# Time Series Analysis Project - Wind Speeds at Valentia Island Weather Station

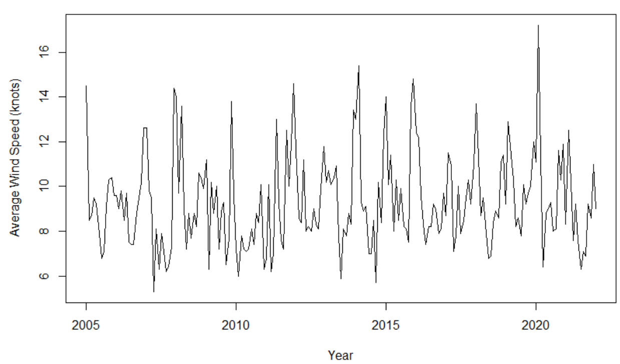
Joseph Claro - 21317623

April 2025

## 1 Initial Exploration

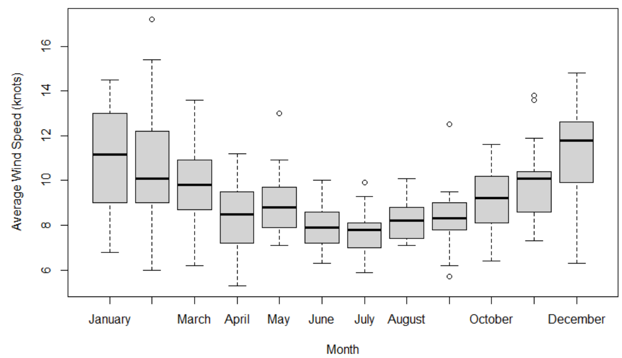
The data worked on for this project are Met Éireann's records of the average monthly wind speeds, in knots (1.852 km/h), measured at the Valentia Island weather station in south Co. Kerry between January 2005 and February 2025. The time series was downloaded from the `met.ie` website. The data originally had 2 `NA` values, which were each filled in with the mean of the values at either side of the empty entry.

Models were fit to the first 85% of the original dataset, with forecasts compared to the final 15%.

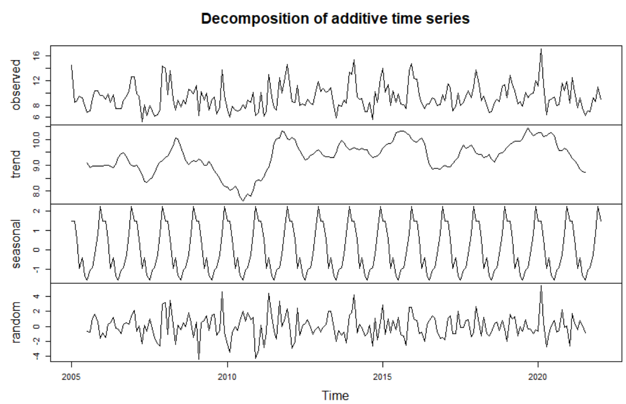



The above plot reveals regular-looking spikes and dips throughout the time series, suggesting there could be seasonal effects at play. Variance is mostly stable and there are no major signs of linear trend. Once any seasonal trends are removed, weak stationarity could be assumed.

A seasonal boxplot for the data captures the time series' expected behaviour in any given year. Summers seem to be less windy on average compared to the

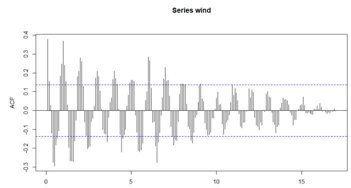


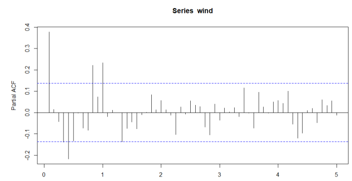

winter months. Data recorded in the winter also appears much less consistent, as is apparent by the wider interquartile ranges and high-valued outliers near January and December.

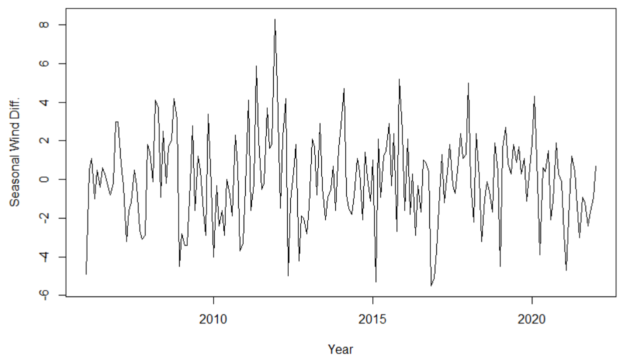



Using the `decompose` function on the data ends up being mostly inconclusive, as the seasonal component picked out by R only ranges between -1 and 2, about the size of the random noise component, compared to the overall series whose values lie in (4,16).

Also shown are ACF and partial ACF plots, measuring autocorrelations in the data at a range of different lags. In the ACF plot, the slow decay and strong sinusoidal evolution of these autocorrelations mean the data must be seasonally differenced to bring the series closer to stationarity, at which point an ARMA model could potentially be fitted. I think the spike (and cut-off for significant autocorrelations) at lag-12 in the Partial ACF plot validates this to some degree.

