# Aspects of High Dimensional Energy Modelling and Forecasting

Dr Jethro Browell
EPSRC Research Fellow

University of Strathclyde, Glasgow, UK
jethro.browell@strath.ac.uk

Isaac Newton Institute, Cambridge, 27 February, 2019

# Contents

- Where do high-dimensional problem emerge?
  - Spatial: generation and demand
  - Temporal: multi-temporal decision-making
  - Modelling: (potentially) huge feature space for statistical learning
  - Multivariate/other: prices, line ratings, reliability…
- Large Feature Spaces and Hierarchies
- Dynamic Temporal Dependency
- Spatio-temporal Forecasting
  - High-dimensionality and sparsity
  - Dynamic dependency structures and atmospheric regimes

# Contents

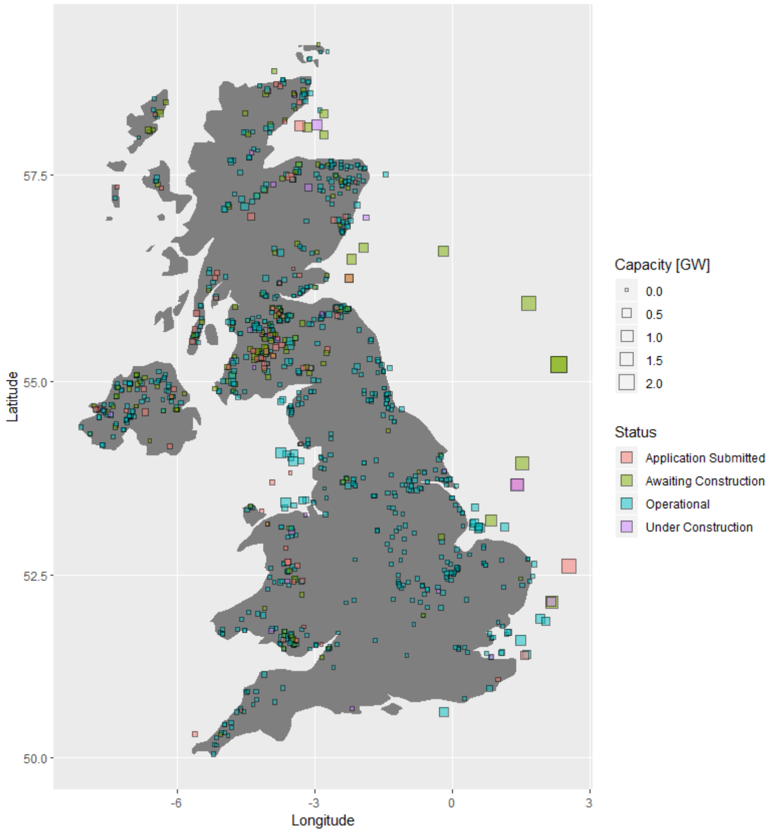
- Acknowledgements:
  - Ciaran Gilbert, David McMillan (Strathclyde)
  - Bri-Mathias Hodge, Tarek Elgindy (NREL)
  - Dan Drew, Kostas Philippopoulos (Reading)

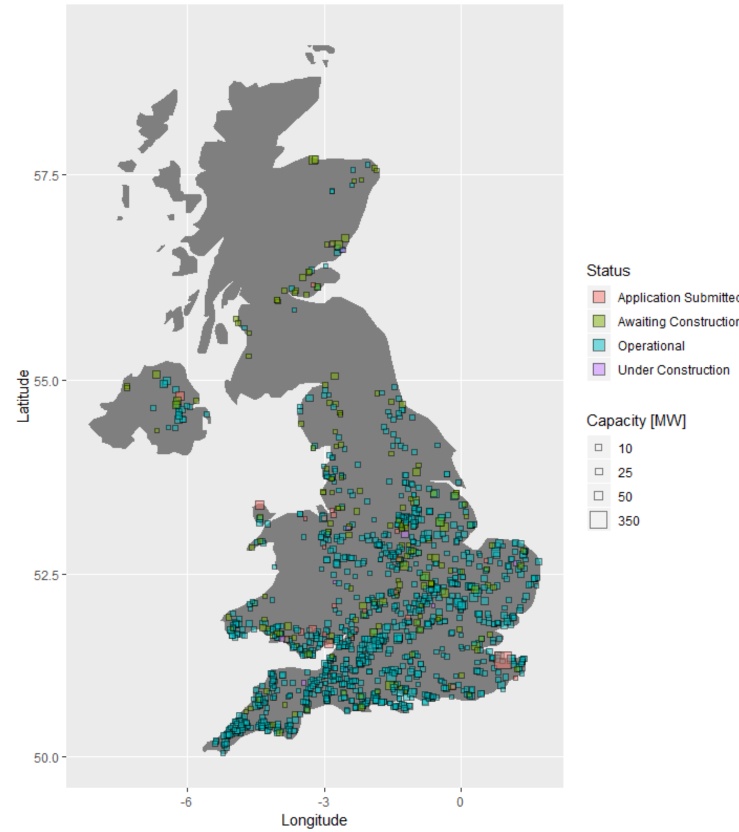
# High Dimension: Spatial

- Generation:

940+ Wind Farms

1300+ Solar Farms (+domestic PV)

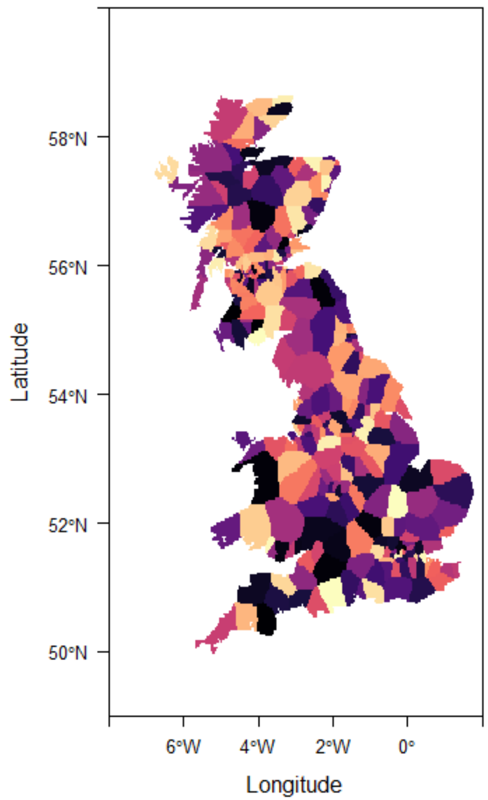


# High Dimension: Spatial

- Demand:

350+ Grid Supply Points/Nodes
400,000+ Substations
40,000,000+ Smart Meters




# High Dimension: Temporal

- Forecast errors often exhibit auto-correlation:
  - Inertia in underlying processes

- Sequential decisions/constraints require multi-temporal modelling:
  - Storage management
  - Maximum/minimum run-times/down-times
  - Cumulative quantities (energy, fuel)

# High Dimension: Features and Hierarchy

- Lots of (potential) explanatory information
  - History
  - Weather (multiple variables on a spatial grid)
  - Other observations (e.g. levels in hierarchy)
  - Engineering large numbers of features from modest numbers of explanatory variables is often beneficial

- Natural hierarchies:
  - Can improve model fidelity
  - Consistency may be necessary for some applications

# Wind Power Forecasting

Leveraging turbine-level data for wind power forecasting

# Large Feature Spaces and Hierarchies

## Motivation:

1. Gather as much useful information as possible to improve forecast skill
   - NWP – multiple models and variables on a grid, ensembles, engineered features…
   - High frequency data and engineered features (especially in very short-term)
   - Other levels of hierarchy
2. Coherency across hierarchy (in some cases)

# Large Feature Spaces and Hierarchies

## Motivation:

1. Gather as much useful information as possible to improve forecast skill
   - NWP – multiple models and variables on a grid, ensembles, engineered features…
   - High frequency data and engineered features (especially in very short-term)
   - ***Other levels of hierarchy***
2. Coherency across hierarchy (in some cases)

# Large Feature Spaces and Hierarchies

- Wind farm power curve is complicated by many factors: layout, terrain etc

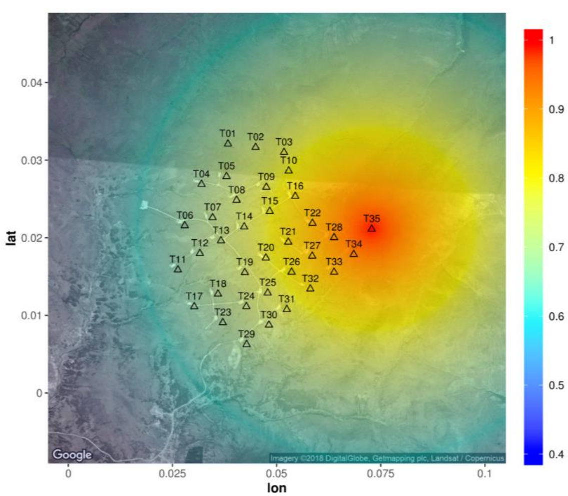
- When fitting a model it is difficult to distinguish between random variation and true processes…

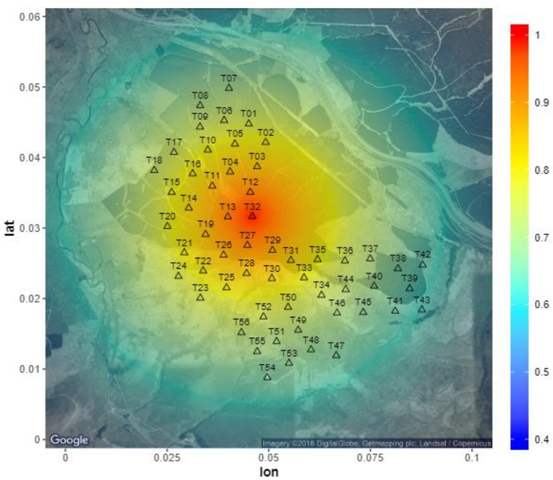
- …Perhaps looking at individual turbines could help!

