Bayesian Non-parametric Modelling of Hourly Wind Speed and Direction

Abderrahmane El Aarfaoui

January 2026

Acronyms

BNP Bayesian Non-parametric. 4

DP Dirichlet Process. 5, 9

DPM Dirichlet Process Mixture. 3, 6, 8, 9

ESS Effective Sample Size. 6, 7

MCMC Markov Chain Monte Carlo. 3, 5, 7, 9

1 Introduction

Wind is a fundamentally bivariate phenomenon characterised by magnitude and direction. These characteristics cannot be modelled separately, and the latter cannot be treated as a circular covariate appended to an otherwise univariate model, as this separation can obscure important physical structures. In other words, wind regimes are more naturally expressed as joint patterns of speed and direction.

In this study, we apply a Dirichlet Process Mixture (DPM) model to hourly wind data from a fixed winter period in Edinburgh to identify the main wind regimes during this time. Bayesian non-parametric mixture models with a conjugate structure are computationally demanding due to high dimensional parameter spaces, long Markov Chain Monte Carlo (MCMC) chains, and identifiability issues. Therefore, limiting the temporal scope keeps inference tractable while preserving sufficient variability to identify latent wind regimes.

2 Data Description

The analysis is based on hourly wind observations recorded between 01-12-2024 00:00 and 02-28-2025 23:00 in Edinburgh corresponding to a continuous winter period. The final dataset comprises $N = 2{,}160$ complete hourly observations, where each observation includes wind speed at 10 meters above ground level (km/h) and wind direction (degrees from North).

3 Vector Representation

To enable coherent joint modelling, wind direction is transformed to radians and combined with speed to form Cartesian wind vectors

$$u = s\cos(\theta), \quad v = s\sin(\theta) \tag{1}$$

where s denotes wind speed and θ denotes wind direction. This transformation embeds circular directional information into a Euclidean $\mathbb{R}^2$ space, preserving physical interpretability where (u, v) represents the eastward and northward wind components, respectively. Specifically, $u > 0$ indicates eastward flow and $u < 0$ westward flow, while $v > 0$ signals northward flow and $v < 0$ indicates southward flow. Representing wind in Cartesian coordinates avoids discontinuities at the $0°/360°$ boundary inherent in raw angular measurements and enables standard multivariate methods (e.g., Gaussian mixtures) while respecting wind's bivariate structure. Therefore, wind speed is recovered as $s = \sqrt{u^2 + v^2}$ and direction as $\theta = \arctan 2(v, u)$.

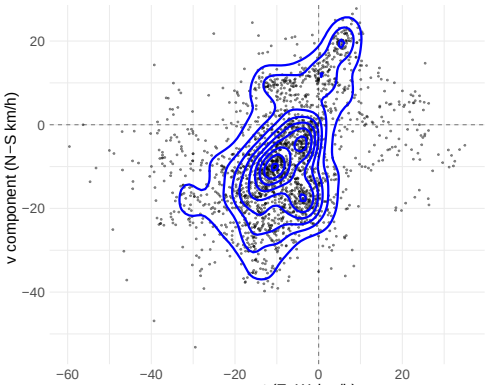


Figure 1: Wind vector scatter plot

4 Exploratory Data Analysis

The hourly wind speed over the observation period is shown in Figure 3. Over the winter period, the average wind speed was 18.9 km/h, as presented in Table 1 with a standard deviation of 9.78 km/h. This indicates that data exhibits high frequency variability at hourly scales, with speeds ranging from near calm to about 60 km/h. Speeds reflect occasional strong synoptic events, ranging from near calm conditions of .30 km/h to a maximum of 61.2 km/h. Furthermore, Cartesian wind components u and v have means of -8.39 and -9.28, respectively which signals a prevailing south-westerly flow, further confirmed in Figure 2. Their respective standard deviations further highlight directional variability and their moderate positive correlation suggests a dependence which motivates the joint modelling of the wind vector.

