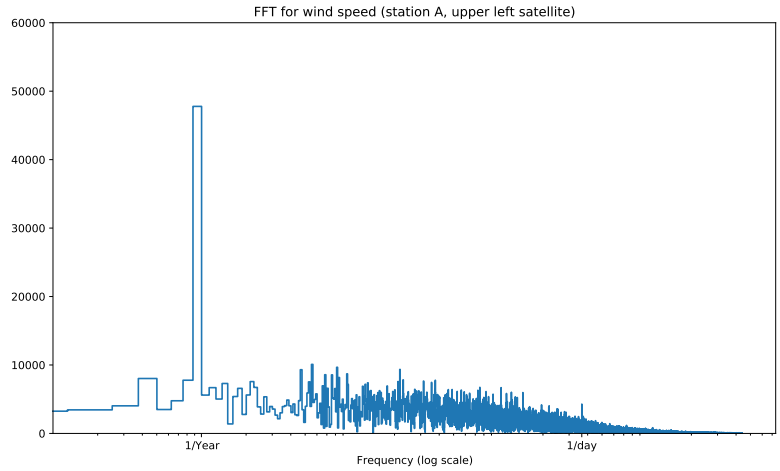


Figure 7: FFT of satellite wind speed time-series for the upper left satellite associated with station A.

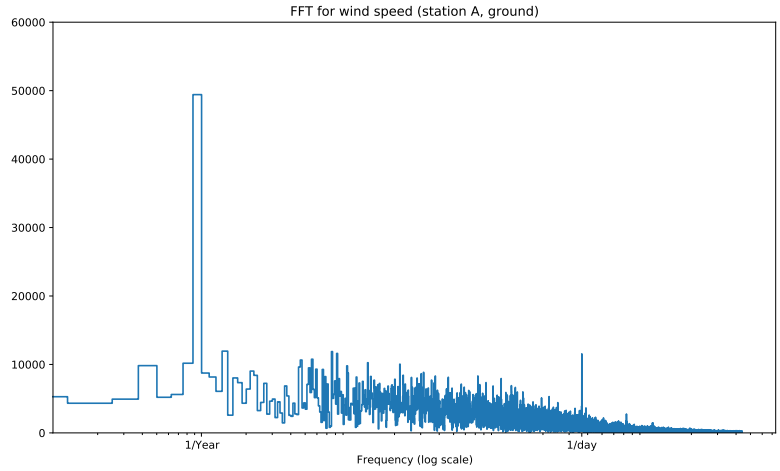


Figure 8: FFT of ground station wind speed time-series for station A.

the satellite measurements are ignored. The resulting histograms are observed in Figure 9.

The analysis of Figure 9 allows observing some characteristics of the data. Firstly, there is a clear correlation between satellites and ground station speeds and directions, although the ground measurements

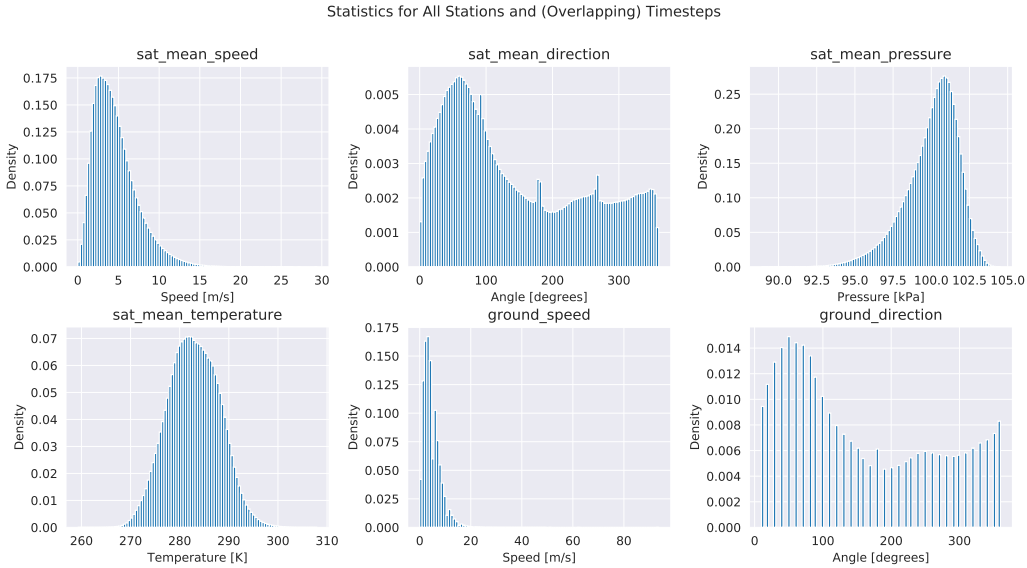


Figure 9: Histogram of all statistics (mean for satellites) when considering overlapping time steps.

3.4 Yearly wind statistics

We are interested in forecasting wind speed and direction over a periods of years, so firstly we observe how the histogram of wind speeds and direction change over periods of one year per station. For this visualisation we selected all samples where satellite data and ground station data overlap and plotted their histograms for each year. We also create an approximated probability density function per feature per year using a Gaussian kernel density estimation. Figures 10 and 11 show the yearly histogram of wind speed and direction, respectively, to station 24125, which was randomly selected among the base stations that had a large number of observations (see Figure 3). Each row in the plot represent a year and the columns represent the feature considered, e.g., ground speed or satellite speed. Note that for each station there are four

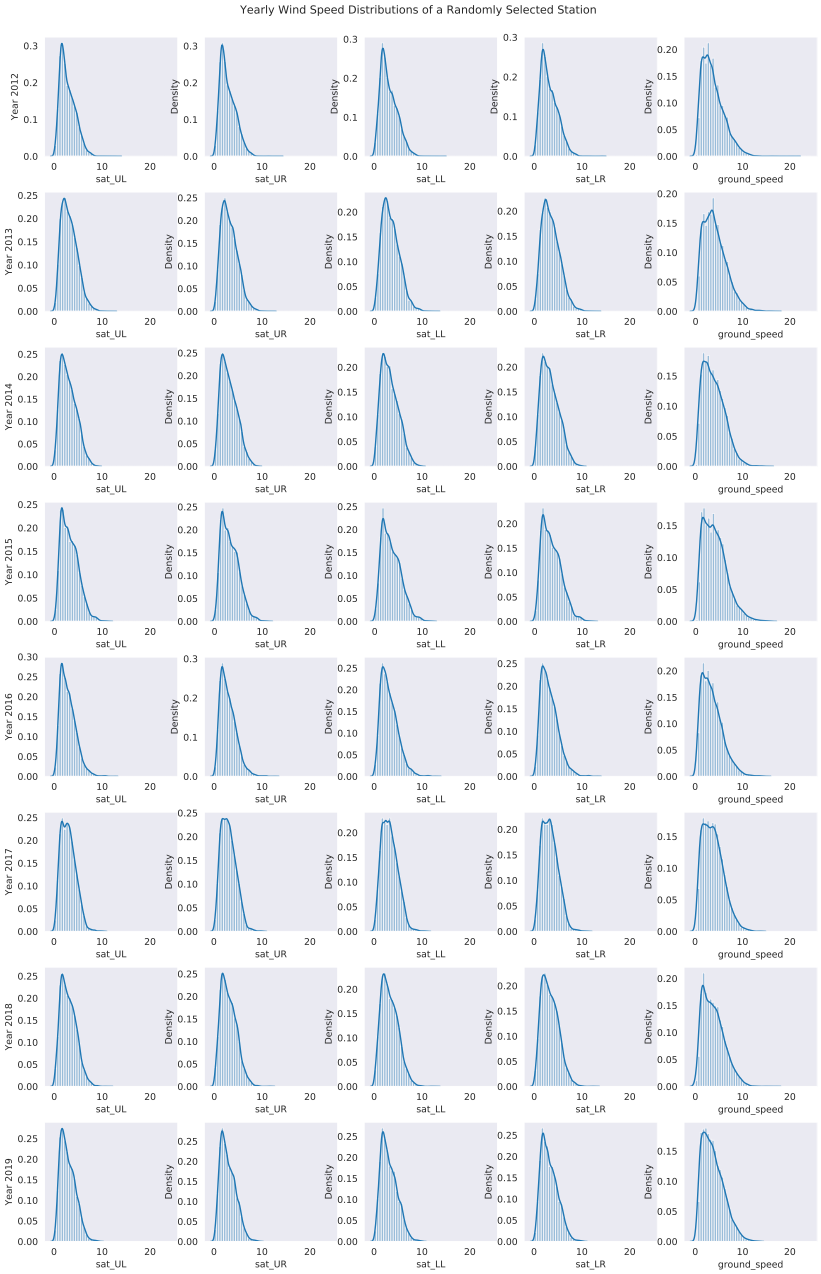


Figure 10: Yearly wind speed histogram for a randomly selected station.

nearby satellites, denoted by Upper Left (UL), Upper Right (UR), Lower

Left (LL) and Lower Right (LR).

The histograms show high correlation among all satellites and base stations measurements, specially for wind speed. The ground station wind direction slightly deviates from the satellites measurements, however follow the general trend. The reason for this deviation could be two-fold. Firstly, the terrain could impact on how the wind flows from the areas measured by the satellite to the station point. Secondly, the ground station wind direction sensor can only measure angles in discrete steps of 10 degrees, while the satellite estimates are continuous. The key observation is that the wind speed and direction distributions are almost static over time, which justifies stationary models assumptions.

3.4.1 Wind Speed and Direction Statistics

We can construct a simple parameterised model of the distribution of wind-speed and direction over a particular time period by assuming a) wind speed and direction are independent, and b) wind statistics between stations are independent.

Since wind speed s is a real-valued, positive quantity, i.e. $s \in \mathcal{R}^+$, one choice of parameterised distribution is the Gamma distribution:

$$P(s_i; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s_i^{\alpha-1} e^{-\beta s_i} \tag{1}$$

where Γ is the gamma function, α is a shape parameter, β is a rate (or inverse-scale) parameter. An advantage of the gamma distribution over other skewed distributions with positive support is that parameter estimation is simple and can be incorporated very easily into Bayesian models with conjugate priors, meaning parameter estimation can be made dynamic.

Wind-direction is a quantity with support over one period of a cosine, conventionally $\theta \in [-\pi, \pi)$. As such it can be modelled by a circular distribution. The most common choice for a circular distribution is the Von-Mises distribution due to its tractability. The Von-Mises distribution is defined as:

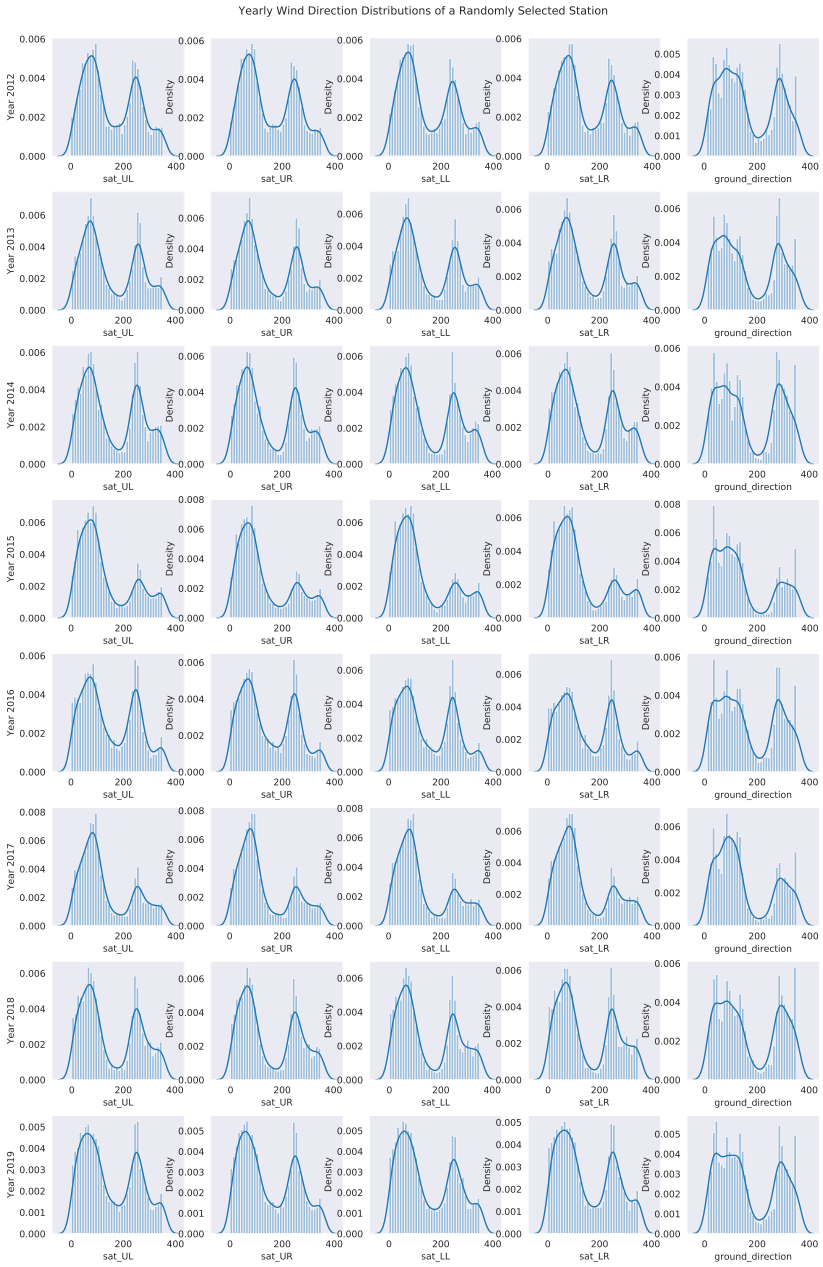


Figure 11: Yearly wind direction histogram for a randomly selected station. The satellite measurements are highly correlated and the distribution of wind speed and direction do not significantly vary over the years.

$$P(\theta_i; \mu, \kappa) = \frac{e^{\kappa \cos(\theta_i - \mu)}}{2\pi I_0(\kappa)} \tag{2}$$

where μ is the mean direction, and κ is analogous to the variance.

Fast maximum-likelihood parameter estimates can be performed with a variety of statistical packages. In our case, we use `scipy.stats`.

3.5 Map visualisation with wind statistics

Observing the histogram of wind speeds and directions can provide intuitive understanding of the behaviour of these variables over time. Similarly, visualising the spatial distribution of the stations and their respective wind statistics allows understanding the behaviour of data in the spatial dimension.

We then aggregate wind statistics to the map visualisation in Figure 12 where each station is represented by a needle marker. The needle is placed over each station coordinates and points towards the the median wind direction, *i.e.* the direction the wind blows to, over all time samples of that particular station. Furthermore, the needle colour indicates the mean wind speed over the respective station, ranging from blue to red, indicating low and high speeds (0 m/s and 15m/s), respectively.

Figure 12 allows observing that the prevalent wind direction (the direction the wind blows to) in the UK is towards the west, which is also in agreement with the ground wind direction histogram (peak at 90 degrees) in Figure 9. It also shows the tendency of having stronger winds towards the coast rather than inland.

4 Experiments

4.1 Metrics

All of our approaches described below make use the following metrics to assess model performance.