# Large Feature Spaces and Hierarchies

**Set up**

- GBM for quantile regression $q^{\alpha} = f_{\mathrm{GBM}}^{\alpha}(\boldsymbol{x}_{\mathrm{NWP}})$
- 2 Wind Farms with 35 and 56 turbines
- NWP inputs plus engineered features
- 30 minute wind farm production
- 30 minute wind turbine production



# Large Feature Spaces and Hierarchies

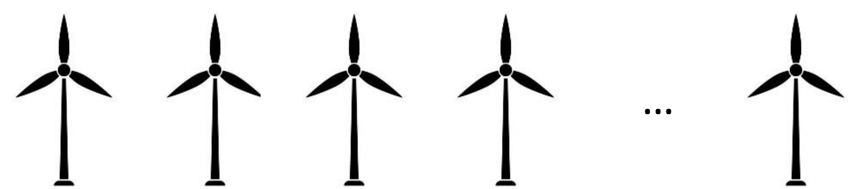
**Method 1 / GBM+T**

1. Produce deterministic forecasts for individual turbines
2. Use these as additional features

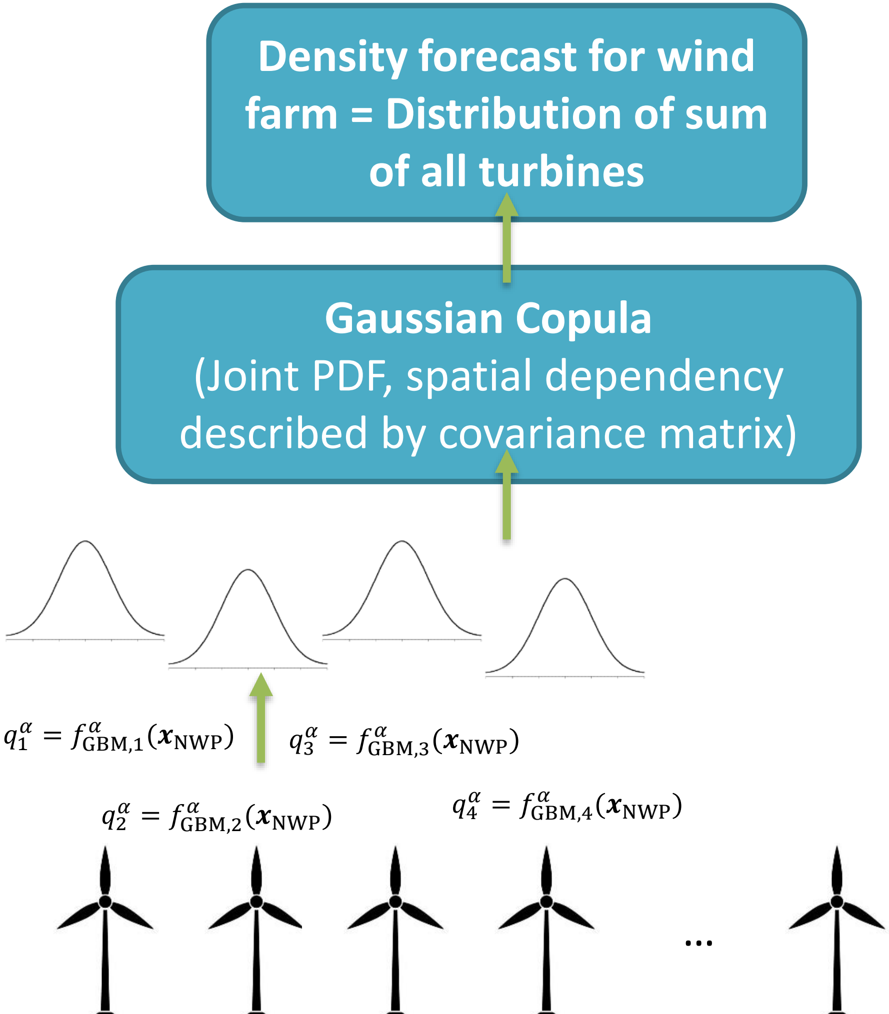
**Density forecast for wind farm**

$$q^{\alpha} = f_{\mathrm{GBM}}^{\alpha}(\boldsymbol{x}_{\mathrm{NWP}}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_N)$$

$\boldsymbol{x}_{(1)}$ $\boldsymbol{x}_{(2)}$ $\boldsymbol{x}_{(3)}$ $\boldsymbol{x}_{(4)}$



# Large Feature Spaces and Hierarchies




**Method 2 / Cop**

1. Produce multivariate density forecast for all turbines
   a) Marginals as before
   b) Gaussian copula for spatial dependency

# Large Feature Spaces and Hierarchies

**Method 1 / GBM+T**

1. Produce deterministic forecasts for individual turbines

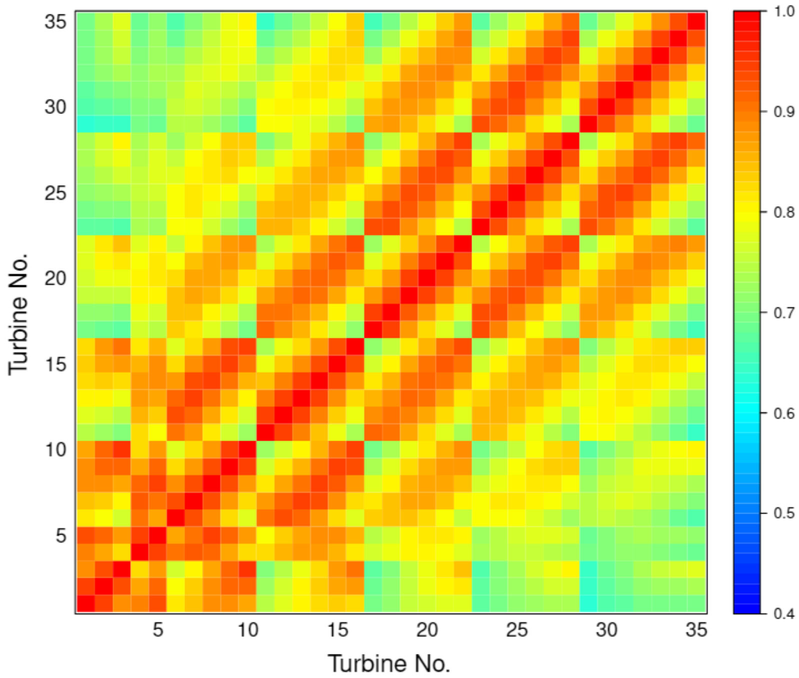
2. Use these as additional features

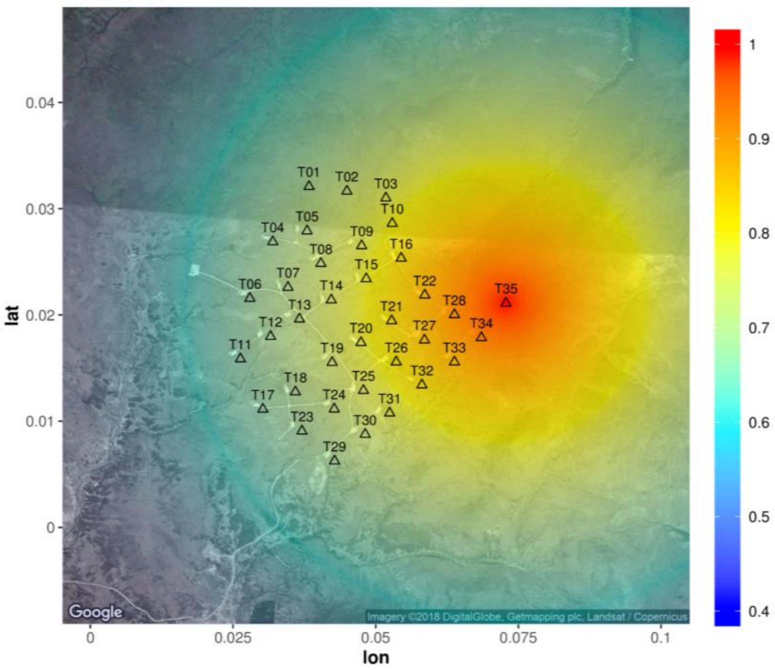
**Method 2 / Cop**

1. Produce multivariate density forecast for all turbines
   a) Marginals as before
   b) Gaussian copula for spatial dependency

