

Wind hazard assessment

W. J. Whiten¹B. Abbasi²S. Z. Hosseinifard³

22 June 2013

Abstract

Airport mean wind and gust speed data is used with past and future simulated data to estimate the maximum gust speed for a 500 year return period. It is demonstrated that the instantaneous speeds lead to the same result as data giving maximums over some time period. The airport mean wind speed was related to the gust speed, and the future simulations related to the past simulations. However, the relation between measured airport data and the simulations is poor. A Weibull distributions fitted to the upper part of the gust data is extrapolated to give the 500 year return period. An alternative approach using techniques similar to some actuarial analyses is also demonstrated.

Subject class: 62-07, 62P12

Keywords: Wind gusts, Extreme distributions, Return periods

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/5905> gives this article, © Austral. Mathematical Soc. 2013. Published July 24, 2013, as part of the Proceedings of the 2012 Mathematics and Statistics in Industry Study Group. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

Contents

1	Introduction	M2
2	Airport wind data	M4
2.1	Relation between gust speed and wind speed	M4
2.2	Extrapolation on return period scale	M14
3	Simulated wind data	M22
3.1	Simulated data at airport locations	M24
4	Relation between airport data and simulated data	M29
5	Three hourly samples of simulated data	M36
6	An actuarial approach	M42
7	Conclusions	M43

1 Introduction

The estimation of the occurrence of extreme events is important for the safety of society. In particular, wind gusts can damage buildings and other structures. To design safe structures a knowledge of the maximum wind gusts is needed. Geoscience Australia, who sponsored this project, estimates extreme events of various types including wind gusts. The maximum wind gust expected in 500 years is used in the building codes. There are three different types of events that cause high wind speeds, at low latitudes cyclonic and thunder storm winds are of concern. At higher latitudes such as Tasmania, the subject of this study, synoptic winds are the cause for concern.

Estimation of maxima over long time periods is difficult. In the case of Tasmania, although some data is available over longer periods, good information

on gust speeds is limited to 10 to 15 years and at only eight airport locations. Extrapolation of return periods (RP) from 15 years to 500 years looks more feasible on a logarithmic scale, but still requires good information on the statistical distribution. Although there is only one point giving the maximum over 15 years, the airport data is recorded every 30 minutes, so that the data goes from 5.8×10^{-5} to 15 years with the requirement to extrapolate to 500 years. On a logarithmic scale the data covers the range -4.2 to 1.2 giving about 5.4 orders of magnitude of data to extrapolate to 2.7 which is an extrapolation of 1.5 orders of magnitude.

Another source of information useful for this project is simulated climate data. Two series of simulated wind speeds were used in this project. One series covered the years 1961 to 1990, and the other, to provide information on future winds, covered 2081 to 2100. Each of these series was generated by CSIRO's CCAM using six general circulation models as input. This data, unlike the point data from airports, was an area average recorded as the maximum over 24 hours.

It was hoped that relations could be found from the airport gust data to the airport mean wind data, then to the historic simulated data, and finally to the future simulated data, so that projections of gust speed could be made for the future over the whole of Tasmania.

The estimated return period wind speed from the observed data at the airport locations differs from that calculated from the simulated data, and there are many more locations of interest beyond the airports for which only the simulated data is available. In addition to the work reported here, an optimisation approach to minimise bias in the simulated data was proposed to obtain the best match between the observed data and the simulated data. Work on this optimisation is continuing and is expected to be reported separately.

The first part of this paper examines the airport data. It is shown how the 30 minute samples relate to daily maxima and maxima over a longer period. A relation between the airport mean wind speed data and the gust data is

found for seven of the eight airports.

Section 3 looks at the simulated data, firstly the past and future data is compared by combining the data over all of Tasmania for each of the six simulations, and then the simulated data at the airport locations is examined. It is found the future data is very similar to the past data, except for a small proportion of more extreme wind speeds.

Section 4 examines the relation between the measured airport data and the simulated wind speed. Section 5 looks at extrapolating using three hour samples from one simulation for 1961 to 1990, and finally section 6 applies some techniques used in actuarial analysis of extreme data.

2 Airport wind data

Data from airports around Tasmania was purchased from the Australian Bureau of Meteorology. The data are recorded every 30 minutes, and had unfortunately been rounded to integer knots then rounded to integer km/hr. For this project it was converted to m/s. The rounding explains the gaps seen in the data. This data included average wind speed over the last 10 minutes of the interval (the so-called mean wind speed) and maximum gust speeds in the same interval from eight airports. The data was edited to remove earlier times where the data is not complete. This gave 7 to 14 years of wind and gust speeds as shown in Table 1. The data for the eight airports differed in size from about 128 000 to 251 000 values.

2.1 Relation between gust speed and wind speed

Figure 1 plots the gust speed versus the mean wind speed for the eight airports. The amount of variation in the gust speed does not increase in proportion to the mean wind speed but instead stays somewhat constant. Note that at

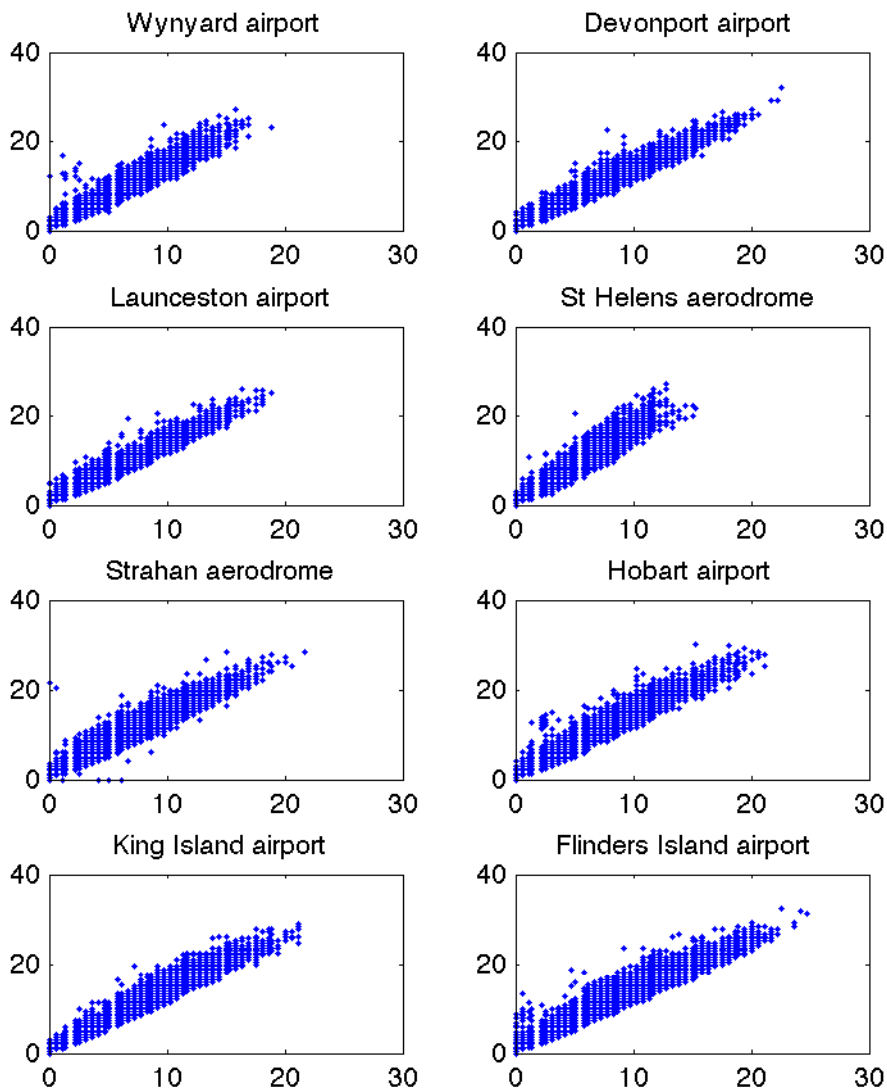


Figure 1: Plots of gust speed (vertical axis) versus wind speed (horizontal axis) at airport locations.

Table 1: Airport names and data range.

Location	Start date	End date
Wynyard airport	07-Oct-1996	15-Dec-2011
Devonport airport	07-Oct-1996	15-Dec-2011
Launceston airport	15-Jul-2004	15-Dec-2011
St Helens aerodrome	11-Jan-2001	15-Dec-2011
Strahan aerodrome	16-Jan-1997	15-Dec-2011
Hobart airport	06-Sep-1997	15-Dec-2011
King Island airport	07-Apr-1998	15-Dec-2011
Flinders Island airport	21-Apr-1998	15-Dec-2011

Table 2: Slopes for relation between gust speed and wind speed.

Location	Slope
Wynyard airport	1.40
Devonport airport	1.29
Launceston airport	1.32
St Helens aerodrome	1.62
Strahan aerodrome	1.40
Hobart airport	1.39
King Island airport	1.37
Flinders Island airport	1.31

lower wind speeds there are many more data points and hence the distribution of gust speeds may appear wider.

The distributions of the gust speed and wind speed are compared in Figure 2. Unlike Figure 1 the association between gust and wind at a given time has been removed, and instead the values are sorted and associated using the order in the sorted values, giving a quantile–quantile plot. These relation are close to a straight lines in each case. Table 2 gives the slopes of these lines.

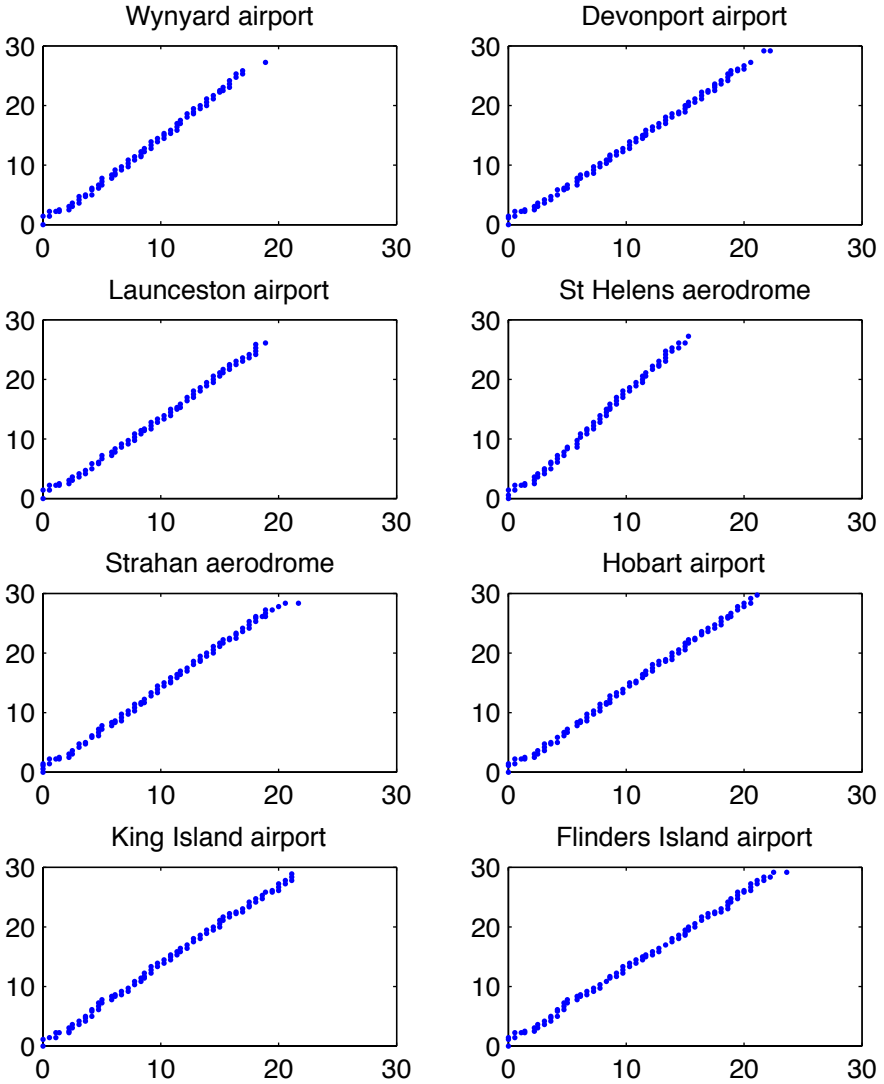


Figure 2: Plots of distributions of gust speed (vertical axis) versus wind speed (horizontal axis) at airport locations.

Table 3: Weibull parameter for airport wind and gust data.

Location	Wind α	Wind s	Gust α	Gust s
Wynyard airport	1.89	4.79	1.70	6.36
Devonport airport	1.77	5.56	1.80	7.42
Launceston airport	1.75	4.84	1.67	6.30
St Helens aerodrome	1.50	3.03	1.52	5.51
Strahan aerodrome	1.55	4.17	1.49	5.68
Hobart airport	1.60	4.84	1.71	7.37
King Island airport	1.98	6.47	2.06	9.18
Flinders Island airport	2.17	7.95	2.37	10.93

So the gust speeds are 30 to 40 percent higher than the mean wind speeds, with St Helens being an apparent outlier.