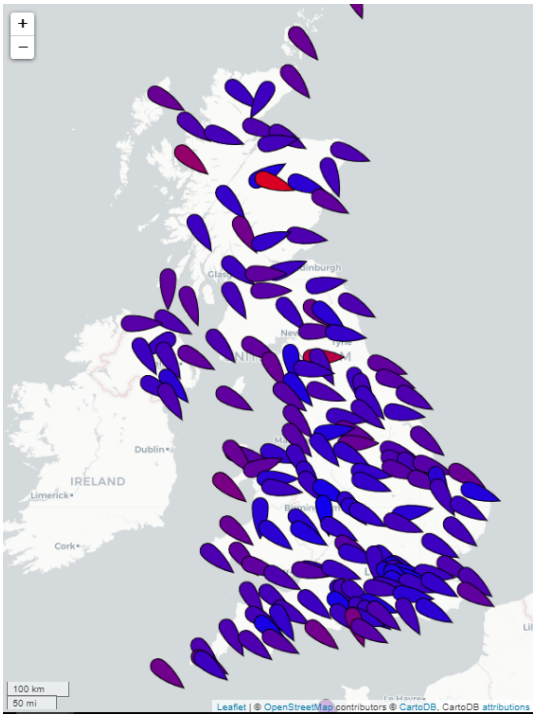


Figure 12: Summary of wind statistics per station. Needles indicating mean wind speed and directions. The needle colours indicate mean wind speed and are interpolated between blue (0 m/s) and red (15 m/s).

4.1.1 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is defined as the ratio in the form:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{3}$$

where A_t is the actual value at time t, F_t is the forecast value, n is the number of samples. This metric is used to assess the accuracy of wind speed prediction. Throughout our report, MAPE is reported in percentage.

4.1.2 Cosine Difference

Cosine difference is defined as 1 minus the cosine of the difference between predicted and actual angles (direction). That is,

$$\text{Cosine Difference}(\theta_{\text{predicted}}, \theta_{\text{actual}}) = 1 - \cos\left(\theta_{\text{predicted}} - \theta_{\text{actual}}\right) \quad (4)$$

This metric is used to assess the accuracy of wind direction prediction, where $\theta_{\text{predicted}}$ and θ_{actual} stand for the predicted angle and actual angle of wind direction respectively. When computing this over a time-series, the median value is reported.

4.2 Baseline approaches

In order to establish a baseline of what was possible to predict in a short timescale, we trialled a three different approaches that could be easily set up. In this section, we describe these approaches, and their accuracies. These are: a satellite mean approach/IDW, a seasonal autoregressive integrated moving average model, and a gaussian process regression method.

4.2.1 Inverse Distance Weighed Interpolation (IDW)

The trivial version of Inverse Distance Weighed Interpolation (IDW) takes the average of wind speed and direction of surrounding satellites to approximate wind of each ground station. We call this approach Satellite Mean Approach. Considering distance, the nontrivial version of this approach takes weighted average as an approximation.

Methodology

Our baseline approach can be summarised as follows:

1. For each station and hourly time instant, use data from both the ground station (only time stamps) and 4 surrounding satellites as input.
2. Take the intersection of time stamps of satellite and ground wind data and truncate both series.

3. Calculate (weighted) average mean speed and direction of satellites to gain approximation for the ground wind.
4. For a specific time frame (e.g., hour, day, week, month, year), take the mean over the time period to generate a final time-series as output.
5. Discard samples with missing values (and samples with small ground speed).
6. Compare the satellite mean/IDW output with the actual ground wind and evaluate the results.

The weighted average approach we use here is Inverse Distance Weighted Interpolation (IDW) [6]. We use different weights for four surrounding satellites to calculate weighted average of wind data to approximate ground wind. That is,

$$x_{\text{pred}} = \sum_{i=1}^{4} w_i \times x_{\text{satellite } i} = \frac{\sum_{i=1}^{4} \frac{1}{d_i^p} \times x_{\text{satellite } i}}{\sum_{i=1}^{4} \frac{1}{d_i^p}} \tag{5}$$

where x_{pred}, x denotes predicted wind data, $x_{\text{satellite } i}$ denotes the satellite wind data from satellite i, d_i denotes the distance from satellite i to the ground station, p is some power, $p = 0, 1, 2, 3, i = 1, 2, 3, 4$. When $p = 0$, this becomes the baseline Satellite Mean Approach, viewing all surrounding satellites as equally important for the ground wind prediction.

4.2.2 SARIMA

Autoregressive Integreated Moving Average, (ARIMA) is one of the most widely used traditional methods for time-series data forecasting which considers the serial correlation among observations. Seasonal ARIMA (SARIMA) is an extension of ARIMA that supports seasonal component of the series. Here we choose SARIMA as the baseline for the temporal part of the problem.

Methodology

The main idea of this method is to firstly use satellites wind speed and direction time-series as input to predict the upcoming time step satellite

data, and then interpolate the satellite predictions to the ground station using spatial interpolation methods (satellite-mean approach, Inverse Distance Weighted approach with different distance powers), as illustrated in Figure 13. To fulfil this target, the SARIMA method applied here is simply to do uni-variate SARIMA time-series forecast for each satellite.

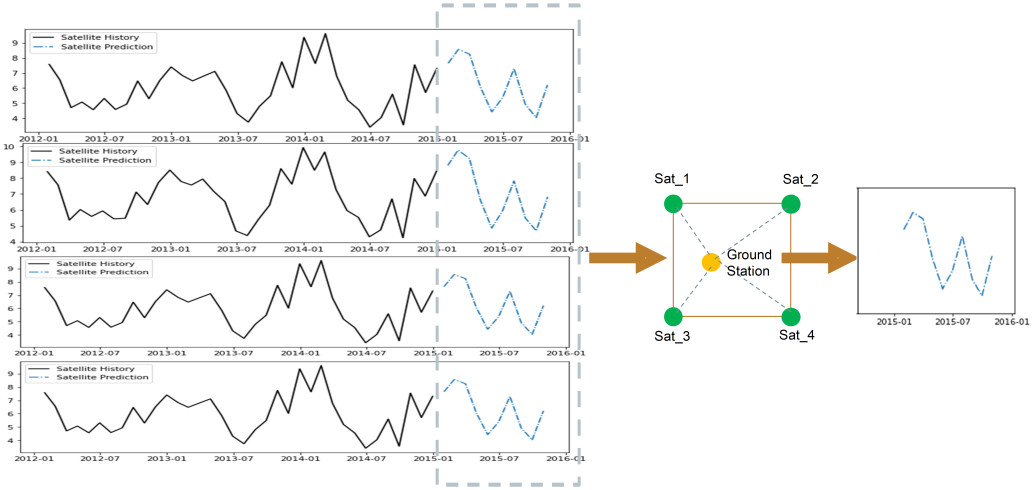


Figure 13: Baseline Approach Flowchart

A concatenated table combining ground station wind speed and direction and also its surrounding 4 satellites wind speed and direction with overlap time index is generated for the time-series prediction.

In this experiment, monthly predictions are modelled. So we firstly re-sample the hourly data to a monthly time-series data and check whether the re-sampled time-series are continuous since time-series model would avoid the discontinuous part of data.

The hyper-parameters for the SARIMA are as follows:

- p and seasonal P: indicate trend and seasonal autoregressive order
- d and seasonal D: indicate trend and seasonal difference order
- q and seasonal Q: indicate number of moving average terms (lags of the forecast errors)
- m: the number of time steps for single seasonal period

- t: indicated seasonal length in the data

For each satellite time-series, we undertook a grid search of these hyper-parameter, and choose the hyper-parameter groups which perform the best for predicting wind speed in the whole dataset using walk forward validation.

As part of the baseline, we choose four representative ground stations among all 173 stations which performs good (station A), moderate (station B), a bit weak (station C) and really bad (station D) respectively in the IDW, to see how the performance change among different methods. The SARIMA hyper-parameter groups for the predictions of chosen stations can be shown in Table 1.

Table 1: SARIMA hyperparameter setting

Station ID	Sat. ID	p [0,1,2]	d [0,1]	p [0,1,2]	P [0,1,2]	D [0,1]	Q [0,1,2]	m [0]	t ['n','c','t','ct']
A	UL	2	0	2	0	0	0	0	't'
A	UR	2	0	2	0	0	0	0	't'
A	LL	2	0	2	0	0	0	0	't'
A	LR	1	0	2	0	0	0	0	'c'
B	UL	0	1	2	0	0	0	0	'n'
B	UR	0	1	2	0	0	0	0	'n'
B	LL	1	0	2	0	0	0	0	'c'
B	LR	1	0	2	0	0	0	0	'c'
C	UL	1	0	2	0	0	0	0	't'
C	UR	2	0	1	0	0	0	0	't'
C	LL	0	1	2	0	0	0	0	'ct'
C	LR	1	0	2	0	0	0	0	't'
D	UL	2	0	1	0	0	0	0	'c'
D	UR	2	0	2	0	0	0	0	'ct'
D	LL	1	1	2	0	0	0	0	'n'
D	LR	2	0	1	0	0	0	0	'ct'

For each prediction model, we also have considerations as follows:

1. Look back time steps. For parameter study, we generated individual models with different look back time steps (12 months, 18 months, 24 months, and 36 months), and compare their results of predicting the upcoming month wind speed and direction.
2. Spatial interpolation methods. Same as the Satellite Mean Approach, the Inverse Distance Weighted interpolation with the power from 0 to 3 are all considered to map the predicted satellite data to ground station.
3. Evaluation matrics. For wind speed prediction in each case, the results are evaluated by the MAPE with the actual ground station, while the results of wind direction are evaluated by the cosine difference between the ground station and prediction.

4.2.3 Gaussian Process Regression for Mean Year Prediction

We have found that, for each of the individual ground stations and satellites, every year of data differs from the mean year of data for that station by less than 8%. This can be seen in Figure 14. Accordingly, one promising, simple approach would be to predict the monthly ground station data for this 'average year', given the monthly annually-averaged wind, temperature, and pressure data for the four corresponding satellites, the digital elevation model of the area surrounding the ground station, and the roughness of the terrain surrounding the station. To accomplish this, we use Gaussian process regression, as it able to provide uncertainties for its predictions and is more versatile than linear regression (see [9] for a detailed overview). This is a simple baseline that only make reasonably accurate prediction at year-level. Note that in this approach, we ignore the wind direction.

Methodology

This approach consisted of the following steps:

1. Firstly, for each station, we calculate the annual mean values of the satellite wind speeds, the ground station wind speeds, the temperatures, and the pressure values for each month of the year.

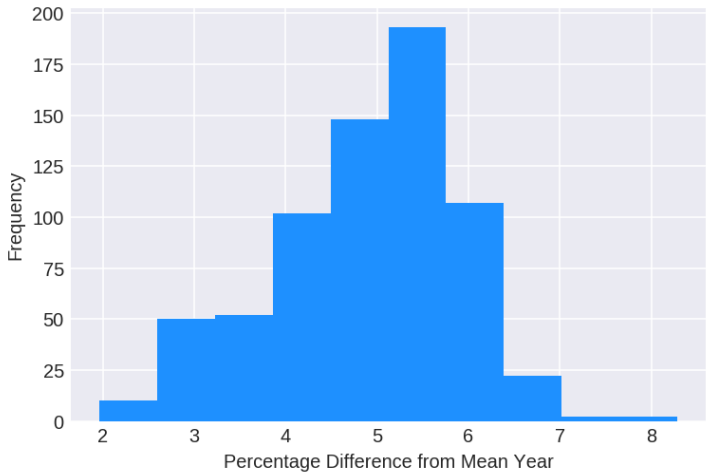


Figure 14: Distribution of the percentage difference from the average year, for all stations.

2. We then create three new features by taking the mean, standard deviation, and maximum value of the digital elevation model maps available for each station. Furthermore, 10 new features are added by taking the number counts of each type of terrain from the roughness map for the station.

3. We then divide up the data set into a training and testing set. The training set contains 157 stations, while the test set contains 15. In order to improve prediction accuracy, the target data, which is the ground wind speed, is scaled to have a mean of zero, and unit variance.

4. Gaussian process regression is then applied to the training set. In this process, the radial basis function kernel is used, and hyperparameters are automatically optimised by maximising the log-marginal-likelihood.

5. Finally, the trained model is used to predict the test set. These predictions are then compared to the true values, and the MAPE is computed.

4.2.4 Baseline results and discussion

We present the results of applying the various baseline approaches discussed above.

IDW : We use MAPE described in Section 4.1.1 as a metric for wind speed and cosine difference for wind direction. To avoid explosion of MAPE to infinity due to tiny ground speed, we either add a small ϵ to the denominator of MAPE, or discard samples with tiny ground speed before performance evaluation. For the rest of the performance evaluation, we consider MAPE discarding zero ground speed.

Table 2: IDW mean MAPE(%) for different time frames and IDW powers

IDW power	hour	day	week	month	year
0	50.0559	33.2832	28.3174	27.2436	25.4940
1	50.9214	33.9991	29.0275	27.9567	26.2262
2	51.6096	34.5760	29.5970	28.5443	26.8429
3	52.0350	34.9435	29.9658	28.9267	27.2901

Table 3: IDW mean cosine difference for different time frames and IDW powers

IDW power	hour	day	week	month	year
0	0.036783	0.041440	0.039183	0.036168	0.037723
1	0.037067	0.041712	0.039315	0.036491	0.038131
2	0.037213	0.042023	0.039562	0.036780	0.038416
3	0.037275	0.042260	0.039744	0.036918	0.038579

For wind direction evaluation, we use Cosine Difference described in Section 4.1.2, which is 1 minus the Cosine Similarity between predicted value and the actual value. We take the median over samples to report the performance, since the distribution over samples is skew. The mean MAPE results and Cosine Difference discarding tiny (zero) ground speed are shown in Table 2 and 3. From these tables, we conclude that in terms of wind speed prediction, this baseline approach performs better in long-term prediction than short-term prediction. In terms of wind direction, time frame does not play a vital role. As IDW power increases, the performance on both wind speed and direction gets slightly worse.

Table 4 and 5 illustrate the minimum (among stations) MAPE results and Cosine Difference discarding tiny (zero) ground speed. Similar to the average performance, in terms of wind speed prediction, IDW has better performance in long-term prediction than short-term prediction for wind speed. However, for both wind speed and direction, the best performance improves as the we move from short term to long term. Besides, as IDW power increases, the performance on wind speed slightly improves. Best performance for wind direction gets slightly worse in short-term prediction (hour, day and week) while better in long-term prediction.

Table 4: IDW min MAPE(%) for different time frames and IDW powers

IDW power	hour	day	week	month	year
0	19.8827	10.0824	5.5641	3.7578	1.1262
1	19.8550	10.0814	5.5616	3.7876	1.0905
2	19.8430	10.0822	5.5630	3.8035	1.2666
3	19.8403	10.0857	5.5647	3.8102	1.3121

Table 5: IDW min cosine difference for different time frames and IDW powers

IDW power	hour	day	week	month	year
0	0.010208	0.007311	0.003168	0.001611	0.000089
1	0.010363	0.007400	0.003395	0.001611	0.000046
2	0.010484	0.007449	0.003532	0.001475	0.000013
3	0.010560	0.007629	0.003613	0.001454	0.000032

The maximum (among stations) MAPE results and cosine difference discarding tiny (zero) ground speed are given in Table 6 and 7. With respect to the worst performance, in terms of wind speed prediction, time frame does not make much difference. As for wind direction prediction, our baseline approaches suffer less extreme bad performance in short term than in long term. Similar to the best performance, as IDW power increases, the performance on wind speed slightly improves. However, IDW power does not play much role in wind direction prediction.