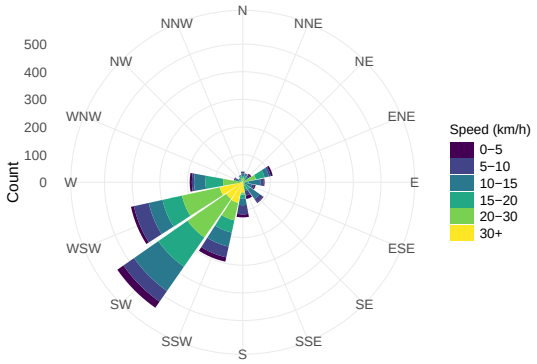


Figure 2: Wind rose - Edinburgh winter 2024-25

Moreover, Figure 3 exhibits no apparent temporal trend across the 3-month period and no extreme outliers or data quality issues with a peak around 60 km/h in mid-January. We notice some periods of sustained high (e.g., mid-January) and low (e.g., early-February) wind activity.

Statistic	Value
N observations	2160
Mean speed (km/h)	18.90
Sd speed (km/h)	9.78
Min speed (km/h)	0.30
Max speed (km/h)	61.20
Mean u (km/h)	-8.39
Mean v (km/h)	-8.28
Sd u (km/h)	11.96
Sd v (km/h)	13.07
Corr(u,v)	0.37

Table 1: Summary statistics

5 Stationarity Assessment

Before fitting the full Bayesian model, we assess the presence of unit roots using the Augmented Dicky-Fuller (ADF) test. As shown in Table 2, there is no evidence of non-stationarity over the winter period as all variables reject the null hypothesis of a unit root at the 5% significance level.

Variable	Statistic	p-value
Wind speed	-5.79	0.01
u	-5.66	0.01
v	-4.85	0.01

Table 2: ADF test for unit root

While short-term autocorrelation may still be present, the absence of a unit root supports modelling the observations as arising from a stationary mixture distribution over this limited time horizon.

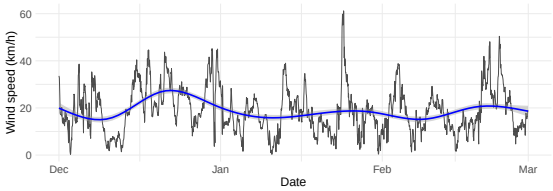


Figure 3: Wind speed time-series

6 Methodology

6.1 Dirichlet Process Mixture

Bayesian Non-parametric (BNP) methods relax the assumptions of traditional finite mixture models where an unknown finite parameter vector indexes a shape-restricted density function. The BNP approach defines random density functions without selecting a priori a specific density. For wind regime analysis, this flexibility is particularly advantageous for wind regime analysis, where the true number of patterns is unknown.

6.2 Bivariate Wind Vector Model

Let (u_t, v_t) denote the Cartesian wind vector at time $t = 1, \ldots, N$ where $N =$ is the total number of hourly observations. We model these vectors arising from an unknown mixture distribution over latent wind regimes.

Likelihood Each wind vector is drawn from a bivariate Gaussian Mixture

$$(u_t, v_t) \mid z_t, \{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k\}_{k=1}^{K} \sim \mathcal{N}_2(\boldsymbol{\mu}_{z_t}, \boldsymbol{\Lambda}_{z_t}^{-1})$$

where $z_t \in \{1, \ldots, K\}$ is the latent regime assignment for observation t, $\boldsymbol{\mu}_k = (\mu_{k,u}, \mu_{k,v})^\top \in \mathbb{R}^2$ is the mean vector for regime k, $\boldsymbol{\Lambda}_k \in \mathbb{R}^{2\times 2}$ is the precision matrix (inverse covariance) for regime k, and K is the truncation level for computational implementation.

Regime assignments Assignments follow a categorical distribution

$$z_t \mid \boldsymbol{w} \sim \text{Categorical}(\boldsymbol{w})$$

where $\boldsymbol{w} = w_1, \ldots, w_K$ are mixture weights satisfying $\sum_{k=1}^{K} w_k = 1$ and $w_k \geq 0$.

Modelling (u, v) jointly via a bivariate Gaussian captures the covariance structure between wind components, which encodes physically meaningful information

- Positive correlation: Wind regimes with consistent directional preference.
- Negative correlation: Variable directional patterns within a regime.
- Near-zero correlation: Isotropic wind patterns.

