

Original Research

Automated Probabilistic Analysis and Parametric Modelling of the Seasonal-Diurnal Wind Vector

Nicholas J Cook *

Independent researcher, 20 Beacon Drive, Christchurch, Dorset, UK; E-Mail: wind@njcook.uk* **Correspondence:** Nicholas J Cook; E-Mail: wind@njcook.uk**Academic Editor:** Andrés Elías Feijóo Lorenzo**Special Issue:** [Progress of Wind Energy Technology and Its Maintenance](#)*Journal of Energy and Power Technology*
2021, volume 3, issue 2
doi:10.21926/jept.2102027**Received:** April 14, 2021
Accepted: May 20, 2021
Published: June 25, 2021**Abstract**

A refined and extended version of the Offset Elliptical Normal mixture model has been developed to parameterise the seasonal diurnal wind vector automatically. Automated using R scripts, the method eliminates any potential risk of confirmation bias posed by the manual supervision in the original method. Refinements to the method include the latest algorithms for clustering of Gaussian mixtures, with Bayesian regularisation to set the number of components and to limit the predisposition to overfit. A new extension uses fuzzy logic to evaluate the probability distributions, autocovariances and spectra of the random perturbations around the mean seasonal-diurnal variations for each component of the mixture. These additional parameters allow the predictions of the OEN model to be validated and its automated application demonstrated using the hourly METAR reports of mean wind speeds at Adelaide, South Australia, showing significant improvements over the previously published analysis. The OEN mixture model is directly applicable to a wide range of wind engineering applications where seasonal and diurnal variation is of importance.

Keywords

Wind vector; wind speed; wind direction; bivariate Gaussian mixtures; wind power resource; OEN model



© 2021 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

1. Introduction

The seasonal and diurnal variations of the wind speed vector is an important consideration in many engineering applications, including the safety and comfort of pedestrians around tall buildings; heating, ventilation, and air-conditioning (HVAC) loads; structural design to resist windstorms; and the availability of wind power generation. “Now-casting” methods are used when the application requires prediction of the wind vector some minutes into the future, eg. for the safety of crane operations or the operational control of wind turbines. Daily forecasts are sufficient for controlling the safety of transient operations, including temporary structures such as grandstands for events. However, the planning and design of engineering projects is almost universally done on the basis of probability of occurrence, as in assessing wind energy potential, or risk of exceedance, as in codes of practice for structural design.

Assessment of the joint probability density function (jPDF) of the wind vector can be done in several ways:

1. As a “raw” or “observed” jPDF by binning occurrences of wind speed and direction;
2. Fitting the observations to a non-parametric model, eg. using kernel density estimation (KDE) methods, which is essentially a smoothing exercise.
3. Fitting the marginal density functions of wind speed and direction observations, separately, to parametric distributions, accepting the resulting loss of information on their joint action;
4. Fitting the observations to a parametric jPDF model, preferably one which is justified by meteorological theory.

The OEN mixture model falls into the fourth category.

It is founded on the physical theory proposed by Brooks and Carruthers [1] and developed later by Crutcher et al [2-5], which predicts each wind generating mechanism produces a bivariate Gaussian jPDF that forms elliptical contours on zonal (W) and meridional (S) axes. The wind climate in any location around the World is recognised to be “mixed”, this is: produced by a mixture of mechanisms, each acting exclusively, that result in a jPDF which is the disjoint sum of the individual component jPDFs.

The OEN mixture model was introduced in 2014 [6] as a novel way of demonstrating why simple meteorological and statistical theory leads to a bivariate Gaussian model for the PDF of the surface wind vector and, from that, the ubiquitous Weibull model for the marginal cumulative distribution function (CDF) of speed irrespective of direction. It was also recognised that OEN had a useful role in separating and parameterising the individual wind components of the mixture that is produced by various meteorological wind-producing mechanisms acting disjointly. Notably, it avoids all the circularity issues that plague other approaches operating on speed and direction e.g. [7], is based on physical meteorological and statistical theory and is globally applicable [8].

The ability of OEN to parameterise seasonal-diurnal variation was first demonstrated in 2015 on the well-studied, mixed-mechanism wind climate of Adelaide [9] where it was demonstrated that each disjoint model component corresponded to the known physical mechanisms. It was then applied to the wind climate of Rome [10], using observations from the Fiumicino and Ciampino airports, and again showed correspondence between model components and mechanisms, and with the differences between sites matching the differences in topography. The method has since

been refined and its application has been extended with additional analyses that were previously thought impractical [11], but which now allow the underlying predictions of the model to be validated. Implementation has been moved from Excel™ into the R statistical computing language to take advantage of R's superior analysis and graphics facilities and to automate the fitting process, as far as is possible, eliminating any risk of confirmation bias posed by manually supervised fitting. Also, in the interim period, an additional six years of METAR reports at hourly intervals have been accrued at Adelaide, so the time is ripe to revisit the previous analysis of the Adelaide wind climate to validate this automated method and to extend the original findings. The principal purpose of this paper is to make the R analysis scripts freely available, allowing the OEN mixture model to be fitted with minimal manual intervention to any other suitable dataset, to parameterise the jPDF of the wind vector at any season and time of day.

2. Materials and Methods

2.1 Meteorological Data

The ability of the OEN mixture model to represent the known physics of the Adelaide mixed-mechanism wind climate is demonstrated by applying it to the hourly METAR reports from Adelaide International Airport from January 1991 to October 2020, extracted from the TD3505 database on the US National Centers for Environmental Information (NCEI) server. The earlier [9] data at 6- and 3-hour intervals, with its reported uncertainty between UTC and local times [12], have been discarded.

Figure 1(a) shows the seasonal-diurnal variation of the marginal probability density functions (PDF) of wind speed, $p(V)$, and direction, $p(\theta)$, obtained by binning the observations, in a format [13] that presents the whole annual cycle in a single chart. Here, the horizontal axis represents the cycle of each hour through the day in twelve panels that represent the cycle of each month during the year. The vertical axes represent the marginal PDF of wind speed (top) and wind direction (bottom) averaged over every day in each month of the 30-year record. This is equivalent to Figure 2 in [9] but now is clearer and reveals more detail in the strong and coherent pattern. The diurnal variation of speed is strongest in the austral summer afternoons. The direction shows a diagonal pattern that indicates an overall net anti-clockwise rotation consistent with Coriolis effects in the southern hemisphere but proceeds in jumps between three preferred directions: $\theta = 230^\circ$, 150° and 20° . The corresponding chart synthesised from the final OEN model is shown as (b) to enable later direct comparison.

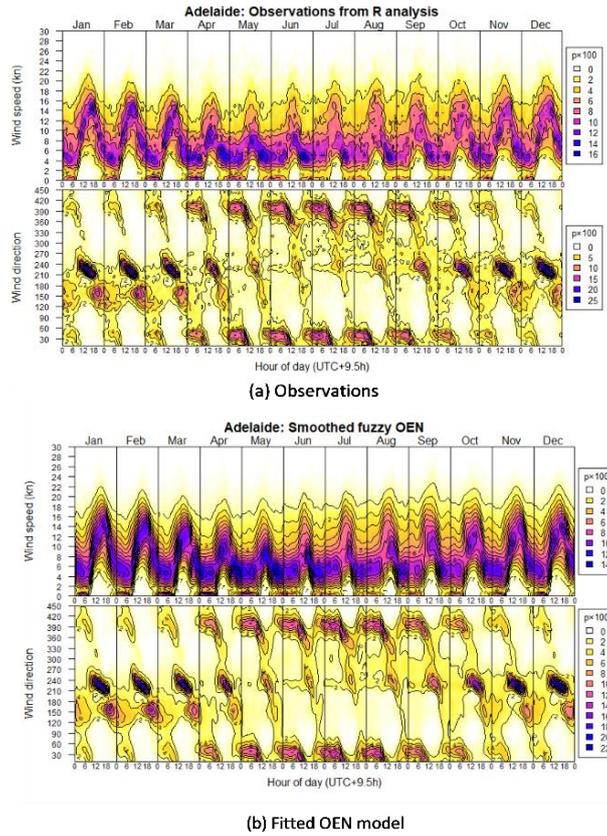


Figure 1 Seasonal-diurnal variation of wind speed and direction marginal PDFs.

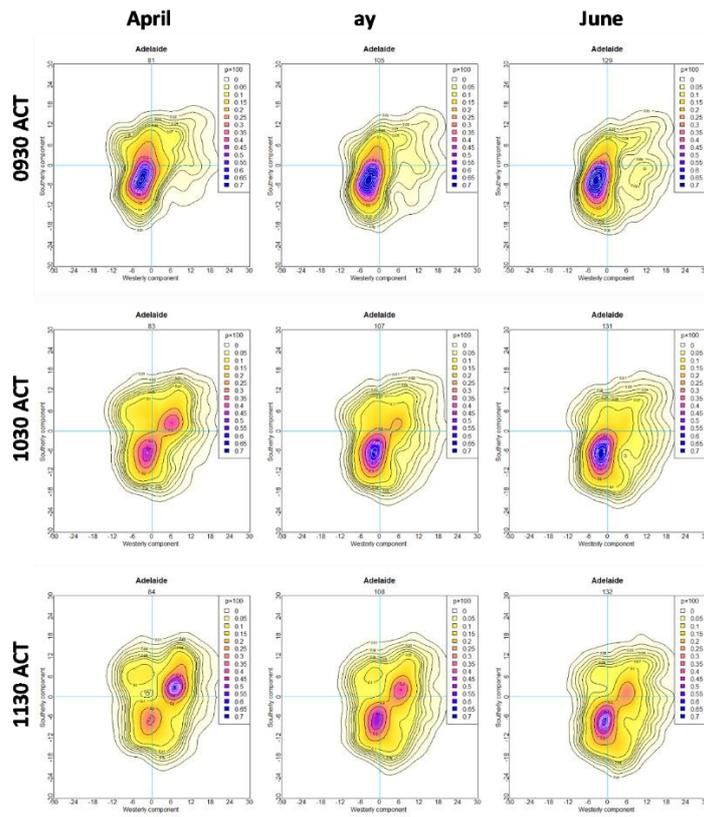


Figure 2 Land-breeze to sea-breeze transition in $p(W,S)$: April to June, 0930 to 1130 ACT.