# Large Feature Spaces and Hierarchies

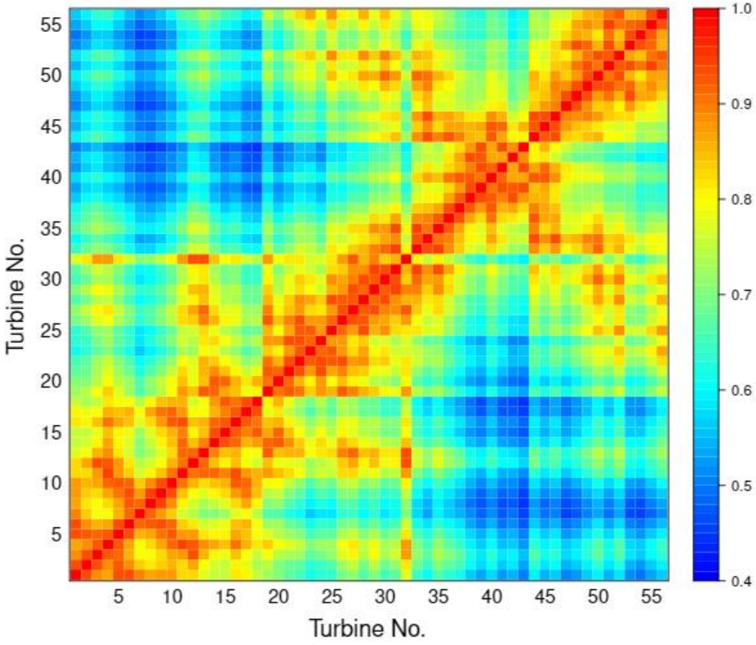
**WF-B**

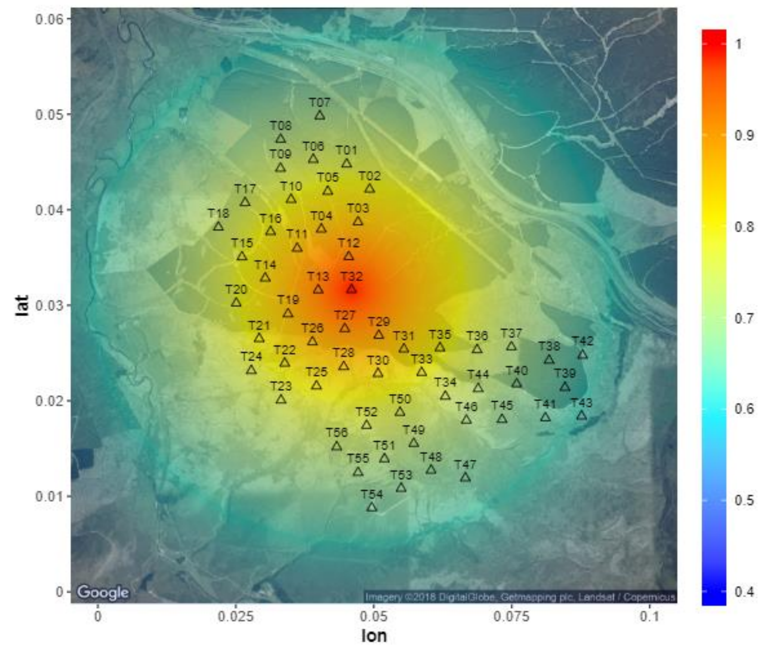


# Large Feature Spaces and Hierarchies

**WF-A**



# Large Feature Spaces and Hierarchies

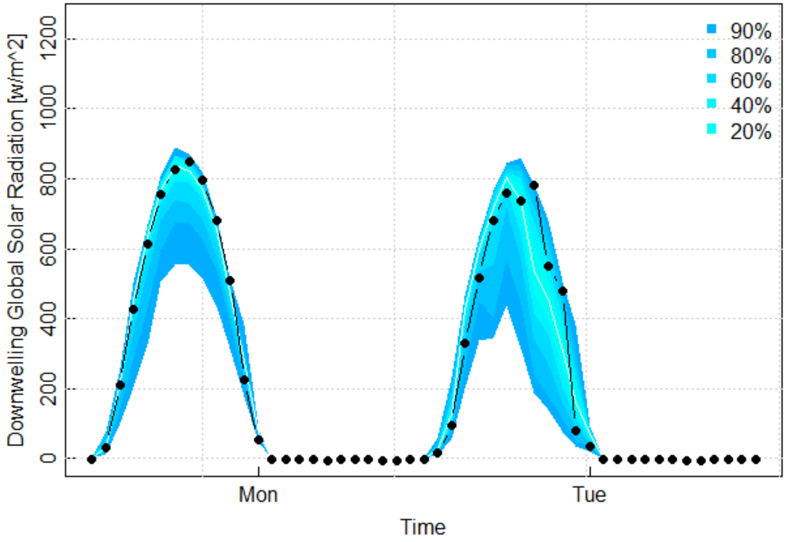
| WF | Score | AnEn | GBM | GBM+T | Cop |
|---|---|---|---|---|---|
| **WF-A** | **MAE** | 9.88 | 9.69 | 9.27 | ***9.11*** |
| | **CRPS** | 7.12 | 7.02 | 6.74 | ***6.66*** |
| **WF-B** | **MAE** | 11.49 | 11.39 | ***11.21*** | 11.26 |
| | **CRPS** | 8.20 | 8.10 | ***8.00*** | 8.02 |

# Solar Power Forecasting

## Dynamic Temporal Dependency

# Dynamic Temporal Structure

- Density forecasts don't give information about temporal structure
- Solar power production looks very different on different types of day…
- …so do forecast errors!

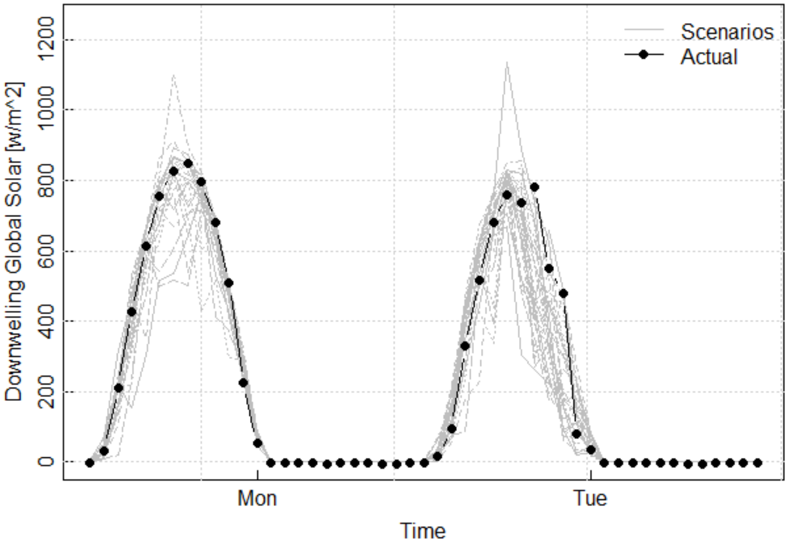


# Dynamic Temporal Structure

- Density forecasts don't give information about temporal structure

- Solar power production looks very different on different types of day…

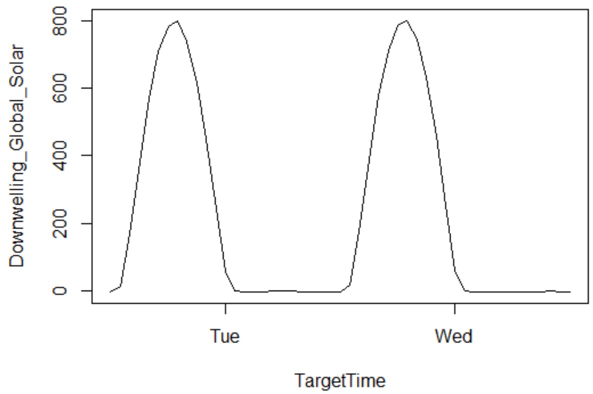
- …so do forecast errors!

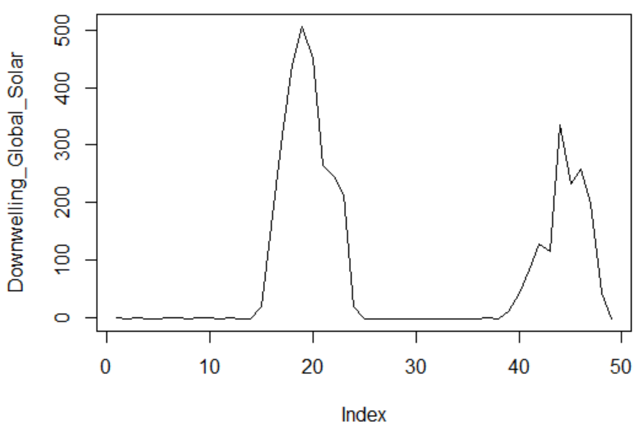


# Temporal Dependency
## Different Day Types

### Different Sources of Error:

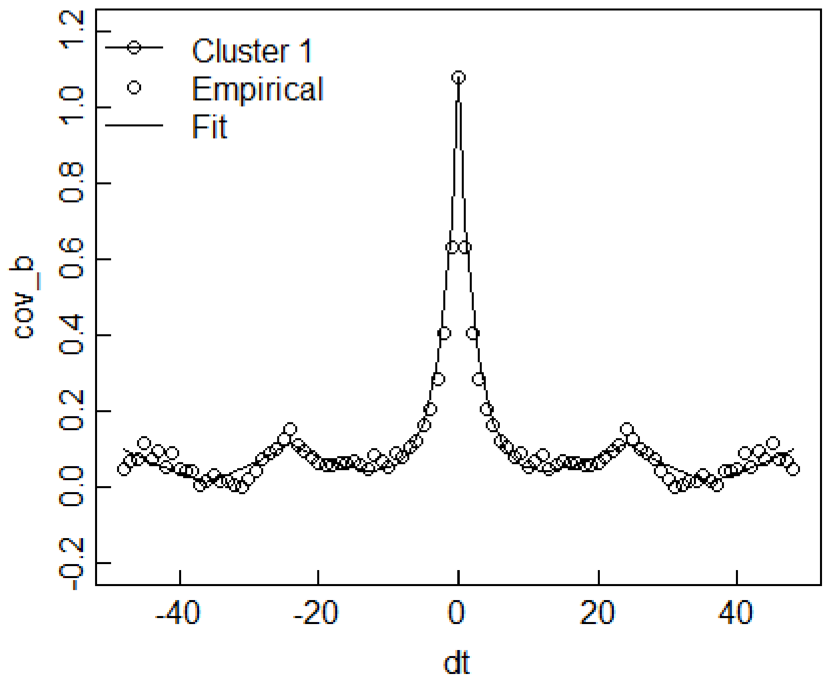
- **Clear Day:** Clear sky estimate (aerosol content etc)
- **Partially Cloudy:** Time and duration of clear/cloudy spells
- **Cloudy Day:** Irradiance penetrating cloud layer(s)