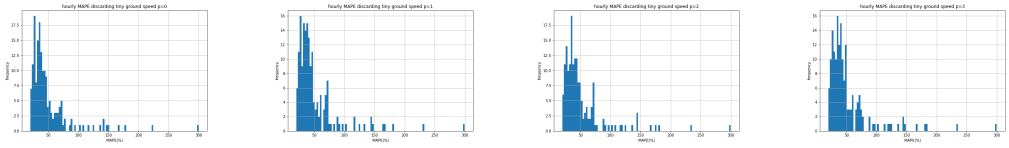
Table 6: IDW max MAPE(%) for different time frames and IDW powers

IDW power	hour	day	week	month	year
0	299.9242	338.0218	319.7430	309.2419	284.6664
1	299.8398	337.8149	319.5934	309.0856	284.6019
2	299.7737	337.6642	319.4809	308.9685	284.5468
3	299.7363	337.5822	319.4184	308.9038	284.5137

Table 7: IDW max cosine Difference for different time frames and IDW powers

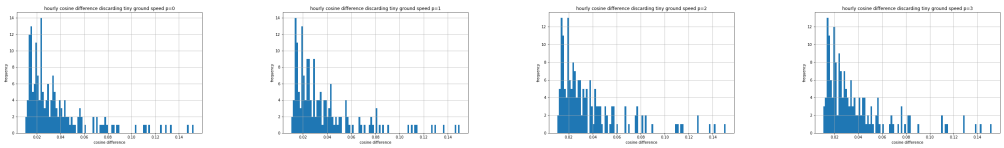
IDW power	hour	day	week	month	year
0	0.152513	0.240515	0.473385	0.468381	0.497541
1	0.149749	0.238789	0.475957	0.472661	0.499384
2	0.150728	0.243671	0.478325	0.476524	0.501052
3	0.152003	0.240573	0.480034	0.479717	0.502422

Histograms of MAPE and Cosine Difference with respect to different IDW powers p and different time frames are shown in Figures 15, 16, 17, 18, 19, 20, 21, 22, 23 and 24. From these plots, we conclude that most stations perform quite well in predicting wind speed or direction. As we increase the length of time frame, even more stations seem to perform well. However, in all cases, there are some stations where our baseline approaches fail to give satisfactory prediction. This may be due to complex terrain or data quality issues. As a baseline spatial interpolation approach, it requires little information, and it does not involve temporal forecasting. However, this model fails to deal with stations with complex terrain, and can not model distinction over time.



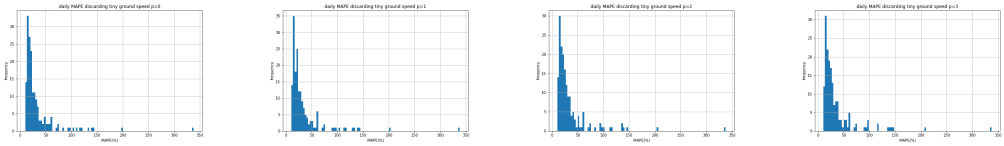
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 15: Hourly IDW MAPE Histograms with Different Powers p



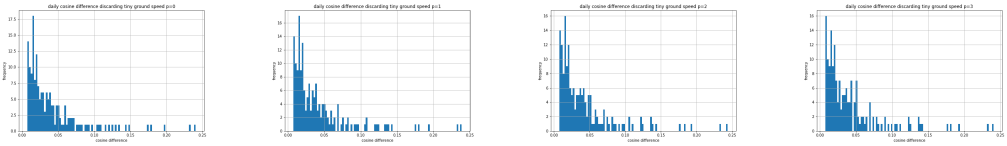
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 16: Hourly IDW Cosine Difference Histograms with Different Powers p



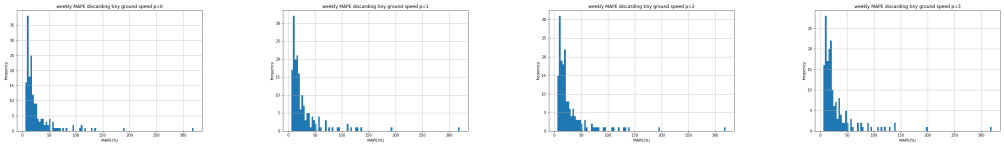
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 17: Daily IDW MAPE Histograms with Different Powers



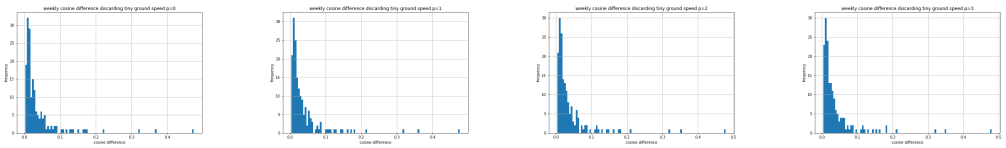
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 18: Daily IDW Cosine Difference with different powers



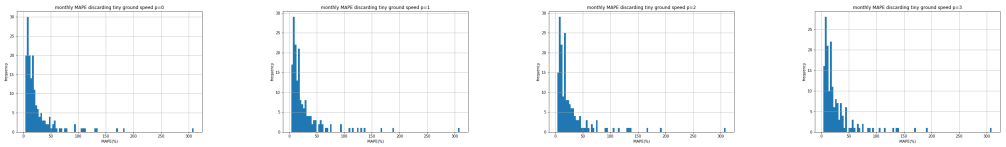
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 19: Weekly IDW MAPE Histograms with Different Powers



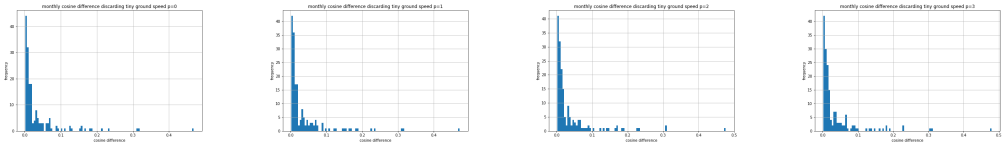
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 20: Weekly IDW Cosine Difference Histograms with Different Powers



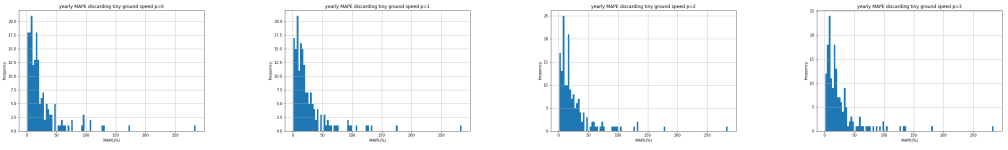
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 21: Monthly IDW MAPE Histograms with Different Powers



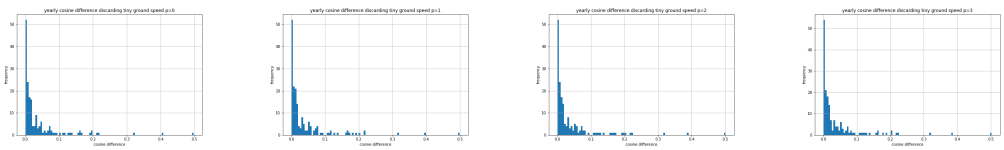
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 22: Monthly IDW Cosine Difference Histograms with Different Powers



(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

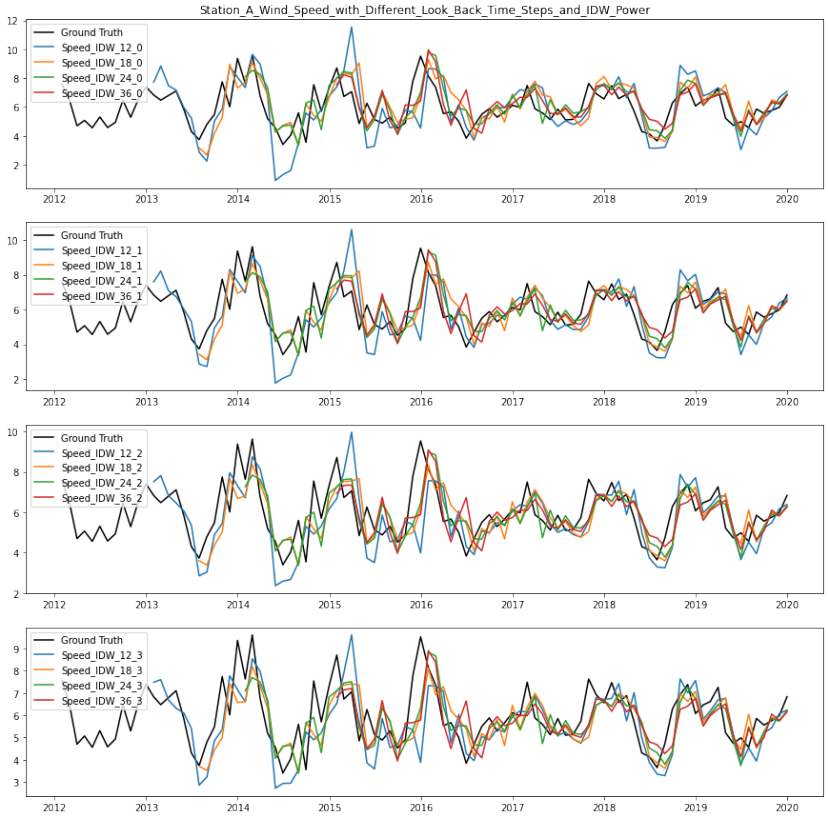
Figure 23: Yearly IDW MAPE Histograms with Different Powers



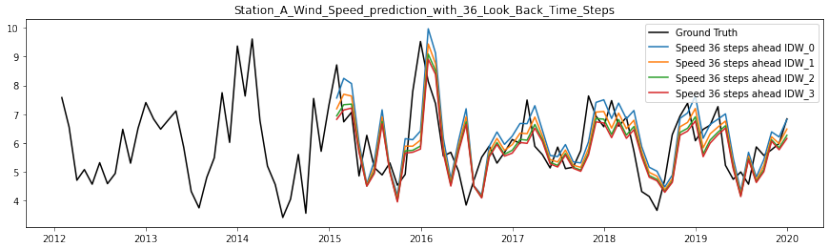
(a) $p = 0$ (b) $p = 1$ (c) $p = 2$ (d) $p = 3$

Figure 24: Yearly IDW cosine difference with different powers

SARIMA: For this model, results are presented in Figures 25 to 32 for various experiments. Figures 25(a) to 32(a) show the difference between the prediction and the ground truth at each station. Similar to the Satellite Mean Approach, the performance of SARIMA vary significantly across the four selected stations. The prediction results presented in Figures 25(b) 32(b) coincide with each other indicating that IDW with different distance powers do not affect the prediction performance. The overall prediction results show that the 12 months look back steps result in an unstable prediction performs.

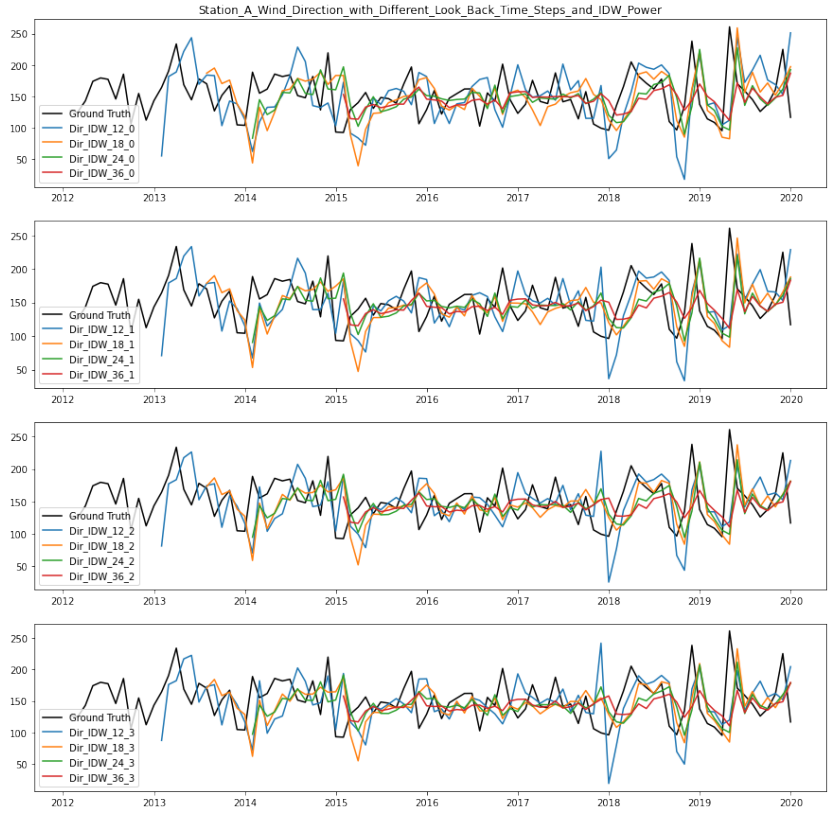


(a) Comparison among different look back steps [x: Time (month), y: Wind Speed (m/s)]

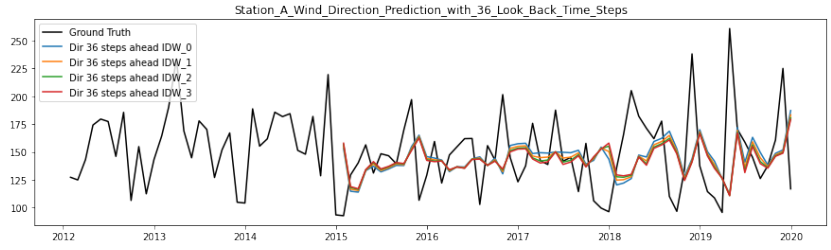


(b) Comparison among IDWs with 36 months look back [x: Time (month), y: Wind Speed (m/s)]

Figure 25: Monthly Wind Speed Prediction for Station A

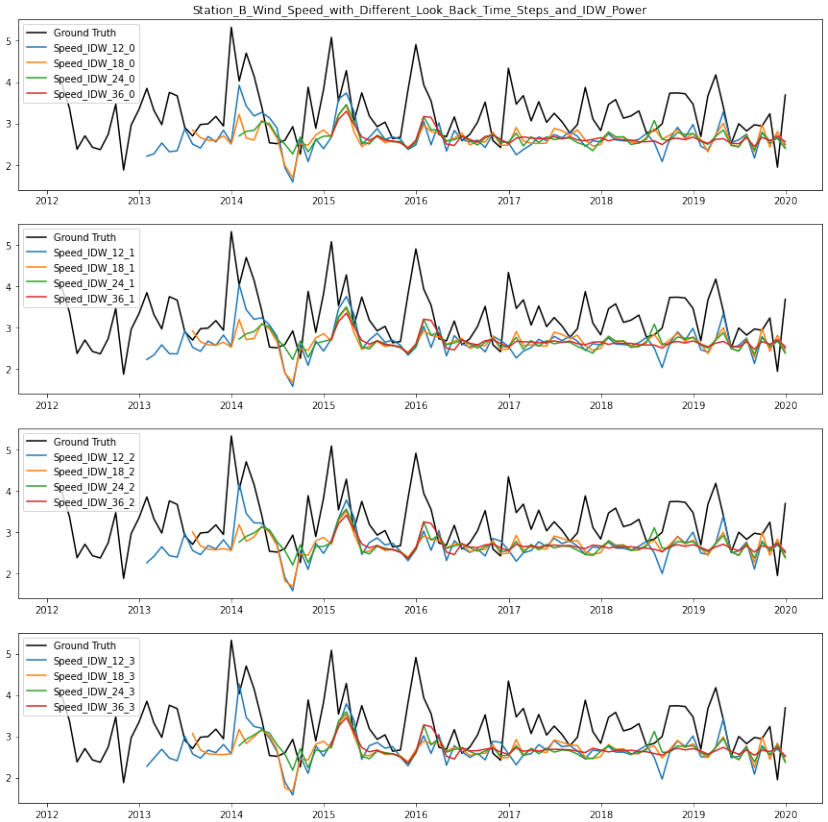


(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Direction (degree)]

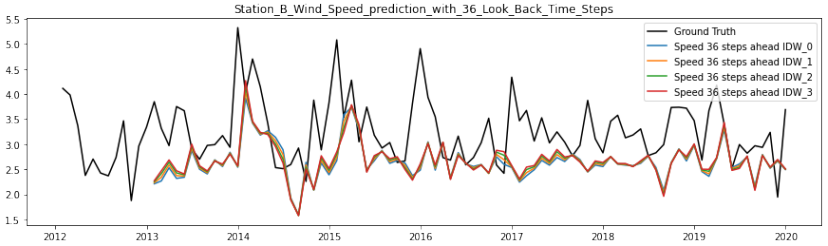


(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Direction (degree)]

