WIND HAZARD USING HIGH RESOLUTION CLIMATE MODELS

L.A. Sanabria [1], R.P. Cechet
Geoscience Australia, Canberra, Australia

1. INTRODUCTION

A new model to assess severe wind hazard in the non-cyclonic regions of Australia ('Region A') using simulated wind speeds is presented in this paper.

The model is especially suitable for regions where there are no wind recording stations. The model uses simulation data produced by a high resolution regional climate model. It compares wind speeds produced by the climate model with observations (*mean* wind speeds) and develops functions which allow wind engineers to correct the simulated data in order to match the observed *mean* wind speeds.

Using the corrected simulated wind *mean* speeds in association with empirical *gust factors*, representative wind *gusts* were generated using a Monte Carlo simulation. These wind *gusts* were used to calculate severe wind hazard for non-cyclonic regions of Australia using extreme value distributions (Coles, 2001).

The approach has been validated in a number of locations where observed records are available.

2. HIGH RESOLUTION CLIMATE SIMULATIONS

The climate simulation data used for this project was obtained from CSIRO's Conformal-Cubic Atmospheric Model (CCAM). Two runs of 49 years were simulated for the period 1951 to 2000, one for the Eastern states and the other one for the Western states of Australia. Hourly maximum wind speed (four lowest levels in the atmosphere) was generated for this study, and the conformal cubic grid was transformed to a regular grid using a 0.2° resolution (Dix, 2008). Here we present results of the CCAM 10-metre height maximum hourly wind speed (maximum of time-step values within each hour) for the 50-year simulation period. The final step in the data preparation process was the transformation of the *mean* hourly speeds to maximum daily *mean* speeds. The R package 'zoo' was used to do this (Zeileis & Grothendieck, 2005).

The observed wind speeds used for this project were acquired from the Bureau of Meteorology (BoM) in 2006. Half-hourly datasets from a number of wind stations in southern NSW were used for algorithm development and testing. These datasets provide maximum wind *gust* and *mean* wind speeds in half-hourly intervals (the actual record has the mean speed and the 3 second maximum gust of the last 10 minutes of the half-hourly interval). For comparison with CCAM-modelled data, maximum daily *mean* wind speeds were calculated from the half-hourly *mean* observed wind speeds.

3. BIAS CORRECTION ALGORITHM

The aim of a wind hazard analysis is to calculate return periods (RP) for maximum wind speeds. Geoscience Australia's Risk and Impact Analysis Group (RIAG) has developed a statistical model to do this. The core of the Statistical model calculates return periods of maximum wind speed using the Generalised Pareto Distribution (GPD) (Coles, 2001). This distribution allows for the extrapolation of wind speed return periods well beyond the range of available data (Sanabria & Cechet, 2007a).

For this project, we use the statistical model to generate RP for both observed and CCAM-generated maximum (max) daily *mean* wind speeds. CCAM-generated wind speeds were extracted from model gridpoints surrounding a given wind recording station. Four standard cases were considered:

Case 1 (nearest gridpoint or 0 km case);
Case 2 (1x1 gridpoints or 20 km case);
Case 3 (3x3 gridpoints or 60 km case);
Case 4 (5x5 gridpoints or 100 km case).

The algorithm also uses the concept of 'super-station' (Holmes, 2002). A super-station for Sydney was constructed by joining the observed wind speed datasets from Sydney, Bankstown and Richmond airports. This super-station was named 'Sydney Region' or 'sydreg'. The four standard cases were defined for this region by joining the corresponding four cases of each of the three wind recording stations.

To explain the bias correction algorithm consider Figure 1a. It shows the observed RP of Sydney Airport max daily *mean* speed and the CCAM wind speeds around the recording station (gridpoints) as explained above. The same results for the Sydney Region are presented in Figure 1b, the full line is the observed RP of wind speeds (GCS is the lat-long or Geographical coordinate system).