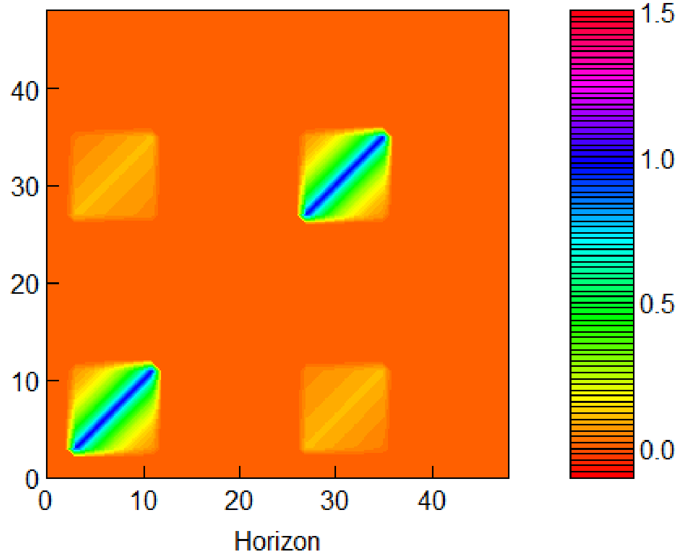


# Temporal Dependency

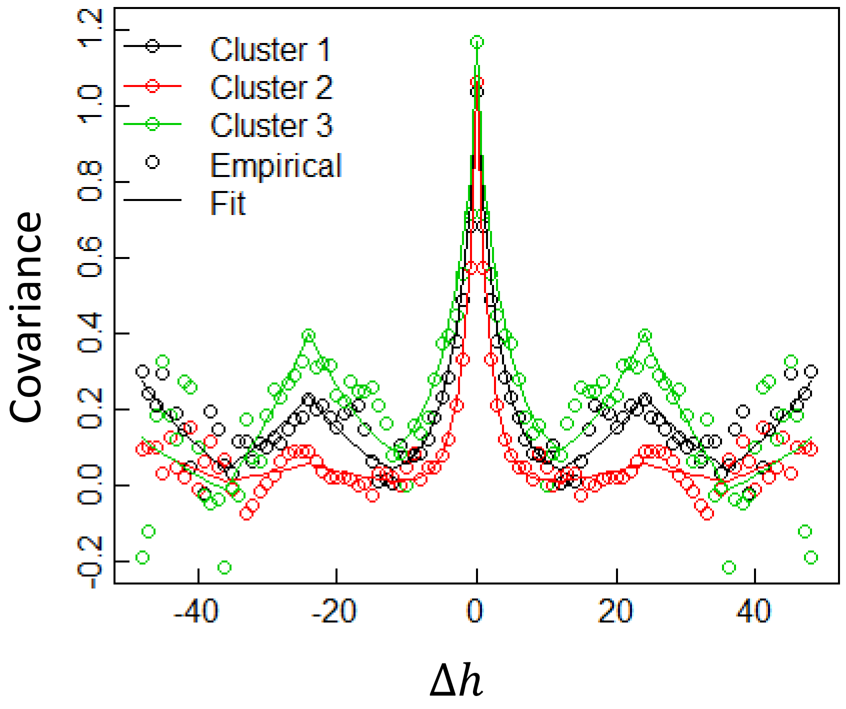
## Single Day Type

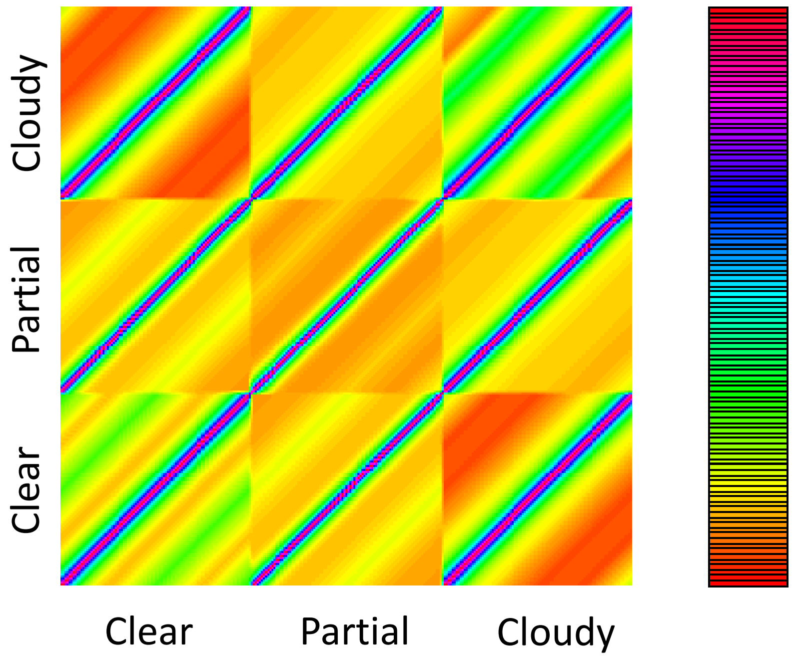


# Temporal Dependency

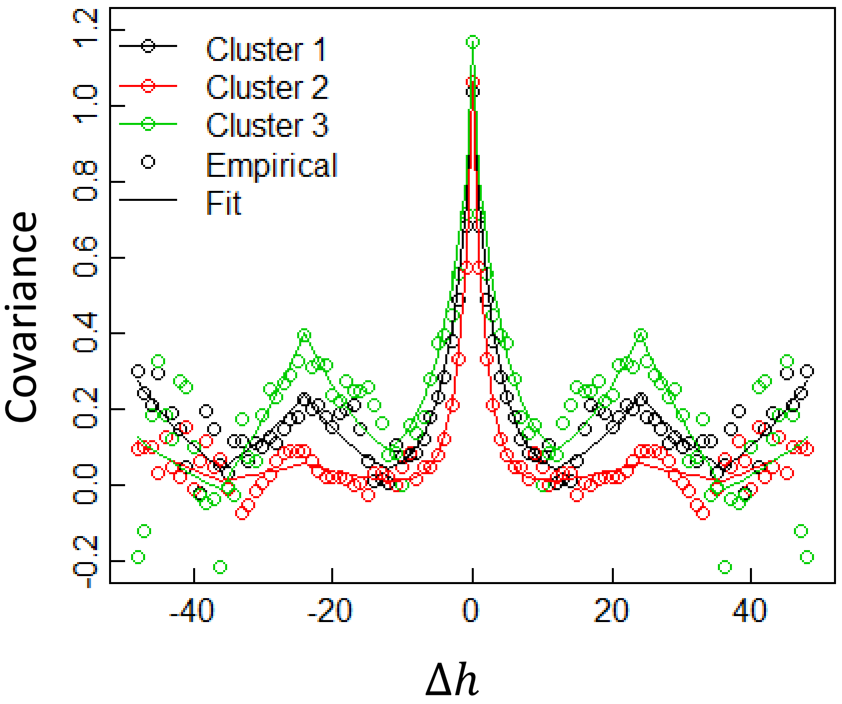
## Three Day Types

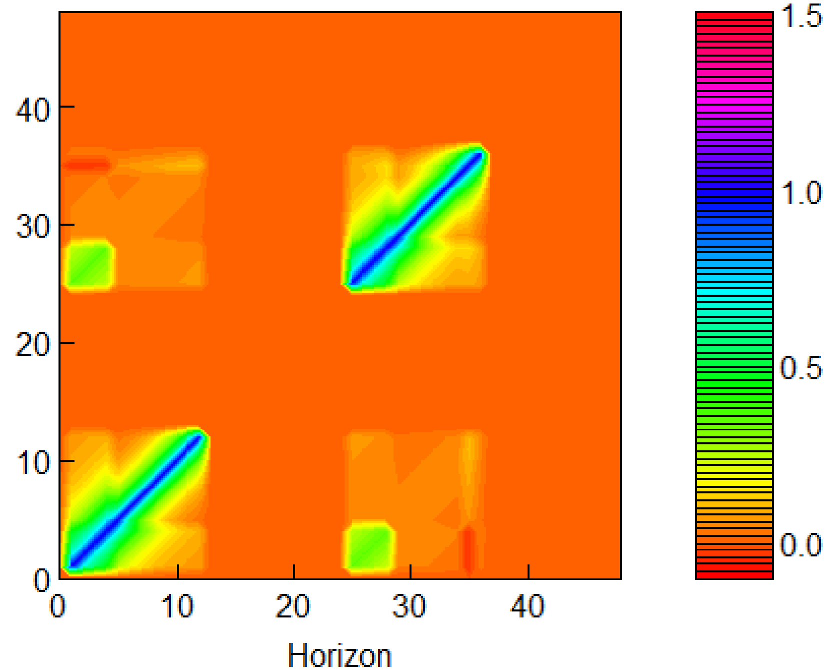


# Temporal Dependency

## Three Day Types

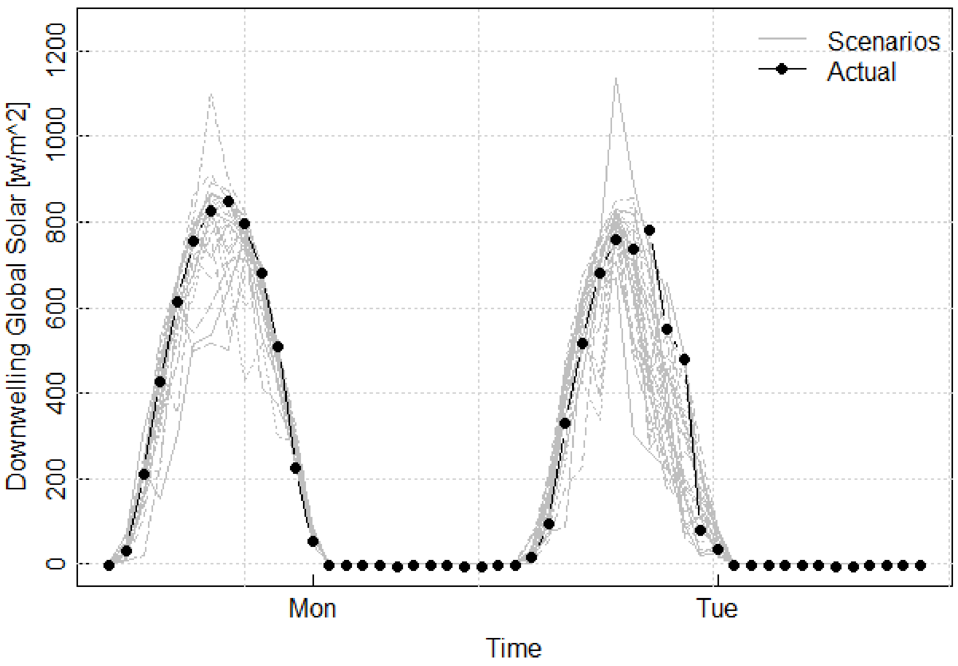


# Gaussian Copula

## Sampling

| Covariance Matrix | MV Energy Score | Variogram Score |
|---|---|---|
| Identity | 419.0 | 27348 |
| Static | 411.9 | 27147 |
| Dynamic | 411.6 | **27087** |



# Wind Power Forecasting

## Very Short-term

# Wind Power Forecasting

## Very Short-term

Vector Auto-regression:

$$\boldsymbol{y}_{t+1} = \sum_{\tau=0}^{p-1} \boldsymbol{A}_\tau \boldsymbol{y}_{t-\tau} + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{y}_t, \boldsymbol{\varepsilon}_t \in \mathbb{R}^N \ , \ \boldsymbol{A}_i \in \mathbb{R}^{N \times N}$$

Full Parameterisation:

$$\boldsymbol{A}_i = \begin{pmatrix} a_{i,11} & \cdots & a_{i,1N} \\ \vdots & \ddots & \vdots \\ a_{i,N1} & \cdots & a_{i,NN} \end{pmatrix} \longrightarrow pN^2$$

parameters to estimate!