Figure 26: Monthly wind direction prediction for Station A

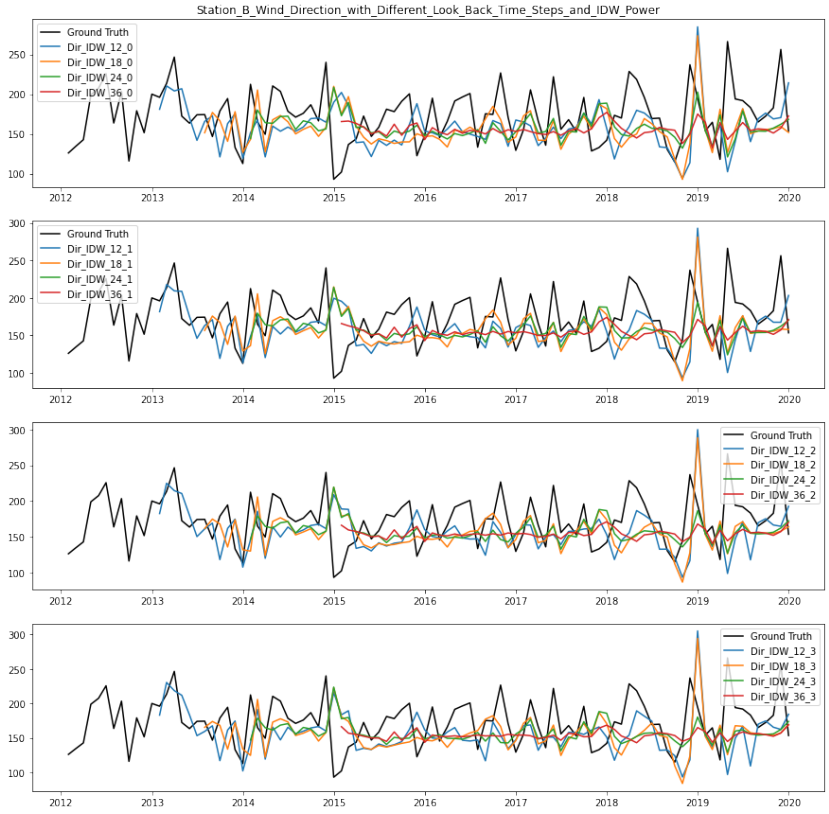


(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Speed (m/s)]

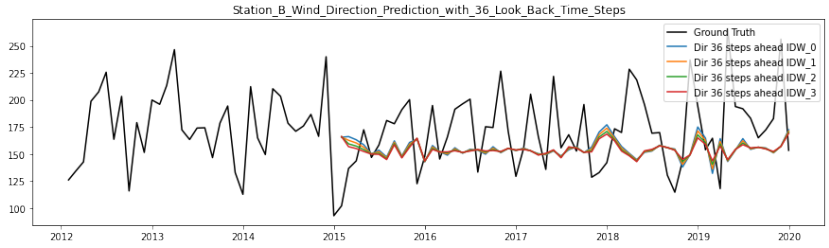


(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Speed (m/s)]

Figure 27: Monthly Wind Speed Prediction for Station B

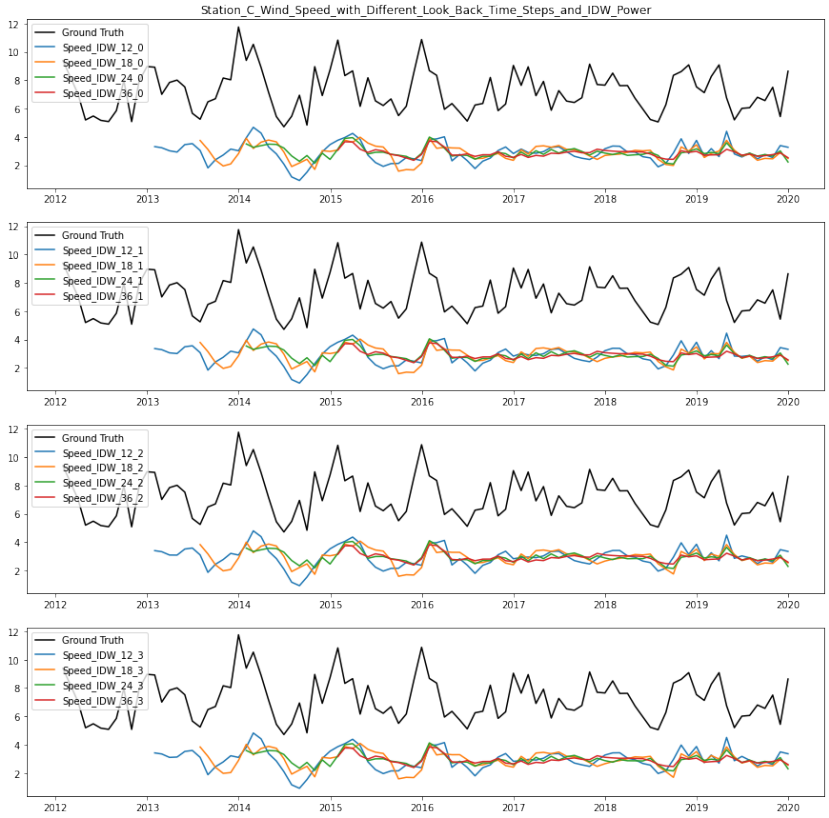


(a) Comparison among different look back steps [x: Time (month), y: Wind Direction (degree)]

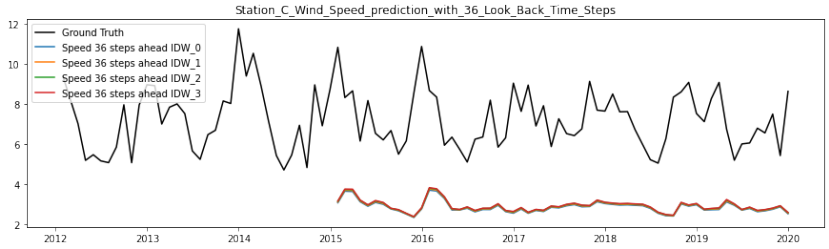


(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Direction (degree)]

Figure 28: Monthly Wind Direction Prediction for Station B

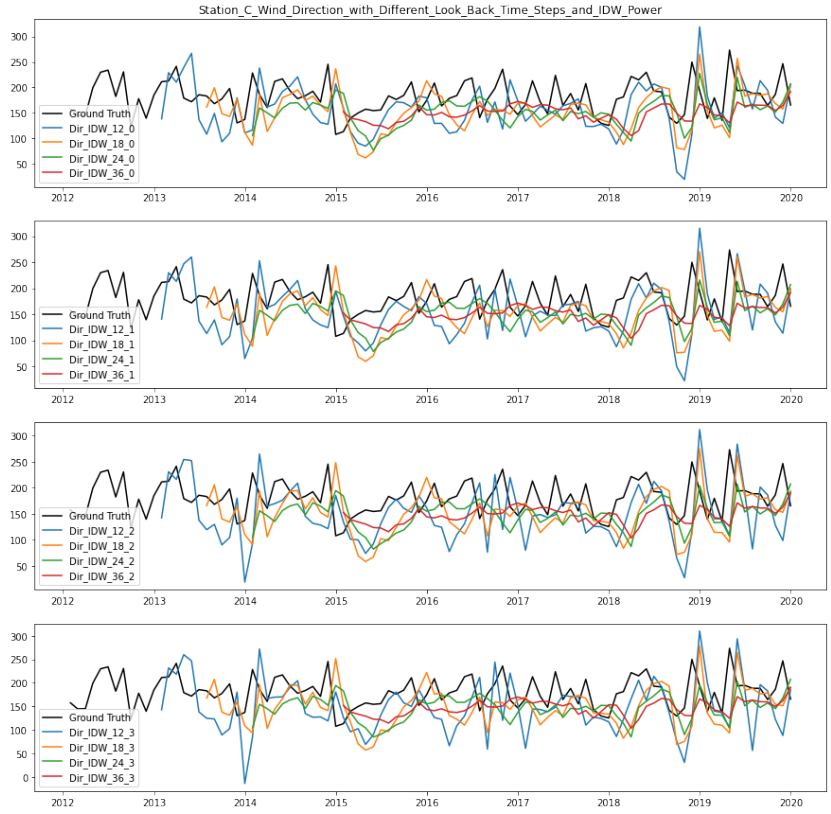


(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Speed (m/s)]

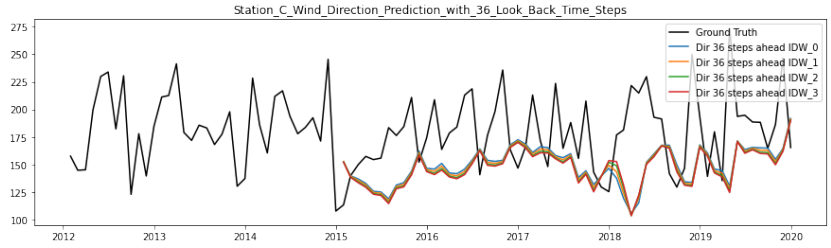


(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Speed (m/s)]

Figure 29: Monthly Wind Speed Prediction for Station C

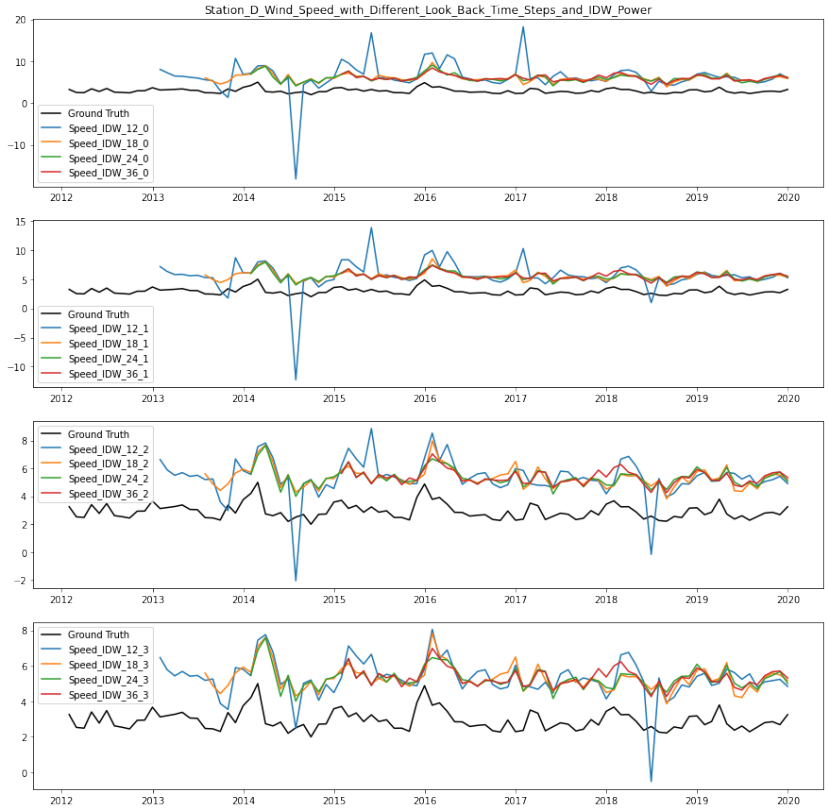


(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Direction (degree)]



(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Direction (degree)]

Figure 30: Monthly Wind Direction Prediction for Station C

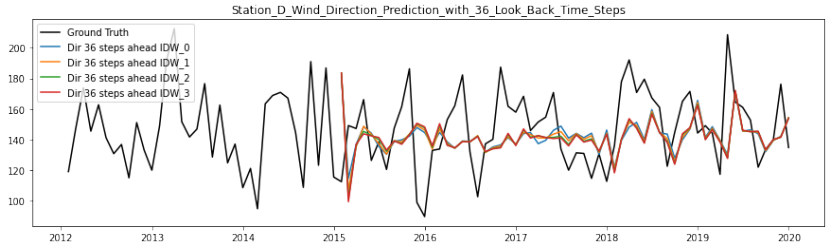


(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Speed (m/s)]

(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Speed (m/s)]

Figure 31: Monthly wind speed prediction for Station D

(a) Comparison among Different Look Back Steps [x: Time (month), y: Wind Direction (degree)]



(b) Comparison among IDWs with 36 Months Look Back [x: Time (month), y: Wind Direction (degree)]

Figure 32: Monthly wind direction prediction for station D

Tables 8 to 15 compare the wind speed prediction with different *lookback steps*. Ideally increasing the look back steps should improve the prediction performance as it has more data to capture the trend in data. However, we observe that this is not necessarily the case here. We found that for predicting wind speed and direction, look back steps of 24 months perform better than that of 36 months.

Table 8: Station A monthly wind speed (MAPE) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	12.3	13.9	11.3	11.2
1	10.9	12.8	11.1	11.8
2	11.5	12.9	11.8	12.6
3	11.7	13.0	12.1	13.0

Table 9: Station A monthly wind direction (cosine difference) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	0.268	0.156	0.074	0.109
1	0.140	0.088	0.071	0.104
2	0.087	0.061	0.072	0.106
3	0.089	0.054	0.073	0.108

Table 10: Station B monthly wind speed (MAPE) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	18.5	19.6	19.2	19.5
1	18.4	19.6	19.2	19.3
2	18.3	19.7	19.2	19.1
3	18.4	19.7	19.3	18.9

Table 11: Station B monthly wind direction (cosine difference) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	0.200	0.132	0.118	0.137
1	0.204	0.125	0.111	0.125
2	0.172	0.118	0.105	0.129
3	0.128	0.101	0.110	0.125

Table 12: Station C monthly wind speed (MAPE) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	56.2	58.7	57.9	58.8
1	55.5	58.1	57.4	58.4
2	54.8	57.6	57.1	58.0
3	54.4	57.4	56.9	57.8

Table 13: Station C monthly wind direction (cosine difference) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	0.234	0.064	0.161	0.093
1	0.291	0.069	0.196	0.098
2	0.337	0.095	0.225	0.103
3	0.373	0.095	0.244	0.107

Table 14: Station D monthly wind speed (MAPE) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	118.3	115.9	114.3	114.2
1	99.7	96.6	96.8	97.8
2	92.9	88.9	91.2	92.1
3	91.7	87.3	90.4	91.1

Table 15: Station D monthly wind direction (cosine difference) for different time steps, look back and IDW powers

IDW power	12 back	18 back	24 back	36 back
0	0.124	0.041	0.019	0.023
1	0.106	0.044	0.030	0.024
2	0.131	0.032	0.047	0.023
3	0.120	0.027	0.053	0.023

Tables 16 to 23 also compare the predicted satellite data with their observations. We now observe that the predicted wind speed and direction for the satellite data (in contrast to station-level prediction) show uniform performance across the four selected locations : around 15% in MAPE and around 0.1 in cosine distance. This indicated that SARIMA can reliably model the temporal dynamics in the satellite data.