The distributions of mean wind and gust were examined using various statistical distributions including a generalised gamma (gamma distribution on a power of the data). The Weibull distribution (in cumulative form ([Wikipedia, 2012e](#)))

$$P(x) = 1 - \exp[-(x/s)^\alpha] \quad (1)$$

gave a good fit. This is a special case of the generalised gamma distribution and is also the simpler exponential distribution on a power of the data. As the interest is in the upper extreme the data was fitted to the upper 1% of the data. Figure 3 gives the fits for the wind distributions and Figure 4 the fits for the gust distributions. In both cases most of the fits are within the quantile bounds. The parameters of the Weibull fits to the upper 1% of the data are given in Table 3.

These Weibull distributions show the wind and gust distributions at each airport have similar exponents α . Equal exponents correspond to the straight lines seen in Figure 2. Figure 5 plots the gust exponents against the wind exponents and shows a trend with the gust exponent approximately equal to the wind exponent.

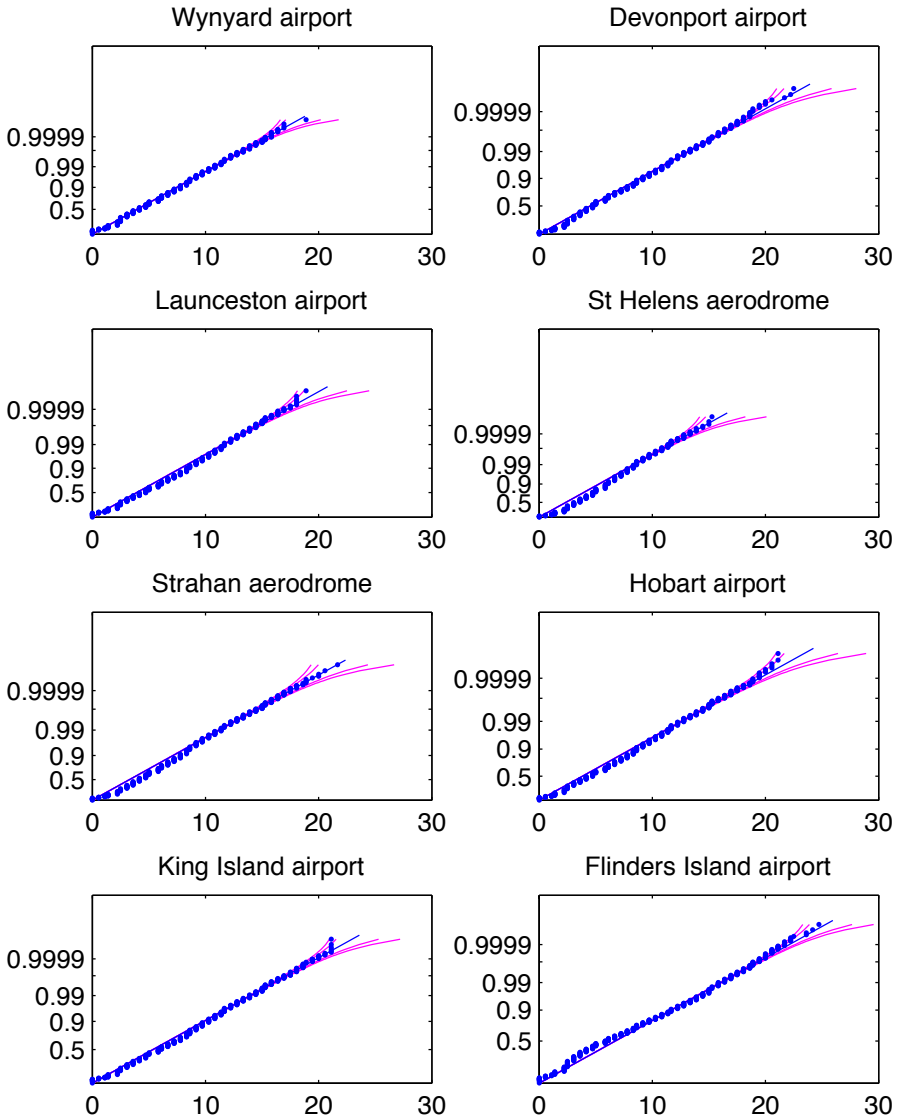


Figure 3: Weibull distribution plots, fitted to the upper 1% of the data, of wind speed at airport locations, with distribution median in blue, and the 0.005, 0.05, 0.95 and 0.995 quantiles in magenta. The horizontal axes have been scaled to the same range, causing the vertical axis probability scale to vary according to the data range.

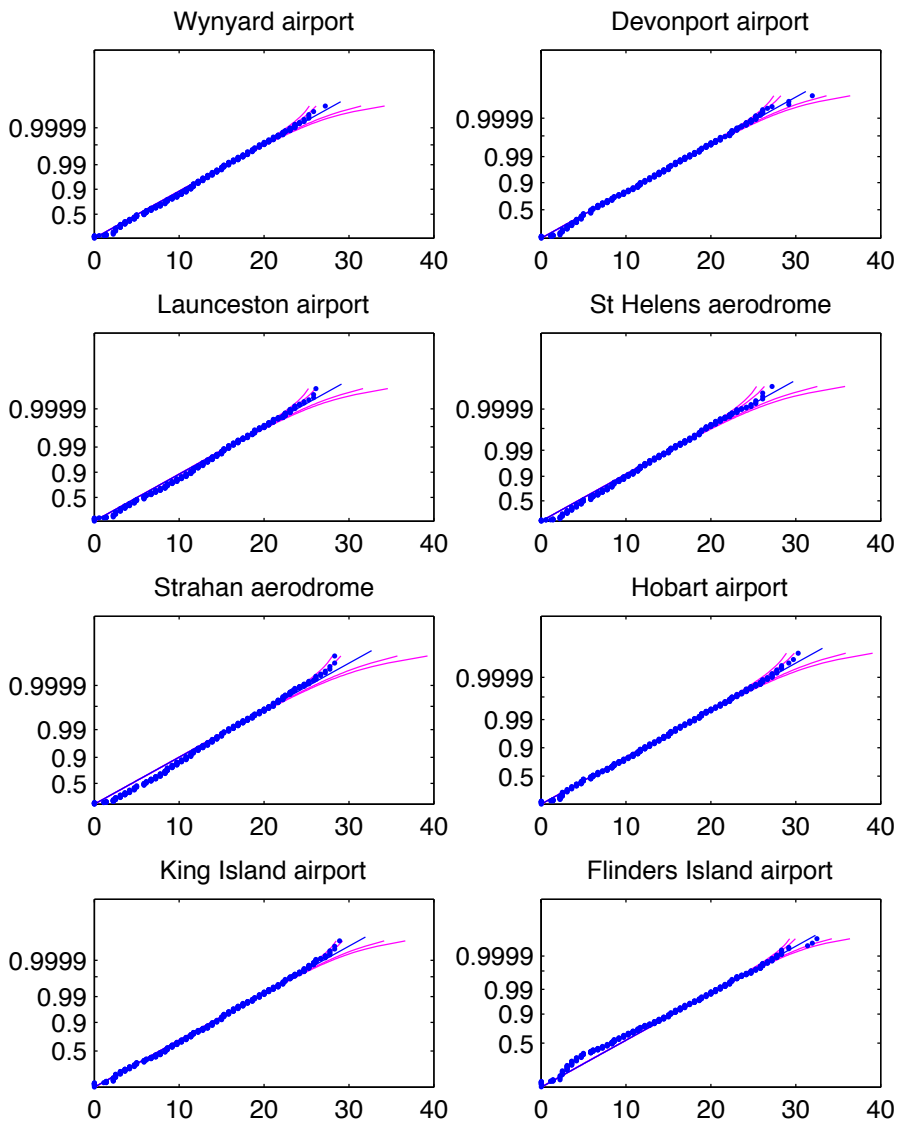


Figure 4: Weibull distribution plots, fitted to the upper 1% of the data, of gust speed at airport locations, with distribution median in blue, and the 0.005, 0.05, 0.95 and 0.995 quantiles in magenta.

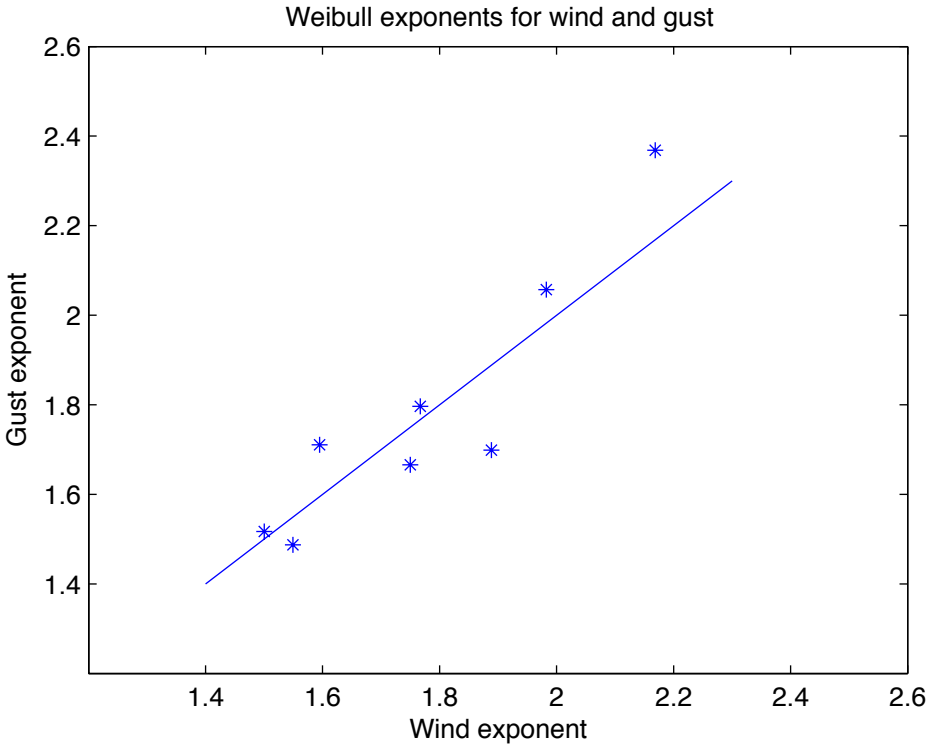


Figure 5: Plot of Weibull gust exponent versus Weibull wind exponent. Blue line is the Gust exponent equal to the Wind exponent.

Assuming the exponent from the Weibull fit to the wind data can be used for the gust distribution the scale factors are compared. It is found that the scale factor raised to the power of the Weibull exponent form a straight line as seen in Figure 6. The scale values calculated using the full distribution gave a similar straight line. Using this line the estimate of the gust scale $s_i^{(g)}$ is

$$s_i^{(g)} = 1.76^{1/a_i} s_i^{(w)}. \quad (2)$$

Now assuming the Weibull exponent of the wind can be used as the exponent

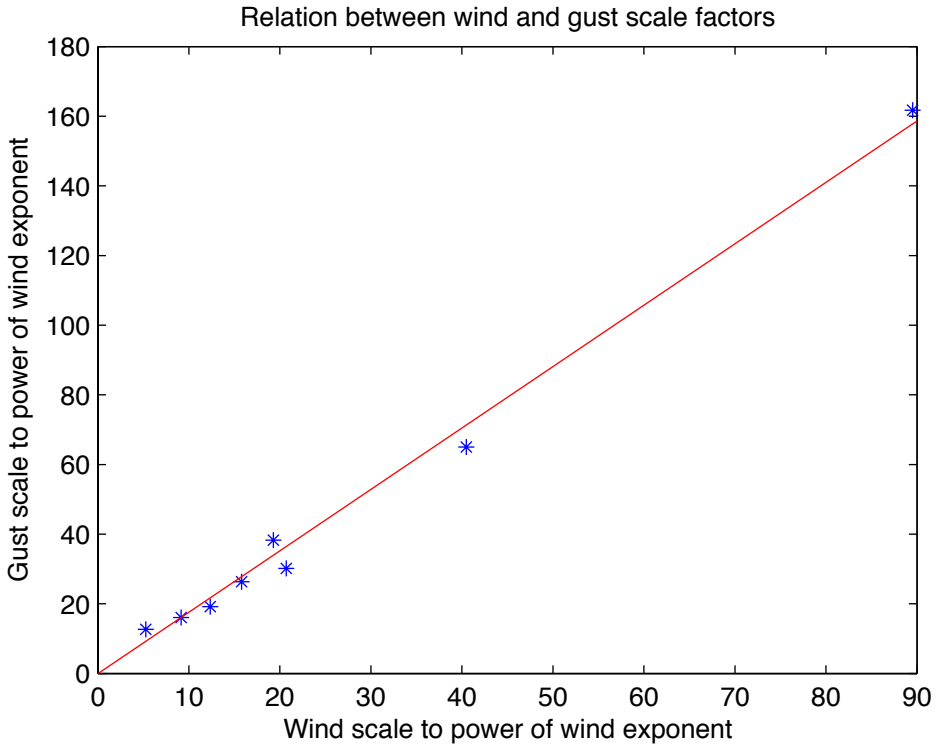


Figure 6: Plot of wind and gust Weibull scale values.

for the gusts (Figure 5) and using the relations given in equation (2) and Figure 6 the gust distribution is predicted from the wind data as seen in Figure 7. With the exception of the St Helens aerodrome which is predicted about 15% too low, the other seven airports have useable projections. Possibly a non-linearity could be introduced into Figure 6 to improve the projection at the low value corresponding to St Helens.