**Quickly becomes impractical for large $N$**

# Wind Power Forecasting

## Very Short-term

Sparse Vector Auto-regression:

Set most parameters to zero…

$$\boldsymbol{A}_i = \begin{pmatrix} a_{i,11} & 0 & 0 & \dots \\ 0 & a_{i,22} & 0 & \\ a_{i,31} & 0 & a_{i,33} & \\ \vdots & & & \ddots \end{pmatrix}$$

**Which ones?**

- Rank by a summary statistic and choose number of parameters that minimises some information criterion
- Penalised Linear Regression

# Atmospheric Regimes

**Spatio-temporal Structure:**

- If VAR parameters are static, we're assuming that the spatio-temporal structure is static

# Atmospheric Regimes

**Spatio-temporal Structure:**

- If VAR parameters are static, we’re assuming that the spatio-temporal structure is static

- It is easy to track changes, albeit with some lag…
- …we also know somethings about the underlying weather!

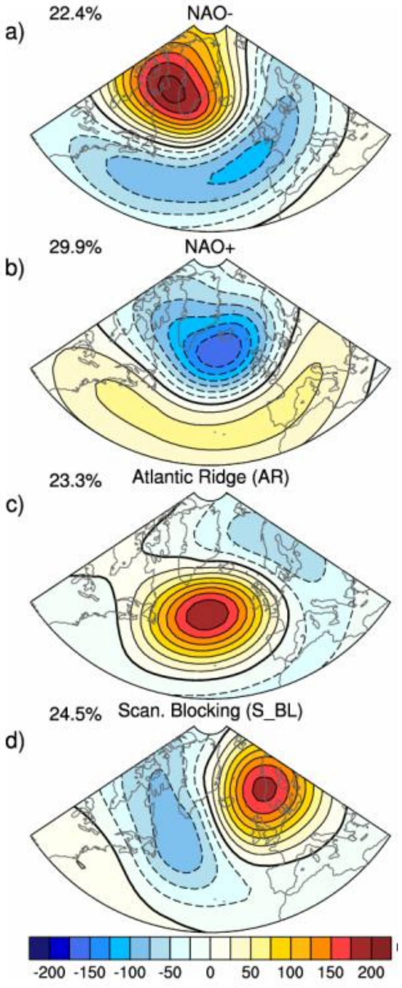
# Atmospheric Regimes

**Large-scale meteorological phenomena:**

- Persist for days, weeks and beyond
- Are associated with particular weather types or anomalies
- Provide seasonal predictability and information about short-term predictability

# Atmospheric Regimes

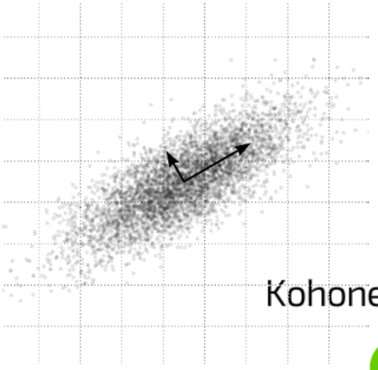
- Large-scale regimes fundamental to seasonal/sub-seasonal predictability
  - E.g. El Nino, North Atlantic Oscillation (right)
- Information: Are we expecting a wet and mild or a cold and dry winter in Europe?

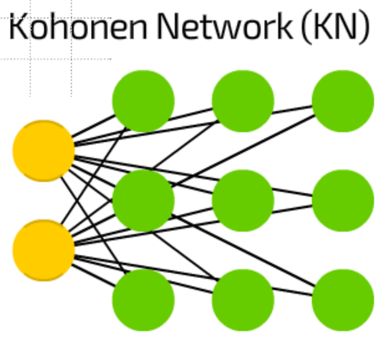



Christophe Cassou, Euro-Atlantic regimes and their teleconnections, ECMWF Seminar on Predictability in the European and Atlantic regions, 6 – 9 September 2010

# Atmospheric Regimes

- Many data-driven approaches to identification:
  - Principal Components
    - Classical dimension reduction
    - Linear in features
  - Kohonen Network/Self-organising Map
    - Unsupervised learning/dimension reduction
    - Non-linear in features
  - Auto-encoder
    - Supervised learning
    - Compression via non-linear features





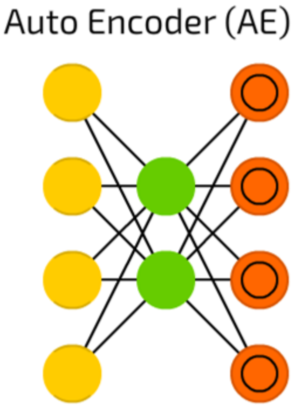




Figure Credit: www.asimovinstitute.org

# Atmospheric Regimes

Applications in short- and very short-term renewable energy forecasting:

1. Spatial correlation useful in very short-term forecasts; correlation structure ***depends on regime***

2. Structure in day-ahead forecast uncertainty; structure ***depends on regime***

# Atmospheric Regimes

1. Very short-term wind forecasting

$$\boldsymbol{y}_{t+1} = \sum_{\tau=0}^{p-1} \boldsymbol{A}_\tau \boldsymbol{y}_{t-\tau} + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{y}_t, \boldsymbol{\varepsilon}_t \in \mathbb{R}^N \;,\; \boldsymbol{A}_i \in \mathbb{R}^{N \times N}$$

Forecasts based on recent observations at spatially dispersed locations

# Atmospheric Regimes

## 1. Very short-term wind forecasting

$$\boldsymbol{A}_i = \begin{pmatrix} a_{i,11} & \cdots & a_{i,1N} \\ \vdots & \ddots & \vdots \\ a_{i,N1} & \cdots & a_{i,NN} \end{pmatrix}$$

Recent advances focuses on structure and dynamics of this matrix:

- Sparsity[1,2] (large scale applications)
- Adaptive Updates[3] (slow dynamics)
- ***Regimes***[4] (switching, fast dynamics)

# Atmospheric Regimes

## 1. Very short-term wind forecasting

$$\hat{\boldsymbol{y}}_{t+1} = \begin{cases} \sum \boldsymbol{A}_{\tau,1}\boldsymbol{y}_{t-\tau} & \text{if Regime} = 1 \\ \sum \boldsymbol{A}_{\tau,2}\boldsymbol{y}_{t-\tau} & \text{if Regime} = 2 \\ \vdots & \\ \sum \boldsymbol{A}_{\tau,M}\boldsymbol{y}_{t-\tau} & \text{if Regime} = M \end{cases}$$

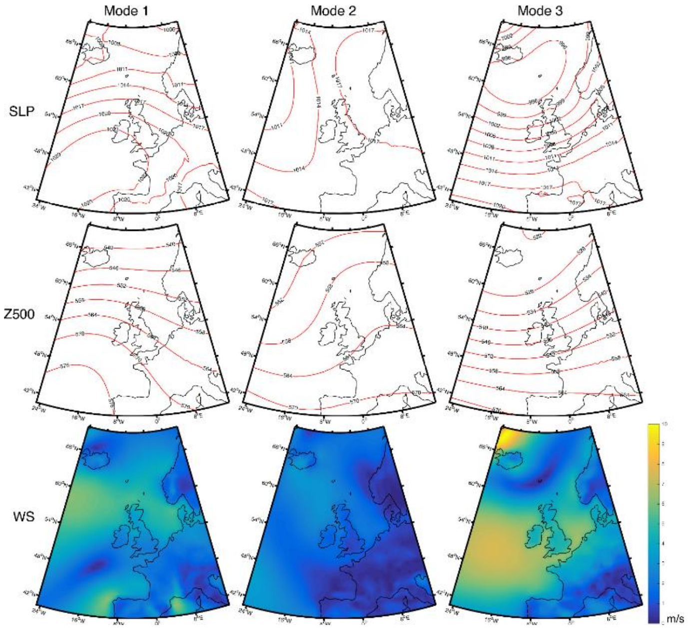



Figure Source: [4]

# Atmospheric Classification
## Self-organising Maps

**Unsupervised learning**
Training is entirely data-driven without using desired output examples
The objective is to find patterns in input data space:
*e.g. Cluster Analysis, Dimension Reduction*

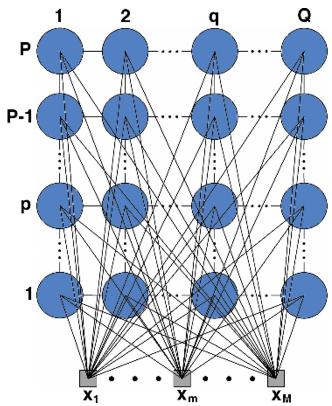
### Advantages of SOM

The SOM are analogous to other clustering algorithms (e.g. k-means) but provide:

- Better visualization – The resulting patterns are part of composites map
- Provides information regarding the relationship of the patterns
- Similar patterns are located close in the SOM – Dissimilar further apart
- Identify transient states between atmospheric patterns

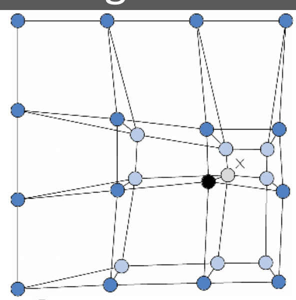
# Atmospheric Classification
## Self-organising Maps

### Architecture

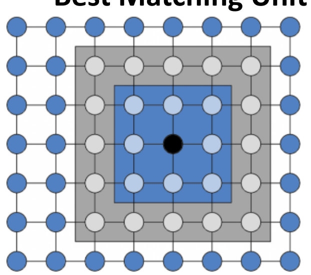



- Two layer network: Input layer & Output layer (*PxQ* neurons lattice)
- Each neuron is fully connected with the input nodes and with its neighboring neurons
- Characterized by its synaptic weights vector ***w*** and by its location at the SOM lattice

### Training – Learning of SOM



● Best Matching Unit



- **Competition**

An input pattern is presented to the network. A metric distance (e.g. Euclidean distance) is calculated for all neurons. The neuron with the smallest distance is the 'winner' (Best Matching Unit - BMU).