Figures 1a and 1b show that CCAM underestimates the RP for wind speeds in both the Sydney Region and the Sydney Airport cases. The bias is best considered in Figure 2 which presents a plot of CCAM RP for the 100 km case of the Sydney Region (Case 4 above) and the observed max daily *mean* speeds. The black points are the corresponding RP of wind speeds for 10, 100, 1000 and 10000 years. It is clear that although they are not equal, there is a strong

---

[1] *Corresponding author address:* Sanabria L.A.; Geoscience Australia; GPO Box 378; Canberra ACT 2601; Australia; e-mail: augusto.sanabria@ga.gov.au.

linear correlation – in the semi logarithmic scale – between the RP of CCAM speeds and the RP of observed speeds. The same strong linear correlation was observed between CCAM-modelled speeds and observed speeds in the Sydney Airport and in Bankstown and Richmond stations. This characteristic of the modelled wind speeds will be used to develop an expression to correct the bias of the CCAM-modelled speeds

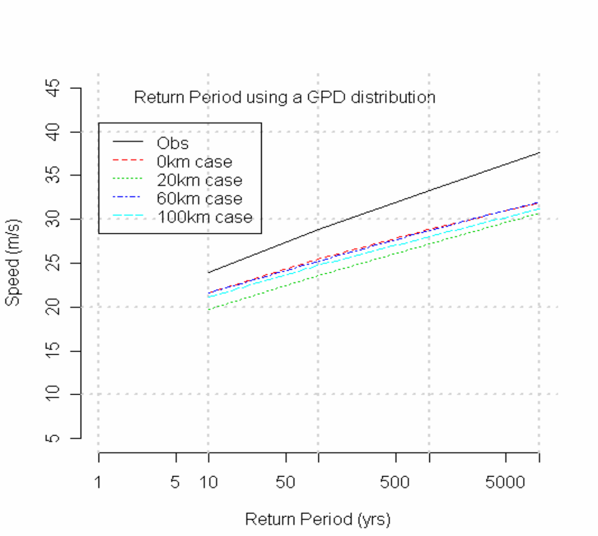

Figure 1a. Sydney Airp. Wind RP

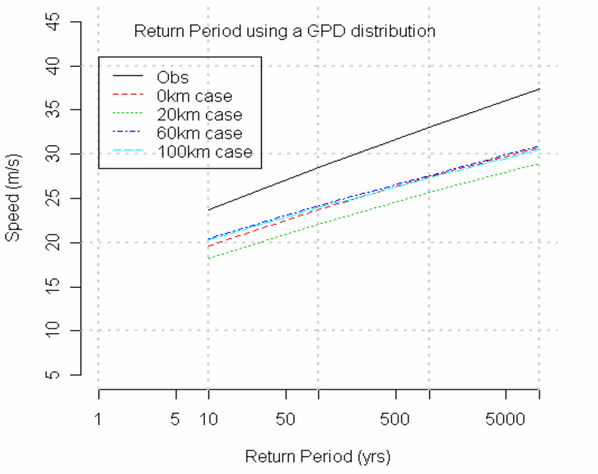

Figure 1b. Sydney Region wind RP

3.1 Comparing RP of CCAM-modelled and observed mean wind speeds

A linear regression (LR) between CCAM modelled *mean* wind speeds and observed *mean* wind speeds was calculated for each one of the 4 cases explained above. Figure 3 shows one of the LR for the Sydney Region example (Case 4 above). The regression expression is shown at the top of Figure 3. The relative insensitivity of the results with regard to the sampling area was surprising (Case 1-4; Figures 1a & b). This may be an indication that the horizontal resolution of the simulation may not be sufficient for this application. As a result we have decided to use an average of the regression lines that were calculated for the sampling areas (cases). The average of the Sydney Region regression lines operating on the average of the RP of the 4 CCAM cases was used to correct CCAM wind speed bias.