Figure 3(a) shows the whole-year joint PDF of the observations, $p(W,S)$, excluding calms and plotted as westerly, W , and southerly, S , components on the zonal-meridional plane. This is equivalent to Figure 4(b) in [9] but is again clearer and more detailed. The ripple in the outer contours is the residual effect of the 10° increments of wind direction after the transformation from polar to cartesian coordinates. Again, (b) and (c) are results, included here for later direct comparison and discussion.

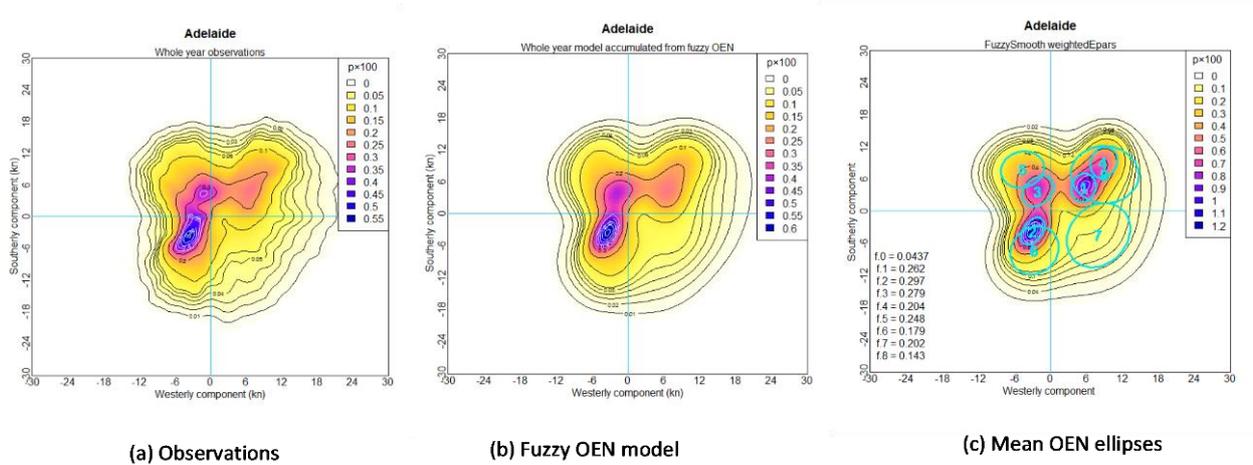


Figure 3 The whole-year joint PDF, $p(W,S)$, in the zonal-meridional plane.

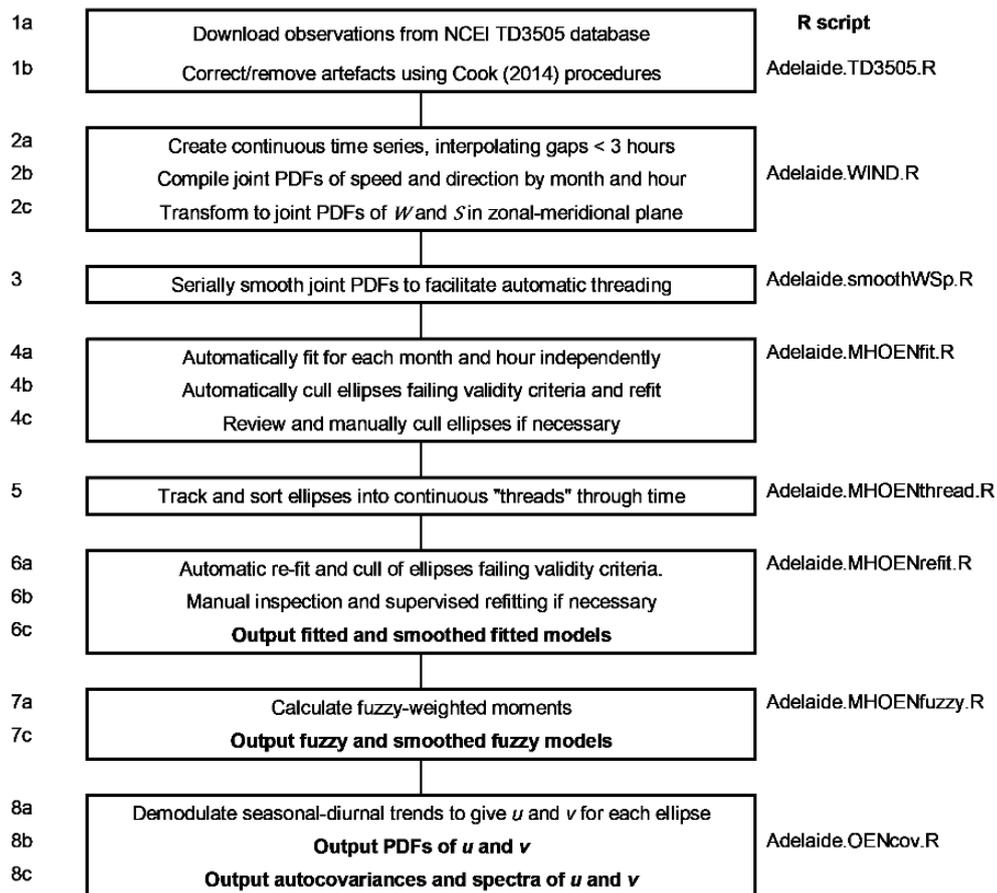


Figure 4 OEN implementation steps.

Most of the data in this study are four-dimensional, depending on the wind vector components, W and S , in the zonal-meridional plane (or their meteorological counterparts, V and Θ), the local time of day, H , in Australian Central Time (ACT = UTC+0930) and the month, M . The most dimensions that can be represented on any two-dimensional chart is three, i.e., as contours or colour-maps, hence compromises must be made in their presentation which result in some loss of information. In Figure 1 the compromise is to show the marginal distributions of V and Θ separately, so omitting their joint characteristics. Figure 3 shows whole-year data in which the seasonal-diurnal variations are averaged out. Revealing these variations in $p(W,S)$ in hourly data requires 288 separate charts which are best laid out in the form of an H by M matrix, as provided by the supplementary online Figure S1. Figure 2 is an extract which shows that the establishment of the sea breeze between 0930 ACT and 1130 ACT from April to June (austral autumn) occurs slightly later in each successive month.

2.2 Basis of the OEN Method

The original OEN mixture model is fully described in [9] and the later refinements and extensions in [10, 11] and its application to whole-year data in [8]. Here, for the convenience of the reader, the representation of the model on the zonal-meridional plane is shown in Figure 5 and the defining equations are reproduced in Appendix A. The model joint PDF of each component of the mixture forms elliptical contours around the mean vector, so it is convenient in Figure 5 to represent each component of the mixture by its one standard deviation contour and refer to this as the “ellipse”.

The theory underpinning the OEN mixture model predicts that:

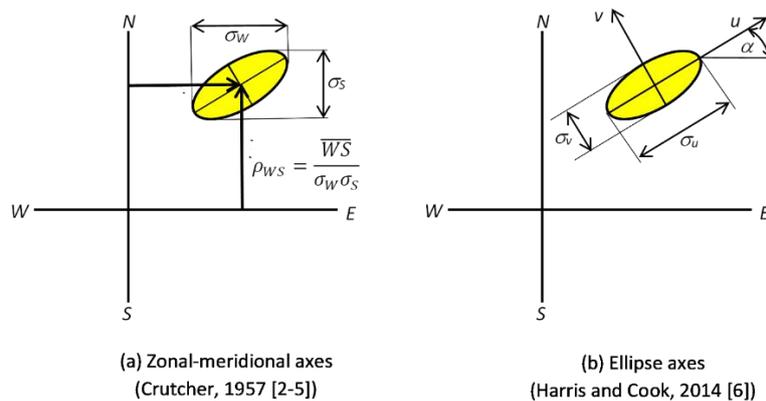


Figure 5 Definitions of the Offset Elliptical Normal model parameters. Note that all wind speeds values in this paper are expressed in the original observed units of knots (1kn = 0.514ms⁻¹) to avoid unit conversion bias when binning to probability density function (PDF) estimates.

1. The ellipses of the mixed climate are exclusive (disjoint), but not necessarily independent. This means that only one mechanism occurs at a given time (exclusive) so that the individual jPDFs are additive in proportion to their frequency (disjoint) in accordance with Eqn. A.1. However, they may not be independent in the sense that they interact, for example: a land breeze generally follows a sea breeze; a sea-breeze cell is strengthened by an overlying synoptic off-shore breeze; and land/sea breezes are inhibited by synoptic winds in excess of ~15 knots.

2. All the non-stationary seasonal-diurnal variation is expressed by the parameters of Equations A.3 and A.4. Hence the location and scale parameters, representing the ellipse centres, sizes, and orientations, vary deterministically with season and time of day.

3. That, by rotating the axes such that $\rho_{uv} = 0$ in Figure 5(b), all the random variation is expressed by stationary, orthogonal, Normally distributed random perturbations u/σ_u and v/σ_v with zero mean and unity standard deviations.

The extension of the fuzzy model in [11] to derive u/σ_u and v/σ_v and to estimate their PDFs, autocovariances and spectra now allows these predictions to be tested.