Thus, separate univariate models for u and v would assume independence, which would not capture these regime specific directional characteristics.

6.3 Dirichlet Process Prior

6.3.1 Stick-Breaking Construction

Rather than fixing K, we use a Dirichlet Process (DP) prior on the mixture weights, which allows the number of active regimes to be learned from the data and employ the stick breaking representation

$$
\begin{aligned}
v_k &\sim \text{Beta}(1, \eta), \quad 1, \ldots, K \\
w_1 &= v_1 \\
w_k &= v_k \prod_{j=1}^{k-1} (1 - v_j), \quad k = 2, \ldots, K
\end{aligned}
$$

where $v_k \in [0, 1]$ are auxiliary stick-breaking variables and $\eta > 0$ is the concentration parameter controlling the expected number of active regimes. Higher η favours more regimes and lower η favours fewer regimes.
The stick-breaking construction can be visualised as sequentially breaking a unit length stick, v_1 determines the length of the first piece with weight w_1, v_2 breaks the remaining stick to determine w_2, and so on. This ensures $\sum_{k=1}^{K} w_k = 1$ automatically.
The concentration parameter has an intuitive interpretation, where a small η signals strong prior belief in few distinct regimes, whereas a large one indicates prior belief that many regimes exist. A moderate value for η is a more balanced, non-informative stance. Therefore, we use $\eta \sim \Gamma(1, 1)$ to learn η from data rather than fixing it.

6.3.2 Truncation Approximation

The DP mixtures have countably infinite components ($K = \infty$). For computational implementation, we use a truncated approximation with $K = 10$ components. $K = 10$ in this case doesn't represent an a priori number of population clusters, but rather an upper-bound on the number of unknown population clusters. This is justified because under stick-breaking prior $\mathbb{E}[w_k]$ decreases exponentially with k, making weights beyond some K negligible. Moreover, posterior inference shows only $K^* \approx 3-4$ active regimes (weight $> 1\%$), well below the truncation level.

6.4 Prior Specification

6.4.1 Regime-Specific Parameters

For each regime $k = 1, \ldots, K$, we specify conjugate priors to enable efficient MCMC sampling

Mean Vectors

$$\boldsymbol{\mu}_k \sim \mathcal{N}_2(\boldsymbol{\mu}_0, \boldsymbol{R}_0^{-1})$$

where $\boldsymbol{\mu}_0 = (\bar{u}, \bar{v})^\top$ is set to the empirical mean of wind vectors, and $\boldsymbol{R}_0 = 0.01 \cdot \boldsymbol{I}_2$ is a vague prior precision (large prior variance). This weakly informative prior centres regimes near observed wind patterns while allowing important variation.

Precision Matrices

$$\boldsymbol{\Lambda}_k \sim \mathcal{W}_2(\boldsymbol{S}, \nu)$$

where $\mathcal{W}_2$ denotes the 2-dimensional Wishart distribution with scale matrix $\boldsymbol{S} = \frac{1}{2}\hat{\boldsymbol{\Sigma}}^{-1}$ where $\hat{\boldsymbol{\Sigma}}$ is the empirical covariance of (u, v), and Degrees of freedom: $\nu = 3$. The Wishart prior is conjugate to the multivariate normal likelihood, which enables efficient Gibbs sampling of precision matrices (Gelman et al., 1995).

Concentration parameter

$$\eta \sim \Gamma(1, 1)$$

This diffuse prior on η allows the data to inform the expected number of regimes. The $\Gamma(1, 1)$ prior has a mean 1 and variance 1, representing moderate prior uncertainty.

6.4.2 Prior Justification

Our prior choices balance three considerations. First, Normal-Wishart priors for $\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k$ enable conjugate Gibbs updates, which improves MCMC efficiency. Moreover, Vague priors on regime parameters allow data to dominate posterior inference. Finally, Centring priors near observed wind statistics prevents regimes in physically unrealistic regions of the parameter space.

7 MCMC Implementation

7.1 Posterior Inference via Gibbs Sampling

We perform Bayesian inference using MCMC implemented in the Nimble probabilistic programming framework.

