Simulating the distribution and cross-correlation of wind farm output

Problem presented by

Tom Allerton

E.ON Engineering

Executive Summary

The problem was to devise a simulation method for the wind speeds at a set of sites, that has the correct autocorrelation, cross-correlation and distributions. The report includes one way of doing this, using a multivariate auto-regressive system, and other comments and observations that may lead to better ways of achieving the aim.

Version 1.0
11 April 2008

Report coordinator

David Allwright (Smith Institute)

Contributors

Violeta Andrade (Instituto Tecnológico Autónomo de México)
Thanos Avramidis (University of Southampton)
Jeff Dewynne (Universities of Oxford and Woollongong)
Sam Howison (University of Oxford)
Andrew Lacey (Heriot-Watt University)
Daniela Licea (Instituto Tecnológico Autónomo de México)
Sandhya Patidar (Loughborough University)
Andrey Pototsky (Loughborough University)
Honora Smith (University of Southampton)
Antoine Tambue (Heriot-Watt University)

ESGI64 was jointly organised by
Heriot-Watt University
The Knowledge Transfer Network for Industrial Mathematics
The International Centre for Mathematical Sciences

and was supported by
Engineering and Physical Sciences Research Council
The London Mathematical Society
The Institute of Mathematics and its Applications
The European Journal of Applied Mathematics

Contents

1 Introduction **1**
1.1 Data available 2

2 Outline of modelling approaches **2**

3 Preliminary data analysis **4**

4 Complex Gaussian model **5**
4.1 Scalar case 5
4.2 Multivariate case 8

5 Simulated data **9**

6 Conclusions and other remarks **9**

7 Appendix: Data file structure **10**
7.1 Ordering time-stamps 11
7.2 Outlier 11
7.3 Midnight zeros 11
7.4 Stuck readings 12

8 Appendix: Generating specified distributions by an AR(1) process **12**

9 Appendix: Complex Gaussian Random Vectors **14**

1 Introduction

Figure 1: A UK wind farm.

(1.1) When wind is used to generate electricity, intermittency can be a problem. Since it is unlikely to be calm simultaneously over a large area, one way to make intermittency less of a problem is to have a portfolio of wind farms spread over a wide geographical area. To run simulations to estimate this effect quantitatively, and for other purposes, E.ON wish to have a method of generating wind speed time series at a set of sites, with the correct statistical properties.

(1.2) One property is that the correlation coefficient R between the wind speeds at 2 sites separated by a distance d should be a decreasing function of d, for instance of the form

$$R^2 \approx \alpha e^{-d} + \beta. \quad (1)$$

(1.3) Another property is that the correlation in time of the wind speed at a single site should be correct. E.ON consider that enough accuracy is provided by an auto-regressive first-order (AR(1)) model of the form

$$y_t = Ay_{t-1} + n_t \quad (2)$$

where y_t is the wind speed and n_t is random noise. If t is measured in half-hours, A is usually at least 0.95.

(1.4) The third property that E.ON wish to have correct is the distribution of wind speed at a single site, which should follow a Weibull distribution with scale λ and shape k, so the density function is

$$p(y) = \frac{ky^{k-1}}{\lambda^k} \exp\left(-(y/\lambda)^k\right), \quad (y > 0). \tag{3}$$

We call this the Weibull(k) distribution. Typical values of k obtained by fitting to observed data are in the range 1.6 to 2.3.

(1.5) If the simulation method can fit the joint spatio-temporal correlation structure as well as the space and time correlations individually, that would be an added advantage.

1.1 Data available

(1.6) The data provided by E.ON for this study was wind speed data averaged over 10-minute intervals from 7 wind farm sites taken at various times during the 4 years 2002–2005. The data (after some cleaning described in Section 7) is displayed in Figure 2. Data for an individual site is present for between 8 and 20 months out of the 4 years, so coverage is somewhat sparse, but there is, for instance, a 3-month period in which there is synchronized data for 5 sites.

2 Outline of modelling approaches

(2.1) Models of various kinds were proposed, and were tested and pursued to various extents. In this section we describe them briefly.

(2.2) One suggested approach was to use a stochastic differential equation model for the wind speed process at a point. These models were closely related to the discrete-time models, broadly taking the form of replacing a discrete-time model of the form (2) by

$$dy = -(1 - A)y\,dt + d\xi(t) \tag{4}$$

where $\xi(t)$ is a noise process.

(2.3) The distribution of the differences of successive values $d_t = y_t - y_{t-1}$ was found to have, very closely, a 2-sided exponential distribution. This led to some suggested models for the wind speed process. In particular the GARCH model, which has a stochastic volatility, appeared to provide many of the desired features observed in the (d_t) time series. Any resulting negative speeds y_t would have to be replaced by 0.

(2.4) Various models that fall under the broad heading of regime-switching were proposed. For instance, one could have an underlying Markov process that