2.3 Implementation of the OEN Method in R

Implementation of the refined and extended OEN analysis method is broken down into eight R scripts, as indicated in Figure 4:

- Steps 1 and 2 to extract the observations, correct/remove artefacts [14], then form the joint PDFs $p(W,S)$ for each M and H , operate identically to the previous implementation [9].
- Step 3 is new and applies a (0.25,0.5,0.25) smoothing filter between the joint PDFs $p(W,S)$ for preceding, current, and succeeding hours, H , and months, M , to reduce ephemeral peaks in the data field and facilitate the automated “threading” process: the pairing of ellipses between successive observation times, H , in step 5. This smoothing has the unwanted effect of reducing the gradients and peak values in the data field which is corrected later by the fuzzy model.
- Step 4 generates the first OEN fit without any manual intervention. It takes advantage of the latest algorithms in the package “mclust” [15] for clustering of Gaussian mixtures [16], including Bayesian regularisation to set the number of ellipses and limit any predisposition to overfit the parameters, as well as R’s standard multiparameter optimisation functions that are equivalent to the Excel™ Solver function used previously. Nominally, the result has the optimal number of ellipses to adequately describe $p(W,S)$ for each M and H independently of the others. In practice, this Bayesian regularisation step is often too aggressive, merging ellipses that are clearly separate in the adjacent H , resulting in a loss of serial continuity between adjacent hours and in values of the coefficient of determination, R^2 , that are unacceptably low. Therefore, whenever R^2 is below the target minimum value of 0.99, additional ellipses are added and re-fitted until the target is reached. Figure 6 compares the R^2 values for the first fit (black circles) with the improved final fit (red crosses). Those few instances where the first fit appears to be better are over-parameterisations, where an extra ellipse appears which is not found in the preceding or succeeding observations and has been removed.

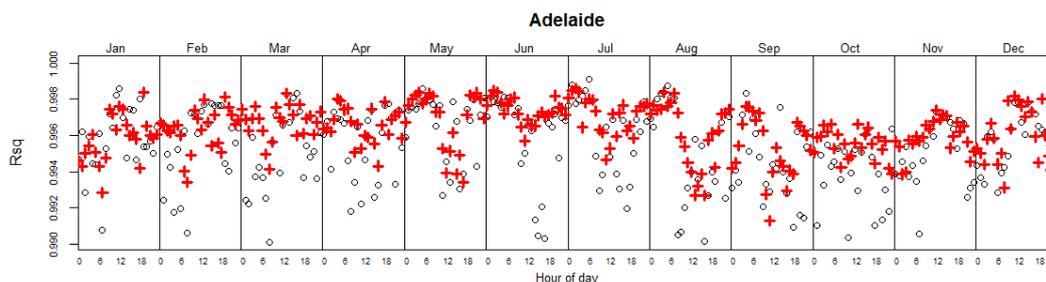


Figure 6 Coefficient of determination, R^2 , for first fit (black circles) and final fit (red crosses).

- In step 5, a simple algorithm replaces the previous visual “threading” process. Each ellipse is paired with the closest-matching ellipse of the previous H and assigned the corresponding index. Ellipses with no counterpart in the previous H are assigned a new ellipse index. Previously found ellipses with no counterpart in the current H are “remembered” so that they can be matched if they reappear later. When two similar ellipses move past each other, the algorithm sometimes erroneously swaps their indices. A manual check and any necessary corrections are made after threading each month and the algorithm learns the pattern of ellipses, so that fewer errors require correction when threading the later months.
- Step 6 permits manual inspection, adjustment and refitting of the first fit. Now that the ellipses are serially “threaded”, refits can be initialised using the previous/next H or M parameters to avoid introducing any potential confirmation bias.
- Step 7 assigns the ellipse probabilities (membership), F , calculated from the final fit and computes the fuzzy OEN model parameters (see §A.2). This stage does not use the pre-smoothed $\rho(W,S)$ fields in Step 3, so eliminates the associated errors.
- Step 8 implements the new fuzzy OEN procedures described in [10, 11] (see §A.3) to remove the deterministic seasonal/diurnal variation of the location and scale parameters by a process of demodulation, so revealing the random perturbations and estimating their PDFs, autocovariances and spectra.

The fuzzy demodulation process in step 8 to obtain u and v for each observation is approximate and has some obvious flaws. For each ellipse, the mean $\bar{W}(M, H)$ and $\bar{S}(M, H)$ is subtracted from every observation to shift the vector origin to the ellipse centre, the axes are rotated by $\alpha(M, H)$ to align with the ellipse axes to give the u and v perturbations, which are then normalised by the standard deviations σ_u and σ_v , respectively. This demodulation is exact only for those observations that belong to the target ellipse, when the resulting perturbations will be small but their weighted contribution from F_i will be large. In observations which belong to the other ellipses, a reverse modulation is induced, generating spurious seasonal/diurnal perturbations, but the weight from F_i will be small. Hence the accuracy of this fuzzy demodulation process is a trade-off between the large probability of only a small error and the small probability of a large error. Some “leakage” from other ellipses is therefore inevitable. Three indicators of the quality of the normalised perturbations, u/σ_u and v/σ_v , are that the means should be zero, the standard deviations should be unity and the mean cross-products should be zero.

The principal difference between this implementation and the original in [9] is that the ellipses attributed to large scale weather (LSW) systems are no longer forced to be non-diurnal by averaging over all H . The improved fitting process no longer needs this precaution to be able to resolve these ellipses. In any case, [10] recognised that the relative frequencies of the LSW ellipses must vary diurnally to maintain $\sum f_i = 1$ in response to the varying diurnal components.

3. Results of Applying the Method to Adelaide

The first fit, final fit and fuzzy OEN parameters are provided in the Additional Materials as .CSV files, allowing the model to be applied and the figures in this paper to be reproduced by the R scripts without having to repeat the analysis and fitting steps. However any figure that includes the raw TD3505 observations will require steps 1 – 3 of Figure 4 to be run to extract them from the NCEI TD3505 database to comply with the NCEI terms of use.

3.1 Overall

The seasonal-diurnal chart for the marginal PDFs, Figure 1(b), and the whole-year joint PDF, Figure 3(b), synthesised from the fuzzy OEN mixture model, are both faithful, smoothed, reproductions of the corresponding observed charts. Figure 3(c) shows the mean locations and sizes of the eight ellipses deduced by the method, weighted by their probability of occurrence, and the $\rho(W,S)$ field generated from them. This is included as a useful aide in interpreting the later results. But the probability field in (c) is not the same as in (b) because the mean of the 288 fields and the field from the mean parameters are not concomitant due to the non-linear nature of the model. In comparison with the five ellipses of the earlier analysis [9]: the sea breeze ellipse now separates into the local, E1, and continental sea breeze, E4; the downslope ellipse separates into a diurnal downslope, E3, and a summer SE ellipse, E5; and the LSW2 ellipse separates into a NE ellipse and a winter NW ellipse. The ellipses in each new pair show markedly different characters, supporting their separation and suggesting that the previous model was under-parameterised.

Figure 7(a) shows the contributions from each ellipse to the overall marginal PDF of wind speed, $p(V)$ by the thin curves labelled with the ellipse index, the overall PDF for all ellipses by the thick (blue) curve and the observed PDF by the circle symbols. It reveals that E8, the SW component, controls the eventual upper tail and therefore the extreme wind speeds. Figure 7(b) shows the corresponding overall CDF plotted on standard axes that transform the Weibull distribution $P(V) = 1 - \exp(-V^k/C^k)$ into a straight line, where the observations are shown by the circular symbols and the thick (blue) curve is the fuzzy OEN model. The OEN model prediction that u and v are Normally distributed leads directly to the expectation that $P(V)$ converges onto a straight line of slope $k = 2$ [6] in the upper tail. The straight (red) line is the best fit with slope $k = 2$ through the upper tail of the data. Figure 7(c) shows the contributions of each ellipse to the overall marginal PDF of wind direction, using the same key as 7(a). Figures 7(b) and (c) are equivalent to Figure 19 (a) and (b) of [9] except that here $P(V)$ includes calms so that the data converges onto the Weibull line from above.

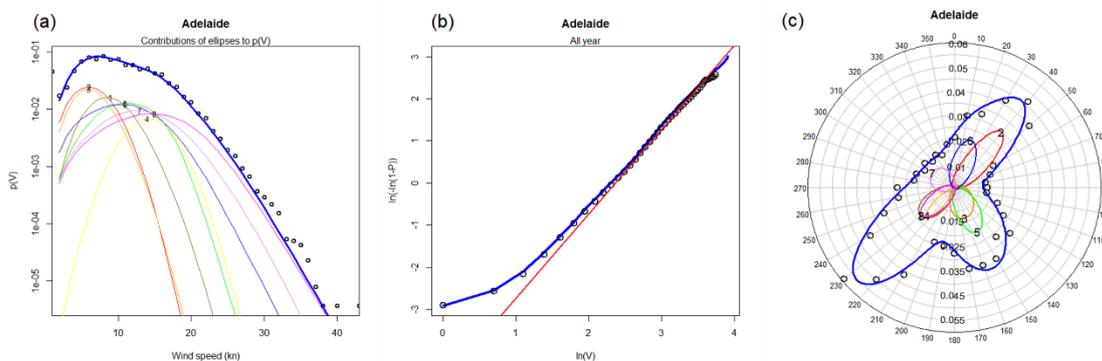


Figure 7 Marginal distributions: (a) $p(V)$ for each and all ellipses; (b) Weibull plot for $P(V)$; (c) $p(\theta)$ for each and all ellipses; compared with observations – see text for key.