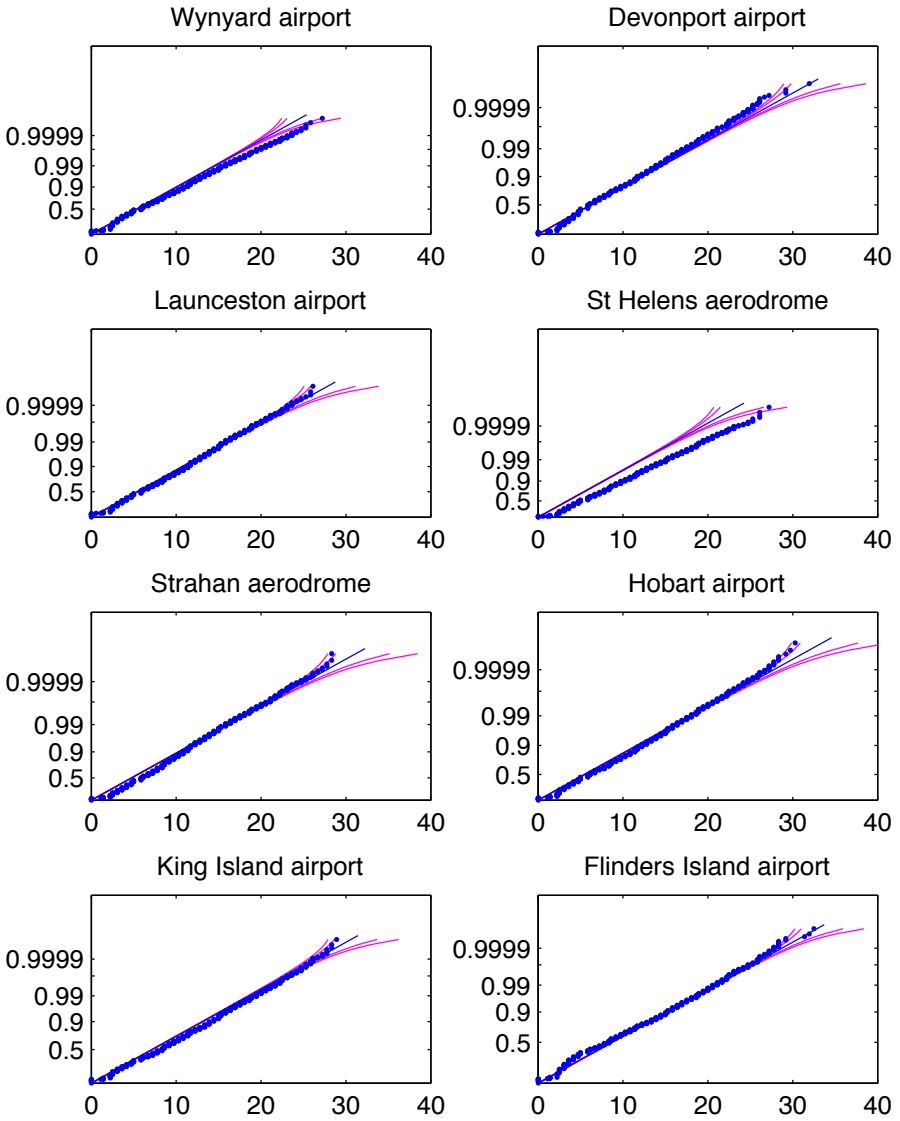


Figure 7: Gust distribution data with estimated gust distribution (blue line, and 0.005, 0.05, 0.95 and 0.995 quantiles in magenta) using wind data at same site.

2.2 Extrapolation on return period scale

The “return period” is defined as $1/(1-p)$ where p is the cumulative probability of events up to that size occurring in a given time. This does not say the extreme event will occur once in each return period, but only that the long term average will be one occurrence per return period. For a given return period of time: zero, one, or more of the extreme events can occur. As noted in the introduction, the estimation of values corresponding to large return periods, such as 500 years, is a significant extrapolation from existing data.

The return period graphs are the same as the cumulative probability graphs with the probability scales relabelled as return periods, but by convention the return period scale is plotted on the horizontal axis. The return period plots are often plotted using a logarithmic scale for the return period axis instead of a probability scale. The logarithmic return period scale generally eliminates the possibility of the distribution having a simple straight line extrapolation as occurs in probability based plots.

For the Hobart gust data the return period plot is given in Figure 8. Estimates of the gust size for a return period of 500 years are read from these graphs. The quantile lines show the fit to the data is as good as could be expected with the extreme data values being within the middle 90% quantile range.

The return period is often calculated from data giving the maximum observed value over a given period. The gust data, which is 0.5 hour samples, is converted to daily maxima (of 48 samples) and for demonstration 48 day maxima. The maximum values are sorted and assigned probability values corresponding to the median of the corresponding order statistic. For the n th ordered value of total m uniform values this is (Kendall and Stuart, 1963; Wikipedia, 2012a)

$$\beta_{0.5}^{-1}(n, m + 1 - n), \quad (3)$$

where $\beta_x^{-1}(n, m + 1 - n)$ is the inverse incomplete beta function (Wikipedia, 2012b). The median is used for the probability plots, as the median, unlike the mean, remains the median when converted to another probability scale.

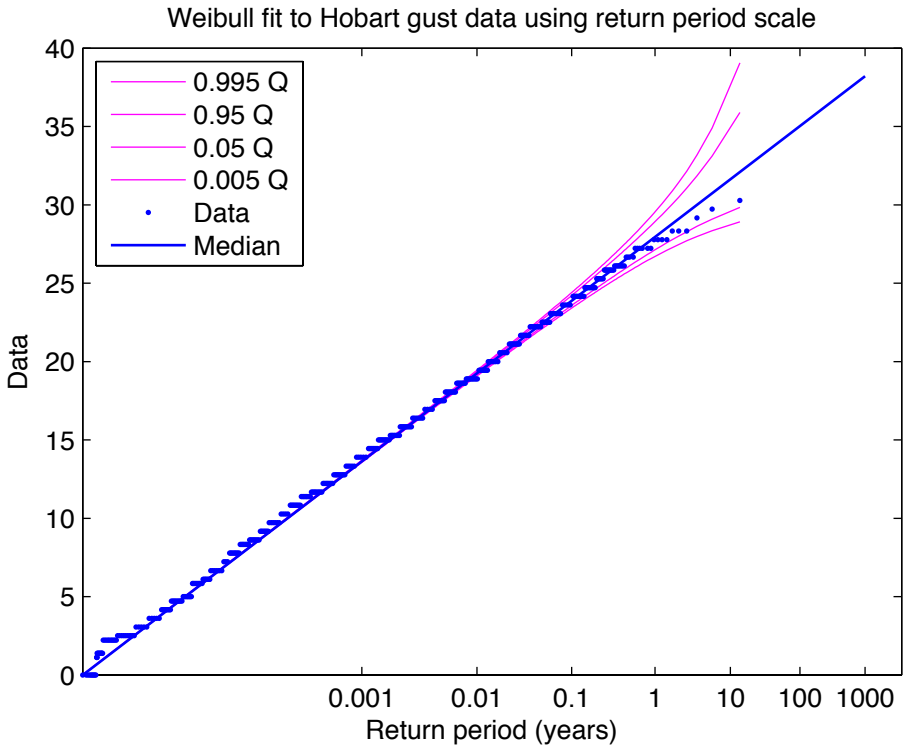


Figure 8: Return period plot of gust data for Hobart airport. Upper 1% of data fitted to the Weibull distribution. The lack of fit for data less than 3 is due to the data having been rounded.

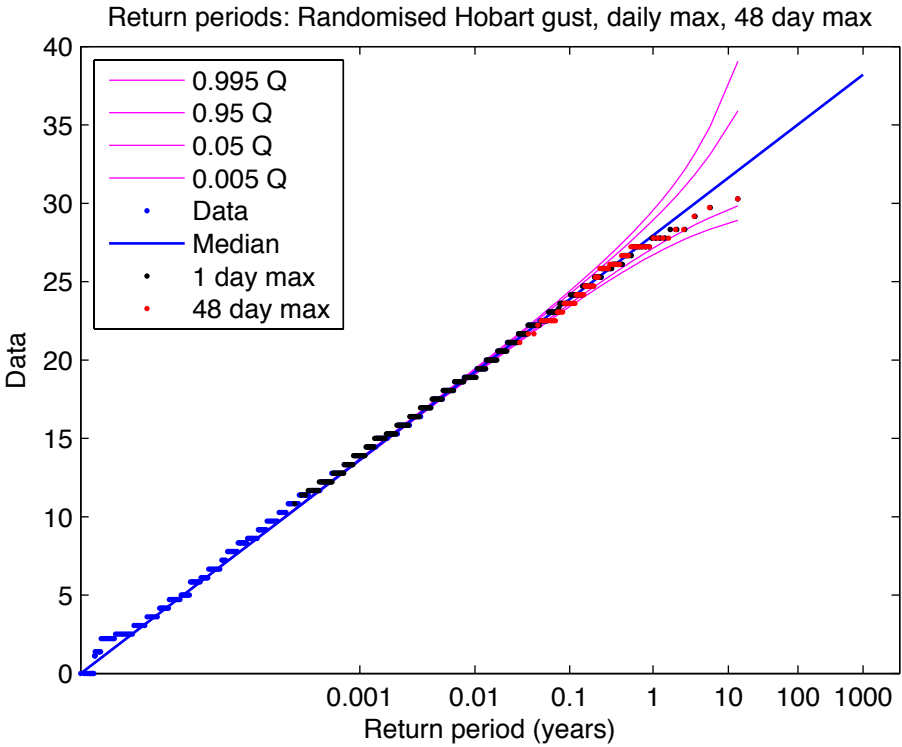


Figure 9: Hobart gust data showing daily and 48 day maximums using randomised order of original data.

These probabilities are raised to the power $1/48$ and $1/48^2$ to account for the number of values over which the maximum was taken. To compare with the effect of autocorrelation Figure 9 shows the Hobart gust data, daily maxima, and 48 day maxima using randomised order of the original gust data to remove autocorrelations. The daily data (black) lies on top of the original values, and the 48 day maxima (red) lies on top of both the original data and the daily data.

When the data order is randomised the maximum values supply similar

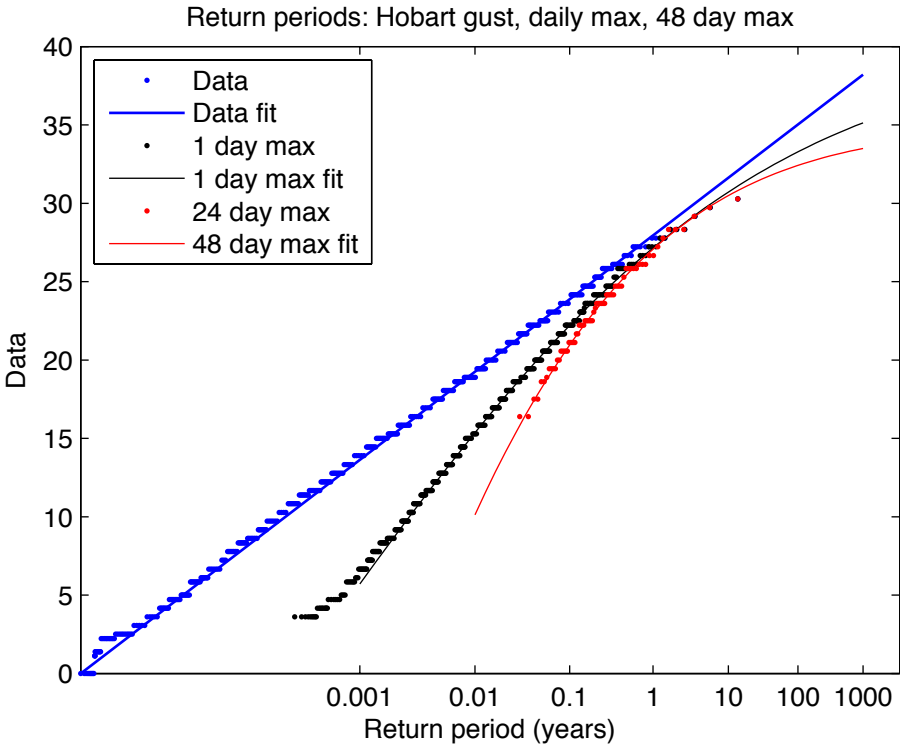


Figure 10: Hobart gust data showing daily and 48 day maximums using original data.

information to the original values, except that the amount of data available for analysis is reduced. Figure 9 shows that with uncorrelated data the original data provides the same information, as the maximum values, but provides more data points for use in the analysis. However, in many cases the calculation of extreme maxima is done using data giving the maximum seen in given periods as this is the only data available.

Figure 10 shows the daily and 48 day maximum data using the data as recorded (the order not randomised). Here due to autocorrelation in the data

the curves diverge but at the high end, as the auto correlation effects reduce, tend to the same values as the original data. The final values of the three data sets coincide. This plot also shows fits to the distributions with the daily and 48 day maximum data being fitted, to all the data, by the cumulative generalised extreme value distribution (Coles, 2001; Wikipedia, 2012d)

$$P(x) = \exp \left\{ - \left[1 + \alpha \left(\frac{x - u}{s} \right) \right]^{-1/\alpha} \right\} . \quad (4)$$

Parameters for this distribution are given in Table 4. In this distribution a negative values of α , as obtained, gives an upper limit to the distribution.