- **Cooperation**

The BMU through a radial basis function determines the topological neighborhood of the 'excited' neurons

- **Adaptation**

The BMU and its neighboring neurons weight vectors are updated towards the input vector.

# Atmospheric Classification

## Case Study and Application

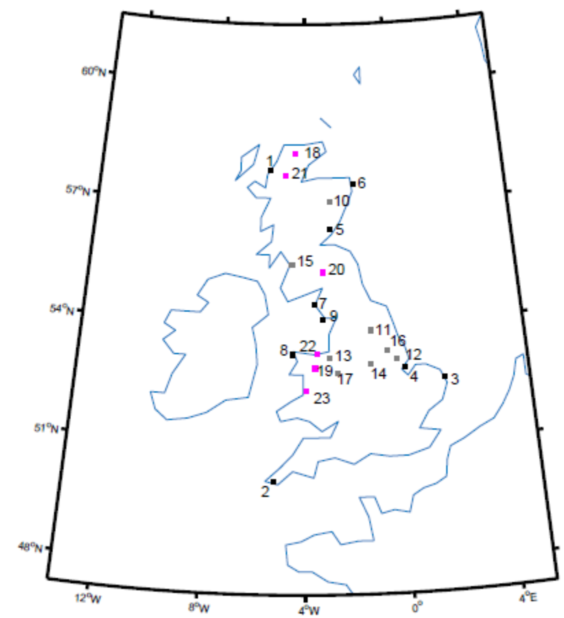
University of Strathclyde Glasgow

### Objective

- Examine the relationship of large-scale circulation and wind speeds over GB
- Identify wind regimes with distinct spatio-temporal characteristics for use in very-short-term forecasting

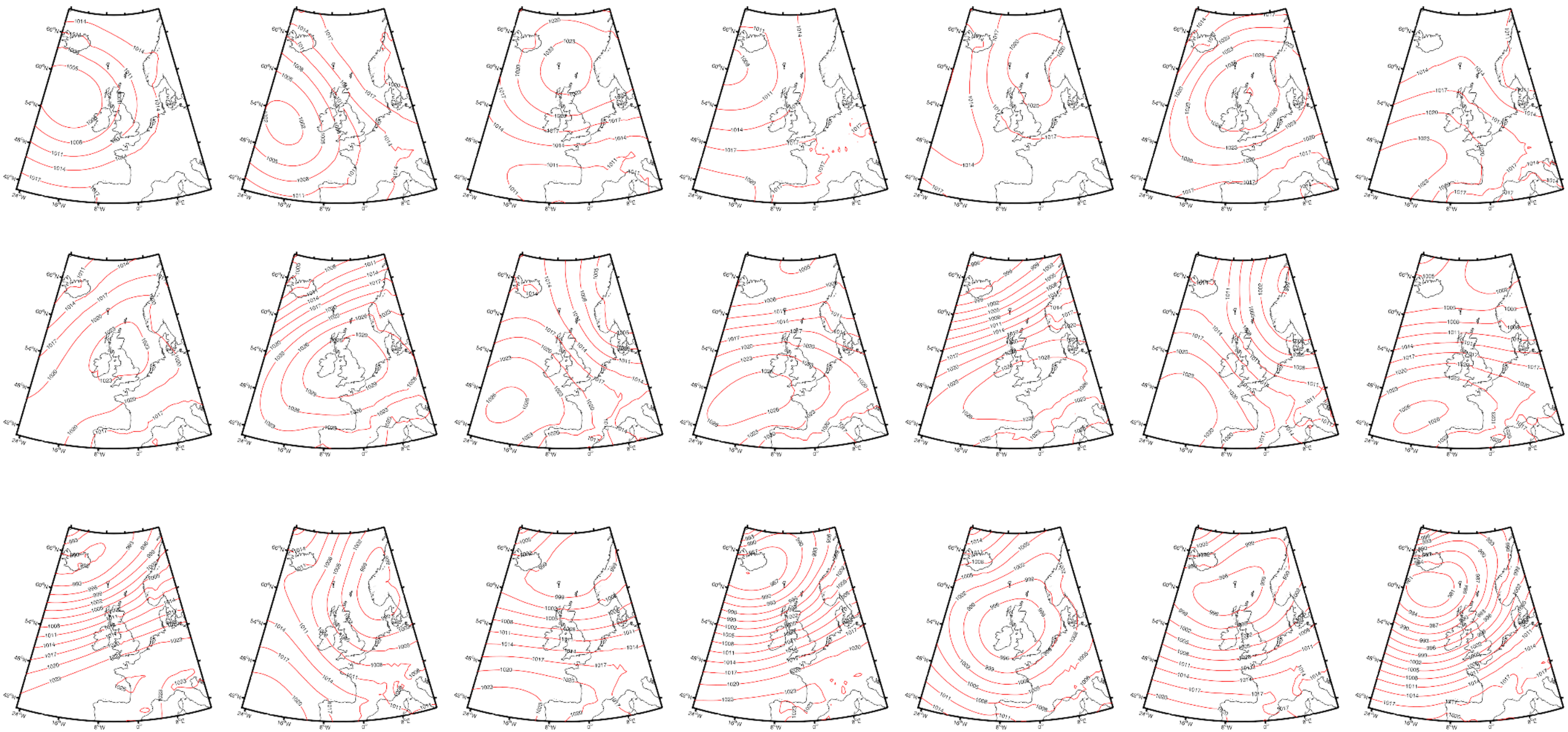
### Data

- Reanalysis Data: MERRA-2
  - SLP, Z500, $U_{10}$, $V_{10}$
  - 1980 to 2014, hourly resolution
  - Interpolated to 0.75°x 0.75° grid over GB
- Measurements from 23 Met Stations
  - 2002-2005 (Training), 2006-2007 (Testing)