Figure 8 shows the autocovariance and spectrum of wind speed for the whole record. It is not possible to obtain the power spectral density function (PSDF) directly using the fast Fourier transform (FFT) because this does not tolerate any missing observations which are inevitably found

in practice. Instead, the classical Blackman and Tukey [17] method has been adopted, whereby autocovariance is first computed from the lagged products, discounting missing observations, then Fourier transformed to obtain the PSDF. A maximum lag of two years is the minimum length that will give a whole number of daily and annual $\frac{1}{2}$ cycles, such that all their harmonics and their sidebands [11, 18] each occupy only a single frequency bin. Figure 8(a) presents the autocorrelation for the first 20 days of lags, showing that:

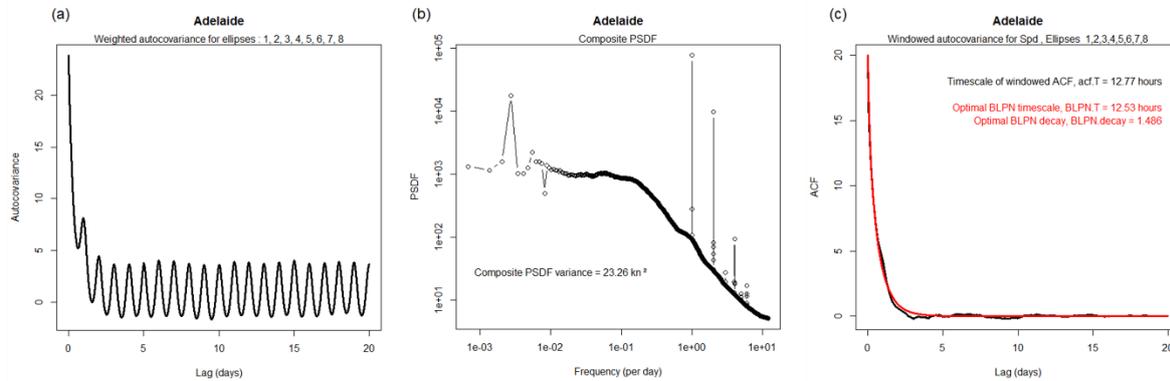


Figure 8 Overall autocovariances and spectrum of wind speed: (a) Autocovariance; (b) Spectrum; (c) Autocovariance of random component.

1. The diurnal cycle does not decay because it is phase locked.
2. The mean of the diurnal cycle at these short lags is above the origin, representing the contribution of the annual cycle ($\approx 2.5 \text{ kn}^2 \text{ pk-pk}$).

Figure 8(b) presents the composite PSDF obtained by the following procedure:

1. The raw PSDF was obtained as the Fourier transform of Figure 8(a). In this, the daily and annual harmonics occupy single frequency bins and present as sharp spikes. Small sidebands representing the annual modulation of the daily cycle, first noted by Oort and Taylor [18], present as smaller peaks spaced at $1/365.25 \text{ day}^{-1}$ intervals either side of them.
2. These spikes and sidebands were removed by heterodyning to leave the raw PSDF of only the random component.
3. The random component was smoothed using a filter bandwidth proportional to frequency.
4. The harmonics and sidebands were reinstated to form the composite spectrum shown.

Filtering the PSDF was essential because the bandwidth-time product [17] $BT = 14.13$ for these data is insufficient for defining the random component but filtering the raw PSDF directly would also artificially widen the bandwidth of the daily and annual harmonics. The single peak to the left is the annual cycle, the peak values towards the right are the harmonics of the daily cycle, while the lower values on these harmonic spikes are the annual sidebands. The underlying random component takes the characteristic shape of the bandwidth-limited pink noise (BLPN) spectrum of Equation A.10.

Figure 8(c) presents the autocovariance of the random component after heterodyne removal of the daily and annual harmonics and the application of a 20-day Hanning window. The time scale obtained by direct integration is $T = 12.8\text{h}$. The fit to the BLPN model gives a good match, $T = 12.5\text{h}$, and the decay constant, $\beta = 1.49$ is reasonably close to the Kolmogorov decay value $\beta = 5/3 = 1.67$.

3.2 Seasonal Diurnal Variation of Individual OEN Ellipses

3.2.1 Relative Frequencies

Figure 9 shows the relative frequency, f , for each ellipse at each H through the year. This is equivalent to Figure 14 in [9] but now reveals how the LSW ellipses have been split and vary diurnally. Note that the previous arbitrary split between “SL” and “DS”, marked “A” in Figure 14(a) in [9] was automatically resolved into f_1 and f_3 by the refined method.

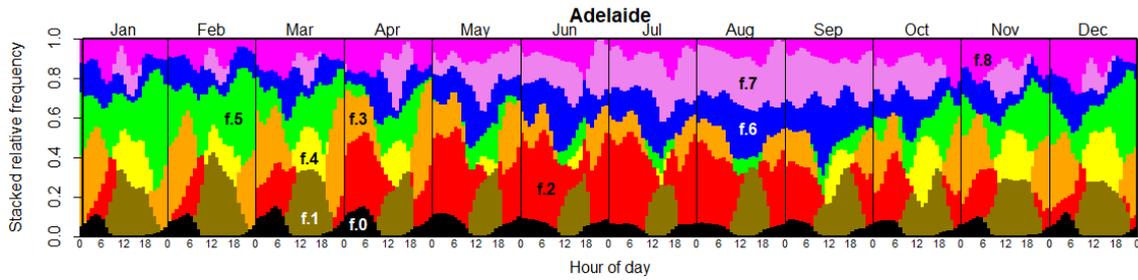


Figure 9 Relative frequencies of the eight OEN ellipses indexed by hour and month.

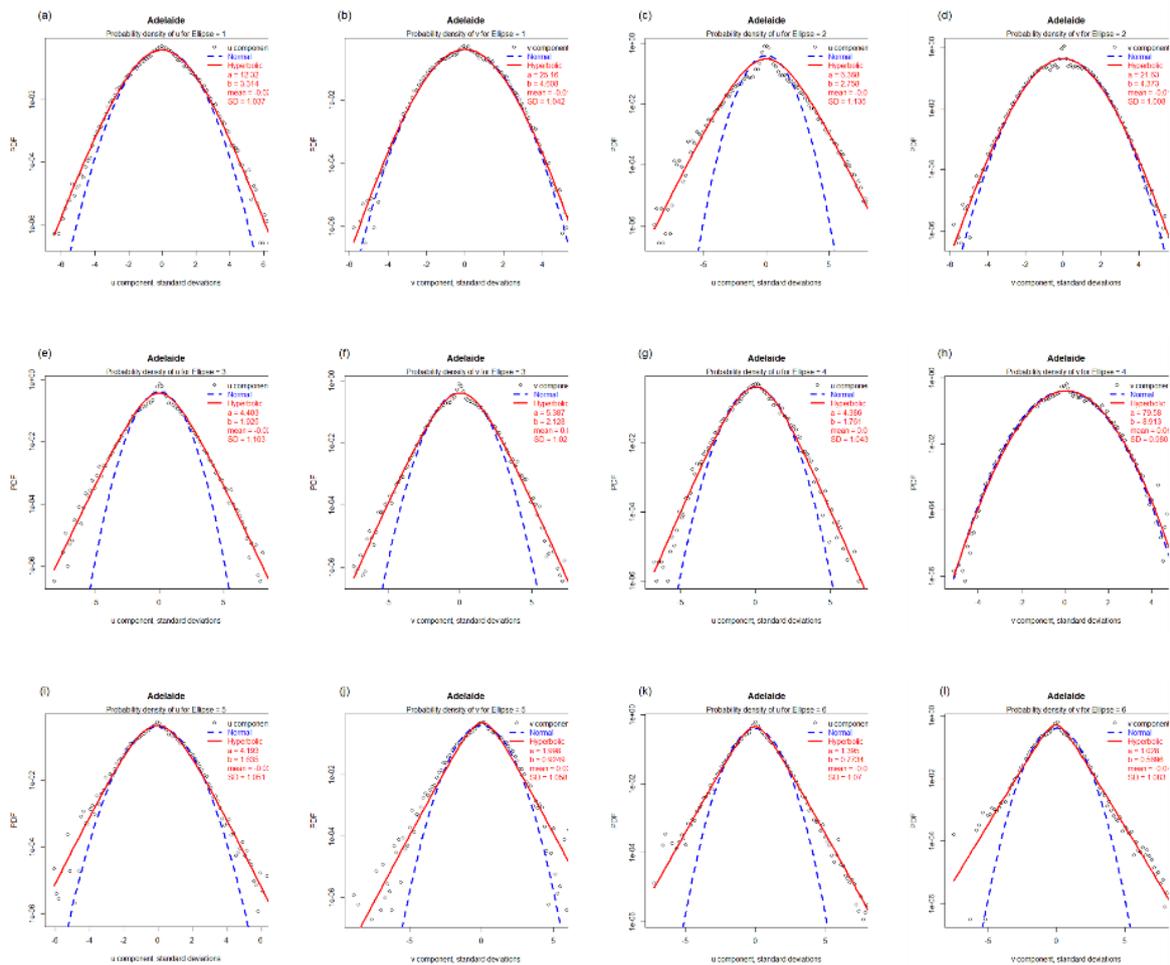


Figure 10 Fuzzy PDFs of the scaled random perturbations u/σ_u and v/σ_v for each ellipse.