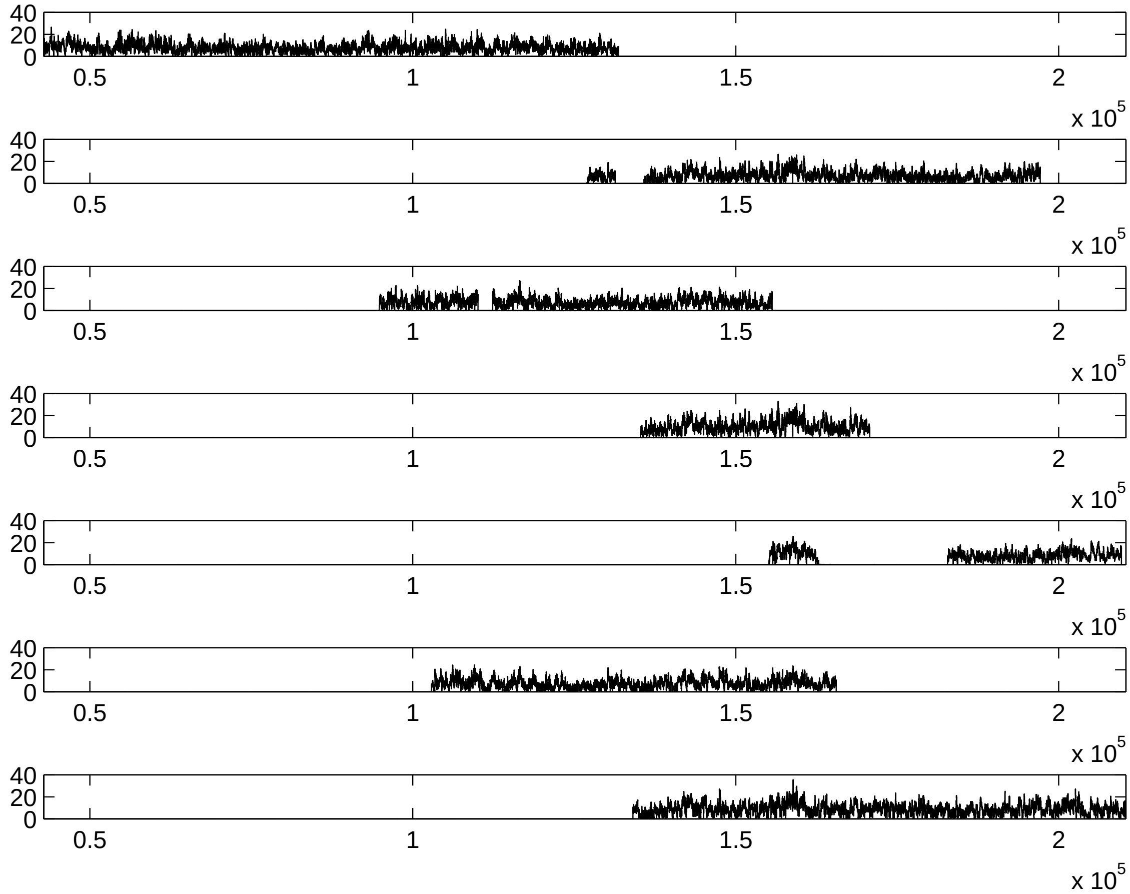

Figure 2: Measured wind speeds in m/s averaged over 10-minute intervals from 7 sites. The horizontal axis is the row index. In the data this runs from 1 to 210384 but here we only display from the start of the main record in column 2. This covers the 38 month period from November 2002 to December 2005.

switches between calm and stormy weather, with certain transition rates, and then a random process for wind speed whose parameters differ depending on the calm/stormy state. In mathematical finance, analogous models are sometimes called mixing models.

(2.5) The Weibull(2) distribution is also called the Rayleigh distribution, and can also be viewed as the square root of the χ_2^2 distribution. In fact, if the wind velocity vector $\mathbf{V}$ in r dimensions has zero mean and independent components of equal variance, then $|\mathbf{V}|^2$ is a χ_r^2 random variable and so $|\mathbf{V}|$ has the square root of that distribution. Hence when $r = 2$, this "Gaussian wind" blowing in 2 dimensions has exactly a Weibull(2) distribution.

(2.6) Some models attempted to bring about a Weibull(2) distribution for y by assuming an underlying Gaussian model for a wind speed vector (u, v) with zero mean and equal variances in the East and North directions. This can lead to some success, as we shall see in Section 3, but is difficult to validate in detail without wind direction data, which at the moment we do not have.

(2.7) A wind farm position that was particularly exposed or sheltered (compared with being on a large flat plain) might be modelled by a distribution like that of $|\mathbf{V}|$ above with a larger or smaller value of r, not necessarily an integer. That distribution could then be approximated by a Weibull(k) distribution, and there would be a correspondence (which could be made precise by moment-matching or in other ways) between values of r and the k. However, we did not have data on how the fitted values of k deviate from 2 at different kinds of sites, so this was not pursued in detail.

(2.8) Lastly, some models attempted to incorporate a prevailing wind velocity into the model of the preceding paragraph, since this is one of the most obvious ways in which it fails to represent reality. In this case, if the mean wind velocity is non-zero and the deviations from that are Gaussian and independent in the 2 directions, then the square of the speed is a non-central χ_2^2 distribution. This is more complicated and we shall here only describe the simpler case where the mean velocity is taken as zero.

3 Preliminary data analysis

(3.1) If we take the section of the data consisting of rows 137001 to 149000, then columns 3, 4, 5, 7, 8 are all active, and so we have a 12000 point (about 3 months) 5-variate time series. This data is plotted in Figure 3.

(3.2) The autocorrelations of these time series are plotted in Figure 4, and the cross-correlations in Figure 5. It is clear that the cross-correlations do not have their peaks at zero lag, *i.e.* the greatest correlation between the wind speeds at different sites is not for simultaneous values, but for one site being slightly delayed or advanced with respect to the other.

(3.3) We fitted higher order auto-regressive models to this multivariate time series, using the Yule-Walker equations. The sum of the variances in the data started at 88, and fitting the first order AR model reduced that to 2.91. Further increase of the order to 5 reduced it to 2.88, and even going to order 256 only reduced it to about 2.5. So, much the main reduction is obtained by the first step, suggesting that it still makes sense to look at AR(1) models for the multivariate data.

(3.4) The comparison of R^2 against distance can be plotted from these, and is shown in Figure 6, but since few sites were involved, and also the location of site 5 was not provided, fitting R^2 by a function of d is inconclusive.