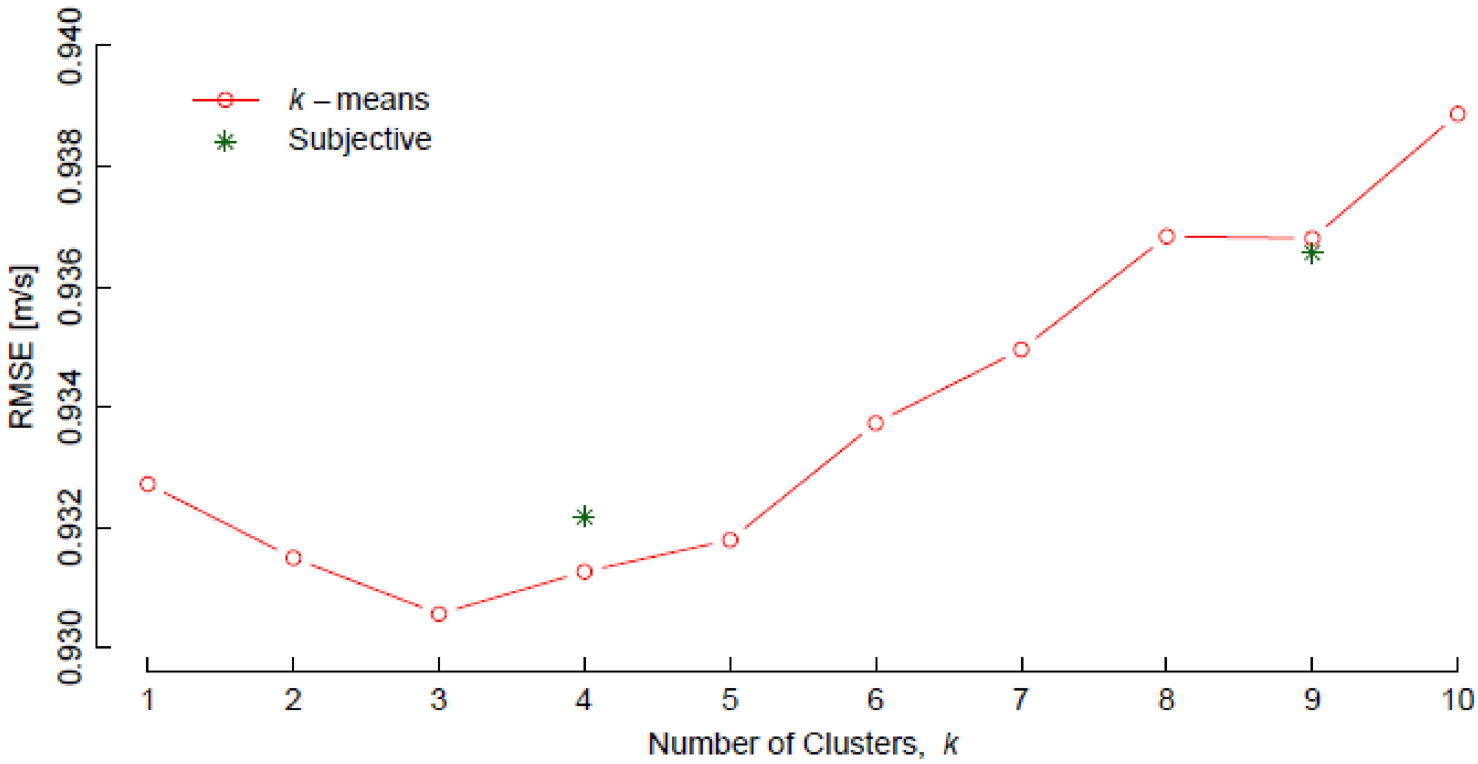
# Atmospheric Classification

## Self-organising Maps



# Atmospheric Classification

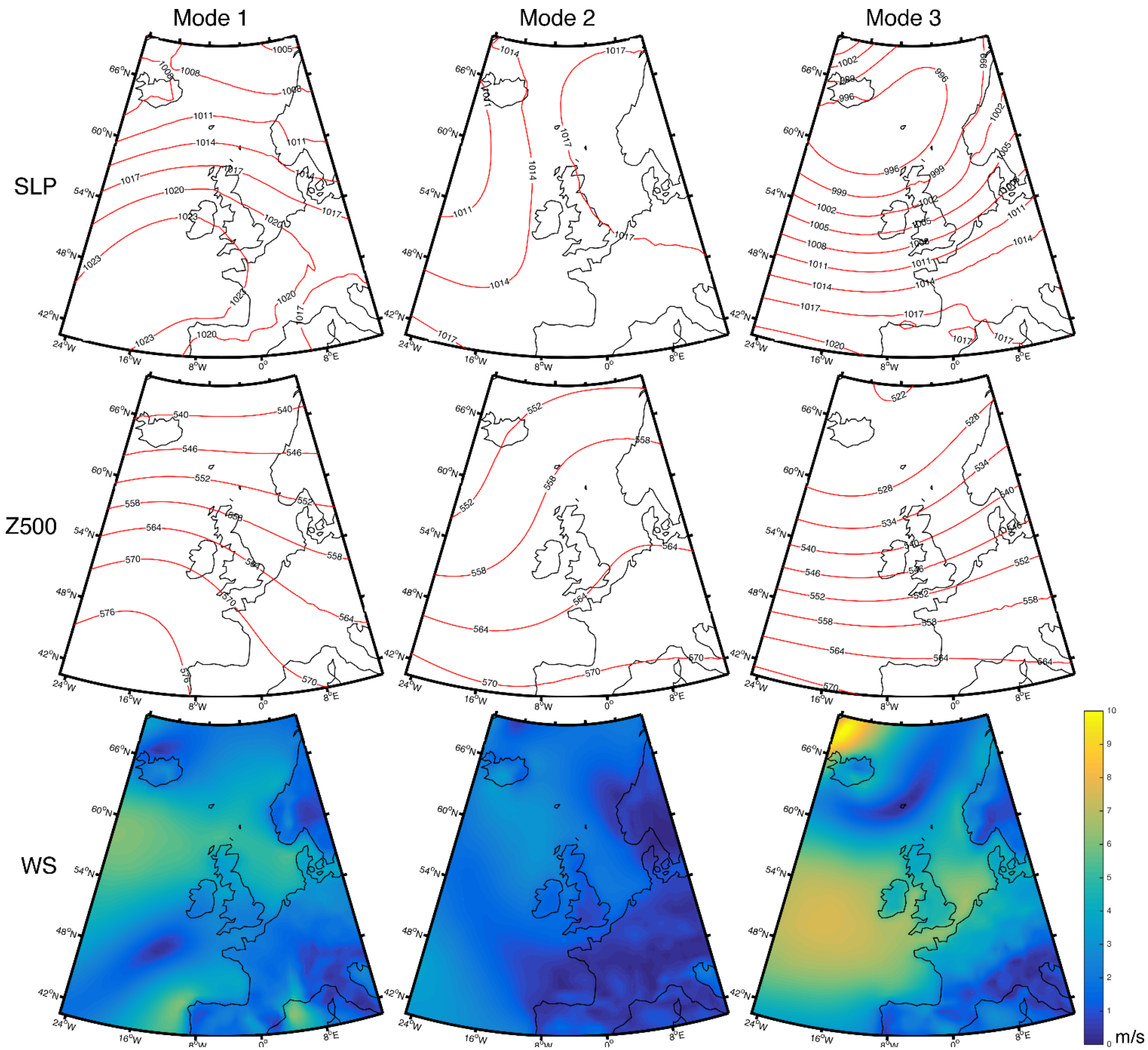
## Clustering for Optimal Forecast Performance