A clearer presentation for the relative frequency, f , of each ellipse is a contour map on hour-month axes, as in Figure 11:

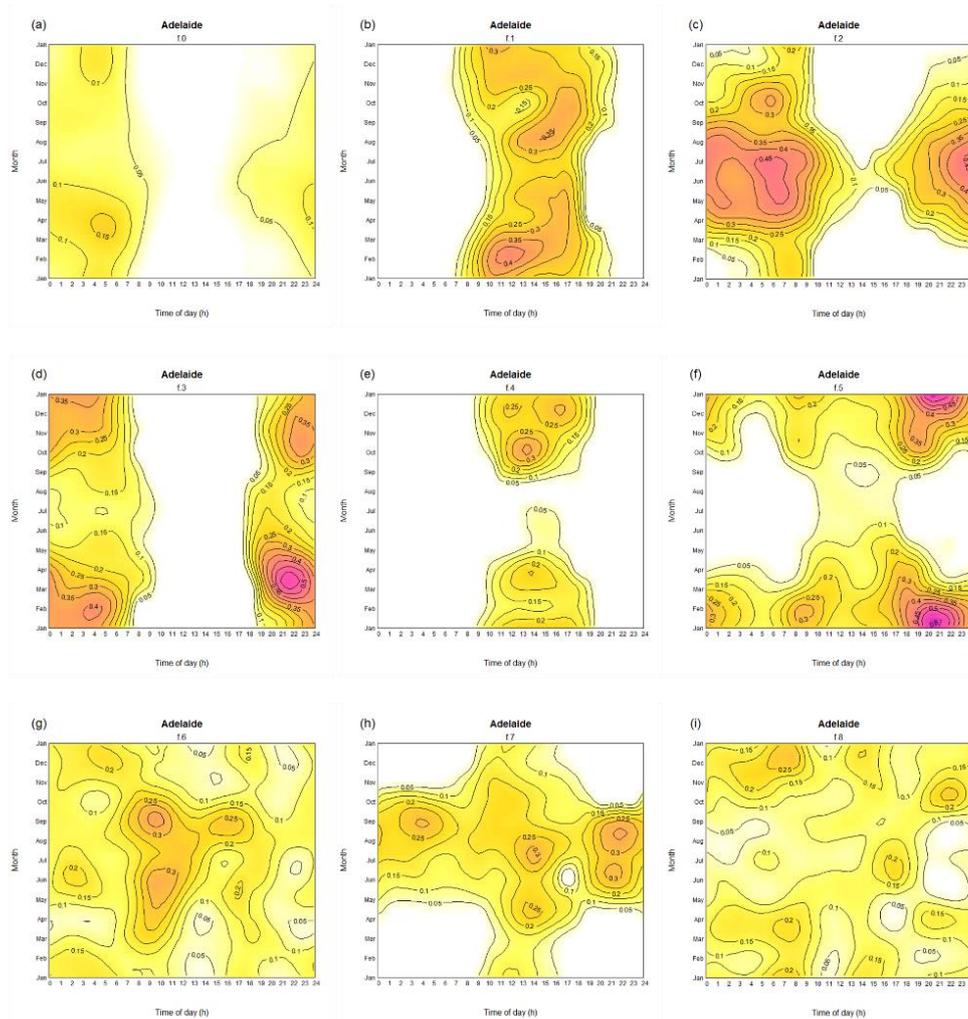


Figure 11 Relative frequencies of the eight ellipses presented as contour maps.

- a. Calms, f_0 , are seen to occur in the early hours throughout the year.
- b. The local diurnal sea-breeze, f_1 , peak occurrence is early afternoon.
- c. The land-breeze, f_2 , is nocturnal with peak occurrence in winter.
- d. The downslope winds, f_3 , are nocturnal with peak occurrence in summer.
- e. The continental diurnal sea-breeze, f_4 , peak occurrence is early afternoon in summer.
- f. Southeast winds, f_5 , occur throughout the day in summer with peak occurrence in the evening.
- g. Northeast winds, f_6 , occur throughout the year with peak occurrence at midday in winter.
- h. Northwest winds, f_7 , occur through the day in winter and around noon in summer.
- i. Southwest winds, f_8 , are nominally non-diurnal throughout the year.

As anticipated, now that the LSW ellipses have not been forced to be non-diurnal, the patterns for (g) E6, (h) E7 and (i) E8 are less coherent than the compact diurnal ellipses but are no longer arbitrarily constrained.

3.2.2 Magnitude of the Mean Vector

The magnitude of the mean vector, $\bar{V} = (\bar{W}^2 + \bar{S}^2)^{0.5}$, for each ellipse is plotted in Figure 12 as contour maps, where the contour labels are in knots and the colour scale is the same for all charts.

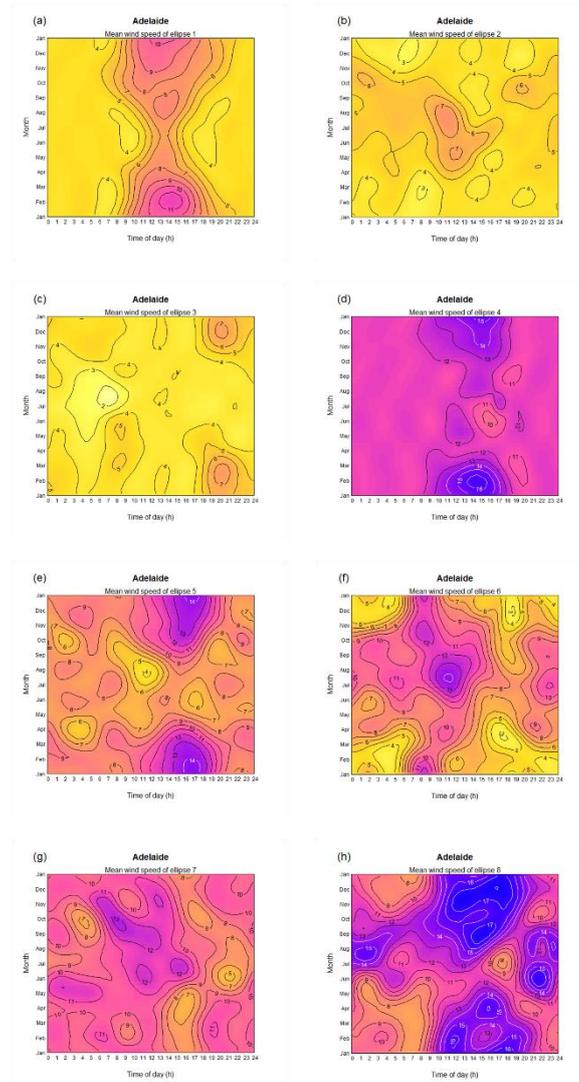


Figure 12 Mean vector magnitude, $\bar{V} = (\bar{W}^2 + \bar{S}^2)^{0.5}$ (kn), of the eight ellipses. (Note that the charts show values where the ellipse does not exist because $f = 0$. The R function specially developed to display these data replaces all missing values with the average value around the periphery of the valid data field to mitigate the Gibbs [19] phenomenon: the “ringing” effect that would otherwise appear next to a jump discontinuity after Fourier smoothing, but this has no impact in the OEN model since $f = 0$.)

These deterministic seasonal-diurnal cycles of the mean vectors represent the regular periodic movement of the ellipse centres around which the random components act. Studies of the sea/land breeze cycle, e.g., those cited in [9], particularly [20], focus strongly on the mean diurnal hodographs formed by these movements, and they speculate on why a particular hodograph does, or does not,

show the sunwise rotation predicted by Haurwitz [21] theory – citing the “delicate balance” between Coriolis force, surface and synoptic pressure gradients, topography, and advection effects. The corresponding whole-year diurnal hodographs for Adelaide, compiled by averaging the Figure 12 fields across all months, are plotted in Figure 13. The individual graphs for each ellipse are labelled by the ellipse index and the small labels indicate the time at 6h intervals. The (black) set labelled “All” is the overall hodograph for all the ellipses combined and corresponds in format to the hodographs in [20]. Here, it describes a sunwise (anti-clockwise) circuit which, as it contains the origin, results in the continuous net rotation of the wind vector noted earlier (Figure 1). However, this net rotation is achieved in jumps between the sea-breeze, downslope, and land-breeze ellipses, repeating in sequence. The hodographs for the local (E1) and continental (E2) sea breezes contradict the observation in [22] that the former’s vector at Adelaide is normal to the coast and the latter’s is southerly: both show a sunwise (anticlockwise) rotation with H consistent with Coriolis forcing, veering from SW towards S, but the continental sea breeze establishes later and is stronger. The largest LSW ellipses, E7 and E8, describe a clockwise circuit which must be driven by the shape of the corresponding geostrophic systems as they are advected past the site. The other ellipses form almost closed loops indicating no Coriolis forcing.

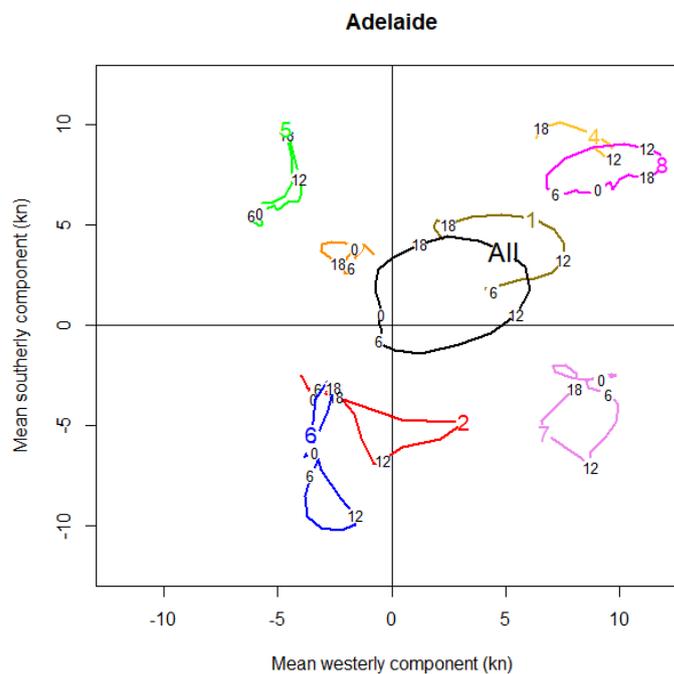


Figure 13 Mean vector diurnal hodographs for the eight individual ellipses and for all ellipses. The large (coloured) numbers indicate the ellipse index and the small numbers the time of day, H .

3.2.3 Scale of the OEN Ellipses

Figure 14 indicates of the scale of each ellipse by $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ which, as u and v are predicted by the OEN model to be orthogonal and uncorrelated, represents an overall standard deviation of the random perturbations. The contour labels are in knots and the colour scale is the same as Figure 12. This shows that the diurnal ellipses E1 to E4 are consistently small, while LSW

ellipses E6 to E8 are large, and E5 is intermediate. The westerly-dominated LSW ellipses are largest during the transition from summer to winter, indicating the most variability in the strongest winds.