The return period plot can also be drawn with the horizontal axis as a log scale instead of the probability scale used above. This no longer presents the original data as a straight line making it difficult to judge the quality of the extrapolation. Figure 11 shows the original wind data for each of the eight airports (blue), the daily maximums (black), and the 48 day maximums (red). The solid lines are the fits to the data: Weibull for the original data; and generalised extreme value for the two maximum curves. The original data shows less curvature than the daily and 48 day maximum data which are thus harder to extrapolate. Figure 12 presents the gust data in a similar manner to Figure 11.

In both Figures 11 and 12 some of the fits to the data have followed the main part of the data and give less accurate fits to the upper part and thus a less reliable extrapolation, thus extrapolation to 500 years gives different values for the original, daily and 48 day maximum data. When there is a good fit over the full range of the data, the three curves give similar projections. However, the generalised extreme distribution has a finite upper limit, so for more extreme projections the Weibull distribution will predict higher speeds.

A Weibull distribution gives an adequate fit to the half hour wind and gust data. These two distributions are related, giving a means to convert from wind distributions to gust distributions. The 1 day and 48 day maximums are fitted with the generalised extreme distribution, but as the maximums

Table 4: Parameter for fit to generalised extreme distribution, equation (4), using all the data, for airport wind and gust, daily and 48 day maxima

Location	Daily maxima			48 day maxima		
	a Shape	u Offset	s Scale	a Shape	u Offset	s Scale
Wind						
Wynyard	-0.193	7.27	2.31	-0.196	13.47	1.46
Devonport	-0.129	8.10	2.82	-0.107	15.93	1.86
Launceston	-0.260	7.74	2.50	-0.120	13.84	1.52
St Helens	-0.141	5.91	1.60	-0.094	10.45	1.31
Strahan	-0.124	7.32	2.25	-0.188	14.49	1.94
Hobart	-0.105	7.66	2.59	-0.173	15.22	2.21
King Island	-0.129	8.81	2.57	-0.240	16.18	2.11
Flinders Island	-0.150	9.17	3.57	-0.177	18.51	1.94
Gust						
Wynyard	-0.173	10.32	3.46	-0.124	20.10	2.22
Devonport	-0.115	10.37	3.86	-0.119	21.24	2.55
Launceston	-0.219	10.22	3.43	-0.092	19.54	2.20
St Helens	-0.104	9.57	3.11	-0.078	18.84	1.98
Strahan	-0.063	9.97	3.20	-0.157	20.56	2.48
Hobart	-0.084	10.55	3.87	-0.246	21.89	3.25
King Island	-0.144	12.04	3.75	-0.133	22.28	2.37
Flinders Island	-0.190	12.31	4.52	-0.206	23.98	2.54

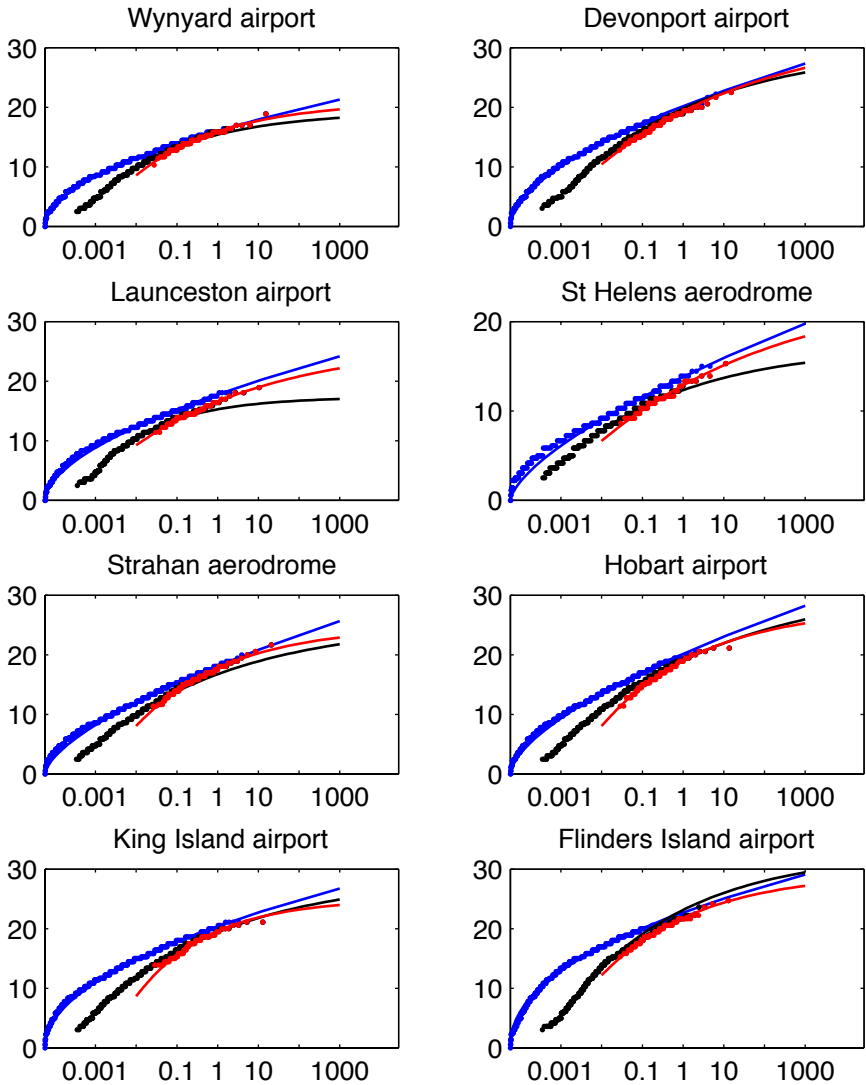


Figure 11: Return period plots for wind data with years on the horizontal axis and metre/second on the vertical axis. Original 0.5 hour data in blue, daily maximum values in black, and 48 day maximums in red. Solid lines give corresponding fitted distribution and their extrapolation out to 1000 years.

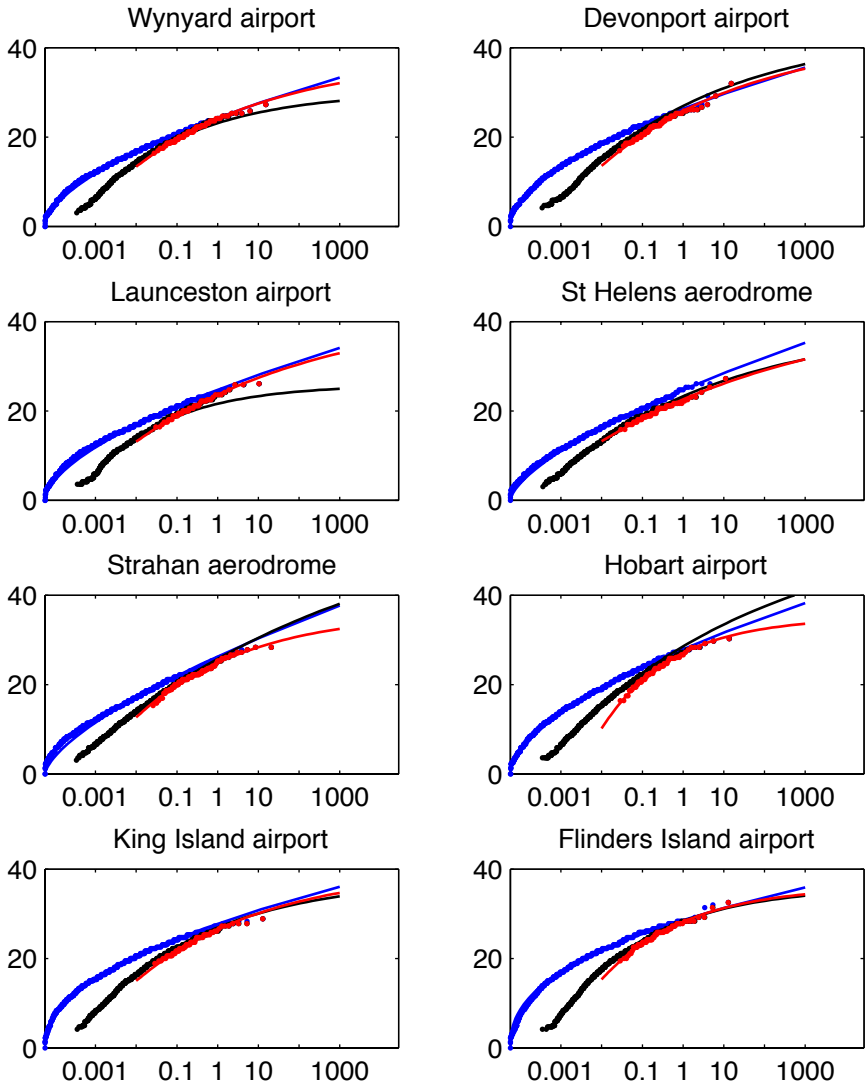


Figure 12: Return period plots for gust data with years on the horizontal axis and metre/second on the vertical axis. Original 0.5 hour data in blue, daily maximum values in black, and 48 day maximums in red. Solid lines give corresponding fitted distribution and their extrapolation out to 1000 years.

Table 5: Driver models used for the finer grid data.

Mk3.5	CSIRO (Australia)
ECHAM5	Max-Planck Institut (Germany)
GFDL2.0	Princeton University/NOAA (USA)
GFDL2.1	Princeton University/NOAA (USA)
MIROC3	Centre for Climate System Research (Japan)
UKhad	Hadley Centre Met Office (UK)

tend to the same values as the original half hour data and the generalised extreme distribution assumes a fixed upper bound, the Weibull distribution may be preferred for the extrapolate to the 500 year point.

3 Simulated wind data

Simulated wind data for 1961 to 1990 and for 2081 to 2100 were obtained from the Climate Futures for Tasmania project ([Climate futures, 2012](#)). The data was generated by CSIRO's Conformal Cubic Atmospheric Model using a grid of 0.1 by 0.1 degrees (approximately 14 kms) using global simulations from six different global circulation models listed in Table 5, which provided a coarse scale input (approximately 200–300 kms) to drive the fine scale simulation. Each of the global simulations used the Intergovernmental Panel of Climate Change scenario A2 ([IPCC Working Group III, 2000](#)). Results from the six simulations were recorded at three hourly intervals and provided to this project as daily maximum wind speeds at each grid location. To avoid confusion with additional results these simulated results will be referred to as data. Unlike the airport data, this data provides information covering all of Tasmania, and hence provides the possibility of extending 500 year gust projections to all of Tasmania.

Figure 13 compares the distribution of all the wind speeds over the period

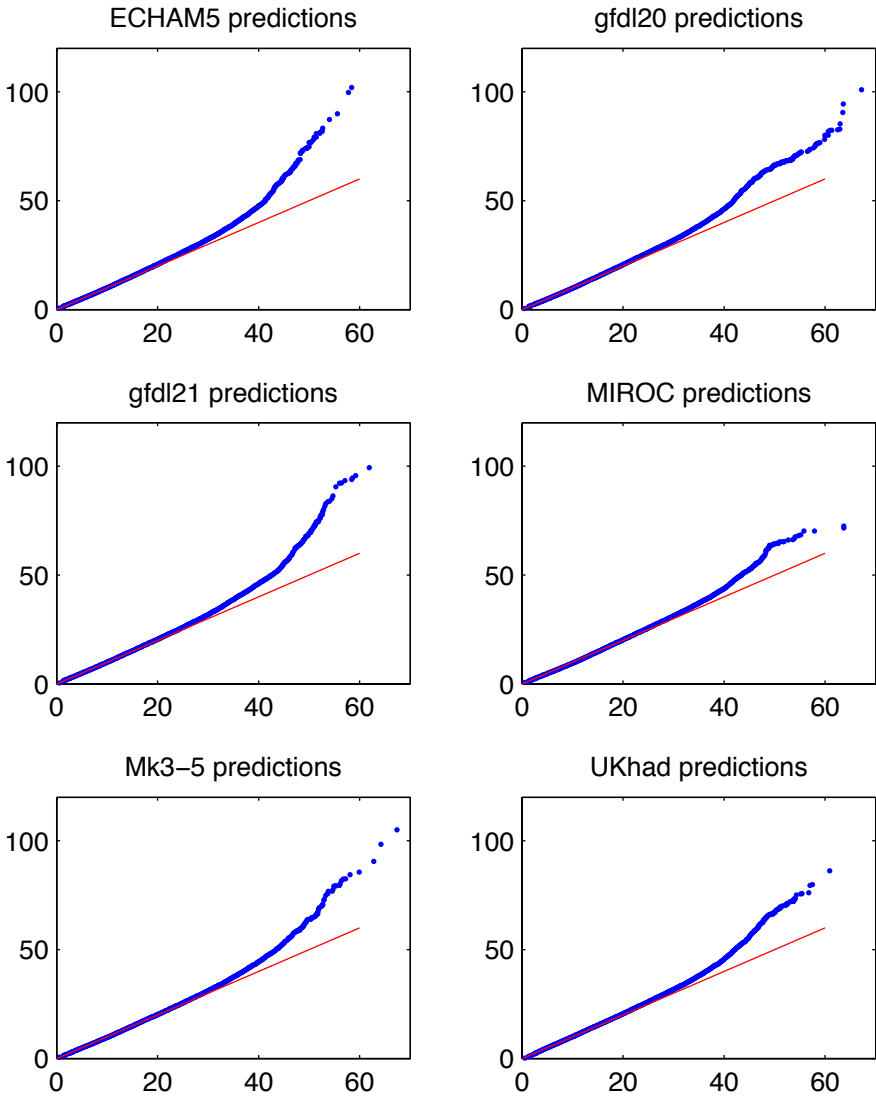


Figure 13: Comparison of maximum daily wind speed distributions 1961–1980 (horizontal axis) and 2081–2100 (vertical axis). Red line is 2081–2100 distribution=1961–1980 distribution. Note only a very small proportion of the data exceeds 40 m/s, see Table 6.