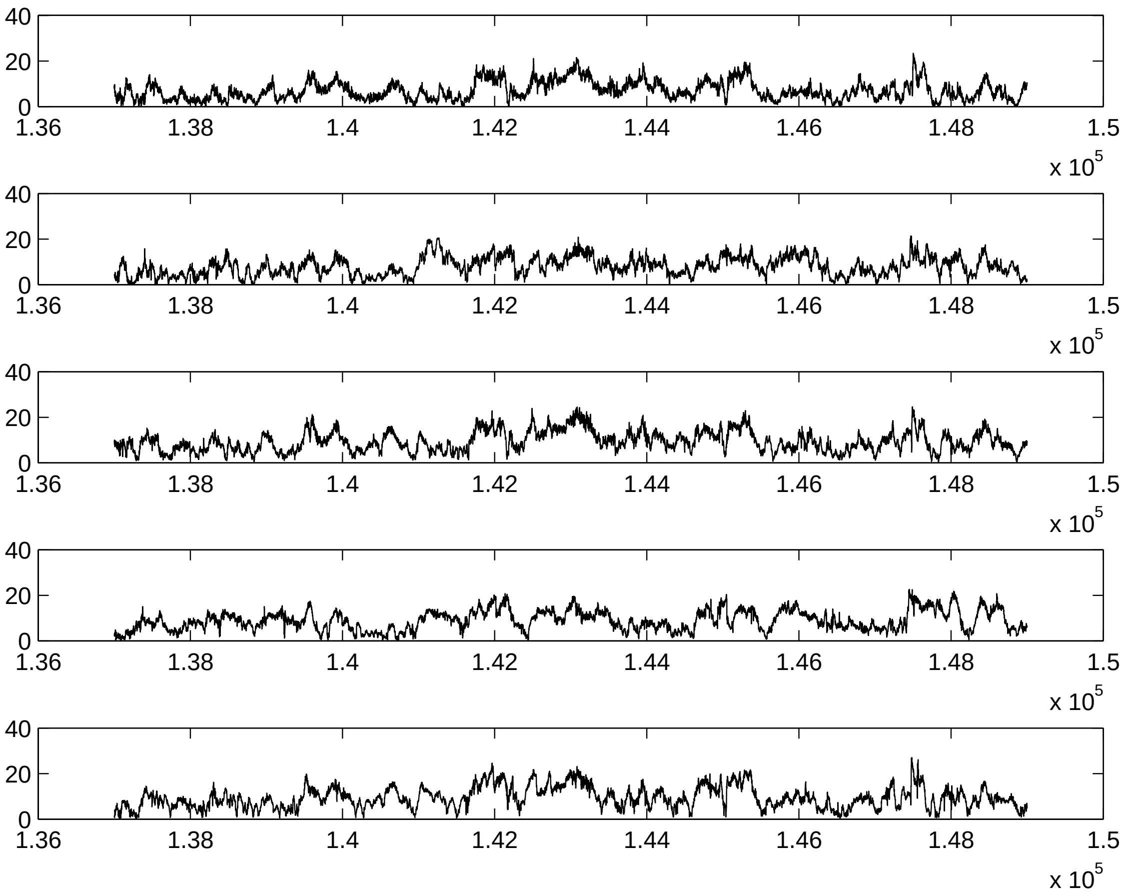

Figure 3: Measured wind speed data in rows 137001 to 149000, in columns 3, 4, 5, 7, 8

4 Complex Gaussian model

4.1 Scalar case

(4.1) We first show the fact mentioned above about the Weibull(2) distribution, so suppose that (U, V) are independent Gaussians with mean 0 and equal variance $\sigma^2/2$, and let

$$Y = \sqrt{U^2 + V^2} = |U + iV|. \tag{5}$$

Then the joint density function of (U, V) is

$$\left(\frac{1}{\sqrt{\pi}} \exp(-u^2/\sigma^2) \frac{du}{\sigma}\right) \left(\frac{1}{\sqrt{\pi}} \exp(-v^2/\sigma^2) \frac{dv}{\sigma}\right). \tag{6}$$

When we go to polar coordinates such that $u + iv = y \exp(i\theta)$, this becomes

$$\left(\exp(-y^2/\sigma^2) \frac{2y\,dy}{\sigma^2}\right) \left(\frac{d\theta}{2\pi}\right). \tag{7}$$

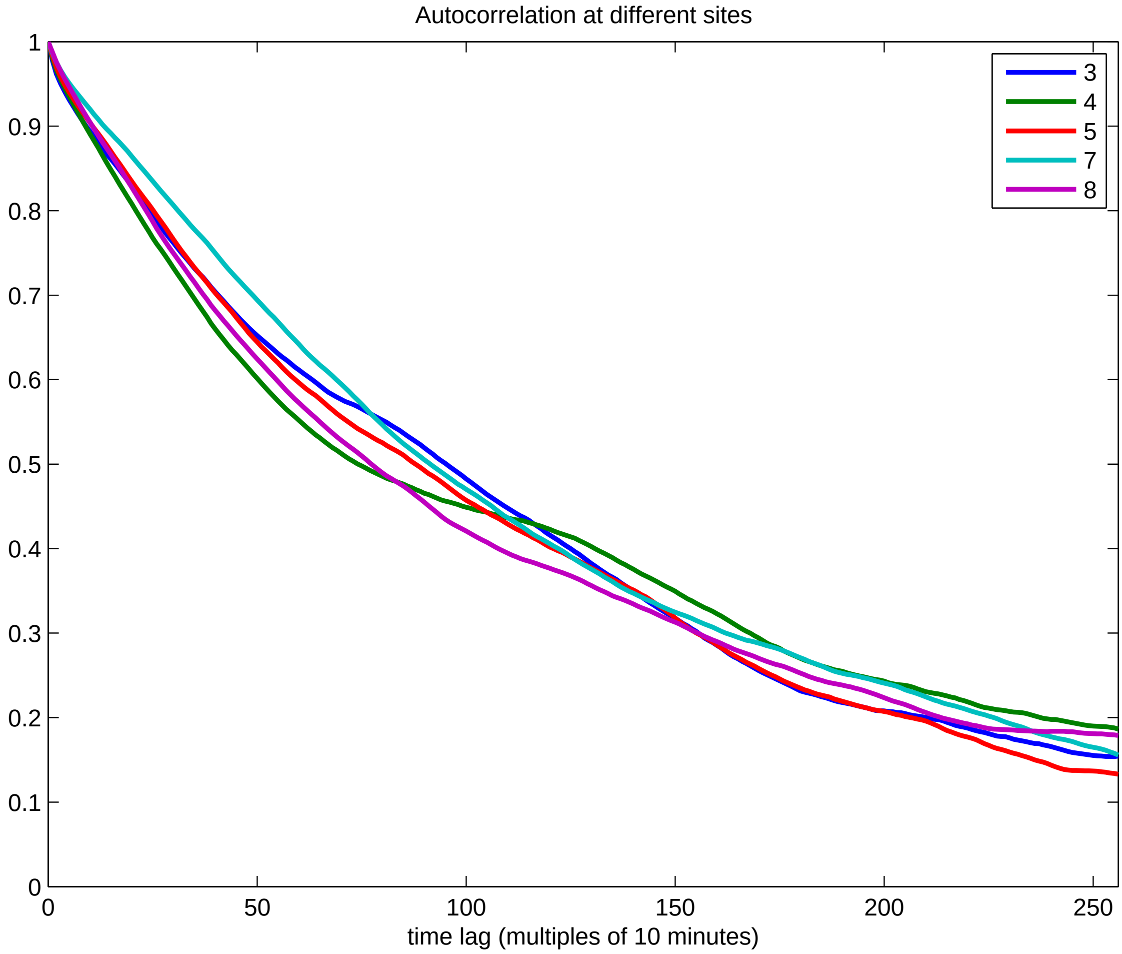

Figure 4: Autocorrelation of wind speeds at sites 3,4,5,7,8.