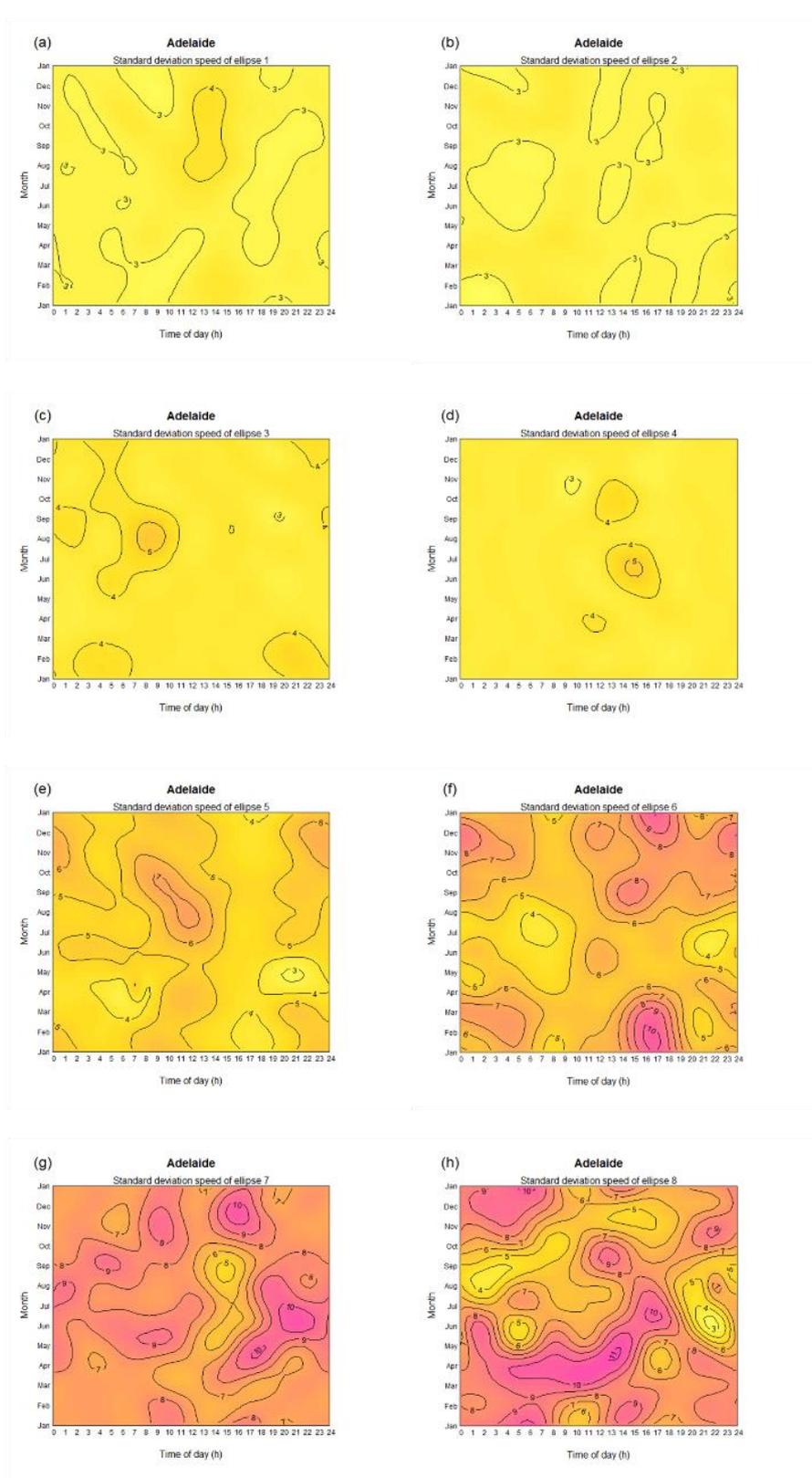


Figure 14 Scale parameter, $\sigma = (\sigma_u^2 + \sigma_v^2)^{0.5}$ (kn), of the eight ellipses as contour maps.

3.3 The Random Perturbation Intensities, u/σ_u and v/σ_v

The OEN model is founded on theory that predicts random, orthogonal, Gaussian perturbations modulated by the deterministic mean seasonal-diurnal cycles. The previous analysis [9] gave no insights into how well the characteristics of the random perturbations matched these predictions. The developments of the fuzzy OEN model in [11], defined in §A.3, enable the predictions and the quality of the fuzzy demodulation process to be tested, which is important if the fuzzy OEN mixture model is to be useful in practice. The extended analysis also provides the parameters needed to synthesise time series of the wind vector with the correct serial correlation as well as the seasonal-diurnal variation [11].

3.3.1 Location, Scale, and Orthogonality

If the fuzzy demodulation process was perfect, then the mean would be zero and the standard deviation would be unity for u/σ_u and v/σ_v of each ellipse. Additionally, if the perturbations were orthogonal, then their mean cross-product $\rho_{uv} = \overline{u/\sigma_u \times v/\sigma_v}$ would be zero. As noted earlier, the demodulation is a compromise between the large probability of a small error in the target ellipse and the small probability of a large error from the other ellipses. Even if an ellipse is orthogonal, leakage from the other ellipses will contribute non-zero values to ρ_{uv} . Table 1 shows the u/σ_u and v/σ_v values for each ellipse – deviations from the expected values are small but, as expected, they are not zero.

Table 1 Test for location, scale, and orthogonality of the OEN random perturbations at Adelaide.

Ellipse	1	2	3	4	5	6	7	8
μ_{u/σ_u}	-0.0289	-0.0129	-0.0236	0.0201	-0.0333	-0.0317	0.0166	0.0242
μ_{v/σ_v}	-0.0130	-0.0142	0.0359	0.0635	0.0314	-0.0501	-0.0259	0.0285
σ_{u/σ_u}	0.9062	1.042	0.9990	0.9301	0.9772	1.027	1.018	1.008
σ_{v/σ_v}	0.8954	0.8534	0.8999	0.8998	1.004	0.9919	0.9846	1.029
ρ	0.0294	0.1442	0.1337	0.0230	0.0636	-0.0074	0.0923	-0.1443

3.3.2 Fuzzy Probability Densities

As the fuzzy PDFs do not depend on the order of the observations in time, any deviations from the model prediction are due solely to imperfect demodulation. Figure 10 presents the fuzzy PDFs of the random perturbations for each ellipse, plotted on logarithmic probability axes to emphasise the tails. On these axes the Normal distribution presents as a parabola, represented here by the dashed (blue) curves.

While the distributions are reasonable fits over ~2 orders of magnitude, they all some show exponential deviations in the far tails. This is a common phenomenon found in Gaussian processes and is caused by non-linearities in the perturbations. It is found in nature at all scales from the distribution of stars in elliptical galaxies down to electrical signals in biological cells [23]. The

parabola is the form of conic section that lies on the boundary between the hyperbola and the ellipse. A model for the PDF corresponding to the best fitting hyperbola on these axes is given by:

$$p(x) = \frac{c}{\sigma_x \sqrt{2\pi}} \exp \left(-a \left(\sqrt{\frac{[x - \bar{x}]^2}{\sigma_x^2} + 1} - 1 \right) \right) \quad (1)$$

where a and b are the parameters of the canonical hyperbola equation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$, and c is the scaling constant that is determined by integrating Equation 1 for each value of a and b . The solid (red) curves in Figure 10 show that this model gives a good fit to the exponential deviations from Normal parabola.

Note that the anomaly around the origin of the PDFs of the smaller ellipses, most evident in (c), is a resolution bias artefact which is introduced by the 10° increments of direction [11] and can therefore be discounted.

3.3.3 Autocovariances

The fuzzy autocovariances functions (ACFs) are dependent on the time lag between observations so additional error from “leakage” through the F_i values of the other ellipses must be expected in addition to that from imperfect demodulation. The autocovariances, $R(\tau)$, of u/σ_u and v/σ_v for each ellipse are presented in Figure 15. With perfect demodulation and orthogonality, the standard deviations in Table 1 would be unity, so the autocovariance and autocorrelation would be identical. Autocovariance is presented here to reveal the imperfections in the process. Here, (d) shows the most scatter because E4 is active for the least time and has the smallest population. Some residual diurnal variation is evident in most of the cases but decays with lag, τ , indicating that this is likely to be a “leakage” error since any real diurnal cycle must show a constant amplitude. The fitted curves are the bandwidth limited pink noise (BLPN) model introduced in [10] and defined by Equation A.10.