7.2 Conjugate Sampling Structure

The model exhibits conjugacy at multiple levels, enabling efficient Gibbs Sampling.

Regime assignments z_t | rest Given current parameter values, the conditional posterior for z_t is

$$p(z_t = k \mid \boldsymbol{u}_t, \text{rest}) \propto w_k \cdot \mathcal{N}_2(\boldsymbol{u}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$

This is a categorical distribution, sampled efficiently via NIMBLE's built-in categorical sampler.

Mean vectors $\boldsymbol{\mu}_k$ | rest The normal prior on $\boldsymbol{\mu}_k$ combined with Normal likelihood yield a conjugate Normal posterior

$$\boldsymbol{\mu}_k \mid \{z_t\}, \{\boldsymbol{u}_t\}, \boldsymbol{\Lambda}_k \sim \mathcal{N}_2(\tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{R}}_k^{-1})$$

where $\tilde{\boldsymbol{R}}_k = \boldsymbol{R}_0 + n_k \boldsymbol{\Lambda}_k$, $\tilde{\boldsymbol{\mu}}_k = \tilde{\boldsymbol{R}}_k^{-1}(\boldsymbol{R}_0 \boldsymbol{\mu}_0 + n_k \boldsymbol{\Lambda}_k \bar{\boldsymbol{u}}_k)$, $n_k = \sum_{t=1}^{N} \mathbb{I}(z_t = k)$ is the number of observations in regime k, and $\bar{\boldsymbol{u}}_k = n_k^{-1} \sum_{t:z_t=k} \boldsymbol{u}_t$ is the empirical mean of regime k. NIMBLE recognises this conjugacy and uses a specialised conjugate *dmnorm_dmnorm* sampler.

Precision matrices $\boldsymbol{\Lambda}_k$ | rest The Wishart prior on $\boldsymbol{\Lambda}_k$ combined with Normal likelihood yields a conjugate Wishart posterior

$$\boldsymbol{\Lambda}_k \mid \{z_t\}, \{\boldsymbol{u}_t\}, \boldsymbol{\mu}_k \sim \mathcal{W}_2(\tilde{\boldsymbol{S}}_k, \tilde{\nu}_k)$$

where $\tilde{\nu}_k = \nu + n_k$, and $\tilde{\boldsymbol{S}}_k = \left[\boldsymbol{S}^{-1} + \sum_{t:z_t=k} (\boldsymbol{u}_t - \boldsymbol{\mu}_k)(\boldsymbol{u}_t - \boldsymbol{\mu}_k)^\top\right]^{-1}$. NIMBLE uses a specialised conjugate *dwish_dmnorm* sampler for this update.

Stick-breaking weights v_k Due to conjugacy between Beta priors and discrete allocation patterns, v_k can be updated via Metropolis-Hastings with efficient proposals. NIMBLE assigns random walk samplers to v_k.

Concentration parameter η Updated via random walk Metropolis due to lack of conjugacy with Beta distributions. This is a single scalar parameter, so mixing is generally good despite non-conjugacy.

8 Model Fitting

The following code defines the Bayesian model in NIMBLE, and implements the DPM via stick breaking.

We specify weakly informative priors centred on observed wind patterns.

Chains are initialised using K-means clustering on (u, v) to provide reasonable starting values.

The MCMC configuration assigned the following samplers

- Random walk samplers (7): eta, v[1 : 6].
- Conjugate samplers (12): mu[1 : 6] (6 multivariate), prec[1 : 6] (6 multivariate).
- Categorical samplers (2160): z[1 : 2160] (one per observation)

This demonstrates the conjugate structure described in Section 7.2, with 12 of 19 parameters blocks using efficient conjugate Gibbs updates.

9 Running MCMC

We run 3 independent chains with 50,000 iteration per chain, of which 10,000 are discarded as burn-in, and thin by 10. Hence we retain 4,000 iterations per chain.

10 Computational Performance

- Hardware: MacBook Air M4, 24GB RAM.
- Configuration time: 16.56 minutes.
- MCMC runtime: 30.43 minutes.