# Atmospheric Classification

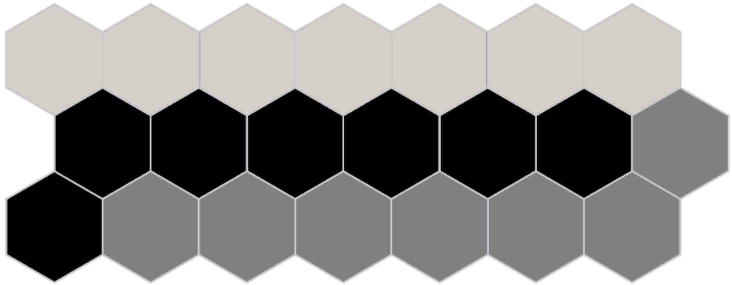
## Final Modes

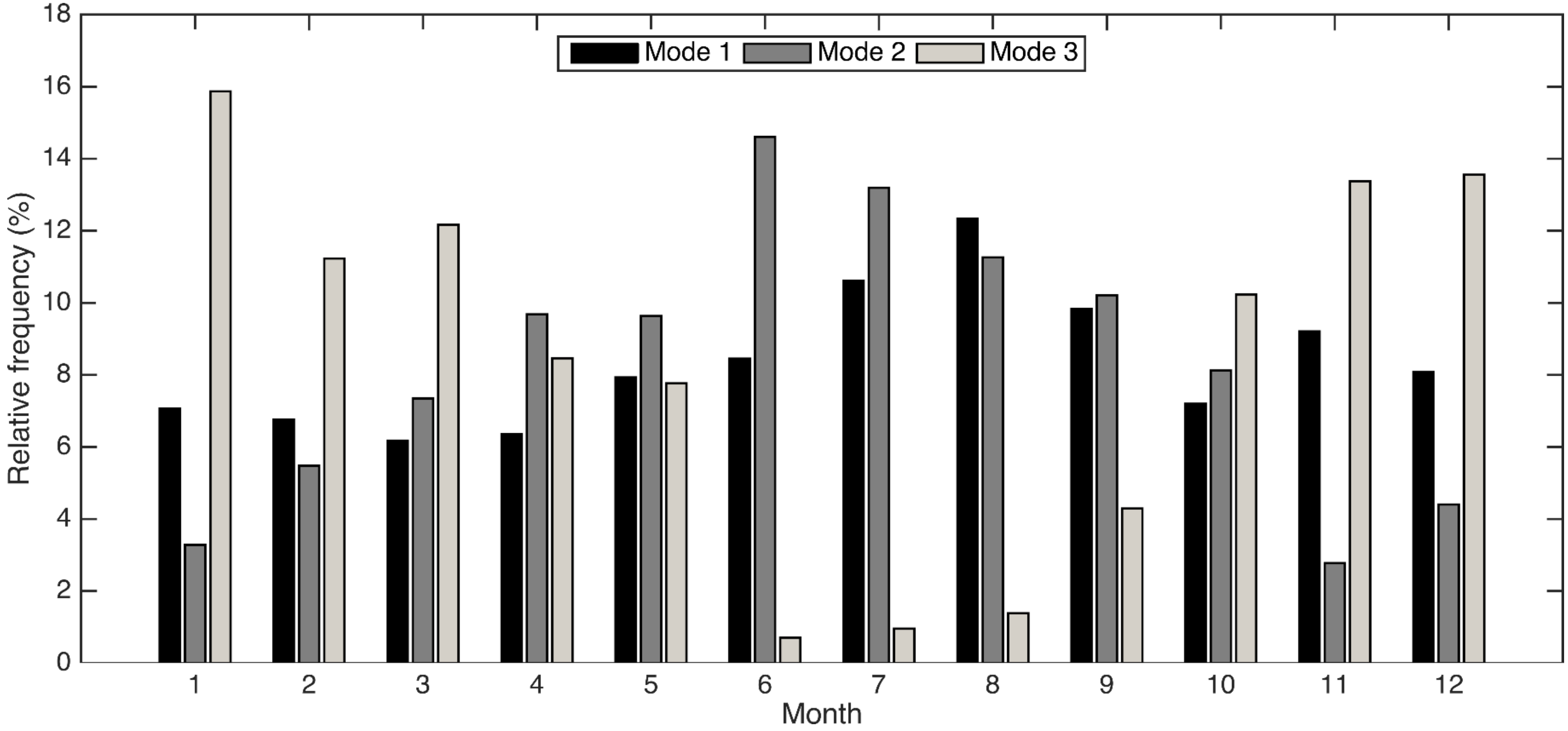
**Mode Centroids:**



# Atmospheric Classification

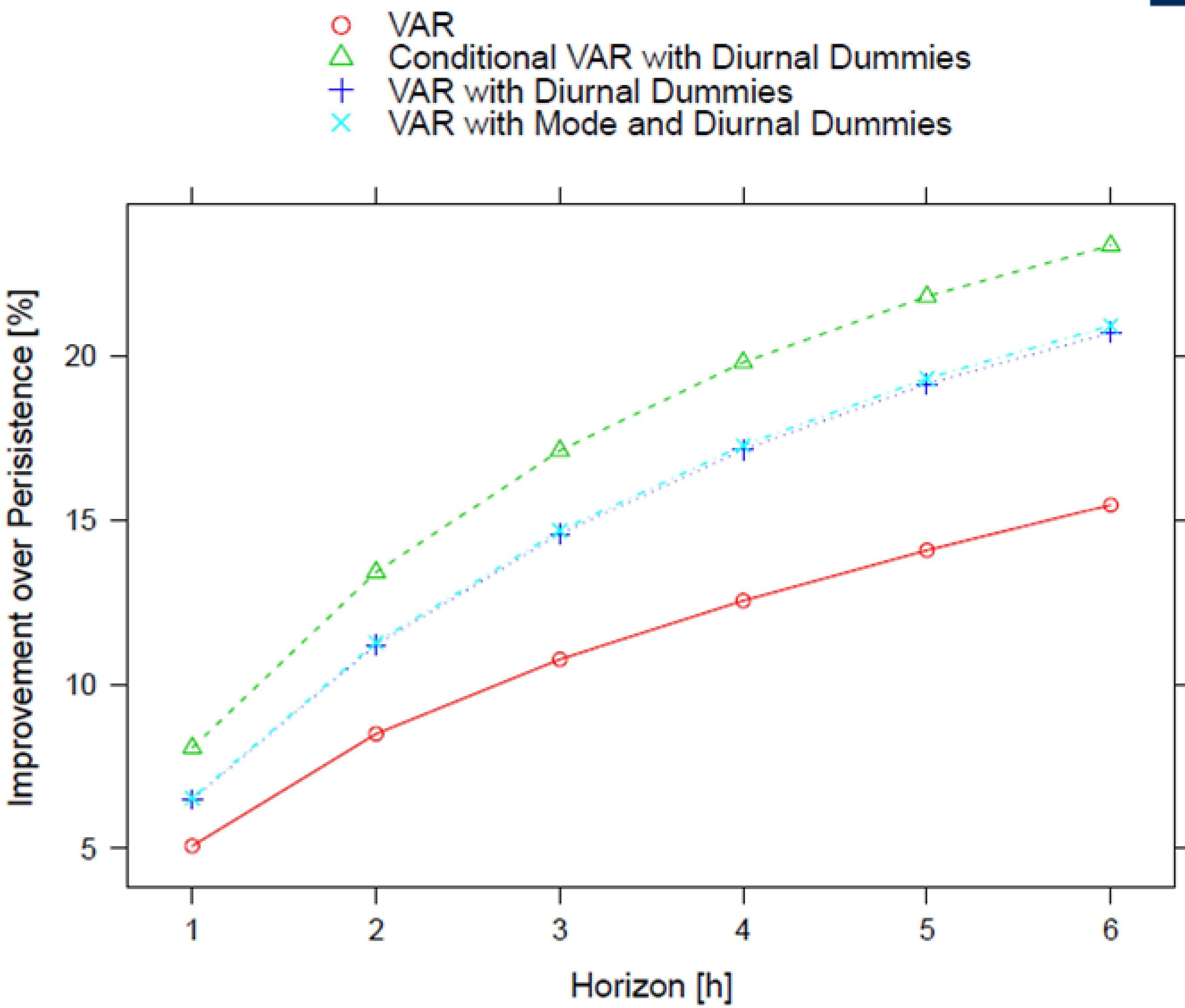
## Final Modes





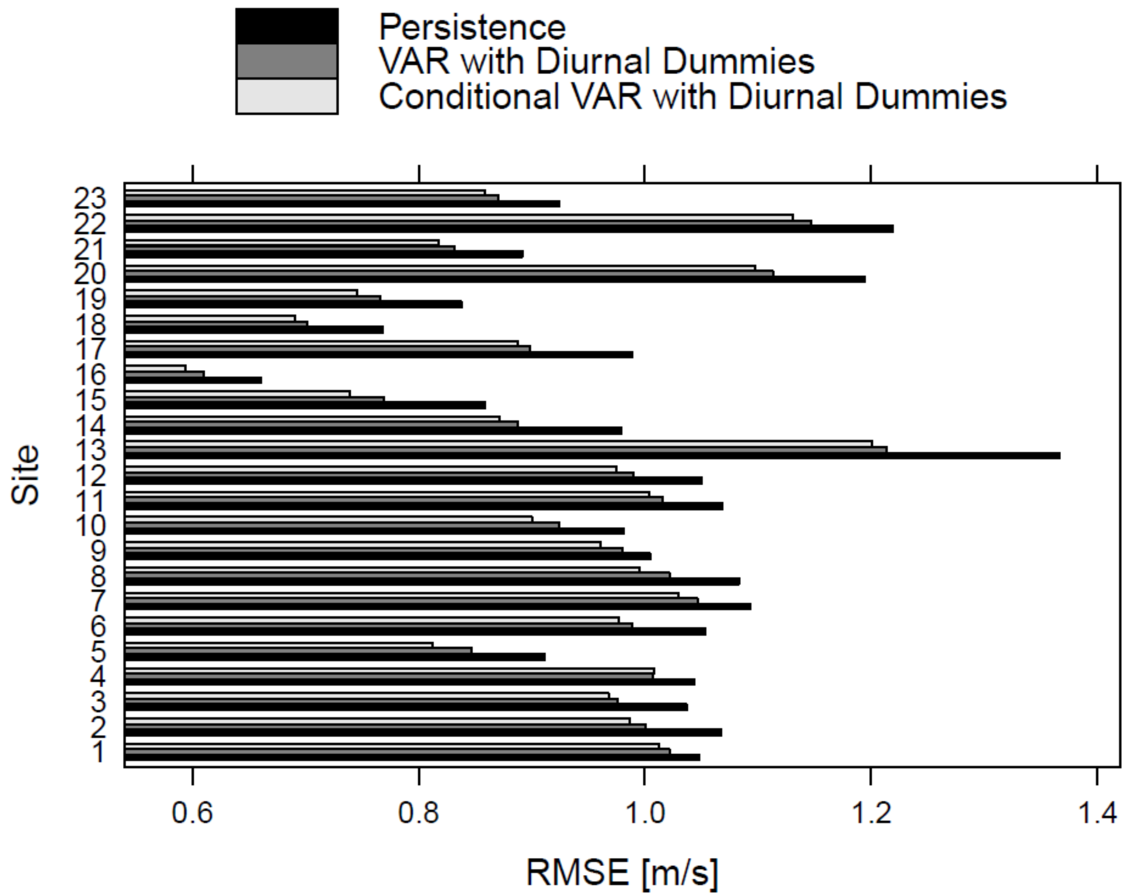
# Atmospheric Classification

## Forecast Performance

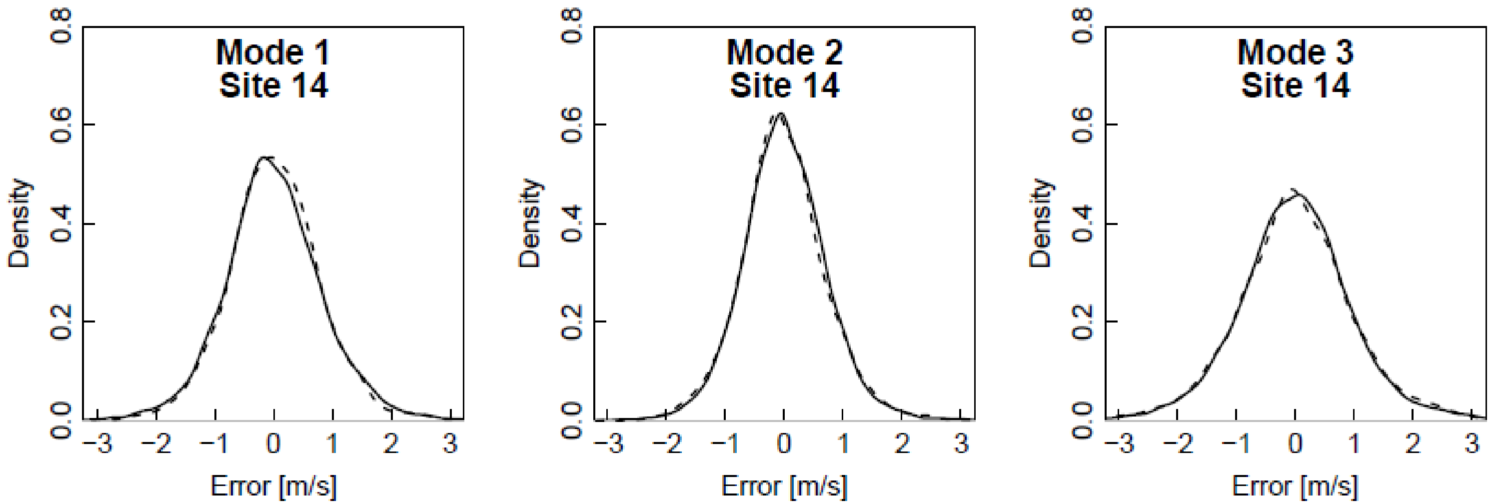


# Atmospheric Classification
## Forecast Performance

- Performance Improved at all 23 locations
- 1-hour-ahead forecast improved by 0.3%-4.1%
- Overall 1-hour-ahead forecast improved by 1.6%
- Overall 6-hour-ahead forecast improved by 3.1%



# Atmospheric Classification
## Forecast Performance




- Distinct error characteristics for each mode:
  - Provide decision-makers with quantified uncertainty information
  - Suggests similar regime-switching approach would be valuable for probabilistic forecasting

# References


- C. Gilbert, J. Browell, D. McMillan, “Leveraging Turbine-level data for improved wind power forecasting,” *submitted,* 2019.
- J. Dowell and P. Pinson, "Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression," in IEEE Transactions on Smart Grid, vol. 7, no. 2, pp. 763-770, March 2016. doi: 10.1109/TSG.2015.2424078
- Cavalcante, L., Bessa, R. J., Reis, M., and Browell, J. (2017) LASSO vector autoregression structures for very short-term wind power forecasting. Wind Energy, 20: 657–675. doi: 10.1002/we.2029
- Jakob W. Messner, Pierre Pinson, Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting, International Journal of Forecasting, 2018. doi: 10.1016/j.ijforecast.2018.02.001.
- Browell J, Drew DR, Philippopoulos K. Improved very short-term spatio-temporal wind forecasting using atmospheric regimes. Wind Energy. 2018; 1–12. doi: 10.1002/we.2207
- Daniel R. Drew, Janet F. Barlow, Phil J. Coker, Identifying and characterising large ramps in power output of offshore wind farms, Renewable Energy, Volume 127, 2018, Pages 195-203. doi: 10.1016/j.renene.2018.04.064
- Bessa, R.J.; Möhrlen, C.; Fundel, V.; Siefert, M.; Browell, J.; Haglund El Gaidi, S.; Hodge, B.-M.; Cali, U.; Kariniotakis, G. Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry. Energies 2017, 10, 1402. doi: 10.3390/en10091402
- J. Dobschinski, R. Bessak, P. Du, K. Geisler, S.E. Haupt, M. Lange, C. Möhrlen, D. Nakafuji and M. de la Torre Rodriguez, Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors, IEEE Power and Energy Magazine, vol. 15, no. 6, pp. 40-49, Nov.-Dec. 2017, doi: 10.1109/MPE.2017.2729100