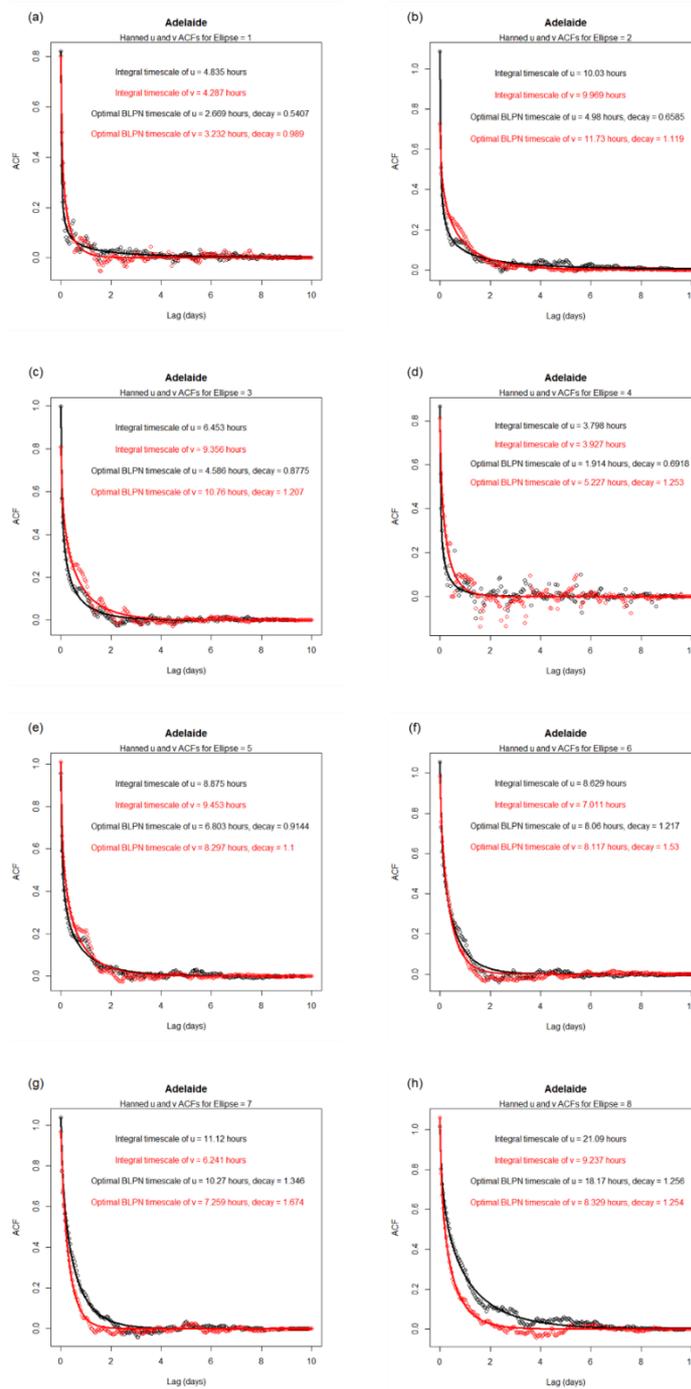


Figure 15 Fuzzy ACFs of the scaled random perturbations for each ellipse.

3.3.4 Improvement in Consistency

Figure 16 compares the RMS error between observed $p(W,S)$ and final OEN model from the refined method of this paper with the corresponding error in the original analysis in [9]. The range of error in the new method is smaller, indicating that it is more consistent. The larger RMS errors of the original method are reduced. However, some smaller RMS errors of the original method are exceeded – this is the small price that is paid for removing over-parameterisation to achieve serially consistent threads of ellipses.

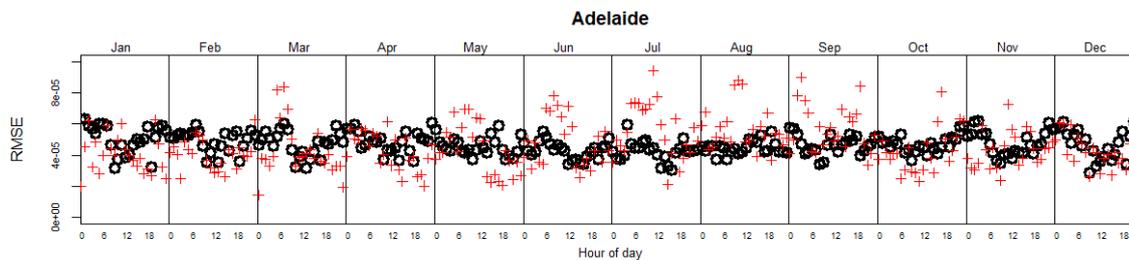


Figure 16 RMS error between observed $p(W,S)$ and OEN model in final fit: (black) circles – current analysis; (red) plus symbols – previously published analysis.

4. Discussion

The earlier analysis [9] presented the OEN model as a simplification of reality, a “black-box” method which met the principal aim of characterising and simulating the Adelaide wind climate sufficiently well for engineering design. The later suggestion that the OEN ellipses represented actual physical mechanisms can now be viewed with more confidence since this extended analysis, and the intervening analyses of the two Rome stations in [10], confirm that the ellipses present characteristics consistent with the wind mechanisms attributed to the two well-studied sites. It is now proposed that the OEN mixture model is a useful tool to identify, separate and characterise the physical meteorology. The extension of the OEN method to estimate the PDFs and ACFs/spectra was introduced specifically to verify the predictions of the OEN model theory; but is also a necessary first step in the development of indefinitely long timeseries simulations of the wind vector which combine the observed seasonal-diurnal variation with the serial correlation of the random components. As noted in [11], this can now be done for each individual ellipse component, but not for the complete mixture until a model for the disjoint transitions between mixture components has been developed.

The extensive review in [8] of previously proposed methods for assessing wind energy potential exposed the futility of seeking a universal single-model distribution for the marginal PDF of wind speed, $p(V)$, based on goodness-of-fit alone. On this basis, the Weibull distribution is inevitably rejected in favour of esoteric distributions with multiple adjustable parameters. The majority of these methods operate on the marginal distribution of $p(V)$, or of $p(V)$ and $p(\theta)$ independently, with the unfortunate consequence of losing all information on their joint action. Further, this approach is essentially an empirical curve-fitting exercise in which the more multiple adjustable parameters there are, the better the apparent fit will always be. It is important to recognise the essential difference between empirical and theoretical models: the former, obtained by curve-fit, is valid only within the range of the fitted data, so cannot be extrapolated, only interpolated¹; whereas the latter is a representation of the known physics from which reliable inferences can be drawn, including extrapolation outside the range of the data.

Only a few studies, so far, recognise that all wind climates are generated by a mixture of mechanisms, so is always best represented by a mixture of distributions [8]. Modelling the marginal

¹ In some curve-fit methods, interpolation *between* data points becomes increasingly less accurate as more adjustable parameters are added to improve the fit *at* the data points, affecting integration.

PDF of direction has the additional complexity of circularity as the direction passes through North: Carta et al [7] and others [8] address this issue using a mixture of von Mises distributions. Even fewer studies attempt to represent the joint characteristics of the wind vector, whether in meteorological polar coordinates (V, θ) [7] or in zonal-meridional coordinates (W, S) as here.

Consider the marginal distributions in Figure 7:

- In (a), the whole-year $p(V)$ evaluated from the OEN model, which fits the observations well, has a flattened region after the mode so is not a good fit in the body to a single Weibull distribution, although converging in (b) to Weibull, $k = 2$, in the upper tail. This overall shape is dictated by the shapes and relative frequencies of the individual ellipse distributions, each of which would be expected to be reasonably well represented by a single Weibull distribution, so leading to a Weibull mixture model as found by others [8].

- In (c), the whole-year $p(\theta)$ evaluated from the OEN model, which fits the observations well, forms three lobes corresponding to NE, SE and SW winds. The NE lobe is formed by ellipses 2 and 6, the SE lobe by 3 and 5, and the SW lobe by 1, 4 and 8, each of which occupy different speed ranges in (a).

Returning to verify the principal predictions of the OEN model:

1. The assumption, that all the seasonal-diurnal variation is encapsulated by the five OEN parameters for each ellipse, is verified in the u and v perturbations by the lack of any organised, non-decaying, diurnal cycles in their ACFs, and any residual variation about the fitted BLPN model is ascribed to “leakage” errors.

2. The values of ρ in Table 1 test the prediction that u and v are orthogonal and uncorrelated but can never actually be zero due to the imperfect demodulation. However, the values are sufficiently low for this to be a good working assumption.

3. The PDFs show that the prediction that u and v are Normally distributed holds well for about ± 2 standard deviations, outside of which the tails tend to become exponential. It is not at all clear whether this is due to non-linear errors or to non-linear behaviour. The former can be discounted but the latter must be addressed, especially if the interest is in the extremes. The hyperbolic distribution, Equation 1, is a good fit to the tails but is a less good fit near the origin, suggesting that a mixture model may be better.

The unique characteristic of the OEN method is that it builds a bivariate Normal mixture model of $p(V, \theta)$ from the observations without incurring any loss of information. The conventional marginal distributions of each V and θ may be subsequently evaluated from the OEN model by marginalising.

5. Conclusions

The merits of the refined, extended, and automated OEN methodology are:

- 1) That previous model constraints have been removed or mitigated.
- 2) That automation of the implementation removes any possibility of unconscious confirmation bias.
- 3) Introduction of Bayesian regularisation to set the initial number of ellipses limits any predisposition to overfit the parameters but can be too aggressive.
- 4) The overall analysis time has been considerably reduced. The automated scripts run unsupervised and manual supervision is confined to:

- a) Checking/correcting the “threading” process which indexes the ellipses for successive observation times, and
- b) Checking/correcting the final fit for serial continuity.

The open-source R scripts to implement the OEN mixture model are offered a useful tool to identify, separate and characterise the physical meteorology at any site with a reasonably long observational record at hourly, or shorter, intervals.

Appendix A. Summary of the OEN Mixture Model Equations

A.1 Original Model in [6] and [8]

The cumulative distribution function (CDF), $P(V)$, of wind speeds including calms for a mixture of N disjoint components of the wind speed, V , is:

$$P(V) = f_0 + (1 - f_0) \sum_{i=1}^N f_i \times P_i(V) \tag{A.1}$$

where f_0 is the relative frequency of calms, f_i is the relative frequency of the i -th wind component and $\sum f_i = 1$.