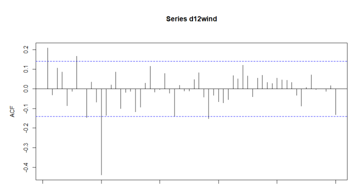


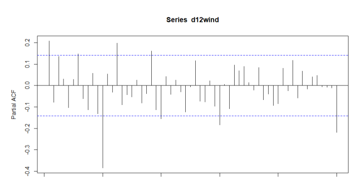




After seasonal differencing, the data still seems to have a regularly spiking pattern, so perhaps differencing hasn't captured all seasonal behaviour.

## 2 Identifying Models

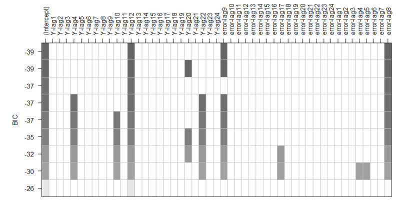



ACF and PACF plots for the differenced series show the oscillating trend through the lags has been removed. High autocorrelation at lag-12 appearing again in the PACF would suggest, in the case that a pure AR model is used, that AR(12) would be the optimal model. This would so far give a seasonal model with an AR(12) component.

The `eacf` function, having been applied to the seasonally differenced data, also suggests an AR(12) component, as the highest 0-value above a lower triangle of significant terms is at (12,0).

```
> eacf(d12wind)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o x o o o o  x  o  o
1 x o o o o o x o o o o  x  x  o
2 x x x o o o o o x o o  x  x  o
3 x x o o o o o o o o o  x  x  x
4 x x x o o o o o o o o  x  o  o
5 x x x x o o o o o o o  x  x  o
6 x x x o x o o o o o o  x  o  o
7 x x x o o o o o o o o  x  o  o
```



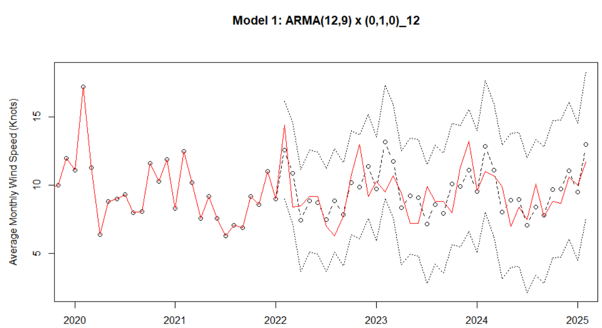
Using `armasubsets` to work out the Bayesian Information Criterion for various model types fitted to the seasonally differenced data gives two strong concise models, whose BIC values only differ by two:

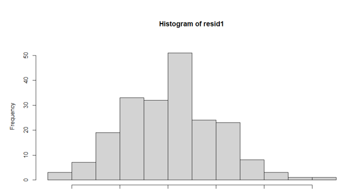
- ARIMA(12,9) $\times(0,1,0)_{12}$ and
- ARIMA(12,8) $\times(0,1,0)_{12}$.

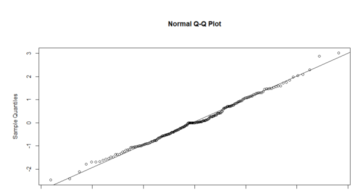
# 3 Model Fitting and Prediction

## 3.1 ARIMA(12,9) $\times(0,1,0)_{12}$

The model, once fitted to the data, seems to behave reasonably accurately, with all measured wind speeds sitting within the 95% confidence interval.



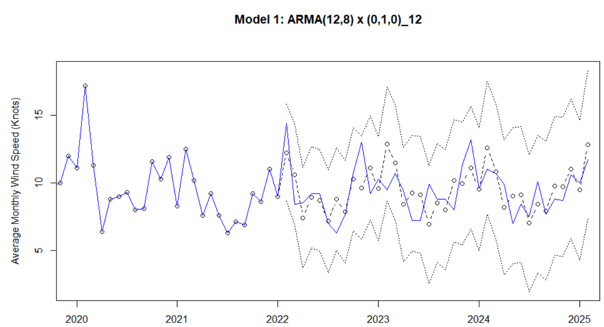



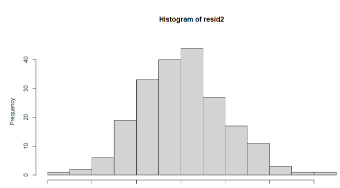



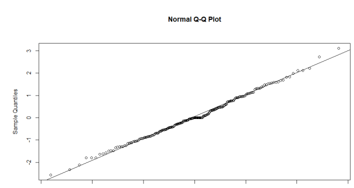
Residuals for the first fitted model seem mostly normally distributed, as per the pictured bell-curved histogram and the density of points lying on the QQ-plot's diagonal.

### 3.2 ARIMA(12,8) $\times(0,1,0)_{12}$

The same can be said for the second model, given how similar the two are in structure. As before, model residuals are normally distributed.




Given that the second model has one less MA term than the first, and seems to fit the measured data excluded from the training set, I'd prefer it to the first.