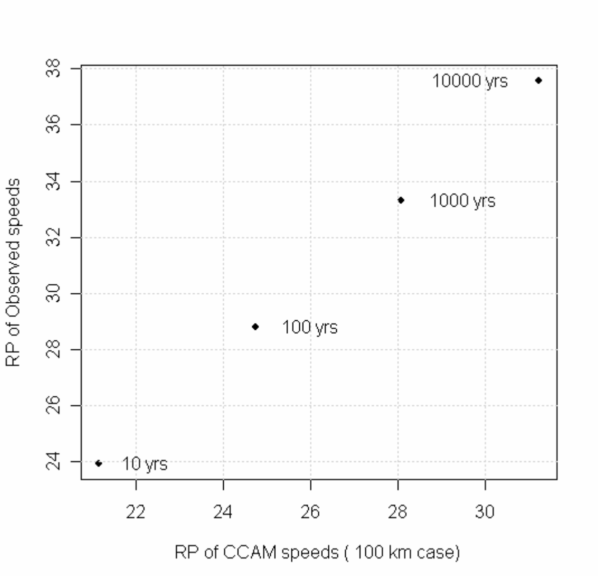

Figure 2. CCAM and observed RP

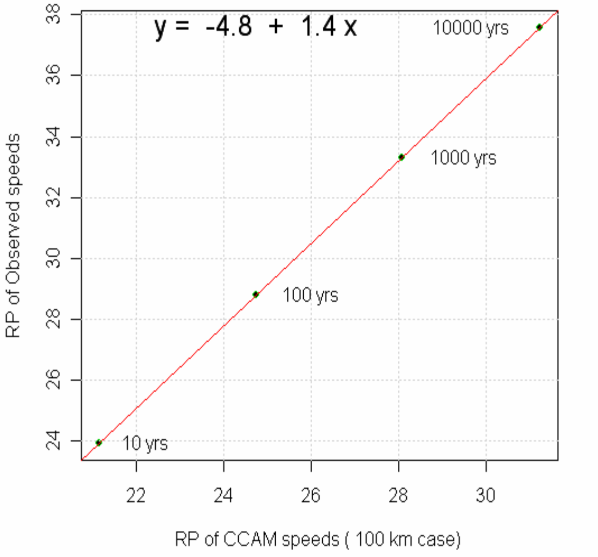

Figure 3. LR of CCAM and obs. wind RP

3.2 RP of corrected CCAM-modelled wind speeds

Figure 4 compares the observed RP of wind speeds with the corrected CCAM-modelled RP of wind speeds for the Sydney Airport example; for this case the average of the LR expressions of the Sydney Region were used to correct the average RP of CCAM-modelled Sydney Airport speeds. Comparing Figures 1a and 4, it is possible to see that the correction has substantially improved CCAM results.

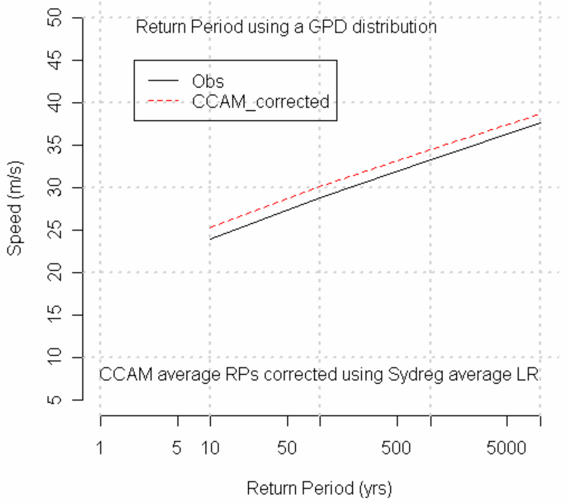

Figure 4. CCAM-corrected and Obs. RP

Table 1 presents a summary of the results. 'Corrected CCAM' is the corrected average of CCAM RP for the 4 standard cases discussed above. The third column presents the 95% confidence interval (CI) for the observed RP. The error between the observed (Obs) and the CCAM corrected RP is defined as:

Error (%) = Abs[(Obs – CCAM)/Obs]*100

Table 1.Wind RP using corrected CCAM speeds

RP	Observed	95% CI	Corrected CCAM	Error (%)
10	23.9	(22.8,25.7)	25.3	5.9
100	28.8	(26.4,32.4)	30.1	4.5
1000	33.3	(29.1,38.2)	34.5	3.6
10000	37.6	(31.5,43.5)	38.6	2.7

The error is small, and in the worst case it amounts to less than 1.5 m/s on a 10 year return period. Notice that in all cases CCAM corrected RP is within the 95% confidence interval (CI).

Figures 5a and 5b compares the observed and CCAM-modelled wind speeds for Bankstown and Richmond stations. The average of the 4 standard cases of wind speed RP of each station was corrected using the average of the LR expressions of the Sydney Region as explained previously. An error analysis of the results show that the maximum error obtained in assessing wind hazard using CCAM wind speed is 21.2%, which occurs for RP of 10 years in Richmond.