The second factor is just the uniform density over θ, and the first is a Weibull(2) distribution of Y with scale σ. In such a case we say that $X = U + iV$ is a complex Gaussian random variable with variance $\sigma^2 = \mathbb{E}(|X|^2)$.

(4.2) One way to produce a Weibull(2) time series Y_t with autocorrelation would therefore be to take a stationary complex Gaussian time series X_t generated by

$$X_t = BX_{t-1} + Z_t \tag{8}$$

and let $Y_t = |X_t|$. We shall want the noise Z_t to be complex Gaussian, and independent of the past history of X, *i.e.* independent of $(X_{t-1}, X_{t-2}, \ldots)$. If the variance of Z_t is σ^2 as above, then taking the variance of (8) gives the variance of the stationary distribution as

$$\mathrm{Var}(X_t) = \mathbb{E}(|X_t|^2) = \frac{\sigma^2}{1 - |B|^2}. \tag{9}$$

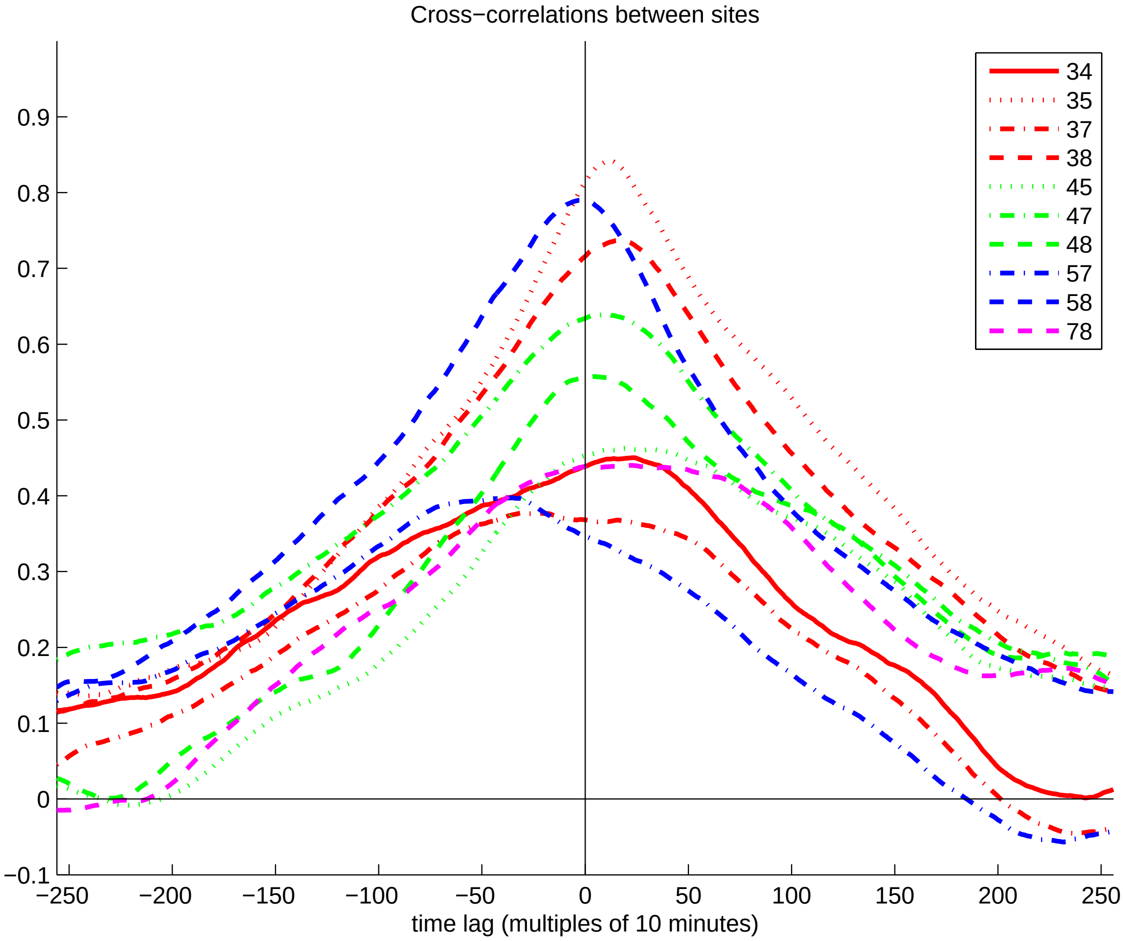

Figure 5: Cross-correlations between sites 3,4,5,7,8.

The autocorrelation structure of X is that

$$\mathbb{E}(X_t\overline{X}_{t-k}) = B^k \mathrm{Var}(X_t) \tag{10}$$

for $k \geq 0$.

(4.3) The question then is how to choose σ and B to produce a given Weibull(2) scale λ and a given autocorrelation in Y. Certainly we must take

$$\frac{\sigma^2}{1-|B|^2} = \left(\mathbb{E}(Y_t^2)\right)_{\mathrm{data}}, \tag{11}$$

where the subscript "data" on the right means the estimate based on the data.

(4.4) To choose B, if we fit to the observed autocorrelation of the process then we would need to ensure that

$$\mathbb{E}(|X_t||X_{t-1}|) = \left(\mathbb{E}(Y_tY_{t-1})\right)_{\mathrm{data}}, \tag{12}$$