11 Results

11.1 Convergence diagnostics

The Gelman-Rubin statistics indicate an acceptable convergence across chains with a maximum $\hat{R}$ value of 1.12. Only 2 of 49 parameters exceeded the 1.1 threshold, which suggests that the three chains have converged. The Effective Sample Sizes (ESSs) confirm low autocorrelation for most parameters. The minimum ESS of 340 is adequate for inference, while the median ESS of 4,304 indicates that most parameters have excellent mixing properties. Importantly, no parameter has critically low ESS values below 100.

Metric	Value
Total Monitored parameters	49
Maximum $\hat{R}$	1.12
Parameters with $\hat{R} > 1.1$	2 / 49
Minimum ESS	340
Median ESS	4,304
Mean ESS	4,406
Parameters with ESS ¡ 100	0 / 49

Table 3: MCMC convergence diagnostics summary

These diagnostics show that the 12,000 retained samples (after burn-in and thinning) provide a solid basis for posterior inference.

12 Label Switching

Bayesian mixture model suffer from label switching, where the posterior distribution is invariant to permutations of component labels because likelihood terms can be arbitrarily reordered. This creates a multimodal posterior with $K!$ identical modes (for K =6, 720 equivalent labellings). As a result, this causes meaningless marginal posteriors, where averaging across label permutations produces uninformative summaries. Furthermore, parameter estimates are uninterpretable. For instance, regime 1 in iteration 100 may become regime 3 in iteration 200.

12.1 Pivotal relabelling algorithm

We implement pivotal relabelling to post-process the MCMC output

- Choose a pivot iteration (we use the median parameter values across all iterations).
- For each MCMC iteration, find the optimal permutation minimising distance to the pivot.
- Relabel that iteration's parameters accordingly.

The pivotal relabelling algorithm reported that 0 of 12,000 iterations required relabelling, with only the identity permutation (1-2-3-4-5-6) observed. The normalised permutation entropy is 0, indicating complete label stability across the chains. Furthermore, the distance matrices shows that each component is substantially closer to its own pivot than to any alternative, which confirms strong separation between regimes. Hence, the posterior is effectively identifiable in practice, which is likely due to the combination of K-means initialisation and well separated mixture components.

12.2 Interpretation of Identified Wind regimes

Table 5 presents the posterior mean characteristics of the six identified wind regimes, ordered by mixture weight. The model identified three major regimes and three minor ones.
Regime 3 shows moderate south-westerly winds ($\sim$226°) with a mean speed of 14.7 km/h and constitutes the dominant winter regime in Edinburgh, which reflects typical Atlantic flow as depicted in Figure 9. This aligns with the wind features in Eastern Scotland, as most winds come from directions between south and north-west, and the strongest winds almost always originate from this range of directions (Met Office, n.d.). The positive correlation ($\rho = 0.40$) indicates a consistent directional preference. While regime 1 also consists of south-westerly winds ($\sim$222°), it exhibits a stronger mean speed of of 21.65 km/h, corresponding to more severe Atlantic depressions, its large standard deviation and near-zero correlation suggest high wind speed variability. Regime 4 is characterised by light north-easterly winds ($\sim$44°) with an average speed of 8.22 km/h, and the negative correlation ($\rho = -0.55$) indicates greater directional variability which could be linked to continental anticyclonic conditions. Regime 6 captures a relatively pure westerly flow comprised of moderate westerly winds ($\sim$ 262°) averaging 18.16 km/h and is tightly concentrated with a low σ_μ. Regime 2, which is potentially linked to cold air outbreaks, features moderate easterly winds ($\sim$ 77°) with a mean speed of 18.15 km/h and a high correlation ($\rho = 0.78$) that reflects a very consistent directional preference. Finally, regime 5 describes light south-westerly winds ($\sim$ 235°) with an average speed of 4.7 km/h, and its very small standard deviations suggest near-calm conditions or transitional periods.

13 Posterior Predictive Checks

To validate the model, we generate posterior predictive samples and compare them with observed data across multiple dimensions.

13.1 Wind Speed Distribution

Figure 4 compares the observed and predicted wind speed distribution. The model captures the overall distribution shape well, with a relative prediction error of 7.5% in the mean (ob-

served: 18.90 km/h, predicted: 20.31 km/h). Moreover, the well matched standard deviations (observed: 9.78 km/h, predicted: 9.53 km/h) indicate that the mixture model successfully captures wind speed variability. The slight overestimation of mean speed (7.5% error) is reasonable given the complexity of the wind distribution and suggests the model provides conservative estimates.