Table 6: Comparison of 1961–1980 data and 2081–2100 data. Each simulation has 20 848 800 data values and 2856 grid points.

Driver	% 1961–1990	Points > 40m/s		Grids>40m/s	
	>2081–2100	1961–1980	2081–2100	1961–1980	2081–2100
Mk3.5	45	714	3807	58	999
ECHAM5	64	1016	3730	110	647
GFDL2.0	48	1418	5424	138	1434
GFDL2.1	88	1011	2574	99	628
MIROC3	86	889	2462	84	680
UKhad	35	1130	3546	131	730

2081–2100 with those over 1961–1980. Equal data sizes made the quantile–quantile plots easy to generate. For each of the six cases the distributions are similar over most of the distribution with the 2081–2100 data being marginally less than the 1961–1980 over most of the range; however, the 2081–2100 extreme values are higher. There are 20 848 800 data values and 2856 grid points for each of the simulations. Table 6 shows the proportion of 1961–1980 data that are larger than the corresponding quantile 2081–2100 values, the number of extreme values (> 40 m/s) in each data set, and the number of grid points with extreme values in each data set. There is only a very small proportion of high wind speed points, but the 2081–2100 data contains considerably more than the earlier data. Similarly the number of grid points with high wind speeds is small in the 1961–1980 data but larger in the future data.

3.1 Simulated data at airport locations

The nearest grid point in the simulated data, to each airport was located, and to reduce the analysis, the six sets of simulated data combined. Figure 14

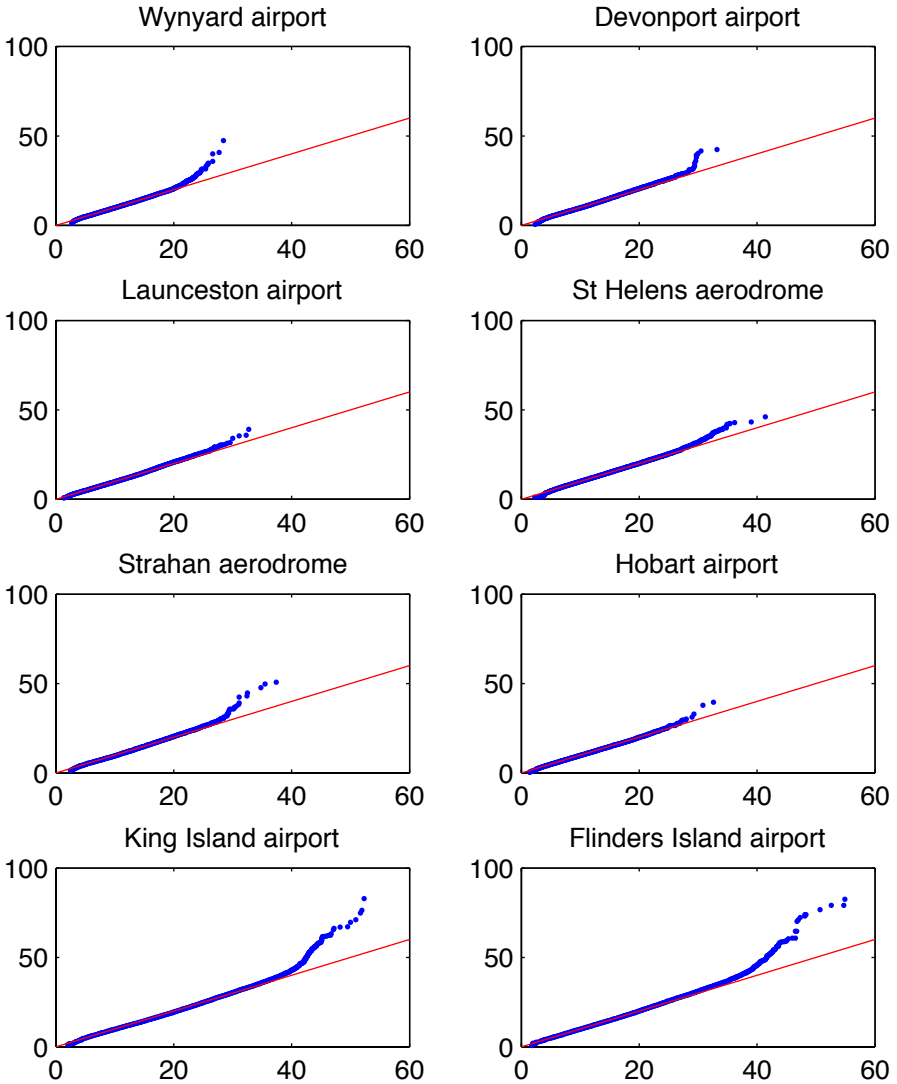


Figure 14: Relation between 1961–1980 data (horizontal axis) and 2080–2100 data (vertical axis) at the eight airports. Data from the six simulations have been combined.

Table 7: Parameter values for simulated data at airports.

Location	1961–1980		2081–2100	
	Scale s	Shape a	Scale s	Shape a
Wynyard airport	6.82	2.99	7.44	2.47
Devonport airport	8.15	2.03	8.83	1.71
Launceston airport	8.21	2.07	8.85	1.75
St Helens aerodrome	10.22	2.26	10.82	2.03
Strahan aerodrome	9.02	2.07	9.88	1.69
Hobart airport	7.72	2.28	7.76	2.14
King Island airport	15.23	1.77	16.99	1.40
Flinders Island airport	15.16	1.65	17.36	1.30

compares the 2081–2100 data with the 1961–1980 data at the airport locations. Again over most of the range the distributions are very close and the 2080–2100 data have a few higher values at the upper extreme of the distributions. In some cases there are very few offset extreme points and these differ by only a small amount. The two island airports in Bass Straight show a large difference in the extreme points.

The simulated airport data was found to fit a cumulative gamma distribution ([Kendall and Stuart, 1963](#); [Wikipedia, 2012c](#))

$$p(x) = \Gamma_{x/s}(a) = \frac{\int_0^{x/s} x^{a-1} \exp(-x) dx}{\int_0^\infty x^{a-1} \exp(-x) dx} \quad (5)$$

on the square of the data as seen for the 1961–1980 data in [Figure 15](#). Generally satisfactory fits were obtained. Distribution plots for 2081–2100 are plotted in [Figure 16](#). These distribution fits are reasonable over most of the range, but unlike the 1961–1980 data deviate to higher values at the extreme top of the distribution. This deviation at the upper extreme makes projection of extreme values difficult.

[Table 7](#) gives the parameters of the fit to the simulated data at the airport

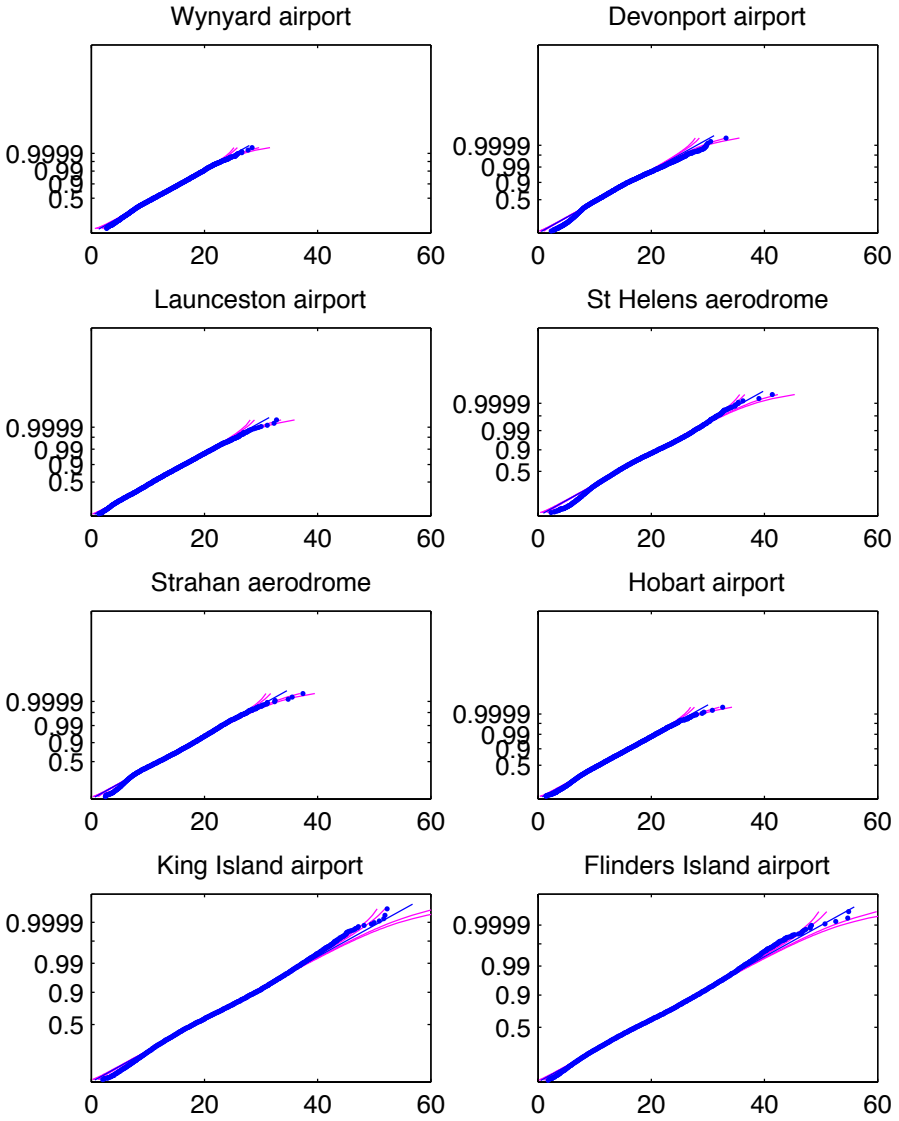


Figure 15: Distribution plots for 1961–1980 data at the airport sites.

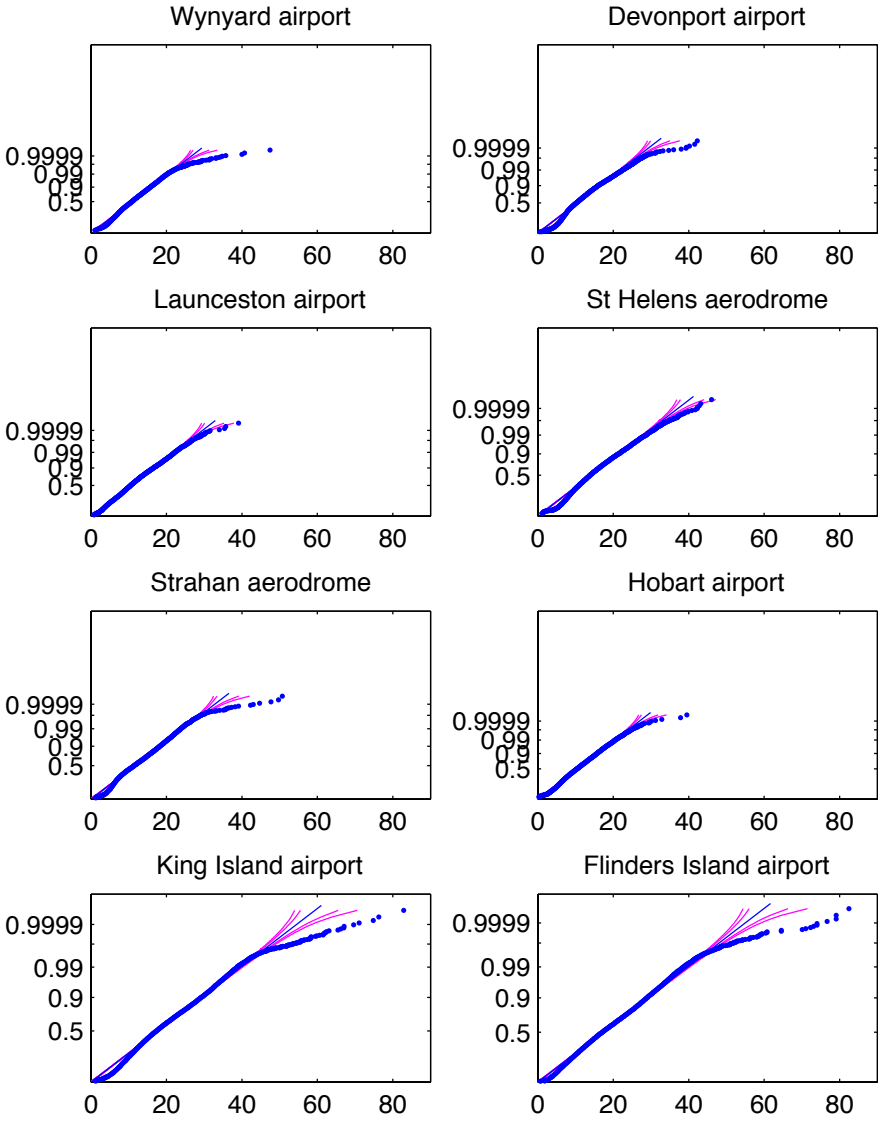


Figure 16: Distribution plots for 2081–2100 data at the airport sites.

locations. There appears to be a relation between the parameters for 1961–1990 and the parameters for 2081–2100 and this could be calculated for all 2856 grid points; however, the lack of fit to the extreme points in the 2081–2100 data makes these relations less useful for predicting extreme behaviour.