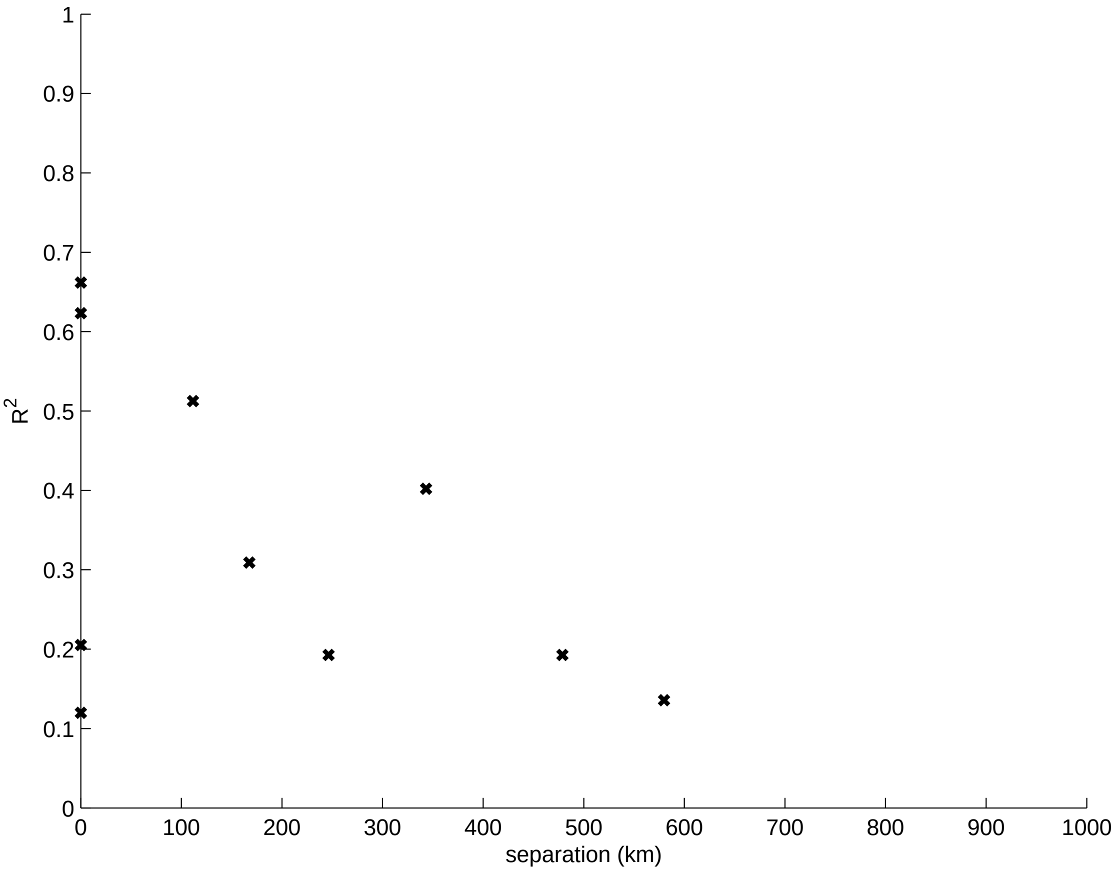

Figure 6: R^2 plotted against separation for pairs of sites.

and the expectation on the left will be $\mathbb{E}(|X_t|^2)$ times a nonlinear function of $|B|$ that could be determined numerically.

(4.5) A simpler, but in some ways less satisfactory, alternative is to fit the fourth moments, since they can be calculated analytically. In fact for $k \geq 0$

$$\mathrm{Cov}(|X_t|^2, |X_{t-k}|^2) = \left(\left|\mathbb{E}(X_t \overline{X}_{t-k})\right|\right)^2 = |B|^{2k}\mathrm{Var}(X_t)^2. \tag{13}$$

So we can fit $|B|$ by estimating the autocorrelation of Y_t^2 at lag 1 from the data and then using this equation. This raises the interesting question whether y_t^2 and y_{t-k}^2 could be *negatively* correlated. We do not see this in the data available, but it cannot be ruled out. If such negative correlations were to occur, one could not fit a model of this type.

4.2 Multivariate case

(4.6) In the multivariate case we can proceed similarly, so we still want the equation (8) to hold, but with X_t and Z_t as complex vectors of height d, the number of sites under consideration, and B a square matrix of order d. Then we shall let $Y_t = |X_t|_c$ where the subscript c means that we take the absolute values *component-wise* to get the vector Y_t of simulated speeds. We shall let V denote the variance matrix of Z, $V = \mathbb{E}(ZZ^\dagger)$ where $(.)^\dagger$ denotes the conjugate transpose. So V is a positive definite Hermitian matrix, and the definition of a multivariate complex Gaussian is that Z has density

$$\frac{1}{\pi^d \det(V)} \exp\left(-z^\dagger V^{-1} z\right) \tag{14}$$

with respect to Lebesgue measure. (Further details of this distribution are collected for reference in the Appendix in section 9.) Then X_t is also a stationary complex Gaussian process, and its variance W is given by

$$W = \mathbb{E}(X_t X_t^\dagger) = V + BWB^\dagger. \tag{15}$$

(4.7) The autocorrelation for $k \geq 0$ is given by

$$\mathbb{E}(X_t X_{t-k}^\dagger) = B^k W. \tag{16}$$

(4.8) One could then fit B by a numerical fit to the observed correlation of the time series. However, as in the scalar case, it is quicker to match the 4th moments. These are given, for $k \geq 0$, by

$$\operatorname{Cov}(|X_{i,t}|^2, |X_{j,t-k}|^2) = \left|\mathbb{E}\left(X_{i,t}\overline{X}_{j,t-k}\right)\right|^2. \tag{17}$$

Hence from the measured data for pairs of sites (i, j) we can estimate the left hand side, and thereby deduce the magnitude of the entries of $B^k W$.

(4.9) The fitting of B and V can now proceed as follows. If we assume that the imaginary parts of B and V are 0, then (17) at $k = 0$ gives us an estimate of the W matrix. Then applying the same equation at $k = 1$, we estimate B. The freedom to use complex B and Hermitian V is not used in this approach. It could perhaps be used to fit the autocorrelation at higher lags than 1, but we have not done so.

5 Simulated data

(5.1) If we follow the procedure above with this data, then we can produce simulated data that will have Weibull(2) marginals and the same cross-correlation and the correct 1-step auto-correlation as the original. A realization of this process is plotted in Figure 7.