Table 16: Station A wind speed (MAPE) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	14.6	16.0	10.4	10.3
UR	14.6	13.2	10.8	10.8
LL	16.4	13.5	10.8	11.3
LR	9.5	10.4	10.9	12.4

Table 17: Station A wind direction (Cosine Difference) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	0.589	0.282	0.168	0.181
UR	0.604	0.304	0.122	0.187
LL	0.452	0.231	0.150	0.145
LR	0.116	0.143	0.088	0.134

Table 18: Station B wind speed (MAPE) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	14.3	14.9	13.3	11.0
UR	14.3	15.0	13.4	11.1
LL	14.6	14.9	10.5	11.4
LR	14.6	14.6	10.1	10.8

Table 19: Station B wind direction (Cosine Difference) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	0.125	0.088	0.061	0.062
UR	0.125	0.077	0.050	0.069
LL	0.246	0.212	0.164	0.141
LR	0.189	0.184	0.155	0.140

Table 20: Station C wind speed (MAPE) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	18.7	15.5	14.5	10.7
UR	18.6	15.1	14.4	10.9
LL	20.0	16.2	14.3	11.0
LR	19.2	16.0	14.4	10.8

Table 21: Station C wind direction (Cosine Difference) between observation and prediction

IDW power	12 back	18 back	24 back	36 back
UL	0.495	0.253	0.159	0.069
UR	0.399	0.204	0.102	0.107
LL	0.350	0.174	0.072	0.084
LR	0.611	0.225	0.164	0.093

Table 22: Station D wind speed (MAPE) between observation and prediction

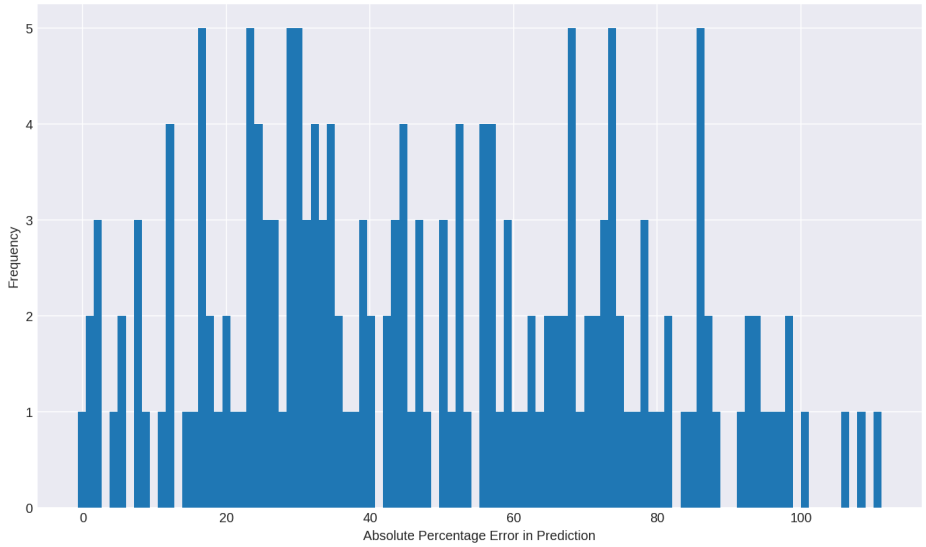
IDW power	12 back	18 back	24 back	36 back
UL	16.3	11.0	12.2	9.8
UR	22.6	19.3	12.7	11.4
LL	16.2	15.0	14.4	12.4
LR	21.8	14.1	12.9	11.2

Table 23: Station D wind direction (Cosine Difference) between observation and prediction

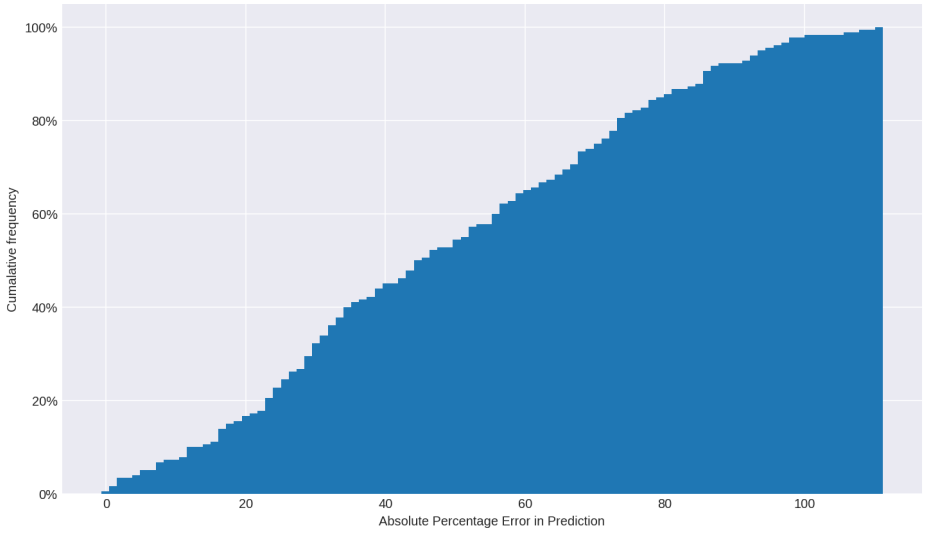
IDW power	12 back	18 back	24 back	36 back
UL	0.124	0.035	0.069	0.050
UR	0.714	0.061	0.059	0.044
LL	0.057	0.045	0.056	0.043
LR	0.041	0.044	0.036	0.071

To summarise, the SARIMA method can capture the trend and seasonality features in the monthly wind time-series, and is suitable as a baseline for further exploration. But each time-series prediction model have not reached the best performance which requires further work on hyper-parameter selection. On the other hand, these baseline approaches can only assume that the spatial correlation among the target area remains simple and stationary, which might cause failure in predicting wind at some stations. The terrain and environmental data are worth considering in further methods to obtain the spatial information for model.

Gaussian Process Regression for Mean Year Prediction: Histograms showing the distribution of the absolute percentage error values for all stations and months in the training set are shown in Figure 33. The MAPE on the training set is 48%. Examining how the absolute percentage error varies with each of the features revealed only two clear trends. The most pronounced of these can be seen in Figure 34. The predictive power worsens as the value of the ground wind speed decreases. A potential strategy for improving this could be to normalise the input variables, or use a different scaling for the output data; further testing is needed to narrow to definitively say what will improve the accuracy. Additionally, a slight seasonality to the MAPE can also be seen in Figure 35. Here, we see that, for this method, the predictions are worse for the Summer months compared to the rest of the year.



(a) Standard Histogram



(b) Cumulative Histogram

Figure 33: Histograms showing distribution of MAPE values for prediction of the mean year using Gaussian process regression. This prediction is carried out on a training set of 15 stations, after training on a set of 157 stations. In both cases, 100 bins are used.

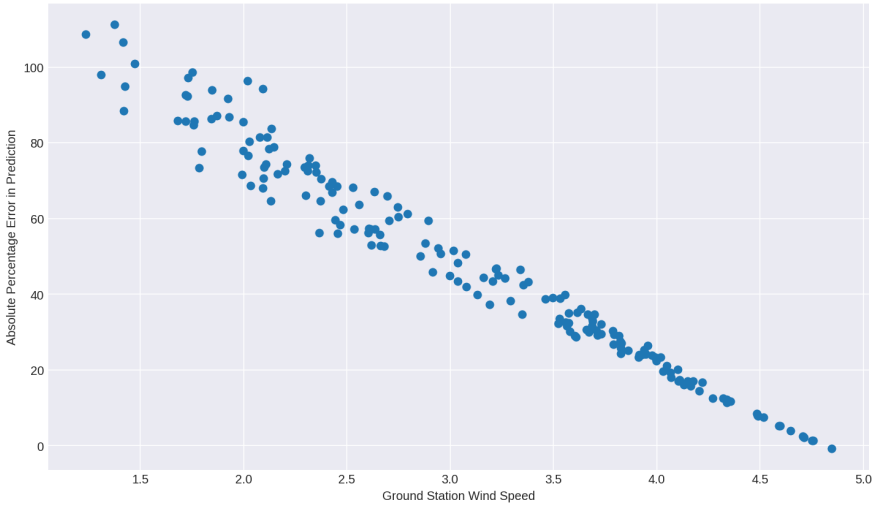


Figure 34: Change in prediction accuracy for training set, for the mean year Gaussian process regression approach, with magnitude of ground station wind speed. In general, the lower the wind speed, the worse the percentage accuracy of the prediction is.

4.3 XGBoost

XGBoost [1] stands for extreme gradient boosting that provides a framework for gradient boosted decision trees designed for speed and performance.

Methodology: For our model, we took each time step at each station as an independent data point and used all the satellite data we had at this time point to predict the ground speed. The way we envisage this approach working is that you would train this model on as much data as possible to expose it to as much variation in features as possible. Then, given any site of interest, you input the features from the 4 nearest satellites over a long period of time. These can then be used as inputs to the fitted model using the distribution of outputs as our prediction of wind speed over time.

Inputs:

- Satellite data for 4 satellites
  - Temperature

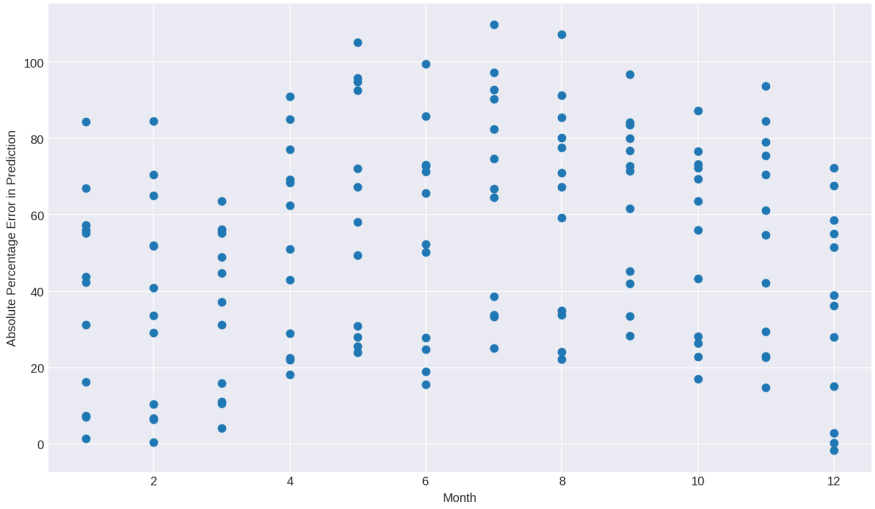


Figure 35: Change in prediction accuracy for training set, for the mean year Gaussian process regression approach, with the month of the year. We see a slight seasonality to the accuracy of prediction, with the Summer months suffering from a decrease in accuracy.

  - Pressure
  - Wind speed
  - Wind direction
  - Distance from ground site
- Elevation
  - Mean, Min, Max and Quartiles
- Roughness
  - Proportion of local grid given of each terrain.
- Time
  - Hour, Day and Month

Outputs:

- Ground wind speed

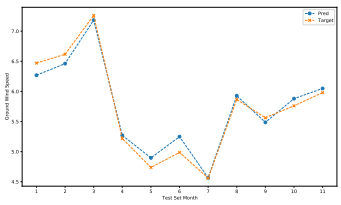
Data preprocessing: Initially, we collated data in a data frame. We removed any time steps where we did not have both ground speed and satellite data. We combined this with the stationary summary statistics about the elevation data from the grid of values given around the ground station and the proportion of the area of the grid of each terrain type. We allowed for behaviour to also be impacted by the hour, day and month; we ignored year as we wanted our model to be able to be used for years which we did not have data for. We also noticed that there was clear periodicity between years (see Section 3.2). We then removed any rows when the ground wind speed was 0 as we believed this to be erroneous.

For our variables that have a cyclic nature (hour, day, month and the wind angles), we experimented with creating two new variables in two ways: (1) taking the sin and cos of scaled versions of these variables, giving a mapping to the x and y coordinates of a unit circle and (2) treating there variables as categorical variables and using one hot encoding. Neither of these results gave any significant improvement in prediction and therefore, for ease of feature importance, the inputs were inputted directly, ignoring their cyclic nature.

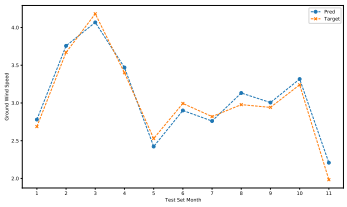
Training procedure: For the training of the model we explored 2 instances: one when the training data included the stations of interest and data from 2019, and the other just not using 2019 data. We fit the model using the default XGBoost parameters with the evaluation metric of mean absolute percentage error. To prevent over fitting the we used early stopping rounds of 10: that if the model was not improved upon in 10 rounds the fitting stopped.

Results: The models were evaluated by looking at their predictions of 4 stations in the year of 2019. The prediction was for the mean monthly wind speed.

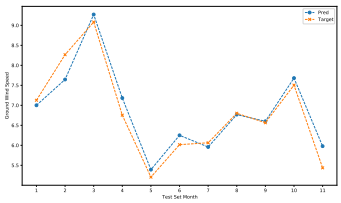
The results of when our model was trained on data which included the stations of interest can be seen in Figure 36. This shows that XGBoost is able to make very accurate average monthly predictions using satellite data of a location it has been trained on even if the time period is different. This is promising as it suggests that once you have some ground data you could be able to make predictions into future just using satellite data.



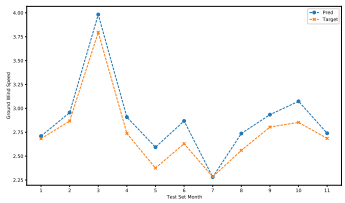
(a) Station A



(b) Station B



(c) Station C



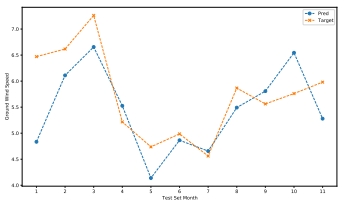
(d) Station D

Figure 36: Comparisons of predictions and true values of monthly mean wind speeds against the for 4 stations when the training set included the test stations.

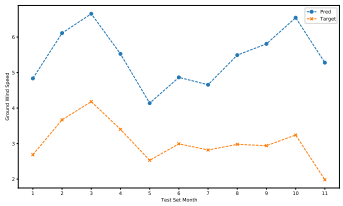
The results of the model when the stations of interest were not included in the training data can be seen in Figure 37. These results are not as good with all of the stations bar station A having a large difference when comparing predictions and true values. These inaccurate predictions could be due to over fitting or not having enough data of stations similar to those which it performed poorly at. These plots do have a positive that for each stations the predictions and true values are correlated which suggests that the model did learn the seasonal trend.

Feature importance XGBoost can give us an understanding of what features are important. This can be done using an F score that sums up how many times each feature is split upon.

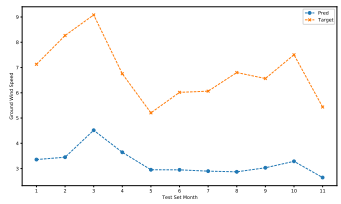
As shown in Figure 38 as expected the wind speeds and wind directions from the 4 satellites were important, we also identifies that hour of the day, month of the year, distance away of the satellite and the presence of cropland. The features of hour of the day and month of the year are expected as we saw that there was periodicity within these values in



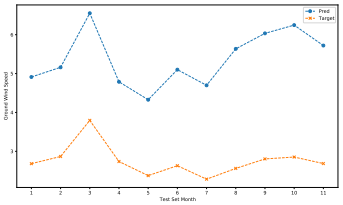
(a) Station A



(b) Station B



(c) Station C



(d) Station D

Figure 37: Comparisons of predictions and true values of monthly mean wind speeds against the for 4 stations.