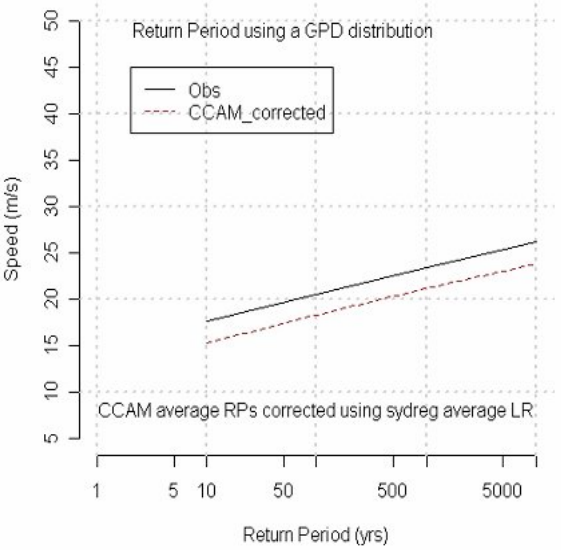

Figure 5a. RP for CCAM Bankstown

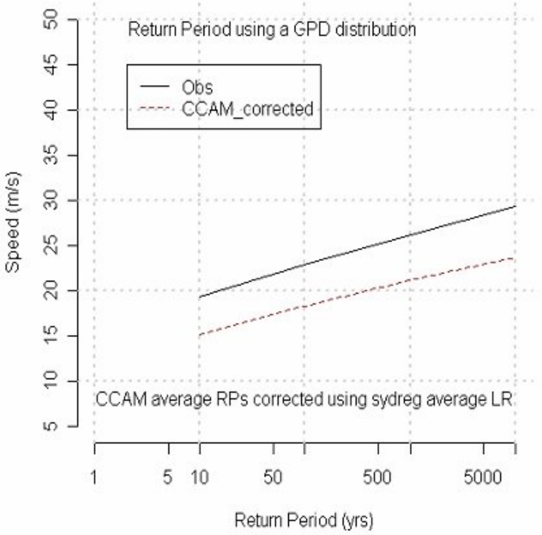

Figure 5b. RP for CCAM Richmond

Although this error appears large in % terms, it represents only a few m/s in a 10-year RP. High-resolution climate models are still in development and there are large sources of uncertainties (apart from horizontal resolution) which can explain differences with observed and modelled results (Déqué et al., 2007). Wind is highly variable both geographically and temporally (meteorological variable with highest variability). This variability persists over a wide range of scales, both in space and time

3.3 Corrected CCAM Wind Outside the Region

A further analysis was carried out to examine the correction algorithm performance on data not used for algorithm development. Consider the wind speeds recorded at Williamtown, a RAAF station 142 km north of Sydney. The CCAM wind speeds for the 4 standard cases were extracted and the average RP for these 4 cases were calculated. Finally using the average LR expression from the Sydney Region, the RP of wind speeds for Williamtown as given by CCAM simulations were corrected and plotted (Figure 6).

The maximum error on CCAM-modelled wind speeds at this station is 18.2% at a RP of 10,000 years. This result is consistent with the previous results, and suggests the algorithm is general enough to be used in stations close to the boundaries of the region used for correction.