6 Conclusions and other remarks

(6.1) We have produced a method that can take measured wind speed data and produce multivariate time series with the correct cross-correlation and 1-step autocorrelation, and with Weibull(2) marginals. If we wish to produce

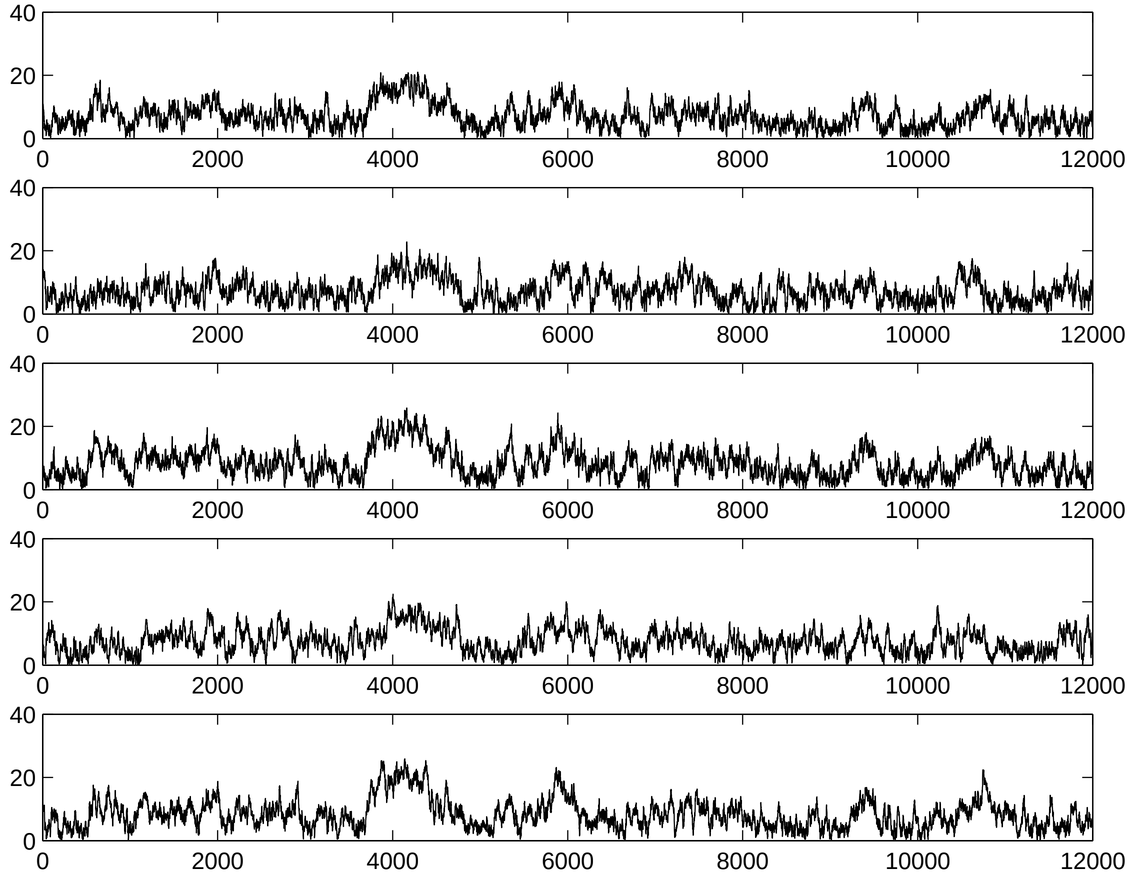

Figure 7: Simulated wind speed data with Weibull(2) marginals and the cross-correlation and auto-correlation fitted to the data in Figure 3

Weibull(k) rather than Weibull(2) then one possibility is to work with $y^{2/k}$ in place of y. (This would require knowing the value of k, which we might do if, for instance, k is known to take different values over land and sea.)

(6.2) We have shown that the goodness of fit between the measured data and Weibull is not good.

(6.3) If we want to use a longer AR process AR(m) for $m > 1$ in place of (8), then we could fit such a process to the data using the multivariate Yule-Walker equations.

7 Appendix: Data file structure

(7.1) The data had been collected into a Microsoft Access file called `tblMaster`, and E.ON exported this to a file `tblMaster.txt` of comma-separated variables. Each row in `tblMaster.txt` is for a single time and contains 8 numbers: a timestamp, followed by the 7 wind speeds in m/s, with unrecorded entries as 0. The timestamp in the Access file is of the form `yyyyddd.hhmm`, where `yyyy` is the year (2002–2005), `ddd` is the day number within the year (001 for 01-Jan through to 365 or 366 for 31-Dec), `hh` is the hour number (00–23) and `mm` is the minute number (00, 10, 20, 30, 40, 50). In exporting to the `tblMaster.txt` file, the minutes were truncated. The file contains 210384 rows: 1461 days, with 144 rows per day.

7.1 Ordering time-stamps

(7.2) The time-stamps in the file were not ordered, so the rows were permuted to put the whole file into time-order.

(7.3) In detail, the file was arranged in the order

(a) 59 rows from 2004282.03[00] to 2004282.12[40]
(b) 145602 rows from 2002001.00[00] to 2004282.02[50]
(c) 885 rows from 2004282.12[50] to 2004288.16[10]
(d) 354 rows from 2004289.02[10] to 2004291.13[00]
(e) 59 rows from 2004288.16[20] to 2004289.02[00]
(f) 413 rows from 2004298.02[30] to 2004300.23[10]
(g) 944 rows from 2004291.13[10] to 2004298.02[20]
(h) 62068 rows from 2004300.23[20] to 2005365.23[50]

The minutes numbers here in square brackets are not in the `tblMaster.txt` file, because of the truncation mentioned earlier, but they are guessed from the numbers of entries present, and confirmed by the fact that when we reassemble the 8 blocks in the order (b), (a), (c), (e), (d), (g), (f), (h), then they provide the full 10-minute sequence through the whole 4 years.

7.2 Outlier

(7.4) At time-stamp 200532.12[10] (row 162362 of the reordered data) the wind speed in column 6 is 730 m/s (faster than Mach 2) so we set this entry to 0. The maximum other speed is 35.45 m/s, and no speeds are reported negative.

7.3 Midnight zeros

(7.5) Sometimes the speeds at midnight (`hhmm`=0000) are reported as zero, even though the speeds next earlier and later (2350 and 0010) are positive. The reason for this is unknown, but we replace each such midnight zero by the mean of the values next earlier and later. This happens 618, 0, 406, 245, 0, 435 and 529 times in columns 2–8 respectively.

7.4 Stuck readings

(7.6) There are also some sections of data in column 6 where the reported speed sticks at a value of about 3.5 m/s, and also at about 0.2 m/s.

8 Appendix: Generating specified distributions by an AR(1) process

(8.1) One question that we considered was the AR(1) process

$$y_t = Ay_{t-1} + n_t \tag{18}$$

with a noise process n_t that is stationary and uncorrelated with the past history of y. How is the stationary distribution of y coupled to that of n, and in particular how do we choose the distribution of n to get a specified stationary distribution of y?