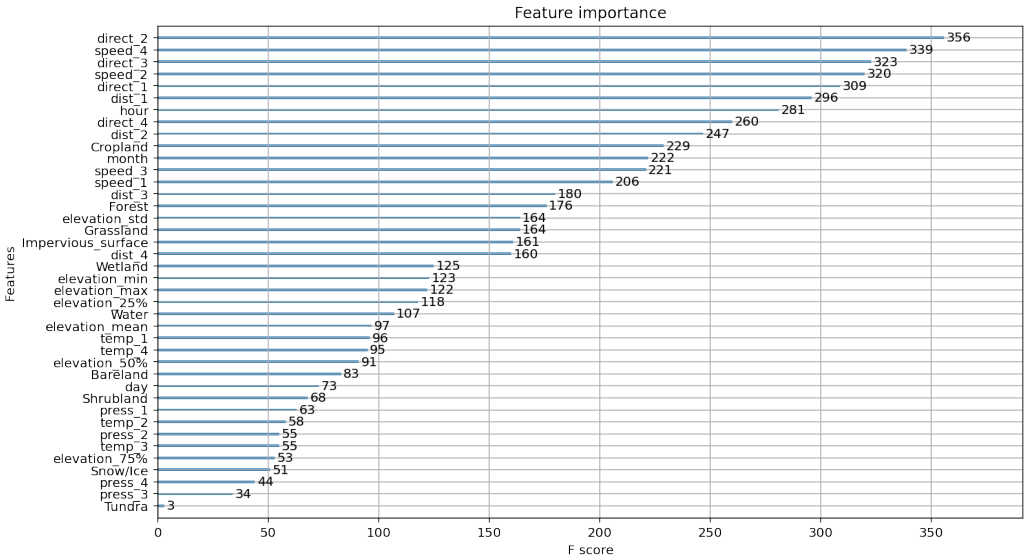


Figure 38: Feature importance

Section 3.2.

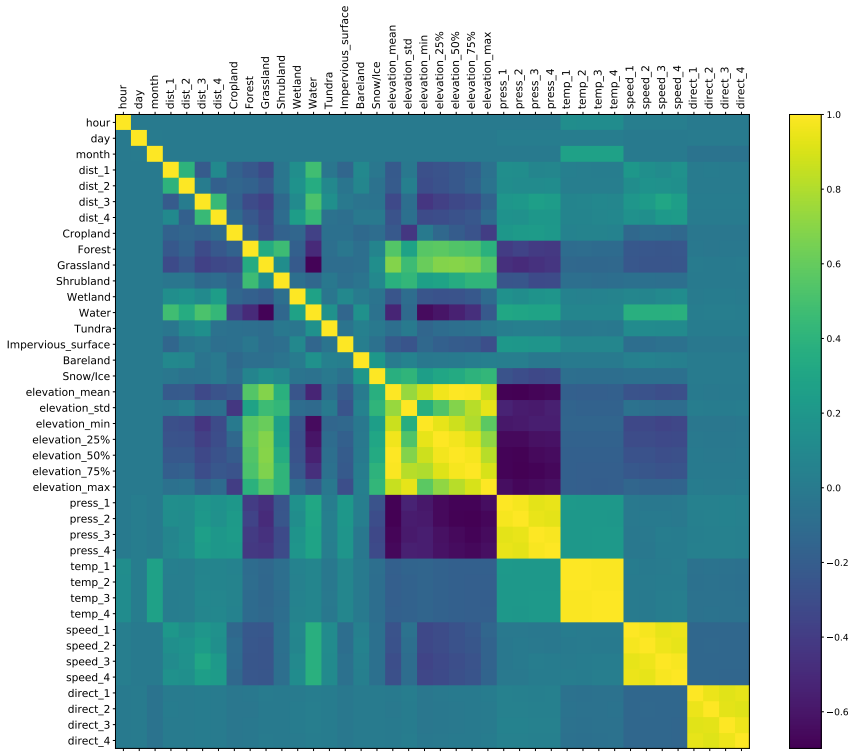


Figure 39: Correlation of all of the variables used in the XGBoost model

Discussion The preliminary fitting on XGBoost has shown some promising signs but there are several ways that it could be improved the first being to tune the hyper parameters which we were unable to do in this project due to time and computational limitations.

The features could have been inputted differently especially elevation and terrain which where taken over the entire grid. This could be improved with more time and data around the satellite locations and instead only considering features between the satellite and ground station.

Another issue with how the features are inputted can be seen in Figure 39 which shows the correlation between all of the inputs of XGBoost. This

shows that every input collected by the four satellites are strongly correlated between then. This is an issue with XGBoost as correlated inputs mean that the wrong features may be picked up by the model. Taking the mean and the standard devaition of these inputs could have been an effective way to overcome this.

This model has the fundamental flaw that the distribution of wind speeds over the period of inputs entered is the same as the current distribution of wind speed over the same time period. The effect of the distribution could be mitigated if we had more longitudinal data so we could perform data analysis over a larger time period.

4.4 Regression-Kriging

Kriging (Gaussian Process Regression) is a geostastical spatial interpolation technique that has been widely applied to predict values of spatial phenomena in unsampled locations (Lam 1983). The spatial and/or spatio-temporal distribution of phenomena are approximated by functions depending on location in multi-dimensional space. Kriging is based on the concept of random functions, where the surface is assumed to be one realisation of a random function with a certain spatial covariance. The main strengths of Kriging lie in the statistical quality of being unbiased in its predictions and quantifying the spatial distribution of the errors. However, the accuracy of this approach depends strongly on the localization of the sampling locations, when in presence of spatial dependence (Zhu and Lin 2010).

Regression-Kriging (RK) extends simple Kriging with a regression model to capture important predictors of a particular phenomenon fig:kriging. The regression equation predicts the values of the target variable in an unknown location using predictor variables that are related. Given the ubiquity and reduced cost of satellite-derived information, including several predictors should not only improve the model fit but also reduce the number of training data required to make predictions. Mathematically, Kriging and regression are closely related but the main difference is that Kriging assumes that the target variable is spatially autocorrlated. Therefore, we assume the error term is a Gaussian process and not just some random Gaussian variable, as in linear regression.

Methodology: We applied the RK model to estimate wind speeds and directions for four selected ground stations using satellite-derived wind speeds and directions, alongside several predictors. Using training data from 2012-2019, we made predictions of both wind speeds and directions at a monthly interval for the year 2019. We took the mean wind speeds and directions from four satellites around each ground station, to make ground estimates of wind speeds and directions at each of the ground stations. The predictors included temperature, pressure and height information from digital elevation models. These were incorporated because they were well correlated with wind speeds and directions and were calculated using a similar procedure to satellite-derived wind speed and directions.

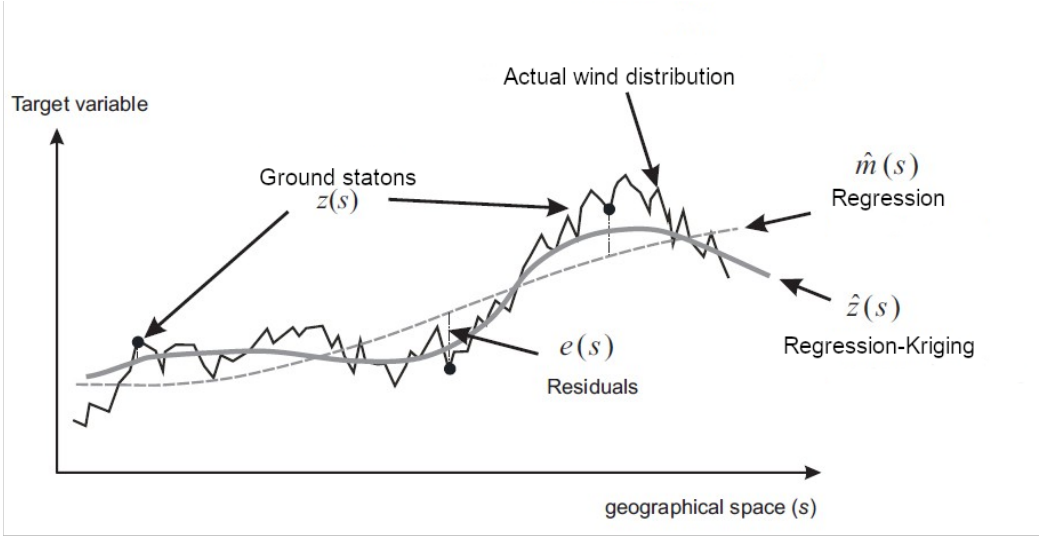


Figure 40: An overview of Regression-Kriging

For the Kriging component, we specified an ”Ordinary Kriging” model and accounted for the three nearest neighbours. We selected the three nearest neighbours over other metrics, which included a fixed distance of 100km. This was because neighbours of some stations were greater than 100km in distance away from the target station. For the regression component, we compared linear regression, support vector machines and Random Forest (RF) models. We found the model fit of the RF model was significantly higher than the other two and we suggest that this is because

of the non-linearity present in wind speeds and direction but also because of the non-linear dependence of the target variables and predictors. We set the number of estimators for the RF model as 100, which allowed us to obtain sub-second estimates for each station.

Results: Overall, the results suggest that the fitted RK model performed consistently well across the four stations. This was particularly true when compared to the baseline models. Despite the model not explicitly accounting for temporal autocorrelation, Figure 41 shows that the the model fitted both the training and test data well, capturing the non-linearity in both wind speeds and direction over time. However, the model failed to capture the structure of wind speeds directions was at the most extreme values.

Figure 41: Prediction of wind speed and direction for each of the four test ground stations using the Regression-Kriging Model

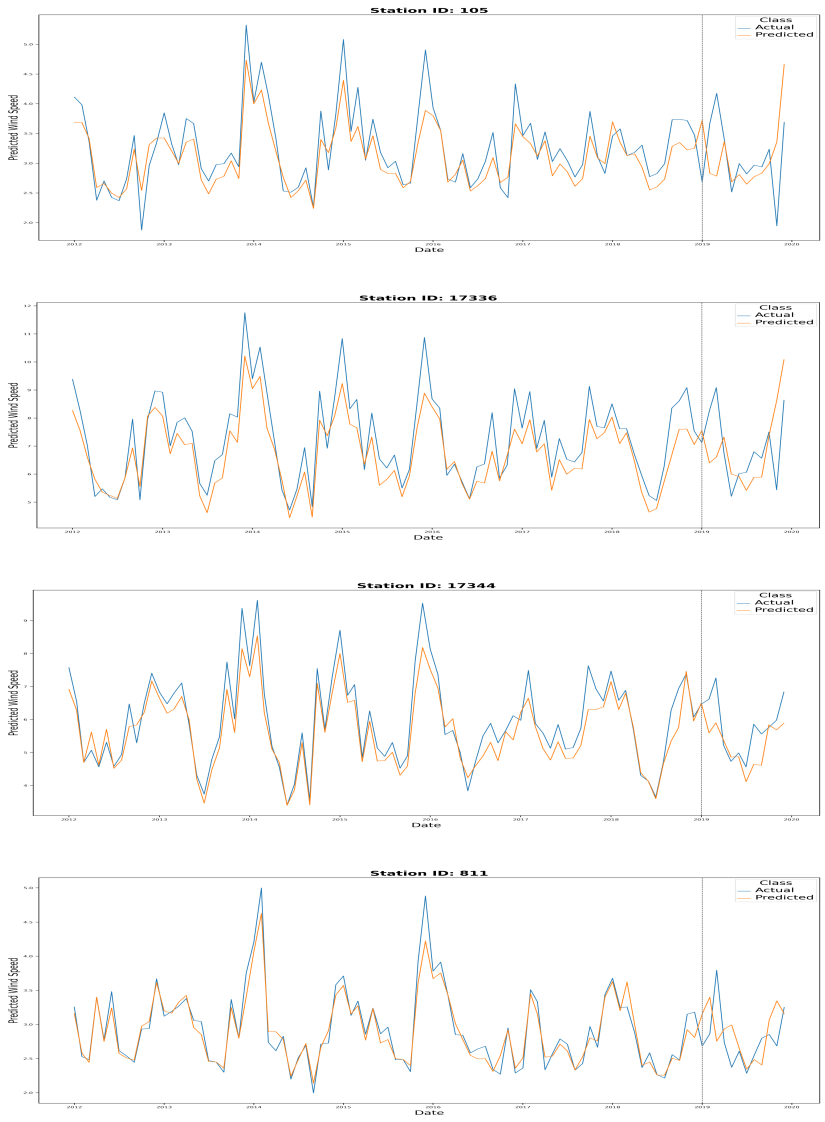


Figure 42: Prediction of wind speed and direction for each of the four test ground stations using the Regression-Kriging Model

The model performed best on station A, which is likely the net result of two components. Firstly, the variability in wind speeds and directions were more consistent over time for this station and secondly, the neighbouring stations were closer in distance compared to the other stations. Information pooled from ”close” neighbours is more likely to be

related than neighbours that are not close and therefore contribute more ”important” knowledge. In contrast, neighbouring stations for station B were greater than 100km in distance, which is likely to have contributed greatly to the poor model fit.

4.5 RNN models

Recurrent Neural Networks (RNN) are designed to learn sequential data efficiently and predict future sequence elements. In this section we focus on using RNNs and their derivatives to predict wind speed and direction time-series.

Compared to standard RNNs, Long short-term memory (LSTM) are a class of RNNs composed of a memory cell - tasked with learning the sequential dependencies - as well as regulating components that control how sequential information is fed into the memory cell. These regulators - or gates - allow the LSTM to “forget” learned weights depending on the next element of the sequence in addition to modulating the activation of the memory cell. Together, these design choices allow LSTMs to outperform RNNs on longer sequences and have led to their wide adoption across the fields of natural language processing and time-series analysis. In our case, we trained LSTM models on historic satellite data to predict wind speed and direction for a future window. Particularly, we predict wind speed and direction at the ground-station in Section 4.5.1 and predict future satellite wind speeds and directions in Section 4.5.2.

Finally, we incorporate contextual information in the form of digital elevation maps and terrain classification information using a Conditional Gated Recurrent Unit (GRU) in Section 4.5.3.

4.5.1 LSTM for ground station data

Feature normalisation In order to train the LSTM models effectively, we chose to scale the input time-series features. This prevented high magnitude variables (such as pressure measured in pascals) from overwhelming variables of lower magnitudes during cell activation. We normalised the features station-wise by subtracting the mean and dividing

by the standard deviation of each feature. Critically, this normalisation was only calculated over the training data, but was applied across the training, validation and test sets. As an example, the normalised features for a given station are shown in Figure 43. As indicated by the figure, each of the features has been scaled to a similar range, and there are few long tails in the normalised features.

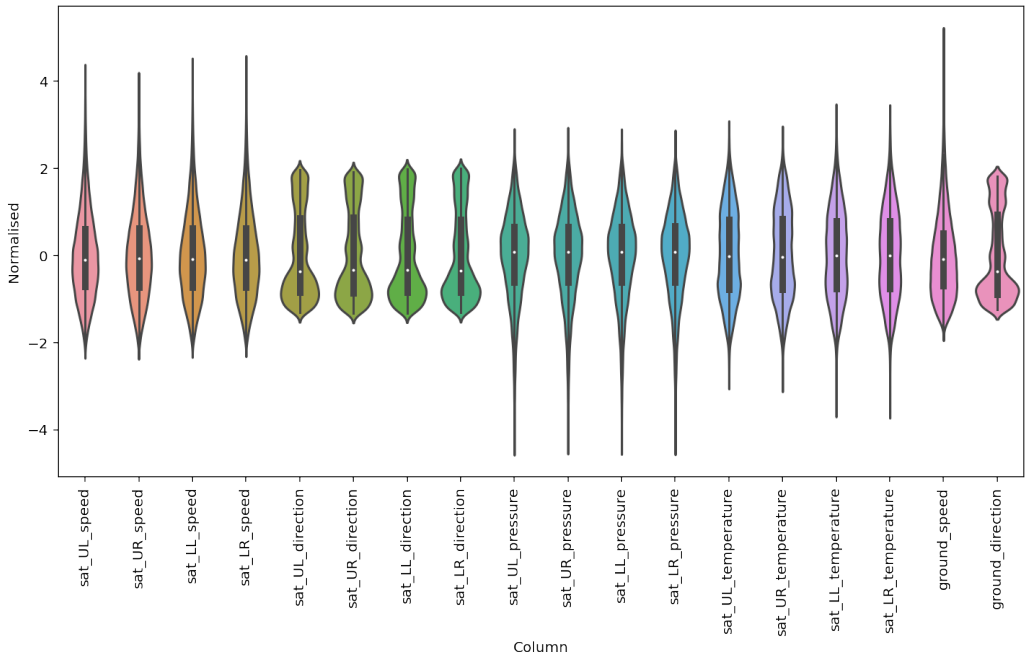


Figure 43: Violin plots of normalised time-series variables in the training data for station A. UL, UR, LL, LR refer to the relative positions of the satellites to the ground station (upper left, upper right, lower left, lower right respectively)

Feature engineering with decomposition

We also investigated the effect transforming input and target variable measurements for wind speed and direction had on predictive performance. We trialled two potential feature transformations:

1. Direction-only decomposition
2. Direction and speed decomposition

For the Direction-only decomposition, measurements of wind direction (for satellites and ground stations) were decomposed to $dir_y = \sin(dir)$ and $dir_x = \cos(dir)$ components, such that each variable for wind direction was replaced in the data set with two new variables dir_x, dir_y. In contrast, for the second transformation both the wind speed and direction variables were decomposed into orthogonal $w_x = speed \cos(dir)$ and $w_y = speed \sin(dir)$ components, converting the magnitude and direction into a 2D vector. In both cases, the predicted values and target labels were reconverted to the original two variables (wind speed and wind direction) prior to the calculation of the evaluation metrics (MAPE and CD).