A quantile-quantile relation between the 1961–1990 and the 2081–2100 data can be calculated as in Figure 14. A plausible assumption is the ratio from these plots can be used to convert quantiles of wind or gust speeds from current data to projections of future speed quantiles. For wind distribution with speed $v^{(o)}(\mathbf{p})$ at fraction quantile \mathbf{p} in 1961–1980 time the corresponding speed $v^{(n)}(\mathbf{p})$ in 2081–2100 is estimated as

$$v^{(n)}(\mathbf{p}) = v^{(o)}(\mathbf{p}) w^{(n)}(\mathbf{p})/w^{(o)}(\mathbf{p}), \quad (6)$$

where $w^{(n)}(\mathbf{p})$ and $w^{(o)}(\mathbf{p})$ are the corresponding quantile values from the quantile–quantile plot, such as Figure 14.

In locations where there are few 2081–2100 outliers and the distribution are essentially the same, such as Launceston and Hobart, this should not be a problem. At other sites this conversion can also be done, but the extreme tail values of the 2081–2100 data are a source of concern, as most grid point have only a small number of these extreme points and in some cases the relation is not smooth.

4 Relation between airport data and simulated data

The airport data is only available at the eight airports, but the interest is in extreme gust values for the whole of Tasmania, so this section looks at the possibility of using the simulated data to predict the airport wind data. Such a projection could then be used to estimate the equivalent of airport wind data for the rest of the Tasmanian grid points and then the gust distributions.

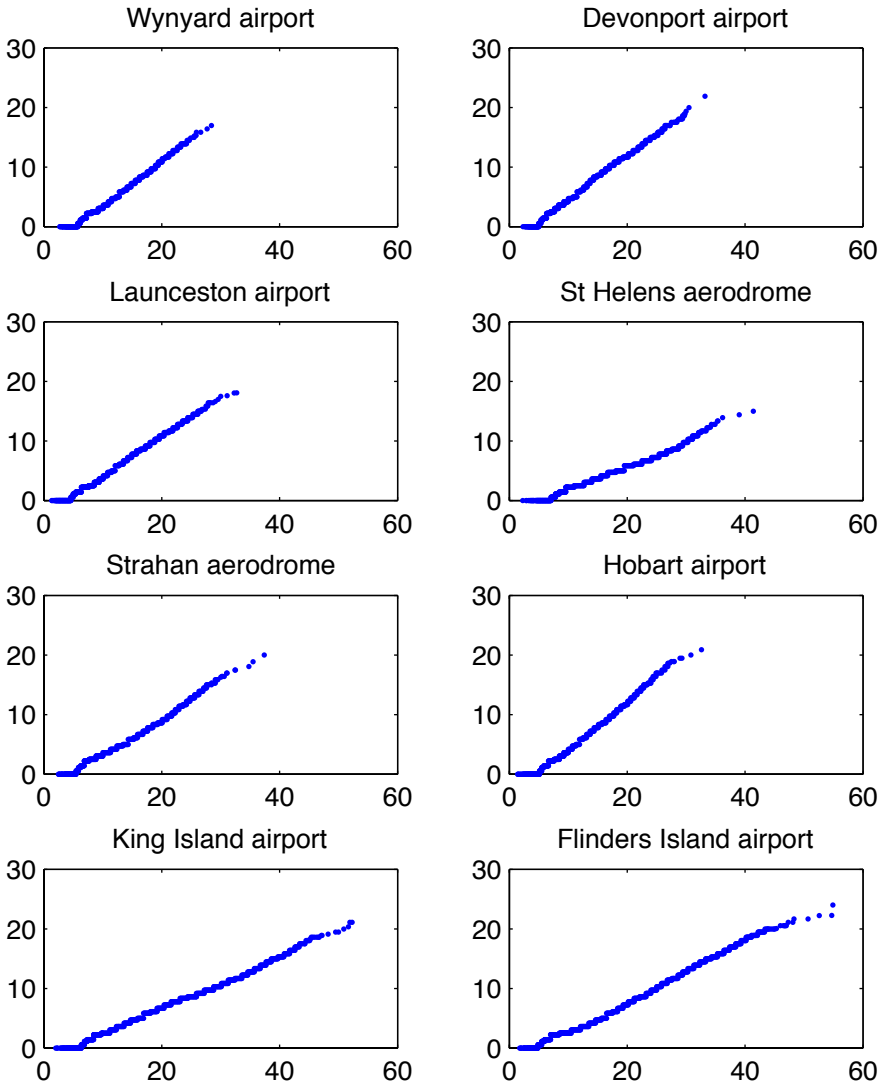


Figure 17: Plots of the distributions of simulated 1961–1990 data (horizontal axis) and airport wind data (vertical axis).

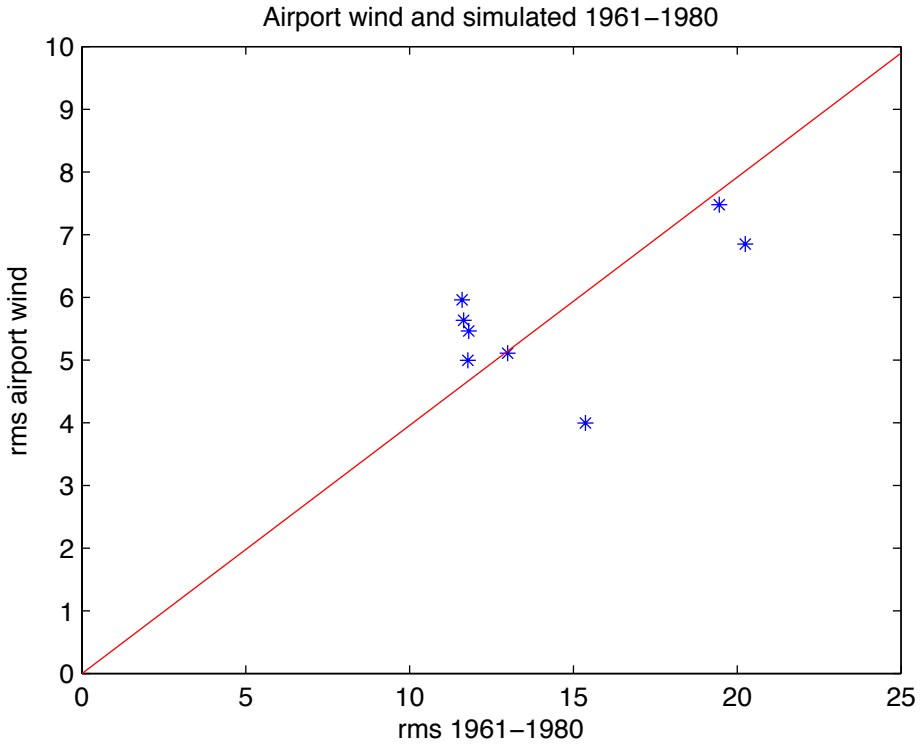


Figure 18: Comparison of root mean square values for airport data and simulated data. Red line is least squares fit through the origin, as a zero airport data is the expected prediction from a zero simulated value.

Figure 17 plots the distributions of the measured data at the airports against the distribution of the simulated data close to the airport site. These graphs show some non-linearity and a range of different slopes. The low region of the graphs shows the airport data, due to rounding, contains many more zero values than the simulated data.

A simple relation between the airport data and the simulated data compares the root mean squares values as in Figure 18. This unfortunately shows

Table 8: Gamma distribution parameters fitted to airport wind data 24 hour maximums.

Location	Scale s	Shape a
Wynyard airport	5.00	2.95
Devonport airport	6.42	2.38
Launceston airport	5.11	3.12
St Helens aerodrome	3.73	3.38
Strahan aerodrome	5.52	2.52
Hobart airport	6.35	2.19
King Island airport	6.05	2.95
Flinders Island airport	7.67	2.20

no trend, and the best that might be derived is a straight line through the origin, with slope 0.38, as shown on the figure. Eliminating the two high values produces a line perpendicular to that expected; such a relation is not considered useful, as the line should go through the origin. This regression explains 80% of the variance.

The above compares 24 hour maximum wind data with 0.5 hour values recorded at airports. To make a possibly more valid comparison, the airport wind data is converted to 24 hour maximum values and a gamma distribution is fitted to the square of these values. Figure 19 shows these fits, which are considered to be good except at very low wind speeds. Table 8 gives the fitted gamma distribution parameters.

Plots showed no relation between the airport wind parameters and those of the 1960–1990 data. To test if a relation can be found for the scale parameter of the 24 hour maximum airport data, the scale and shape parameters of the simulated data are used to regress the scale of the 24 hour airport data. Figure 20 shows the results of this regression. The two predictors explained only 30% of the variance and replacing the predictors by random values it was found 40% of random variables cases gave a more accurate projection

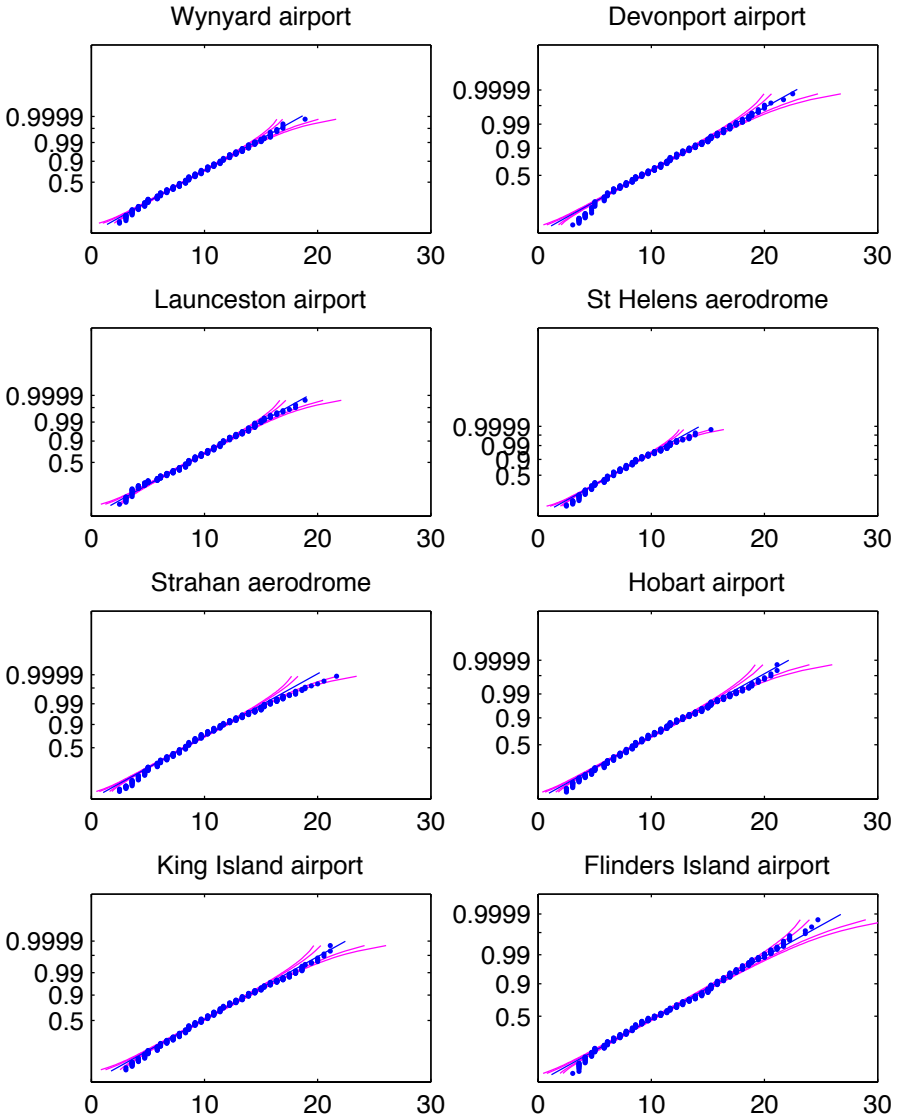


Figure 19: Fits of maximum 24 hour value for airport data.