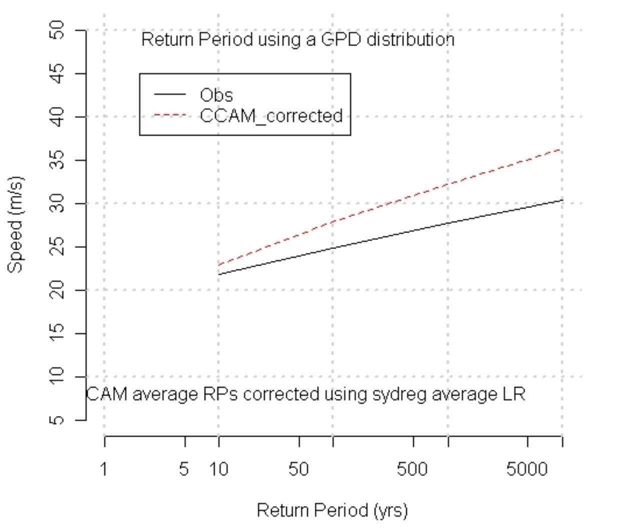

Figure 6. RP for CCAM Williamtown

4. PARAMETERS OF CORRECTED GPD

The correction of the RP calculated using the GPD requires a recalculation of the parameters of the corrected GPD. Figure 4 (dotted curve) shows that the corrected GPD is shifted upwards by the correction algorithm. In mathematical terms the shifting is due to the convolution of the random variable (rv) of wind speed and the linear regression expression. The corresponding parameters of the shifted GPD can be calculated using the method of 'moments' (Kendall & Stuart, 1963). The method of moments provides an easy way to do convolution of random variables. As explained in Kendall & Stuart (1963) the convolution of rv X and Y to produce rv Z corresponds to the sum of their central moments up to order 3, i.e.

$$cm_r(Z) = cm_r(X) + cm_r(Y) \qquad r = 1,2,3 \qquad \text{eq. (1)}$$

where cm_r = central moment or order 'r'

Z, X and Y are random variables

The expression can be extended to r > 3 using cumulants instead of moments (Kendall, 1963).

In our case we need the first three central moments of the shifted GPD to calculate its parameters (location μ, scale σ and shape ξ). The central moments were calculated from the CCAM wind speed dataset using expressions provided in Kendall & Stuart (1963). Expressions to calculate the GPD parameters from moments and vice versa are given in Singh (1995),

$$G = [2(1 - \xi)\,\mathrm{sqrt}(1 + 2\xi)\,]/(1 + 3\,\xi) \qquad \text{eq. (2)}$$
$$\hat{w} = \mu + \sigma/(1 + \xi) \qquad \text{eq. (3)}$$
$$V = \sigma^2/(1 + \xi)^2\,(1 + 2\xi) \qquad \text{eq. (4)}$$

$$\sigma = V\,(1 + \xi)\,\mathrm{sqrt}(1 + 2\xi) \qquad \text{eq. (5)}$$
$$\mu = \hat{w} - \sigma/(\,\sigma + \xi) \qquad \text{eq. (6)}$$

where G = skewness ($cm_3/cm_2^{\,3/2}$)
$\hat{w}$ = mean
V = variance (cm_2)

μ = location
σ = scale
ξ = shape

The variables G, $\hat{w}$ and V are calculated from the central moments using the expressions shown within brackets. To calculate the GPD parameters from moments solve eq, (2) using the bisection procedure (Press et al., 1986) and then solve eq. (5) and (6). Notice that $\hat{w}$ is the given sample mean (or m1, first moment with respect to 0).

The LR expression to shift the GPD for the Sydney airport case has the general form,

$$Z = aX + b \qquad \text{eq. (7)}$$

where X is the original (not corrected) GPD
Z is the corrected GPD (see Figure 4).

and hence applying the expression to the central moments of the original GPD (X) we can get the central moments of the corrected GPD (Z) as,

$$ccm_1(Z) = a\,cm_1(X) + b \qquad \text{eq. (7)}$$
$$ccm_2(Z) = a^2\,cm_2(X)$$
$$ccm_3(Z) = a^3\,cm_3(X)$$

where ccm_r is the corrected central moment of order 'r'.

Once the parameters of the corrected GPD have been calculated it is possible to generate its *mean* wind speeds by using the definition of GPD (Hosking, 1987),

$$g(x) = (1/\sigma)(1 - \xi(x - \mu)/\sigma)^{(-1+1/\xi\,)} \qquad \text{eq. (8)}$$

In this work the package 'evd' of the R programming environment was used (Stephenson, 2004).

Table 2 presents a summary of the CCAM GPD moment/parameter calculation.

Table 2. Corrected GPD parameter calculation

Moments	m_1	cm_2	cm_3
Original GPD	7.22	3.66	15.74
Corrected GPD	7.72	5.95	32.63
GPD parameters	μ	σ	ξ
Original GPD	6.2	1.77	-0.037
Corrected GPD	6.7	2.26	-0.037

5. WIND GUST GENERATION USING MONTE CARLO SIMULATION

The results presented in Section 3 provide an algorithm for correction of the CCAM-modelled *mean* wind speeds. In severe wind hazard, however, the variable of interest is *gust* speeds. In this Section a model to generate *gust* speeds from *mean* speeds is presented.