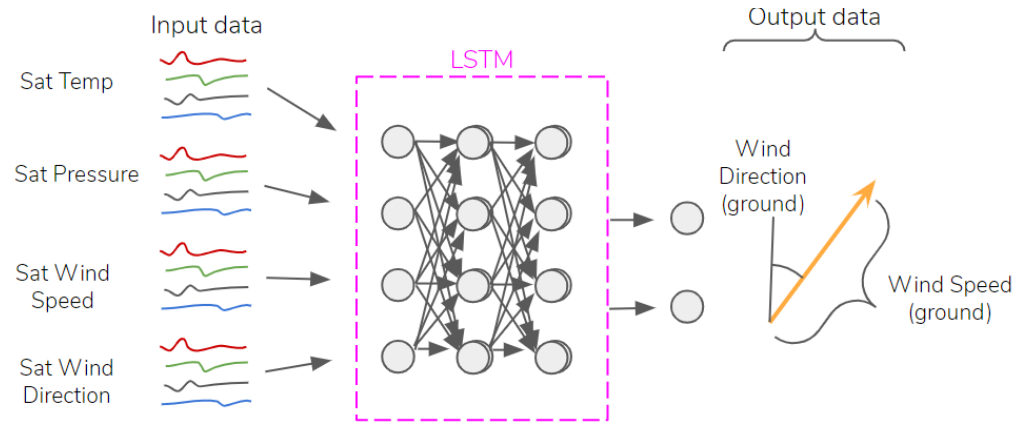
The resulting models trained and evaluated on the transformed features showed similar performance for a number of trial stations. Whilst the exact performance varied owing to the random uniform sampling of the intial LSTM state, Table 24 contains representative results for single instances of the LSTM models.

Table 24: Effect of speed and direction feature transformations on LSTM model performance, calculated over monthly estimates for 2019 at station A. Target label predictions were reconverted prior to calculating the MAPE and cosine difference.

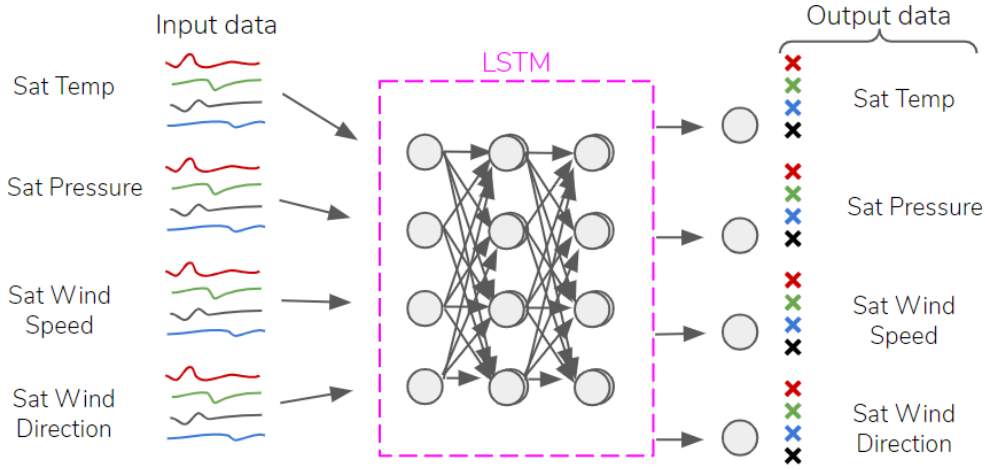
Station ID	Feature transformation	MAPE	CD
A	None	**7.58**	**0.272**
A	Angle-only	11.1	0.281
A	Angle and direction	14.0	0.278

Model architecture We used a single layer vanilla LSTM with 64 hidden units and input sequence of 12 elements, where each element represents a monthly sample. Each sequence element contains measurements of wind speed, direction, pressure and temperature for each of the four satellites, resulting in a total of 16 features per sequence element. For this LSTM model, we used a history of the 12 previous months of satellite data to predict the ground wind speed and direction for the following month. Following the 64 hidden units we use a fully connected layer to predict the output - two or three-dimensional depending on the wind

direction parametrisation. Figure 44a presents the model architecture.



(a) Architecture of LSTM model predicting ground station wind direction and wind speed.



(b) Architecture of auto-regressive LSTM model predicting satellite data.

Figure 44: Diagrams of LSTM models architecture highlighting the input and output features for each model.

Training procedure A separate model is trained for each station independently. We split the time-series from each station into non-overlapping - but contiguous - training, validation and test sets, each with length of 72, 12 and 12 months, respectively. We use the L1 loss function, also known as Mean Absolute Error (MAE), between predicted

and target ground wind speeds/directions. For the optimisation we use batch gradient descent with Adam [3] optimiser using a learning rate of 10^{-3} and default PyTorch [8] parameters. We perform optimisation for a maximum of 50 epochs and stop the optimisation if the validation loss does not improve after 2 epochs. This technique is called early-stopping and is used to prevent over-fitting.

Model performance The evaluation is carried out on the independent, withheld test set for each station. The results are reported in Table 25. Figure 45 shows the model predictions and target (ground-truth) results for Station A. There are two limitations in the evaluation of this model. Firstly, we were only able to evaluate it on a small subset of ground stations that had at least 24 months of continuous monthly data available (to satisfy the test and validation sequence length requirements). Secondly, the model prediction window is for a single month, thus, we need satellite measurements including the month before the prediction. We address this limitation by creating another LSTM model (Section 4.5.2) that predicts satellite wind speeds in an auto regressive fashion, which could be coupled with this model to provide long term ground wind speed and direction predictions.

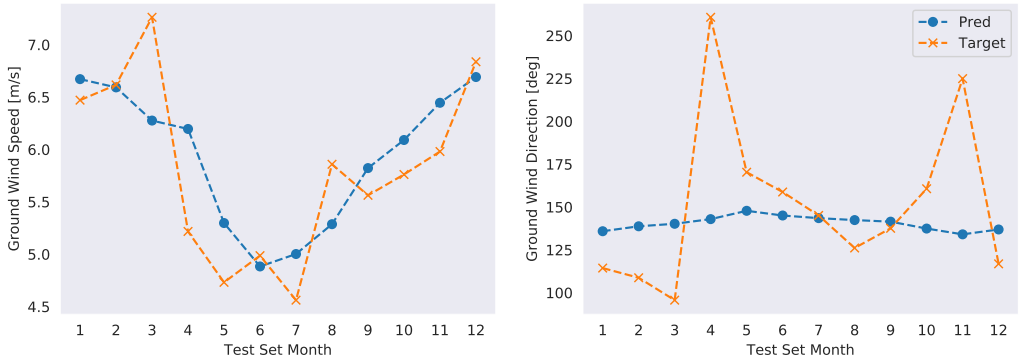


Figure 45: LSTM Ground Wind speed and direction predictions for Station A.

4.5.2 Autoregressive LSTM for satellite data

The previous LSTM model required satellite data up from one month before the prediction target date. However, this is often unfeasible when performing long term forecasting. We address this limitation by creating a new LSTM model that can predict the satellite features (wind speed/direction, temperature and pressure) in an auto regressive fashion. The model forecasts these satellite features for the upcoming month given a history of 12 prior months of data. By feeding back the prediction for the upcoming month into the input we can generate predictions over any time window by effectively sliding the LSTM model's training window forward with each new prediction. As suggested, the satellite predictions from this model could be coupled with the previous LSTM model for ground wind speed and direction to estimate ground wind speed and direction arbitrarily far into the future.

Feature normalisation We have used the same feature normalisation procedure as described in section 4.5.1. Note again that the mean and standard deviation is computed over the training sequences only before being applied to the full data set.

Model architecture The model architecture is effectively the same from Section 4.5.1 except that the output layer has 16 dimensions, corresponding to the four satellites wind speed, direction, temperature and pressure estimates. The output matches the dimensions of the each of the input sequence, and thus can be fed back to generate long term predictions. Figure 44b illustrates the model architecture.

Training procedure This model only requires satellite data as input, thus the number of samples in the data set is much higher, as observed in Figure 3. Furthermore, we train a single model over all stations' satellites measurements. We follow the same train/validation/test sets split in this model, where for each station we hold the last year of data as test and second to last year as validation, the remaining data is used for training. In total we have 172 stations resulting in 11366 training, 2064 validation and 2064 testing sequences (each with 12 months). The model is trained over the L1 loss function for a maximum of 200 epochs and training is interrupted if the validation loss does not decrease after 3 epochs. During training and evaluation the one year estimates are obtained in an auto

regressive fashion.

Model performance We report the performance of this model in terms of the histogram of wind speed and direction absolute errors over the test set considering all 172 stations in Figure 46. The MAPE of satellite wind speeds over all stations for the last year (2018-2019) was 15.35%, the median and mean Cossine Difference (CD) of satellite wind direction was 0.121 and 0.292, respectively.

To give an insight of the MAPE for the predicted satellite wind speeds over a year in each individual station we created a map visualisation where each station is coloured according to the MAPE obtained for the satellites within the respective station boundaries. The visualization show that the model performs poorly on Scottish stations. Our intuition about this behaviour is that the geographic location of stations is an important feature to explain future wind behaviour. Since this LSTM model is agnostic to geographical locations or any discrimination between stations, i.e. the model is trained over all stations independently, it tends to fit to the most common pattern, which in this case are the stations in central and south England, where we find the highest density of stations across the UK. As such, the model fails to generalise to stations which have unique different patterns of wind behaviour.

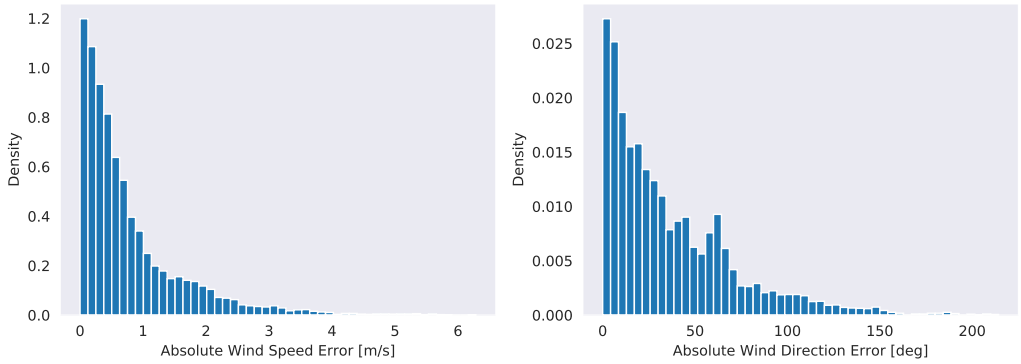


Figure 46: LSTM Satellite wind speed and direction error distributions over all stations.

4.5.3 Conditional GRUs

This model was proposed by viewing the problem in the following way: satellite measurements of wind provides strong estimates of ground-wind minimally affected by terrain and other contextual factors. Therefore we need to find the correct transformation to apply to the distribution of satellite wind measurements to accurately model the ground wind distribution.

Data preprocessing Satellite wind-direction, wind-speed, temperature, and air-pressure measurements over 01-01-2012 to 31-12-2019 were preprocessed by binning into monthly histograms, with 36, 48, 42, 42 bins in each histogram respectively. Since the ground wind data was already discretized, we used the same binning between the ground wind and satellite wind with one small modification: the tail of the wind-speed distribution was collected into one bin $>$47 knots ($\approx$ 24.16 m/s).

The terrain data was comprised of elevation measurements and roughness categorisations for patches of land. Each array was cropped to a 64x64 patch around the ground-station. In most cases it was possible to centre on the ground station, but there were cases where this was not possible. In future, this could be a possible data augmentation (random NxN patch with ground station randomly located within). Elevations were normalized to between 0 and 1, by dividing through by the maximum elevation in the array. Since there are 10 roughness categories, this array was one-hot encoded. Collecting the roughness and elevation data together, each ground station provided an 11x64x64 tensor of spatial information. 35 stations were excluded from training due to missing ground-station data, leaving 137 stations. The remaining stations were then split into 105:32 subsets (training:test).

Model Architecture We model the transformation of distributions using a conditional Gated Recurrent Unit (c-GRU, Figure 47) taking not only monthly distributions of satellite wind measurements, but also distributions of air-pressure and temperature, which are chaotic variables that could potentially modify the transformation. Additionally, the terrain information is passed through a deep convolutional neural network (CNN), and the output is fed into the GRU as input at each time step. The output of the GRU is was then passed through a linear layer and a

softmax layer so that the output vector at each time point summed to 1.

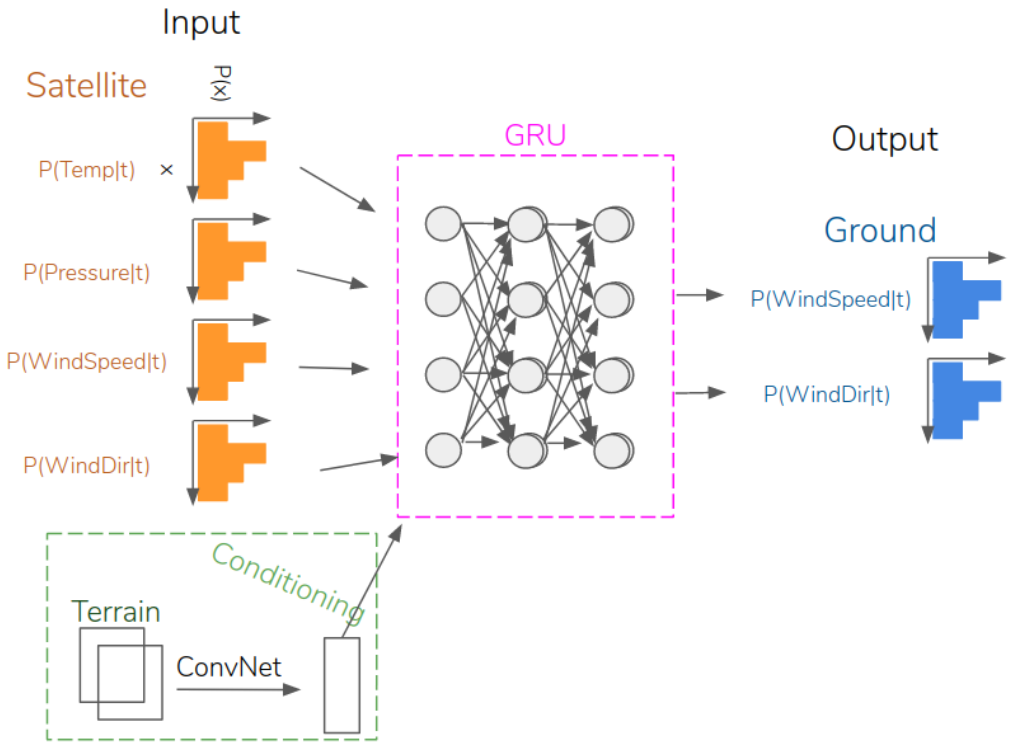


Figure 47: The conditional GRU is comprised of a CNN (green) processing spatial data (Terrain) to pass as input to GRU as contextual input. CNNs were comprised of a series of blocks comprised of (in order) Batch-Norm, 3x3 Convolution, 2x2 Max-Pooling, Leaky ReLU, and 50% Dropout layers. The CNN we used consisted of 4 blocks with 16, 32, 64, and 128 channels. The output of the CNN was then transformed into 32 features through a dense layer. The GRU comprised of 256 hidden units, and was followed by a dense layer and a softmax transformation to output a probability distribution

The model was trained to minimize KL divergence between the target distribution $\mathcal{P}$ and the model output distribution $\mathcal{Q}$,

$$KL = \sum_{i=1}^{N} \mathcal{P}_i \log \frac{\mathcal{P}_i}{\mathcal{Q}_i} \tag{6}$$

where $\mathcal{P}_i$,$\mathcal{Q}_i$ are the discretized probabilities in each bin. The total loss $\mathcal{L}$ was therefore

$$\mathcal{L} = KL_{speed} + KL_{direction} \tag{7}$$

The model was trained using the ADAM optimizer [3], with a batch size of 5 stations. The model was implemented in pytorch v1.6 [8].