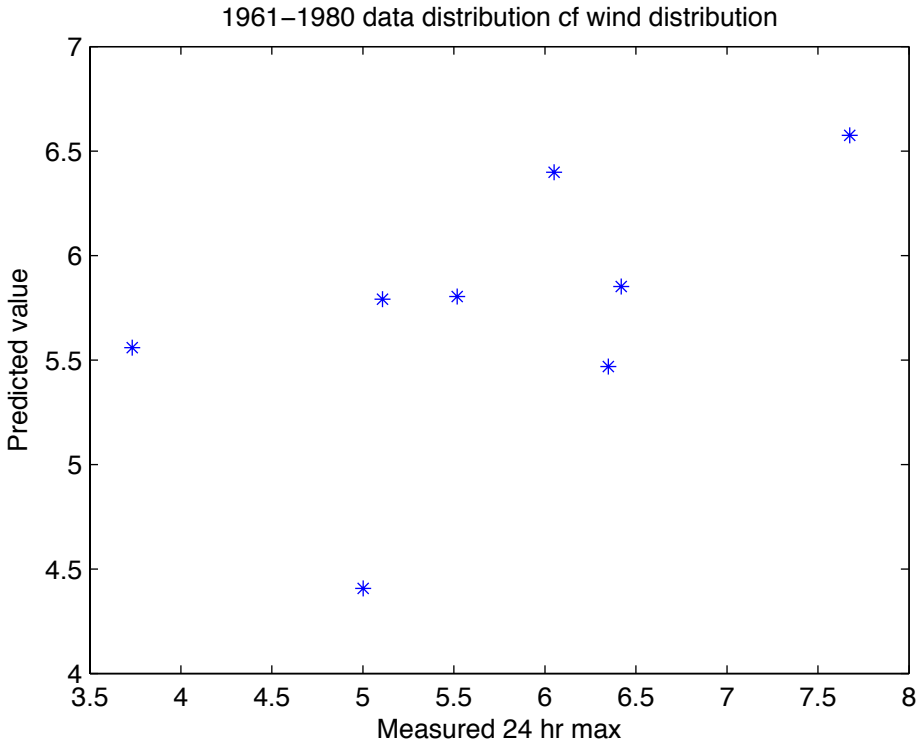


Figure 20: Regression fit to airport 24 hour maximum scale parameter using simulated data parameters—indicating no useful relation.

and thus the slope coefficients of the two variables are not significant. We conclude there is no useful relation of this form relating the airport data to the simulated data.

Figure 21 shows the 24 hour maximum wind and gust speeds, with the simulated data which is also 24 hour maxima. In some cases there is agreement between the 1961–1980 data and the gust speed (Wynyard, Devonport, Launceston, and Strahan), For the Bass Straight islands the simulated data is much higher than the airport data, and the extremes of the 2081–2100 data are

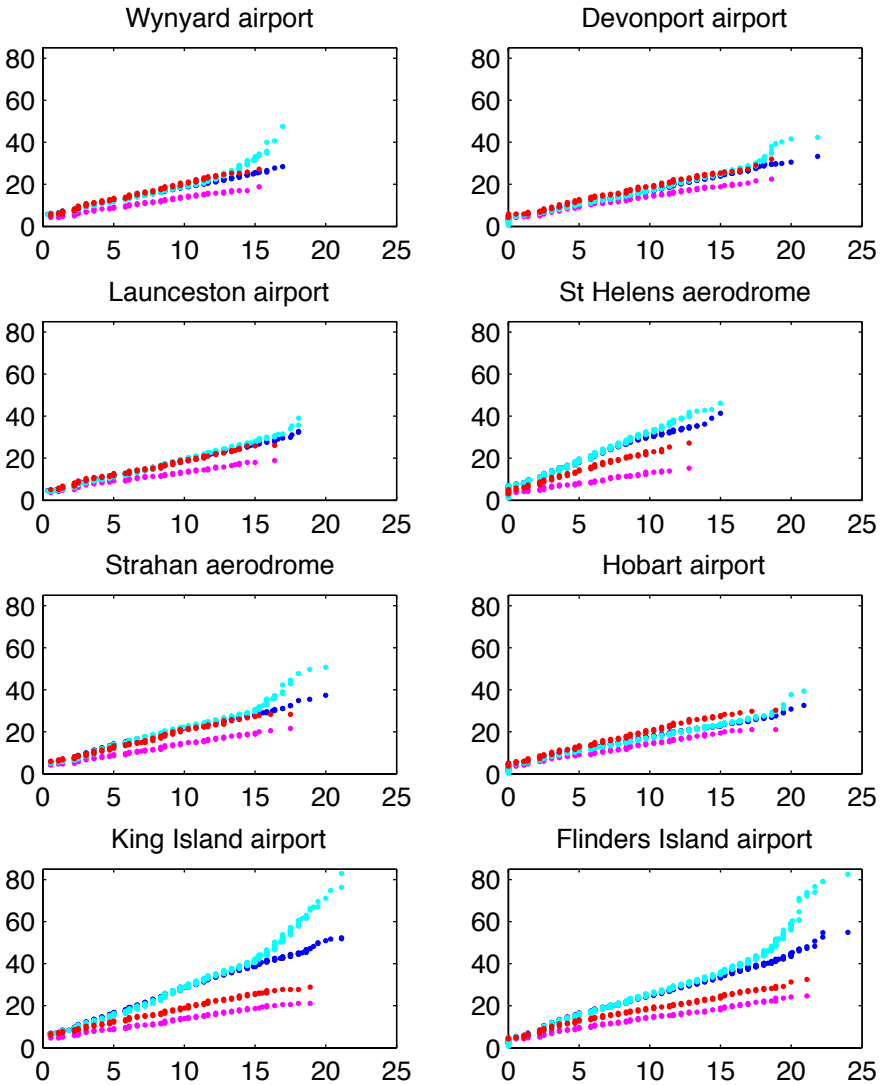


Figure 21: Comparison of 24 hour maximum wind (magenta), 24 hour maximum gust (red), 1961–1980 simulated data (blue), and 2081–2100 simulated data (cyan), plotted as a quantile quantile plot against wind speed (horizontal axis).

Table 9: Mean and standard deviation of squared data for fits in Figure 22.

	Mean	SD
Wynyard airport	30.9	25.7
Devonport airport	20.1	22.0
Launceston airport	6.9	7.4
St Helens aerodrome	14.4	12.6
Strahan aerodrome	44.3	41.5
Hobart airport	13.1	13.8
King Island airport	35.3	34.8
Flinders Island airport	26.6	27.6

more than three times those of the measured wind speed.

5 Three hourly samples of simulated data

One set of simulated data was obtained, giving samples from the simulation at three hourly intervals, rather than daily maxima, for the period 1961–1990 using the ECHAM5 simulation as the base. Extrapolating from the original three hour data was compared with extrapolating from maxima over different periods.

Figure 22 shows fits to this data, at the airport locations, using an exponential distribution on the square of the data. In some cases the data wanders away from the distribution and extrapolation does not look accurate. Table 9 gives the mean and standard deviation of the squared data. The mean is the scale parameter for the exponential distribution and for an exponential distribution it is expected that the standard deviation will be close to the mean.

Figure 23 uses only the upper 7600 of the 87 600 data values to fit the exponential to the square of the data. As expected this gives a better fit to

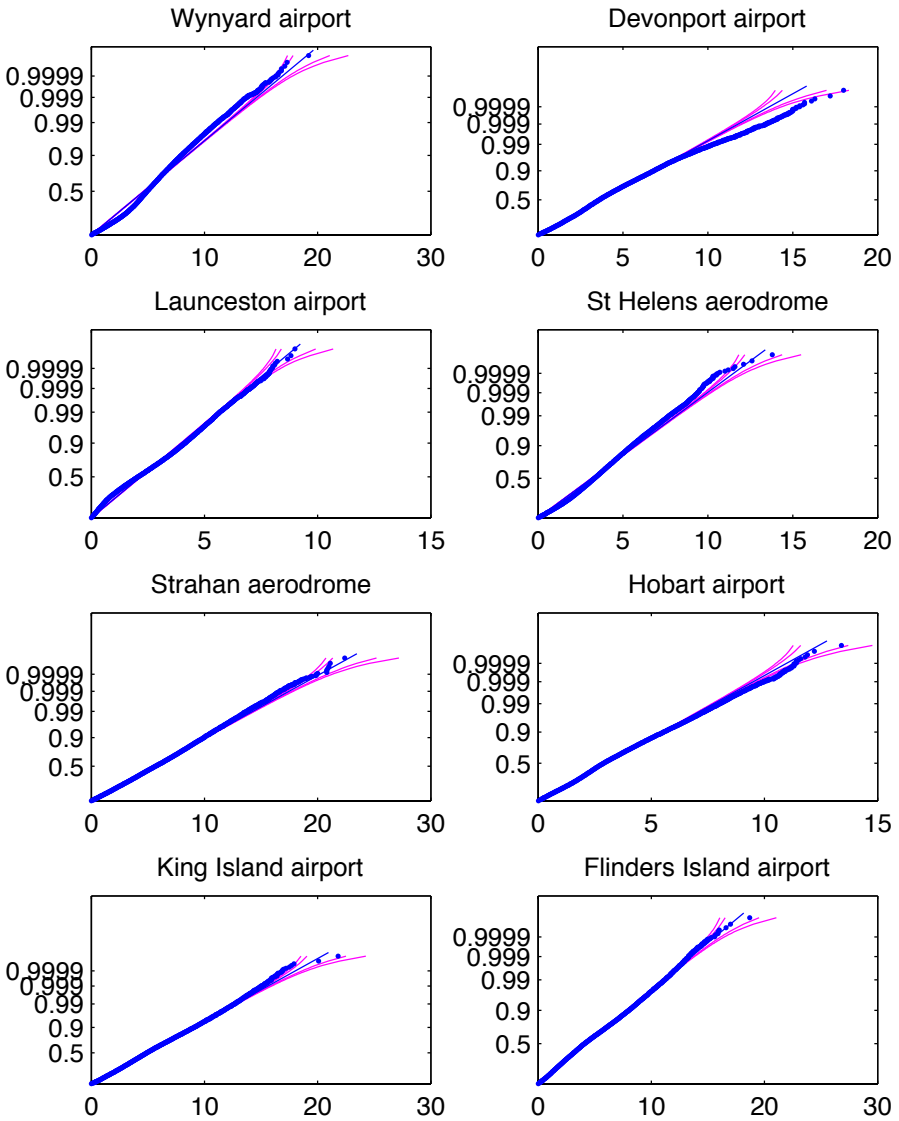


Figure 22: Exponential distribution fitted to square of wind speed using three hour simulated data.

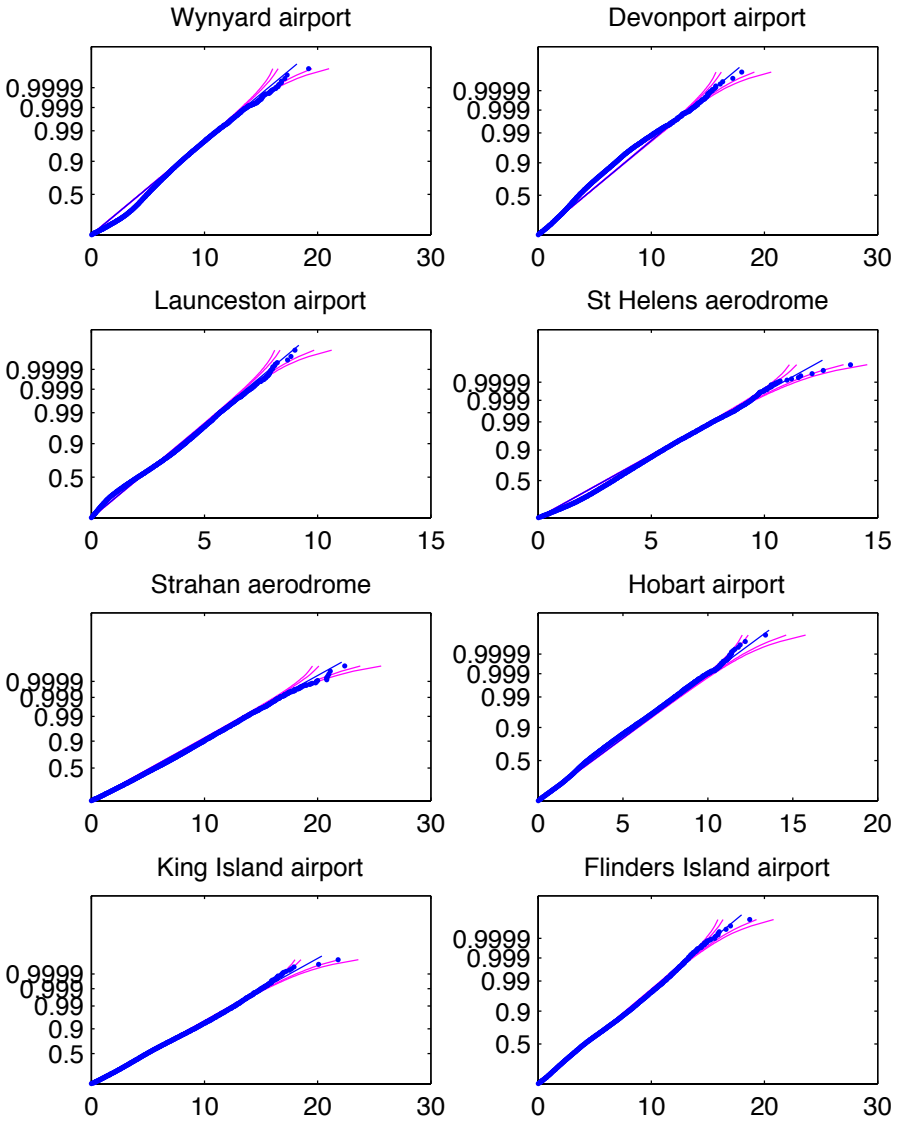


Figure 23: Exponential distribution fitted to upper values of square of wind speed using three hour simulated data.

Table 10: Exponential distribution mean of squared data for fit to upper part of distribution, used for plots in Figure 23.

	Mean
Wynyard airport	26.5
Devonport airport	25.5
Launceston airport	6.8
St Helens aerodrome	12.7
Strahan aerodrome	39.3
Hobart airport	14.9
King Island airport	33.4
Flinders Island airport	26.0

the upper parts of the data. Table 10 gives the parameter values used for this case.