The model simulates the physics of wind generation using Monte Carlo sampling. It assumes that surface wind *gusts* result from the deflection of air parcels flowing higher in the boundary layer, which are brought down by turbulent eddies (Brasseur, 2001). The method takes into account the *mean* wind and the turbulent structure of the atmosphere. Turbulence is represented by the *gust* to *mean* ratio, termed the *gust factor* in wind engineering (Halfdan & Haraldur, 2004).

The model consists of the numerical convolution of max daily *mean* wind speed and the distribution function of *gust factors*. It produces a max daily wind *gust* speed dataset, the geophysical parameter considered in severe wind hazard assessment.

5.1 Gust factors (GF)

The gust factor (GF) is defined as the ratio of maximum wind speed (*gust*) and the *mean* wind speed for the same time period. As an illustration, consider the half hourly speed dataset from Sydney Airport. As explained before the dataset contains the *mean* value of wind speed and the maximum wind speed (3 second *gust*) of the last 10 minutes of the period.

It is recommended to split the range of *mean* wind speed into a number of intervals, so to better capture the sensitivity of *gust factors* (high tale of the distribution) to the wind speed. For this example the *gust factors* were calculated in the intervals [5, < 15] and [≥ 15] m/s as shown in Figure 7. A minimum *mean* wind speed of 5 m/s was chosen, as *gust factors* associated with light winds are of no concern with regard to the assessment of wind hazard.

Using the histograms of Figure 7 the empirical cumulative distribution function (CDF) of the GF was calculated; this is the probabilistic function used for the sampling process in the Monte Carlo simulation.

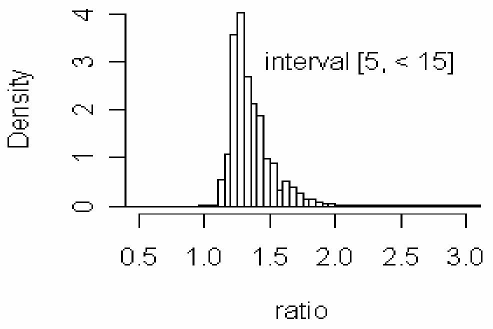

Figure 7a. Gust factor in [5,<15]

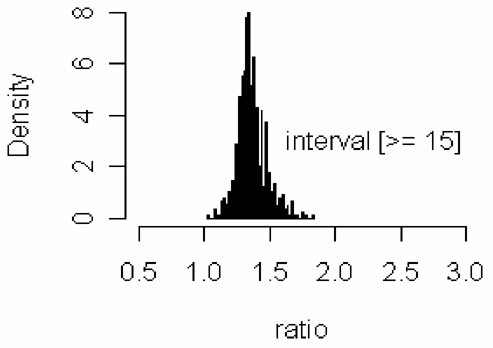

Figure 7b. Gust factor in [15, max.]

Gust factors for a number of wind stations were calculated using the R programming environment (Sanabria & Cechet, 2008). Then using Monte Carlo sampling a numerical convolution of CCAM corrected max *daily mean* wind speed and the distribution function of the *GF* was carried out. In practice a sample from the max *daily mean* wind speed was selected in an ordered fashion, then a sample from the *gust factor* CDF was selected at random (using a uniform random distribution) and the two samples were multiplied together to produce a sample for the max daily *gust speed* dataset.

6. RESULTS

Using eq. (8) a vector of 5000 samples, representative of CCAM *mean* speeds was generated, and hence the max *daily gust* speed, calculated by MC simulation, had the same number of elements. To obtain a large dataset for statistical analysis the process was repeated 200 times to produce 200 max daily *gust* speed vectors which were combined into one large vector. This vector had 936,878 wind speed elements corresponding to 2567 years of equivalent *gust* speed data.

Some of the 200 synthetic max. daily *gust* speed datasets were compared against the actual observed *gust* wind speed dataset. Datasets are similar if the residuals are normally distributed with a mean of zero. The residuals are defined as,

residuals = synthetic – observed datasets

The comparisons show that the residuals follow the normal distribution very closely except for the tail of the distribution (values with speeds greater than 10 m/s). The MC simulation can generate synthetic datasets close to the observed except in the tail of the distribution. This is because an insufficient number of values were generated in the tail. Unfortunately in wind hazard analysis, values in the upper tail region of the distribution are very important and hence this limitation of the method had to be removed.