The corresponding probability density function (PDF) is:

$$p(V) = \frac{dP(V)}{dV} = (1 - f_0) \sum_{i=1}^N f_i \times p_i(V) \tag{A.2}$$

The Crutcher [2, 3, 4] model for the PDF of the wind vector in the zonal-meridional plane is the bivariate Normal distribution:

$$p(W, S) = \frac{\exp \left[-\frac{1}{2(1 - \rho_{WS}^2)} \left(\frac{(W - \bar{W})^2}{\sigma_W^2} - \frac{2\rho_{WS}(W - \bar{W})(S - \bar{S})}{\sigma_W\sigma_S} + \frac{(S - \bar{S})^2}{\sigma_S^2} \right) \right]}{2\pi\sigma_W\sigma_S\sqrt{1 - \rho_{WS}^2}} \tag{A.3}$$

in terms of the westerly, W , and southerly, S , components defined in Figure 5(a). (See also [2-4])

The Offset Elliptical Normal (OEN) model of [6] simplifies (A.3) into:

$$p(u, v) = \frac{\exp \left[-\frac{1}{2} \left(\frac{(u - \bar{u})^2}{\sigma_u^2} + \frac{(v - \bar{v})^2}{\sigma_v^2} \right) \right]}{2\pi\sigma_u\sigma_v} \tag{A.4}$$

where u and v are on new reference axes centred on the mean vector and rotated by the angle α at which $\rho_{uv} = 0$, so that u and v are the uncorrelated perturbations from the mean vector defined in Figure 5(b).

A.2 Fuzzy Model in [10] and [11]

The fuzzy probability F_i of an observation belonging to the i -th ellipse is given by:

$$F_i(W, S) = \frac{f_i p_i(W, S)}{\sum_i f_i p_i(W, S)}, \text{ where } \sum_i F_i = 1 \tag{A.5}$$

and is sometimes called the “membership ratio”. It represents the proportion of observations for any wind vector value that belong to the i -th ellipse. Without additional information it is not known to which ellipse any single individual observation belongs.

The fuzzy model parameters in (A.3) are the moments of W and S weighted by the fuzzy probability, F . For example, the mean W parameter for the i -th ellipse at a given month, M , and hour, H , is:

$$\bar{W}_i(M, H) = \frac{\sum_t [W(M, H) \times F_i(W, S, M, H)]}{\sum_t F_i(W, S, M, H)} \tag{A.6}$$

where \sum_t represents summation of the observations matching M and H .

A.3 The u and v components in [10] and [11]

The u and v components are found from W and S at any given M and H by:

$$\begin{aligned} u_i &= (W - \bar{W}_i) \cos \alpha_i + (S - \bar{S}_i) \sin \alpha_i, \\ v_i &= (S - \bar{S}_i) \cos \alpha_i - (W - \bar{W}_i) \sin \alpha_i \end{aligned} \tag{A.7}$$

assuming every observation contributes to the i -th ellipse through its fuzzy probability F_i .

The fuzzy PDFs of u and v for the i -th ellipse may be obtained by accumulating the F_i values instead of the simple count of observations.

The fuzzy autocovariances are given by the weighted lagged products. E.g., for u :

$$R_i(u, \tau) = \frac{\sum_t [(u_i(t) \times F_i(t) \times u_i(t + \tau) \times F_i(t + \tau)]}{\sum_t [F_i(t) \times F_i(t + \tau)]} \tag{A.8}$$

where τ is the time lag and \sum_t represents summation over the whole record, excluding missing observations.

Analogous to (A.5), the overall autocovariance of a disjoint mixture is:

$$R(\tau) = \sum_i \frac{[R(\tau) \cdot \sigma_i^2]}{\sigma^2} \tag{A.9}$$

The fuzzy spectra must be obtained from the Fourier transform of the autocovariances because conventional Fast Fourier Transform methods do not tolerate missing observations.

The generalised spectral model for bandwidth-limited “pink” noise (BLPN) introduced in [11] is:

$$\frac{S(n \leq n_{\max})}{\sigma^2} = 4T(1 + (AnT)^2)^{-\frac{\beta}{2}} \times \sin^2(\pi nt)(\pi nt)^{-2} \tag{A.10}$$

where T is the integral time scale, β is a decay constant and t is the averaging time of the observations: here, $t = 600s$ for SYNOP/METAR mean wind speeds. A is a normalising constant that ensures $\int_{n=0}^{\infty} S(n)dn = \sigma^2$, the value of which must be obtained by optimisation. The \sin^2 term is the Ogura [24] spectral window representing block-averaging over time t .

Acknowledgments

Use of the METAR observations for Adelaide is permitted under the general not-for-commercial-use terms of the NCEI TD3505 database.

Additional Materials

The following additional materials are uploaded at the page of this paper.

1. Figure S1: Observed distributions of $p(W,S)$ laid out by hour of day, H , in rows and month, M , in columns.
2. Figure S2: Final fitted distributions of $p(W,S)$ laid out by hour of day, H , in rows and month, M , in columns and showing the active OEN ellipses.
3. Figure S3: Fuzzy OEN parameters for each ellipse by hour and month.
4. The R scripts listed in Figure 4 in text files: Adelaide.TD3505.R.
5. The R scripts listed in Figure 4 in text files: Adelaide.Wind.R.
6. The R scripts listed in Figure 4 in text files: Adelaide.smoothWSp.R.
7. The R scripts listed in Figure 4 in text files: Adelaide.MHOENfit.R.
8. The R scripts listed in Figure 4 in text files: Adelaide.MHOENthread.R.
9. The R scripts listed in Figure 4 in text files: Adelaide.MHOENrefit.R.
10. The R scripts listed in Figure 4 in text files: Adelaide.MHOENfuzzy.R.
11. The R scripts listed in Figure 4 in text files: Adelaide.OENcov.R.
12. The OEN parameters in .CSV files: Adelaide.OEN.Rfit1.csv – The first stage fit before threading.
13. The OEN parameters in .CSV files: Adelaide.OEN.Rfit1.csv – Adelaide.OEN.Rfit3.csv – The final, threaded fit.
14. The OEN parameters in .CSV files: Adelaide.OEN.Rfit1.csv – Adelaide.OEN.Fuzzy.csv – The fuzzy OEN model.

Note: All the R scripts require the open-source R scripts that accompany the open access e-book *Wind Engineering Cookbook: Recipes in R*, freely downloadable from www.njcook.uk. The additional materials are also available for download from the Mendeley archive at URL: <http://dx.doi.org/10.17632/sf4tmcg6d2.1> where any future updates might also be found.

Author Contributions

The author is solely responsible for this work.

Funding

This is the result of unfunded, pro bono, curiosity-driven research.

Competing Interests

The author has declared that no competing interests exist.

References

1. Brooks CEP, Durst CS, Carruthers N. Upper winds over the world: Part I, the frequency distribution of winds at a point in the free air. *Q J R Meteorol Soc.* 1946; 72, 55-73.
2. Crutcher HL. On the standard vector-deviation wind rose. *J Meteorol.* 1957; 14: 28-33.
3. Crutcher HL, Baer L. Computations from elliptical wind statistics. *J Appl Meteorol.* 1962; 1: 522-530.
4. Crutcher HL. Separation of mixed data sets into homogeneous sets. Washington DC: NOAA Technical Report EDS 19, Dept. of Commerce; 1977.
5. Crutcher HL, Joiner RL. Another look at the upper winds of the tropics. *J Appl Meteorol.* 1977; 16; 426-476.
6. Harris RI, Cook NJ. The parent wind speed distribution: Why Weibull? *J Wind Eng Ind Aerodyn.* 2014; 131: 72-87.
7. Carta JA, Ramirez P, Bueno C. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Convers Manag.* 2008; 49: 1309-1320.
8. Cook NJ. The OEN mixture model for the joint distribution of wind speed and direction: A globally applicable model with physical justification. *Energy Conv Manag.* 2019; 191: 141-158.
9. Cook NJ. A statistical model of the seasonal-diurnal wind climate at Adelaide. *Aust Meteorol Oceanogr J.* 2015; 65: 206-232.
10. Cook NJ. Parameterising the seasonal-diurnal wind climate of Rome: Fiumicino and Ciampino. *Meteorol Appl.* 2020; 27: e1848.
11. Cook NJ. Implications of the OEN mixture model of the mean wind vector for the generation of synthetic timeseries and for the assessment of extremes. *J Wind Eng Ind Aerodyn.* 2021; 208: 104424.
12. Jakob D. Challenges in developing a high-quality surface wind-speed data-set for Australia. *Aust Meteorol Oceanogr J.* 2010; 60; 227-236.
13. Cook NJ. Visualising seasonal-diurnal trends in wind observations. *Weather.* 2015; 70: 117-121.
14. Cook NJ. Detecting artefacts in analyses of extreme wind speeds. *Wind Struct.* 2014; 19: 271-294.
15. Fraley C, Raftery AE, Scrucca L, Murphy TB, Fop M. Mclust: Gaussian mixture modelling for model-based clustering, classification, and density estimation. Available from: <https://mclust.org.github.io/mclust/>.
16. Fraley C. Algorithms for model-based Gaussian hierarchical clustering. *SIAM J Sci Comput.* 1998; 20: 270-281.
17. Blackman RB, Tukey JW. The measurement of power spectra. New York: Dover Publications; 1958.
18. Oort AH, Taylor A. On the kinetic energy spectrum near the ground. *Mon Weather Rev.* 1969; 97: 623-636.
19. Gibbs JW. Fourier's series. *Nature.* 1899; 59: 200.
20. Moisseeva N, Steyn DG. Dynamical analysis of sea-breeze hodograph rotation in Sardinia. *Atmos Chem Phys.* 2014; 14: 13471-13481.
21. Haurwitz B. Comments on the sea-breeze circulation. *J Atmos Sci.* 1947; 4: 1-8.
22. Physick WL, Byron-Scott RA. Observations of the sea breeze in the vicinity of a gulf. *Weather.* 1977; 32: 373-381.

23. Holzer M, Siggia E. Skewed, exponential pressure distributions from Gaussian velocities. *Phys Fluids A*. 1993; 5: 2525-2532.
24. Ogura Y. The influence of finite observation intervals on the measurement of turbulence diffusion parameters. *J Meteorol*. 1957; 14: 176-181.



Enjoy *JEPT* by:

1. [Submitting a manuscript](#)
2. [Joining in volunteer reviewer bank](#)
3. [Joining Editorial Board](#)
4. [Guest editing a special issue](#)

For more details, please visit:

<http://www.lidsen.com/journal/jept>