The three hourly simulated data was converted to maxima over a day and 48 days and fitted to a gamma distribution (equation (5)). Table 11 gives the gamma distribution parameters. Although the fit was considered quite good, the values of the parameters are quite extreme. The same maxima data were also fitting to the generalised extreme distribution, equation (4), with the parameters given in Table 12. Figure 24 gives return period plots of the original simulated three hour data with the exponential fit to the data squared, and the daily and 48 day maxima fitted to the generalised extreme distribution. Comparing this with Figures 9, 10, 11 and 12 it appears that the simulated data is not as strongly correlated as the airport data. When the generalised extreme distribution fits the data well, extrapolation to 500 years return period is similar to the extrapolation from the original data. In the case the 48 day maxima five of the eight cases set the shape parameter positive giving curve turning upwards and poor extrapolation. The airport data, Figures 11, 12 and Table 4, gave only negative values for the shape parameter.

Table 11: Gamma parameters (equation (5)) for fits to daily maxima and 48 day maxima calculated from three hour ECHAM5 1961–1990 data.

	Daily maxima		48 day maxima	
	Scale s	Shape a	Scale s	Shape a
Wynyard airport	0.46	17.58	0.17	79.79
Devonport airport	0.77	8.65	0.15	85.82
Launceston airport	0.33	12.37	0.07	91.56
St Helens aerodrome	0.32	17.59	0.09	103.53
Strahan aerodrome	0.64	15.32	0.18	90.17
Hobart airport	0.50	10.79	0.15	60.23
King Island airport	0.63	14.15	0.12	123.65
Flinders Island airport	0.58	13.40	0.10	127.52

Table 12: Generalised extreme distribution parameters (equation 4) for values calculated from three hour ECHAM5 1961–1990 data.

Location	Daily maxima			48 day maxima		
	Shape	Offset	Scale	Shape	Offset	Scale
	a	u	s	a	u	s
Wynyard	−0.088	9.38	1.69	−0.111	14.82	14.82
Devonport	−0.163	8.06	2.26	0.028	14.17	14.17
Launceston	−0.239	4.88	0.95	−0.096	7.42	7.42
St Helens	−0.202	6.80	1.16	0.124	9.73	9.73
Strahan	−0.155	11.74	2.05	0.153	17.43	17.43
Hobart	−0.219	6.49	1.52	−0.222	10.26	10.26
King Island	−0.245	10.80	1.98	0.044	15.73	15.73
Flinders Island	−0.188	9.71	1.57	0.056	13.87	13.87

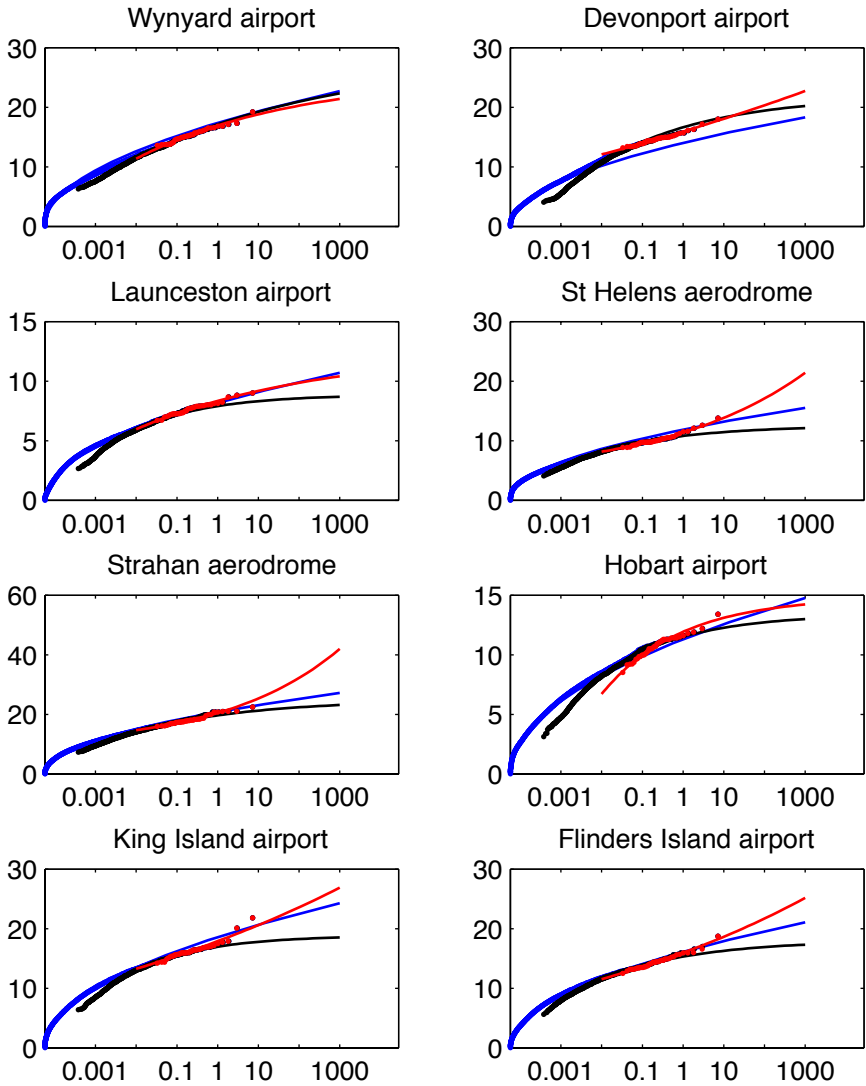


Figure 24: Exponential distributions fitted to upper values of square of wind speed using three hour simulated data (blue), and extreme values distribution fitted to maximum of 24 hour (black) and 48 day (red) periods Horizontal axis scale is return periods.

The problems fitting the generalised extreme distribution indicate the fit to the original data may be easier to work from. Fitting the exponential distribution to all the data or to the upper tail of data is particularly easy as only one parameter, the mean, is needed.

6 An actuarial approach

This approach is demonstrated using wind data from 2 Dec 1984 to 17 Jun 2002 recorded at the Hobart airport. It assumes an exponential distribution fit the upper tails of the data. For the cumulative distribution this is (Brown, 2012; Bühlmann, 1970)

$$P(x) = 1 - \alpha \exp[-(x - b)/s] \quad \text{for } x > b, \quad (7)$$

where α is determined by b as the fraction of the data greater than the threshold value b .

For instance using the daily maximum wind data for Hobart airport, taking the data above 12 m/s ($b = 12$) which gives $\alpha = 0.1158$ (of the total of 6407 values, 748 are above 12 m/s). This portion of the data gives mean 13.83, standard deviation 1.65, skewness 1.69, and kurtosis 3.79.

For the shifted exponential distribution, the scale value s is estimated by the standard deviation of the tail data, or by the mean minus the threshold which in this case is the larger, so taking $s = 1.83$. The probability value for 500 year return period point is

$$P(x) = 1 - 1/500/365.25 = 1 - 0.00000548. \quad (8)$$

Solving equation (7) for $(x - b)/s$ gives 9.96, and substituting $b = 12$ and $s = 1.83$ gives the 500 year wind estimate

$$x = 12 + 9.96 \times 1.83 = 30.23. \quad (9)$$

To check the sensitivity of this estimate to the threshold, an alternative threshold value is used. Using a threshold value of $b = 12.2$ gives $\alpha = 0.1158$ as due rounding of the data the number of points above 12.2 is the same as above 12. However, in this case the standard deviation gives the larger estimate for s of 1.65. Hence the estimate of the 500 year return period wind speed is

$$x = 12.2 + 9.96 \times 1.65 = 28.14. \quad (10)$$

To see the sensitivity of the estimate to the data, an extra value equal to the maximum data value is added to the data set. With the threshold value of $b = 12.2$ this gives mean 13.84, standard deviation 1.68, skewness 1.76 and kurtosis 4.23.

The estimate of the 500 year return value is now

$$x = 12.2 + 9.96 \times 1.84 = 30.53. \quad (11)$$

As the standard deviation is less than the mean minus the threshold, and the skewness and kurtosis are both less than the theoretical values of two and six, the estimates above are conservative being possibly high. However, the calculation of probabilities ten standard deviations from the base is always fraught with danger. In any practical situation of this type, there will always be a lack of sufficient observations and the choice of a suitable model becomes critical.

The model demonstrated in this example is simple and the parameters easily calculated. The shifted exponential distribution has a self similarity property in that the standard deviation does not depend on the threshold value, and thus the selection of the threshold is not critical.

7 Conclusions

The extrapolation beyond the range of the data is always troublesome: however, on a logarithmic scale the extent of extrapolation looks more reasonable.

It seems that often the original data samples, compared with the maxima over a time interval, have a simpler distribution, have more points available, and thus are easier to extrapolate. Also an extrapolation from a good fit to the whole distribution gives more confidence than fits to the upper extreme values.

An aim of this project was to relate the measured gust data recorded at airports to the future simulated data for 2081–2100 over all of Tasmania, so that estimates of future extreme gust data could be made. This was approached by looking at three stages needed to do this:

- The relation between the airport wind data and the airport gust data.
- The relation between the 1961–1990 data and the airport wind data.
- The relation between the 2081–2100 data to the 1961–1990 data.

Using these three stages it would be possible to relate the 2081–2100 data to corresponding measured gust values. The simulated data covered Tasmania but the airport data was from only eight locations, which is not many to obtain a robust relation from.

The airport wind and gust data at a single location were found to share the same Weibull distribution shape parameter. A relation was found for the scale parameter that gave usable projections of the gust data in seven of the eight cases. For the one case with low wind values it predicts too low.

The best relation found between the airport wind data and the 1961–1990 data was a simple straight line giving an average behaviour (Figure 18). This is not very accurate, but attempts to find a better relation failed. The fact that the simulated data covers 1961–1990 and the airport data 2000–2011 is not considered a major problem; however, differences in topography surrounding the airports may be important. The limited number of airport locations is also a serious problem. There are other sites that record wind speed, but these are considered of lower quality than the airport data and do not include gust speeds.

The 2081–2100 data is similar to the 1961–1990 data except for a very small proportion of larger extreme values. In some locations this difference was minimal, and in others there were large differences. Given the small number of extreme data points at each location, projection of these differences does not look promising. However, as the simulated data is available at all the grid points covering Tasmania, conversion using quantile–quantile plots, that for a given quantile, give the ratio of wind speeds, is possible. Over most of the range the distributions are essentially the same with the only difference being a small number of more extreme points in the 2081–2100 data. These points will have a large effect on the extreme gusts predicted for 2081–2100.

A simple alternative approach to determining the return period value followed an actuarial practice by fitting an exponential distribution to the upper part of the data. The exponential distribution having only a scale parameter is particularly easy to fit.

In summary the data analysed has two major problems: the limited number of airports recording actual data; and the small number extreme points in the wind speeds predicted for 2081–2100. The large differences between measured data and simulated data for the Bass Strait islands is a concern.

An optimisation approach to minimise bias is possible. If this work is successful, then it will be reported in due course. An investigation of suitable goodness of fit criteria for the distributions, taking into account the emphasis on the upper values, would be a useful adjunct to this work.

Acknowledgements The moderators thank the industry sponsor, Geoscience Australia for providing the problem, and John Shepherd and his team for organising for a successful MISG week of problem solving. We are grateful to Augusto Sanabria (<mailto:Augusto.Sanabria@ga.gov.au>) of Geoscience Australia who came to the MISG week, described the problem, and provided much useful information on possible methods on analysis. Others that took an active part in this project were Sherin Ahamed, Shuang Wang, Lynne McArthur, Ash Khan, and Jannah Baker. Alan Brown provided the appli-

cation of the actuarial methods to the data analysis. Several other MISG attendees also took an interest in the project.

References

- Brown, A., 2012. Personal communication. M42
- Bühlmann, H., 1970, *Mathematical Methods in Risk Theory*, Chapter 1; Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Vol. 172, Springer Verlag: Berlin. ISBN 978-3-540-05117-6 M42
- Climate futures, 2012. http://www.dpac.tas.gov.au/divisions/climatechange/adapting/climate_futures M22
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer: New York. ISBN 978-1-85233-459-8 M18
- International Panel on Climate Change, Working Group III, 2000. <http://www.ipcc.ch/pdf/special-reports/spm/sres-en.pdf> M22
- Kendall, M.G. and Stuart, A., 1963. *The Advanced Theory of Statistics, Vol. 1*, Griffin: London. ISBN: 978-0-470-66530-5 M14, M26
- Wikipedia, 2012a. http://en.wikipedia.org/wiki/Order_statistic M14
- Wikipedia, 2012b. http://en.wikipedia.org/wiki/Beta_distribution M14
- Wikipedia, 2012c. http://en.wikipedia.org/wiki/Gamma_distribution M26
- Wikipedia, 2012d. http://en.wikipedia.org/wiki/Generalized_extreme_value_distribution M18
- Wikipedia, 2012e. http://en.wikipedia.org/wiki/Weibull_distribution M8

Author addresses

1. **W. J. Whiten**, 4 Magnet Close, Riverhills, Brisbane, 4074, AUSTRALIA.
<mailto:W.Whiten@uq.edu.au>
2. **B. Abbasi**, Mathematical and Geospatial Sciences, RMIT University, Melbourne, Victoria, 3001, AUSTRALIA.
<mailto:babak.abbasi@rmit.edu.au>
3. **S. Z. Hosseinifard**, Mathematical and Geospatial Sciences, RMIT University, Melbourne, Victoria, 3001, AUSTRALIA.
<mailto:hoseinifard@gmail.com>