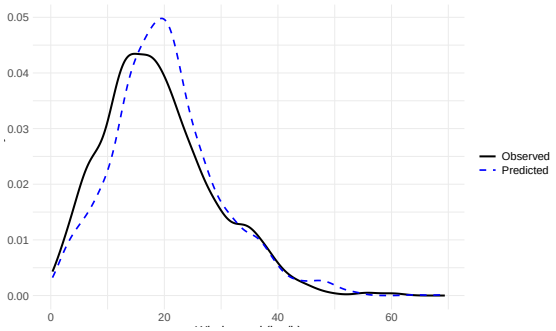


Figure 4: Posterior predictive wind speed

13.2 Wind Direction Distribution

Figure 5 shows the posterior predictive check for wind direction. The predicted distribution closely matches the observed distribution.

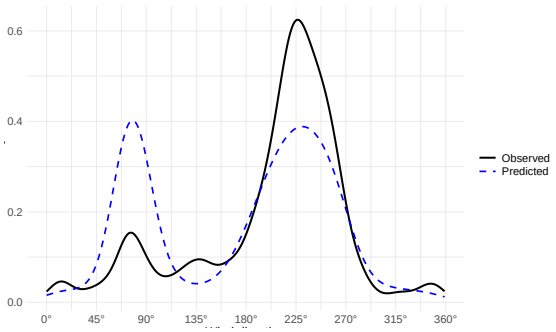


Figure 5: Posterior predictive wind direction

13.3 Joint Wind Vector Structure

Figure 6 displays the joint structure of wind vectors in (u, v) space, with observed data shown as black points, predicted data as blue points, and coloured arrows representing the regime centres. The arrows point in the direction of each regime's mean wind vector, with length proportional to wind speed, and colour indicating direction. The Figure confirms that the model identifies distinct clusters in vector space corresponding to different wind regimes. Furthermore, the regime centres (arrows) align with high-density regions of observed data. We also notice that the predicted samples overlap relatively well with the observed data.

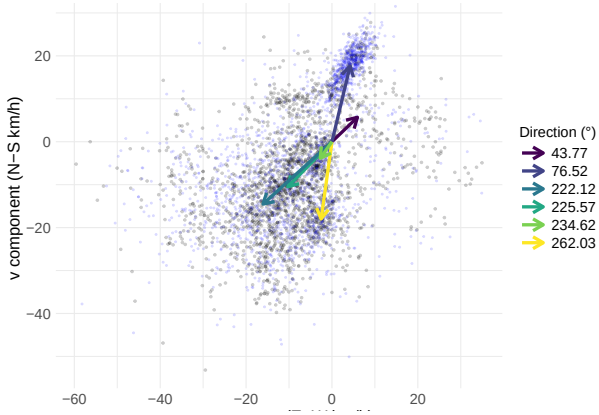


Figure 6: Joint wind vector structure

13.4 Wind Rose

Figure 7 presents a posterior predictive wind rose of the joint distribution of wind direction (angular sectors) and speed (colour bands). The plot indicates that the dominant wind direction is from the south-west (180°-240° sector), consistent with Edinburgh's exposure to Atlantic weather systems. Higher wind speed of 20-30+ km/h occur most frequently in the south-west to west sectors, whereas speeds of up to 10 km/h are observed across all directions but are comparatively less common. Overall, the directional structure appears to be well captured by the mixture model, with appropriate representation of both the primary south-westerly flow and secondary wind directions.