(8.2) One way of looking at the relationship of the stationary distributions of n and y is by means of the Laplace or Fourier transform. In fact, if we use the Laplace transform, let $L_y(s)$ denote the Laplace transform of the p.d.f. of y so

$$L_y(s) = \mathbb{E}(e^{-sY}). \tag{19}$$

Then since n_t is independent of y_{t-1} we have

$$L_y(s) = \mathbb{E}\left(e^{-sY}\right) = \mathbb{E}\left(e^{-s(AY+n)}\right) = L_y(As)L_n(s). \tag{20}$$

So the forward expression, giving the resulting distribution of y in terms of that of n is

$$L_y(s) = L_n(s)L_n(As)L_n(A^2s)\cdots, \tag{21}$$

while the inverse expression, giving n in terms of y is

$$L_n(s) = \frac{L_y(s)}{L_y(As)}. \tag{22}$$

(8.3) An alternative approach is to relate the moments of n and y, and this is easiest done by the cumulants, κ_j. These are given, for a general random variable X, by

$$\begin{aligned}
\kappa_1(X) &= \mathbb{E}(X) = \mu, && (23)\\
\kappa_2(X) &= \mathrm{Var}(X) = \mu_2 = \mathbb{E}\left((X-\mu)^2\right), && (24)\\
\kappa_3(X) &= \mu_3 = \mathbb{E}\left((X-\mu)^3\right), && (25)\\
\kappa_4(X) &= \mu_4 - 3\mu_2^2, \quad \ldots, && (26)
\end{aligned}$$

and can be defined in general by (28) below. These have the property that the cumulants of a sum of independent random variables are the sums of the cumulants. Hence for each $j = 1, 2, 3, \ldots$,

$$(1 - A^j)\kappa_j(y) = \kappa_j(n). \tag{27}$$

These equations are closely related to (20) since

$$\kappa_j(X) = (-1)^j \left(\frac{d}{ds}\right)^j (\log L_x(s))\Big|_{s=0}, \tag{28}$$

so the j-th logarithmic derivative of (20) gives (27)

(8.4) If we attempt to produce a Weibull(k) distribution for y by a process of this form, we run into a difficulty that can be demonstrated either from the Laplace transform approach or the cumulant approach. In the Laplace transform approach, we have the result (22), where L_y is the Laplace transform of a Weibull(k). For any p.d.f. $p(y)$ with a behaviour

$$p(y) \sim cy^{k-1} \quad \text{as } y \to 0 \tag{29}$$

the Laplace transform obeys

$$L_y(s) \sim \frac{c\Gamma(k)}{s^k} \quad \text{as } s \to \infty. \tag{30}$$

Hence (22) shows that

$$L_n(s) \to A^k \quad \text{as } s \to \infty \tag{31}$$

and therefore the density of n has a point mass A^k at the origin.

(8.5) If we subtract that out, then the p.d.f. $\gamma(n)$, say, can be written

$$\gamma(n) = A^k\delta(n) + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left(\frac{L_y(s)}{L_y(As)} - A^k\right) e^{sn}\, ds, \tag{32}$$

by a Bromwich inversion integral. Now for k in the range 1.5 to 2.5 (which covers the values of interest) $L_y(s)$ is an entire function so the singularities of the integrand in (32) are when $L_y(As) = 0$. For k in this range, the rightmost zero of $L_y(s)$, say s^*, can be computed numerically and has non-zero imaginary part. For instance if $k = 2$ and we scale the Weibull to have $\lambda = 1$ then $s^* \approx -2.5 + 5i$. Hence the most slowly decaying term in $\gamma(n)$ as $n \to \infty$ will be proportional to $\mathrm{Re}(K \exp(s^*n/A))$, and since s^* has non-zero imaginary part, this is a decaying *oscillatory* term, and therefore the p.d.f. $\gamma(n)$ would be negative at suitable large values of n, which is impossible.

(8.6) This can also be found in a more elementary way from the approach via the cumulants. For if we are given the y distribution and A then we can find the cumulants of the n distribution as above. However, the cumulants are not arbitrary, since they are connected by certain inequalities. In fact, if we let the central moments of a general random variable X be denoted by

$$\mu_j(X) = \mathbb{E}\left((X-\mu)^j\right) \tag{33}$$

then the matrix

$$\begin{pmatrix} 1 & 0 & \mu_2 & \mu_3 & \mu_4 & \cdots \\ 0 & \mu_2 & \mu_3 & \mu_4 & \cdots & \cdots \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \cdots & \cdots \\ \mu_3 & \mu_4 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{34}$$

must have all its leading principal minors positive. This fails to be the case if A is close to 1 and y is Weibull(k) with k in the range of interest.

(8.7) Thus we see that the Weibull distribution (with k in the range of interest) cannot be produced by any AR(1) process of the type considered, assuming as always that the process (n_t) in (2) is stationary and independent of the past history of y. This may be related to the problems that the VARTA software was found to encounter, although those problems did seem to be more specifically related to multivariate data, whereas this difficulty already arises in the scalar case.

(8.8) The Weibull distribution with $k = 1$ is simply the exponential distribution with mean λ, and this *can* be produced by the AR(1) process (2) by letting n be 0 with probability A, or (with probability $1-A$) an exponential variable with mean λ.

(8.9) It is worth noting another observation from this argument. If we wish y to have any density function $p(y)$ that tends to 0 as $y \to 0$ proportional to some power y^{k-1}, then we have seen that the noise n must have a point mass A^k at the origin. So the time series for y will consist of steps where $y_t = Ay_{t-1}$, with probability A^k, and other steps where $y_t = Ay_{t-1} + n_t$ with $n_t \neq 0$, with probability $1 - A^k$. This is somewhat implausible for a physical process: for instance, it would mean that if one were to plot the points (y_{t-1}, y_t) on a scatter diagram, then a positive proportion A^k of them will lie exactly on the line $y_t = Ay_{t-1}$. The process can produce the stationary distribution correctly, and the autocorrelation, but without producing realistic-looking time series.

9 Appendix: Complex Gaussian Random Vectors

(9.1) If X and Y are n-dimensional random column vectors such that (X^T, Y^T) is multivariate Gaussian ($2n$-dimensional) with mean 0, then we say that $Z = X + iY$ is a complex Gaussian random vector if any of the following equivalent conditions holds:

(a) for any real θ, $e^{i\theta}Z$ has the same distribution as Z;

(b) iZ has the same distribution as Z;

(c) $\mathbb{E}(ZZ^T) = 0$;

(d) $\mathbb{E}(XX^T) = \mathbb{E}(YY^T)$ and $\mathbb{E}(XY^T) = -\mathbb{E}(YX^T)$;

(e) the density of Z is

$$\frac{1}{\pi^n \det(V)} \exp\left(-Z^\dagger V^{-1} Z\right), \tag{35}$$

where $V = \mathbb{E}(ZZ^\dagger)$ is the positive definite Hermitian variance matrix of Z, the density is with respect to the standard Lebesgue measures $dx_1\, dy_1\, dx_2\, dy_2\, \ldots\, dx_n\, dy_n$, and we are using $(.)^\dagger$ to denote the complex conjugate transpose of a vector or matrix.