4.6 Comparison of approaches

4.6.1 Performance on selected stations

To directly compare the approaches, we evaluated the quality of monthly wind speed and direction predictions for the year 2019 made by each method across four representative stations [Table 25]. We see that the simple interpolation using IDW exhibits good performance at station A for both MAPE and CD, 4.1% and 0.003 respectively when setting p = 0. However, this promise fails to generalise across stations B through D, where we observe considerable deterioration in MAPE and CD scores for IDW across values of p - up to 145.2% MAPE for station D.

The SARIMA results also follow this trend, performing best on station A (11.2% MAPE, averaged over instances of parameter p and fixing look-back at 36 months) and deteriorating for stations B-D. We would expect performance across stations for IDW and SARIMA to be similar in this way as SARIMA incorporates the same distance weighting approach.

Similarly, the GPR mean year performs better for stations A and B, than it does for stations C and D. However, this model performs best on station B, rather than station A. Overall, the spread of values between stations is lower than in the SARIMA and IDW cases, and this model has the

Table 25: Comparison of different approaches, showing the MAPE [%] and median Cosine Difference (CD) scores computed on a prediction of the mean monthly wind speed and direction for different stations over all of 2019. Stations were chosen to provide a representative sample of difficulty and ordered as such. With the exception of IDW which can only make forecasts for ground conditions at the same time points for which it has satellite data - these models were trained on data up to 2018 inclusive and their predictions tested on held-out monthly ground station wind speed and direction data for 2019 at each of the stations.

		Station							
		A		**B**		**C**		**D**	
Method		**MAPE**	**CD**	**MAPE**	**CD**	**MAPE**	**CD**	**MAPE**	**CD**
IDW	p=0	**4.1**	.003	17.9	.046	60.0	.120	113.9	**.013**
	p=1	6.7	.003	18.1	.045	60.4	.117	131.7	.018
	p=2	8.7	**.002**	18.2	.045	60.6	.117	141.0	.021
	p=3	9.6	**.002**	18.3	.045	60.8	.116	145.2	.022
SARIMA	p=0	11.2	.109	19.5	.137	58.8	.093	114.2	.023
	p=1	11.8	.105	19.3	.125	58.4	.098	97.8	.024
	p=2	12.6	.106	19.1	.129	58.0	.103	92.1	.023
	p=3	13.0	.108	18.9	.125	57.8	.107	91.1	.023
Regression-Kriging		7.9	.041	19.6	.049	16.53	.086	14.9	.062
GPR Mean Year		23.8	N/A	15.9	N/A	71.3	N/A	59.9	N/A
XGBoost		8.4	N/A	75.8	N/A	47.8	N/A	88.3	N/A
LSTM		**7.4**	.273	16.3	.038	13.9	.052	12.0	.028
Conditional GRU		11.7	**.015**	**5.4**	**.015**	**11.6**	**.001**	**10.0**	**.015**

advantage that it can still give a prediction for individual months, whereas the other two approaches have no temporal dynamics.

The RK model performed consistently well across all stations with an average MAPE of 14.7% and CD of 0.06 for wind speed and direction. Like the other models, the RK model performed best on station A, which is likely the result of two components. Firstly, the variability in wind speeds and directions were more consistent over time for this station and secondly, the neighbouring stations were closer in distance compared to other stations. Information pooled from ”close” neighbours is more likely to be related than neighbours that are not close and therefore contribute

more ”important” knowledge. In contrast, neighbouring stations for station B were greater than 100km in distance, which is likely to have contributed greatly to the poor model fit.

The XGBoost results were not as accurate with it comparing poorly compared to most of the other methods which could be a result of over fitting or not having any knowledge of the types of conditions around the sites.

Considering the LSTM approach, we see that the LSTM models predict both wind speed and wind direction with good performance across the four sample stations, reporting an average MAPE and CD of 12.4% and 0.098 respectively when averaged across the stations. It is reassuring to see more consistent model performance for wind speed and direction across the sample stations compared to the IDW and SARIMA models, which may be a result of the LSTM's ability to learn station-unique historical trends from past time steps.

Lastly, the conditional GRU exhibited consistently strong performance across the sample of stations. This approach - building on the RNN time-series-fed model adopted by the LSTM by explicitly incorporating fixed spatial information - performed the best for stations B, C and D compared to all other approaches. Only for station A was the conditional GRU outperformed by both the simpler LSTM model and the IDW approach. In the case of station C, the conditional GRU was able to return a median CD across the 12 wind direction samples of 0.00, predicting the wind direction over the year with excellent accuracy.

4.6.2 Performance across all stations

Within the project time frame, were were also able to evaluate a subset of the methods across all available stations. For the IDW baseline, discussed in detail in section 4.2.4, using simple satellite mean (p=0) returned the best monthly prediction performance, with the average MAPE across all stations was 27.2% for monthly predictions of wind speed and the average median CD for monthly predictions of wind direction was 0.0361 [Tables 2 and 3]. We fail to identify any clear nationwide trends affecting the performance of the IDW approach, but do note that stations along the coast perform marginally better and posit that

poorly performing stations may have complex underlying terrain. A more systematic evaluation of these claims may prove fruitful.

Comparatively, the LSTM approach generating *satellite* wind speed predictions demonstrates clear geographic trends in prediction quality. As discussed in section 4.5.2, we posit that the single LSTM model trained across all stations learns average patterns of wind behaviour which fail to generalise for stations with potentially unique or deviant joint distributions of wind speed and direction. Owing to the lower number of stations recorded in coastal Scotland, the model may have over-fit to the more correlated behaviours present in the rest of the UK, and thus fail to make appropriate predictions for these outlying stations. The MAPE of satellite wind speeds over all stations for the last year (2018-2019) was 15.35%, the median and mean Cossine Difference (CD) of satellite wind direction was 0.121 and 0.292, respectively.

When including terrain information in recurrent network models (i.e, the contextual GRU), MAPE improves to 14.1% across all held-out stations. This value was computed using the mode of the output and target distributions. For some stations, the MAPE computed for the mean of output and target distributions was lower than computed for the mode. This is likely due to multimodality in the target distribution, although there was insufficient time to explore this thoroughly. The final MAPE could potentially be improved by choosing between mode- and mean-computed MAPE for each station depending on the modality of the target distribution. The median cosine difference across all stations was 0.015.

5 Future work and research avenues

We have a number of suggestions for future work that we were unable to explore in the given time but that we believe hold promise. We briefly present these here, classifying the suggestions into model extensions, new models and interpretability/explainability.

5.1 Model extensions

Generalised satellite mean approach. One could extend the satellite mean approach by learning the best weight for each satellite in a supervised learning setting.

SARIMA. One could extend the SARIMA with more complex spatial interpolation model to generate a spatial-temporal model baseline.

Temporal correlation extension to Regression-Kriging model. To improve the model fit of the Regression-Kriging model further, we recommend modifying the covariance to capture temporal autocorrelation, either by bounding the covariance kernel hyperparameters or adding white noise.

LSTM models with spatial features. We believe the performance of the LSTM models could improve if they were conditioned on the geographical location of stations, which would allow the models to access spatial correlation between stations.

Hybrid models. A hybrid structure that concatenates Conditioned Bayesian Deep models [5] with AR-LSTMs results would allow to capture the stationary features' impact on wind prediction. This takes advantage of the AR-LSTM output of future satellite measurements and can generalise to an arbitrary point in the vicinity of four satellites.

Ensemble models. Given that different models have different performance across the stations it could be possible to create an ensemble of models that exploit the models with best predictive power for each station.

XGBoost. Experiment with using different varying numbers of neighbouring satellite points to predict a ground station from 1 to many more than 4. It would also be worthwhile to consider other ways inputting features to avoid any correlation. We could also input the roughness and elevation features by taking into account these features between the satellite and ground station.

5.2 New models

5.2.1 Deep neural networks

The deep neural network approaches explored in this report can be developed further in a number of ways. Given that the problem identified appears to be one of transforming distributions of satellite wind to ground wind, this is a problem naturally suited to flow-based generative models, e.g., inverse autoregressive flows or GLOW [2, 4]. The basic idea of such models is that they find an invertible transformation that smoothly transforms one probability distribution into another. They are typically used to enhance the flexibility of variational autoencoder approaches, permitting the use of sampling from arbitrarily complex distributions with a simple origin distribution. For these data, it would be necessary to find a flow conditioned on the terrain.

Another method to explore along the probabilistic approach taken here is to replace the histogram method used to describe the distribution of wind data with a more efficient distributional code [10]. One possibility is to use the method of expectiles, or centiles, in which the distributions is described either along a key set of basis functions in the case of the former, or by effectively discretizing the cumulative density function in an efficient way in the case of the latter.

Finally, contrastive approaches to learning probability distributions have been hugely successful in recent years [7]. The crucial component of contrastive approaches is that they learn to maximise mutual information between contextual variables that best distinguish different datapoints. A particular advantage of these approaches is that they have been shown to be very effective forecasters and can generalize well. However, they require far more data than was available to us. With data collected worldwide, it may be feasible to use such an approach.

5.2.2 Others

Another approach that would be worth pursuing for forecasting ground wind speeds and direction, is to apply flexible Bayesian hierarchical models, where both temporal and spatial effects can be incorporated. In particular, a Stochastic Partial Differential Equations (SPDE) approach

with Integrated Nested Laplace Approximation (INLA) should be fitted, for fast estimations of the posterior distribution.

5.3 Interpretability and explainability

Finally, we propose to focus on interpretability and explainability studies in future work.

A first approach will be to investigate which terrain features correlate with better performances on baseline models. For locations with “easy” terrain feature, baseline models provide confident predictions, and even better prediction than complicated models. A comprehensive study along this line, leveraging the whitebox nature of these baseline models, warrants further studies.

On the other hand, more sophisticated approaches to explore interpretability and explainability in more complex models involve performing LIME and SHAP analysis. Using extensions of these methods for neural network models would help to identify important features as well as providing explainability on how the models perform in various types of stations.

6 Team members

Alphabetical order by last name

Eduardo Arnold is a PhD candidate at University of Warwick working on machine learning for autonomous driving perception applications. Eduardo contributed to the Dataset class, data exploration & analysis and co-developed the LSTM methods.

Jiaxin Chen is a PhD candidate at University of Exeter working on using met-ocean data for autonomous marine system decision-making. Jiaxin developed the Baseline SARIMA method, contributed to Data Exploration and Data Preparation of Kriging method.

Ivan Croydon-Veleslavov is a PhD candidate at Imperial College London developing interpretable machine learning methods to intuit stem cell dynamics. He implemented fourier transforms of the time-series

variables, contributed to the feature engineering by decomposition and co-developed the LSTM models.

Anurag Deshpande is a PhD candidate at University College London applying Bayesian statistics and machine learning techniques to Cosmology. He guided the team as one of the facilitators for the project, and developed the Gaussian process regression for mean year approach.

Tanaya Guha is an Assistant Professor in the department of Computer Science, University of Warwick. She served as the PI of this challenge.

Yixuan He is a PhD candidate at University of Oxford in Statistics. She contributed to building Satellite Mean Approach and Inverse Distance Weighted Interpolation and generating results from these approaches. She also proposed making comparisons on four typical stations as well as overall performance across stations.

Ben Moseley is a PhD candidate at the University of Oxford researching the use of physics-informed machine learning for simulation and inversion in physics, as well as applied ML for space exploration. He guided the team as one of the facilitators for the project, and helped generate map visualisations of the dataset.

Gim Seng Ng is a postdoctoral researcher at IPhT CEA Paris-Saclay. He co-developed the satellite mean/IDW baseline approach and contributed to data exploration by investigating data quality.

Luke Prince is a postdoctoral researcher at McGill University and Mila (Quebec AI Institute) developing latent variable models of data in neuroscience. Luke contributed a Dataset class to read the data. He also designed and built the probabilistic conditional-GRU model.

Thomas Statham is a PhD candidate at the University of Bristol working on developing novel methods of small area estimation of population counts in urban areas. His contribution to this data study group was implementing the Regression-Kriging model.

Peter Strong is a PhD candidate at the University of Warwick working on developing models of the movement of human populations. Peter contributed to the group by implementing the XGBoost model.

Greenvest Solutions (www.greenvestsolutions.com), represented by technical lead, **Giambattista Salvi** is on a mission to accelerate renewable energy adoption worldwide. The start-up provides strategic technology to plan, monitor, and assess clean energy projects globally. Our proprietary solution combines Machine Learning (ML) and Big Data analytics on multispectral satellite data to model and predicts renewable energy plants' performances.

References

[1] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). DOI: 10.1145/2939672.2939785.

[2] Diederik P. Kingma, Tim Salimans, and Max Welling. "Improving Variational Inference with Inverse Autoregressive Flow". In: *CoRR* abs/1606.04934 (2016). arXiv: 1606.04934. URL: http://arxiv.org/abs/1606.04934.

[3] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[4] Durk P Kingma and Prafulla Dhariwal. "Glow: Generative Flow with Invertible 1x1 Convolutions". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio et al. Curran Associates, Inc., 2018, pp. 10215–10224. URL: http://papers.nips.cc/paper/8224-glow-generative-flow-with-invertible-1x1-convolutions.pdf.

[5] Charlie Kirkwood and Theo Economou. "Bayesian deep learning: a new era for'big data'geostatistics?" In: *arXiv preprint arXiv:2008.07320* (2020).

[6] Lubos Mitas and Helena Mitasova. "Spatial interpolation". In: *Geographical information systems: principles, techniques, management and applications* 1.2 (1999).

[7] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. “Representation Learning with Contrastive Predictive Coding”. In: *CoRR* abs/1807.03748 (2018). arXiv: 1807 . 03748. URL: http://arxiv.org/abs/1807.03748.

[8] Adam Paszke et al. “Pytorch: An imperative style, high-performance deep learning library”. In: *Advances in neural information processing systems*. 2019, pp. 8026–8037.

[9] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.

[10] Mark Rowland et al. “Statistics and Samples in Distributional Reinforcement Learning”. In: ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, 2019, pp. 5528–5536. URL: http://proceedings.mlr.press/v97/rowland19a.html.

turing.ac.uk
@turinginst