To remove the tail bias in the speeds produced by the simulation it was necessary to use a stratified sampling technique (Liu, 2001): the axis of the *observed* max daily *gust* was split into 5 m/s 'bins' and the percentage of values falling within each bin was calculated. The Monte Carlo distribution was then sampled within each bin to give the same percentage of values per bin as in the observed dataset. Figure 8 compares the residuals against the normal distribution before and after the tail bias was removed. Notice the improved agreement with the normal especially in the tail of the distribution in the bottom plot. Quantitative tests of normality, using Anderson-Darling, Shapiro-Wilks and Pearson algorithms (Gross, 2008) failed to produce meaningful results because of the large number of samples.

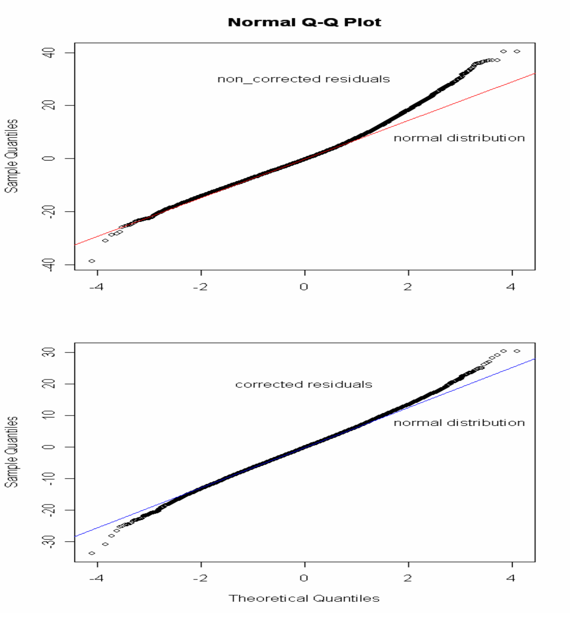

Figure 8. Residuals without improved tail (top) and with improved tail (bottom).

Figure 9 shows the RP produced by the MC simulation with tail bias removed, and the RP of the observed wind *gusts*. The dotted line is the observed RP plot.

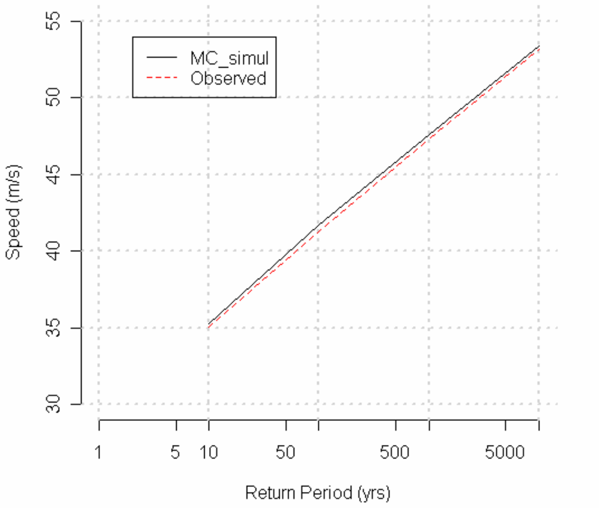

Figure 9. MC simulated and observed RP

Statistics of the results were obtained by repeating the process 250 times. Each repeat generated 200 synthetic datasets, which were combined into a single dataset to plot the RP curve. Figure 10 shows the 250 RP curves. A summary of the simulation results is presented in Table 3.

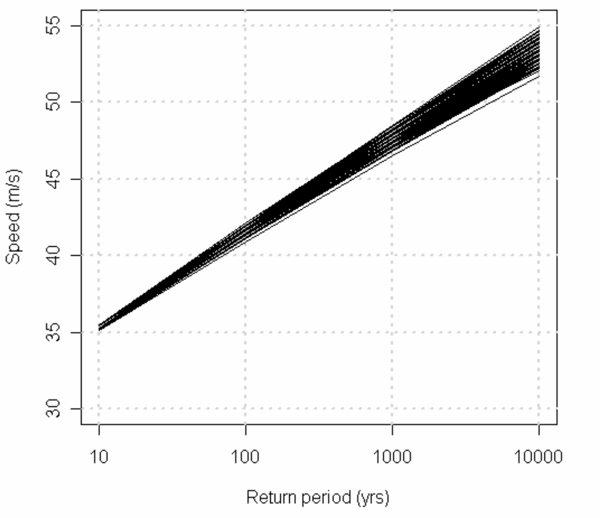

Figure 10. RP of the 250 simulations

Table 3. Summary of MC results (m/s)