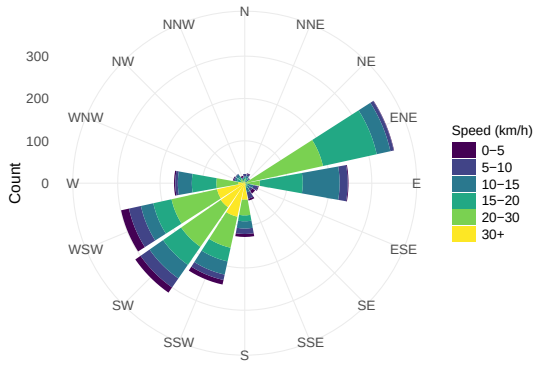


Figure 7: Posterior predictive wind rose

13.5 Model Fit Summary

The posterior predictive checks show that the DPM model provides an excellent fit to the observed wind data

Quantity	Observed	Predicted	Error
Mean speed (km/h)	18.90	20.31	7.5%
SD speed (km/h)	9.78	9.53	2.6%
Mean u (km/h)	-8.39	-6.53	22.2%
Sd u (km/h)	11.96	13.23	10.6%
Mean v (km/h)	-8.28	-2.94	64.5%
Sd v (km/h)	13.07	16.64	27.3%
Cor(u, v)	.37	.52	

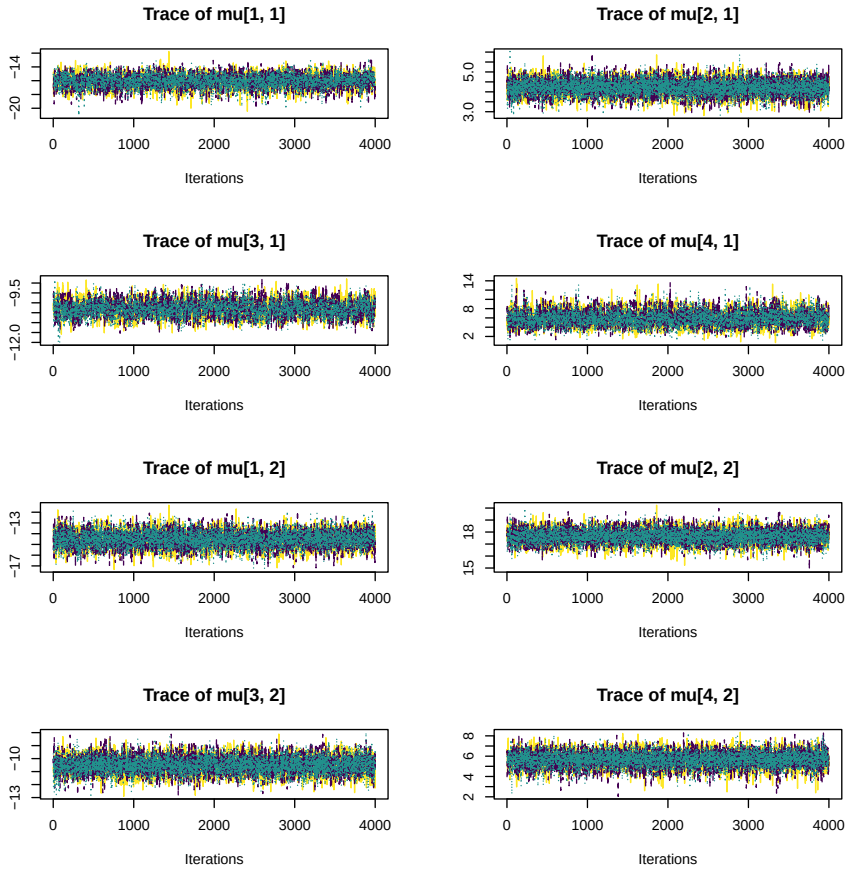
Table 4: Observed vs. predicted summary statistics

The larger errors in the wind vector components u and v compared to speed are expected because vector components are more sensitive to directional biases than speed magnitude, and small systematic biases in direction amplify when projected onto Cartesian axes. Additionally, the mixture model balances fit across all regimes rather than optimising marginal component means. However, despite these componentwise discrepancies, the model successfully captures the essential features of the wind climate, i.e. dominant directional modes, the range of wind speeds, and the multimodal structure of the joint distribution.