(9.2) First note that $V = \mathbb{E}(ZZ^\dagger)$ is certainly Hermitian and positive definite for any random complex vector Z (since given a constant complex vector b, $b^\dagger V b = \mathbb{E}(b^\dagger Z Z^\dagger b) = \mathbb{E}(|b^\dagger Z|^2) \geq 0$), so if we write it as $V = A + iB$ in terms of its real and imaginary parts then A is symmetric and B is skew.

(9.3) To prove equivalence of the five conditions, the first implies the second by taking $\theta = \pi/2$. If the second, then $\mathbb{E}(ZZ^T) = \mathbb{E}\left((iZ)(iZ)^T\right) = -\mathbb{E}(ZZ^T)$, hence the third. If the third, then $0 = \mathbb{E}((X^T + iY^T)(X + iY)) = \mathbb{E}(XX^T - YY^T) + i\mathbb{E}(X^TY + Y^TX)$, hence the fourth. If the fourth, then the variance matrix $A + iB = V = \mathbb{E}(ZZ^\dagger) = \mathbb{E}(XX^T + YY^T) + i\mathbb{E}(YX^T - XY^T)$ and so

$$\operatorname{Var}\left(\begin{pmatrix} X \\ Y \end{pmatrix}\right) = \mathbb{E}\left(\begin{pmatrix} X \\ Y \end{pmatrix}(X^T\, Y^T)\right) = \begin{pmatrix} A/2 & -B/2 \\ B/2 & A/2 \end{pmatrix}. \tag{36}$$

Hence the joint density of (X, Y) is

$$\frac{1}{\sqrt{\det\left(2\pi\begin{pmatrix} A/2 & -B/2 \\ B/2 & A/2 \end{pmatrix}\right)}} \exp\left(-(X^T\, Y^T)\begin{pmatrix} A & -B \\ B & A \end{pmatrix}^{-1}\begin{pmatrix} X \\ Y \end{pmatrix}\right). \tag{37}$$

The determinant here we can simplify, since for any square matrices A, B of the same size,

$$\det\begin{pmatrix} A & -B \\ B & A \end{pmatrix} = \det\begin{pmatrix} A+iB & -B+iA \\ B & A \end{pmatrix} \tag{38}$$

$$= \det\begin{pmatrix} A+iB & 0 \\ B & A-iB \end{pmatrix} \tag{39}$$

$$= \det(A+iB)\det(A-iB), \tag{40}$$

(the first step by adding i times the bottom half of the matrix to the top, the next by subtracting i times the left half from the right). If A is symmetric and B skew, then $\det(A - iB) = \det((A - iB)^T) = \det(A^T - iB^T) = \det(A + iB)$ so the constant factors in (35) and (37) agree. For the other factor, let

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix}^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} P \\ Q \end{pmatrix}, \tag{41}$$

so $X = AP - BQ$, $Y = BP + AQ$, in other words $X + iY = (A + iB)(P + iQ)$. Hence

$$(X^T Y^T) \begin{pmatrix} A & -B \\ B & A \end{pmatrix}^{-1} \begin{pmatrix} X \\ Y \end{pmatrix} = X^T P + Y^T Q \quad (42)$$

$$= \mathrm{Re}\left((X + iY)^\dagger (P + iQ)\right) \quad (43)$$

$$= \mathrm{Re}(Z^\dagger V^{-1} Z) = Z^\dagger V^{-1} Z, \quad (44)$$

and so the exponential factors in (35) and (37) also agree. Finally, the fifth condition implies the first, since the density function is unchanged on replacing Z by $e^{i\theta} Z$.

(9.4) In the case $n = 1$, Z is just a complex number whose real and imaginary parts X and Y are Gaussian, independent, each of mean 0, and of equal variance. The general case can be transformed into products of such cases by writing $V = U^\dagger \Lambda U$ with U unitary and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i > 0$. For then if $Z = U^\dagger \zeta$, and we write $\zeta = \xi + i\eta$ then the density function (35) becomes

$$\frac{1}{\pi^2 \lambda_1 \lambda_2 \cdots \lambda_n} \exp\left(\frac{\sum_i |\zeta_i|^2}{\lambda_i}\right) \quad (45)$$

with respect to the measure $d\xi_1\, d\eta_1\, d\xi_2\, d\eta_2 \cdots d\xi_n\, d\eta_n$. So $\xi_1, \eta_1, \xi_2, \eta_2, \cdots, \xi_n, \eta_n$ are independent Gaussians with mean 0 and with variances $\lambda_1/2$, $\lambda_1/2$, $\lambda_2/2$, $\lambda_2/2$, $\cdots$, $\lambda_n/2$, $\lambda_n/2$.

(9.5) The fourth moments of the distribution are given by

$$\mathbb{E}(Z_i \overline{Z}_j Z_k \overline{Z}_l) = \mathbb{E}(Z_i \overline{Z}_j)\mathbb{E}(Z_k \overline{Z}_l) + \mathbb{E}(Z_i \overline{Z}_l)\mathbb{E}(Z_k \overline{Z}_j) = V_{ij} V_{kl} + V_{il} V_{kj}. \quad (46)$$

In fact in general the moments of order $2k$ are given by

$$\mathbb{E}(Z_{i_1} Z_{i_2} \cdots Z_{i_k} \overline{Z}_{j_1} \overline{Z}_{j_2} \cdots \overline{Z}_{j_k}) = \sum_{\sigma \in S_k} V_{i_1, j_{\sigma(1)}} V_{i_2, j_{\sigma(2)}} \cdots V_{i_k, j_{\sigma(k)}}, \quad (47)$$

where the sum is over all $k!$ permutations σ of $\{1, 2, \ldots, k\}$. These can all be calculated from the moment generating function, which takes the form

$$\mathbb{E}(\exp(s_1^\dagger Z + Z^\dagger s_2)) = \exp(s_1^\dagger V s_2). \quad (48)$$