RP (years)	Mean	Standard deviation	Min	Max
10	35.2	0.09	35.1	35.5
100	41.6	0.30	40.9	42.1
1000	47.5	0.43	46.5	48.5
10000	53.3	0.61	51.7	54.9

The 250 simulation results show a narrow band of RP wind speeds with a small standard deviation, indicating a very stable process. The synthetic datasets with tail bias correction contain a significant greater number of extreme values which better define the extreme upper tail of the distribution and allow a higher threshold value to be selected (optimum) for the fitting of the GPD. The results presented here and other results show that the MC simulation provides upper and lower bounds for the observed RP (Sanabria & Cechet, 2007b).

The Sydney airport wind speeds at RP of 10, 100, 1000 and 10000 years produced by the 250 Monte Carlo simulations compare well with the statistical model, as seen by comparing Figures 9 and 10.

7. CONCLUSIONS

A new model to assess severe wind hazard in Australia's region A has been presented in this paper. The model integrates three models developed in Geoscience Australia in recent years: a statistical model to calculate RP of wind speeds, a model to extract and correct wind speeds from a high resolution regional climate model, and finally a Monte Carlo method to generate representative *gust* speeds from the simulated speeds.

Comparison of model results with observations show that the model is robust enough to match observation RP of *gust* speeds and hence it can be used to assess severe wind hazard in locations where there are no wind recording stations. In these cases, *gust factors* from similar stations can be used in the MC process.

Work is currently under way to replace the computer-demanding MC method by an analytical convolution of wind speeds and *gust factors* to substantially reduce computer time.

8. REFERENCES

Brasseur O. (2001). Development & Application of a Physical Approach to Estimating Wind Gusts. Monthly Weather Review. Vol. 129, Jan 2001. 5-25.

Coles S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer series in statistics. Springer

Déqué M., Rowell D.P., Lüthi D., Giorgi F., Christensen J.H., Rockel B., Jacob D., Kjellström E., M. de Castro and van den Hurk B. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Climate Change, 81:53-70.

Dix M. (2008). CSIRO's Program 'cc2hist'. Personal communication. April 2008.

Gross J. (2008). The 'Nortest' package to test for normality. . The R project for statistical computing. http://www.r-project.org/).

Hosking J.R.M and Wallis J.R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. Technometrics. Vol. 29, No.3, August 1987.

Holmes J.D. (2002). A Re-analysis of Recorded Extreme Wind Speeds in Region A. Australian Journal of Structural Eng. Vol. 4, No. 1, 2002

Halfdan A. & Haraldur O. (2004). Mean Gust Factors in Complex Terrain. Meteorologische Zeitschrift. Vol 13, No. 2, 149-155 (in English).

Liu J.S. (2001). Monte Carlo Strategies in Scientific Computing. Springer Series in Statistics. Springer.

Press W.H., Flannery B.P., Teukolsky S.A. and Vetterling W.T. (1986). Numerical Recipes. Cambridge University Press, 1986.

Sanabria L.A. and Cechet R.P. (2007a). A Statistical Model of Severe Winds. Geoscience Australia, Record 2007/12.

Sanabria L.A. and Cechet R.P. (2008). Development of Wind Gust Factor Distributions for Use in Monte Carlo Simulation of Wind Gust Datasets for Region A of AS/NZS 1170.2. 15th National AMOS Conference. Geelong, Victoria, Jan-Feb 2008.

Sanabria L.A. and Cechet R.P. (2007b). Monte Carlo Modelling of Severe Wind Gust. Int. Congress on Modelling and Simulation 'MODSIM07'. Christchurch NZ, December 2007.

Singh V.P. and Guo H. (1995). Parameter Estimation for 3-parameter Generalized Pareto Distribution by the Principle of Maximum Entropy. Hydrological Sciences, 40.2. April 1995.

Stephenson A (2004). A User's Guide to the 'EVD' Package (Version 2.1). Department of Statistics. Macquarie University. Australia.

Zeileis A. and Grothendieck G. (2005). The "zoo" package V1.0-3.Dec. 2005. http://www.r-project.org/).