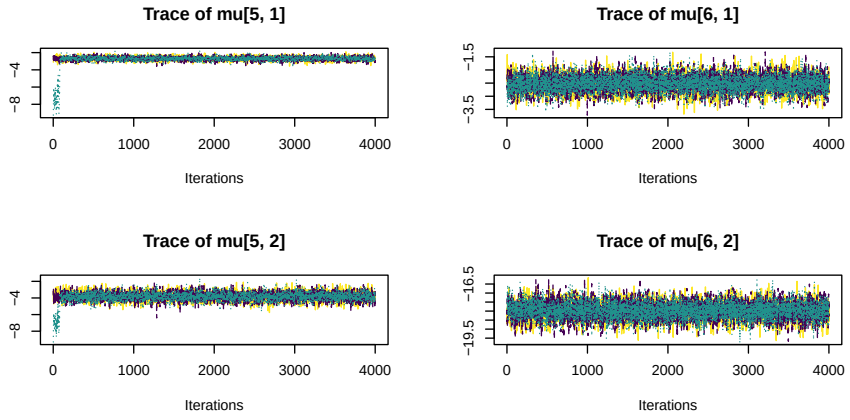
14 Conclusion, Limitations and Future Work

This work demonstrates that Bayesian nonparametric DPM models provide a principled and effective framework for identifying latent wind regimes from bivariate wind vector data. Applied to hourly winter observations in Edinburgh, the method successfully four major wind regimes representing distinct meteorological patterns, with the dominant regime corresponding to moderate south-westerly Atlantic flow. The conjugate structure of the model enabled efficient MCMC sampling with satisfactory convergence properties, while the DP prior allowed automatic determination of regime complexity. Posterior predictive checks confirmed that the model captures features of speed distribution, directional patterns, and multimodal structure. However, several limitations warrant consideration. The current model treats hourly observations as independent, ignoring autocorrelation and persistence in wind regimes. A Hidden Markov Model extension would allow regime transitions to follow a Markov process, capturing temporal dynamics. Furthermore, the analysis was limited to winter months for computational reasons. Extending to a full year would require either seasonal covariates or separate analyses for each season. While $K = 6$ proved adequate, sensitivity analysis on the truncation level would strengthen conclusions. Moreover, incorporating data from multiple locations could regional wind patterns and spatial coherence of regimes. Finally, while bivariate Gaussian components worked well, von Mises or wrapped Gaussian distributions might better capture directional characteristics of certain regimes.

A List of Figures



(a) Trace plot of μ_1 to μ_4



(b) Trace plot of μ_5 to μ_6

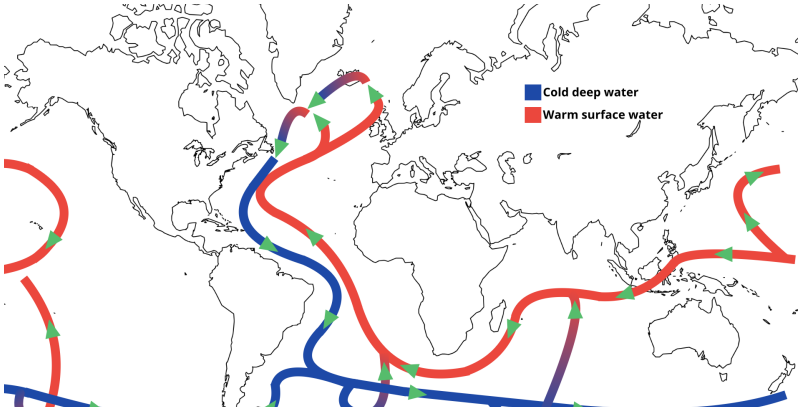


Figure 9: Ocean circulations

B List of Tables

Regime	Weight	μ_u	μ_v	σ_u	σ_v	ρ	Speed	Dir (°)
3	37.48%	-10.29	-10.50	5.12	11.12	0.40	14.70	225.57
1	25.81%	-16.06	-14.52	14.07	9.79	-0.08	21.65	222.12
4	10.09%	5.94	5.69	12.65	7.28	-0.55	8.22	43.77
6	5.44%	-2.52	-17.98	2.17	3.26	0.33	18.16	262.03
2	5.26%	4.23	17.65	3.08	4.33	0.78	18.15	76.52
5	4.62%	-2.72	-3.83	1.52	3.08	0.56	4.70	234.62

Table 5: Posterior mean characteristics of identified wind regimes