Monte-Carlo Modelling of Severe Wind Gust

L.A. Sanabria and R.P. Cechet

Risk and Impact Analysis Group, Geoscience Australia, Canberra ACT.
Email: augusto.sanabria@ga.gov.au

Keywords: *Severe wind hazard, return periods, Monte Carlo simulation.*

EXTENDED ABSTRACT

Severe wind is one of the major hazards facing the Australian continent. While cyclonic winds are the major source of wind hazard in the northern states, non-cyclonic winds driven by synoptic lows, thunderstorms and tornadoes affect the southern states. Severe winds are responsible for about 40% of damages to Australian residential buildings.

Geoscience Australia's Risk and Impact Analysis Group has developed a statistical model to assess severe wind hazard. The model has been tested using observational data from wind stations located in South Eastern Australia and has shown to match the results of the Australian/NZ standard for wind loading of structures (AS/NZS 1170.2) using a more efficient, fully computational method (Sanabria & Cechet, 2007).

The main limitation of the statistical model is its dependency on the quality of the recorded data used. In particular, gust speed recordings are unreliable as instruments are calibrated for *mean* wind speeds; the procedure does not determine their transient response. The transient response is instrument dependent. This is a serious limitation because of the possible inconsistency and reduced accuracy between instrument types. These instruments have been developed over a period of more than 50 years and some studies have shown that differences occur with regard to their ability to measure short wind *gusts* (say *gusts* of 1-3 sec duration). These short duration wind *gusts* are the cause of wind-related damage and thus crucial in wind hazard studies.

The second limitation of the statistical model involves the calculation of confidence intervals for the wind hazard return period. This interval increases in proportion to the return-period (years) considered. For observing stations with record lengths of 20 to 30 years, it has been found that the derived return periods beyond about 500 years are too unreliable for use in practical applications.

Geoscience Australia has also developed an alternative wind hazard estimation method; this alternative method is fully based on physical considerations. The main motivation for such an approach is to overcome the problems of the statistical model. The proposed approach assumes that surface *gusts* result from the deflection of air parcels flowing higher in the boundary layer, which are brought down by turbulent eddies. The method takes into account the *mean* wind and the turbulent structure of the atmosphere (*gust to mean ratio*) and produces statistical distributions from observational data for these components (*mean* and *gust-to-mean ratio*). Monte Carlo sampling provides a range of likely *gust* magnitudes which can be used to evaluate the quality of the original gust observations.

This Monte Carlo approach has been used to create synthetic datasets that attempt to describe the full range of possible outcomes for *peak wind gusts*. In some cases a shorter dataset than the observed *peak wind gusts* was used for generation of the synthetic dataset. It is assumed that the shorter record retains the level of variability expected in the longer record. This is true for the majority of mid-latitude observing stations (say for return-periods less than 500 years) where *peak wind gusts* are mainly the result of relatively frequent thunderstorms and intense synoptic weather systems (i.e. not caused by rare tropical cyclone or small-scale tornado activity).

Results from the Monte Carlo simulation show the algorithm is robust and can produce long records equivalent to thousands of years of data. Examination of subsets of these long records shows a stable process (narrow band of peak wind gust return periods with small standard deviations). The confidence intervals calculated from these longer datasets are narrower than those from the statistical model. In examples shown here, we observe that the results from the statistical model lay within the narrower 95% confidence interval calculated from the Monte Carlo approach. This adds to our confidence in the *peak wind gust* datasets being utilised for wind hazard studies. The technique has been found to be useful for data quality checking of problematic datasets and assists in determining whether we use all or part of the dataset for hazard determination.

1. INTRODUCTION

The aim of this paper is to present a Monte Carlo (MC) technique for the generation of synthetic *peak wind gust* datasets (with thousands of years of data) for comparison with observed datasets which typically have a record length of only 20 to 50 years. The observed *peak wind* datasets have been obtained using a variety of instruments, calibration procedures and recording systems (paper record and digital) which introduce inconsistencies (data quality issues are discussed in Section 2). The synthetic datasets can be used for consistency checking to build confidence in the observed datasets.

Severe wind is one of the major hazards facing the Australian continent. While cyclonic winds are the major source of wind hazard in the northern states, non-cyclonic winds driven by synoptic lows, thunderstorms and tornadoes affect the southern states as well as the southern part of both the east and west coasts. Severe winds are responsible for about 40% of the damage to Australian residential buildings (Chen, 2004).

Geoscience Australia's Risk and Impact Analysis Group (RIAG) has developed a statistical model to assess severe wind hazard (Sanabria & Cechet, 2007). The model produces return periods (RP) for *gust wind speeds*. It has been tested using observational data from wind stations in southern NSW including the Sydney region.

The main limitation of the model is its dependency on the quality of the observational *peak wind gust* data. A second limitation of the statistical model involves the calculation of a confidence interval for the RP based on a short data record length (say 20-50 years). The interval increases in proportion to the RP (years) considered. Results show that return periods beyond about 500 years are too unreliable for use in practical applications (Sanabria & Cechet, 2007).

To overcome these problems RIAG has developed an alternative approach to calculate wind hazard. The proposed approach assumes that surface *gusts* result from the deflection of air parcels flowing higher in the boundary layer, which are brought down by turbulent eddies. The method takes into account the *mean* wind and the turbulent structure of the atmosphere (*gust to mean ratio*) and produces statistical distributions from observational data for each of these components. Section 4 describes the generation of the *peak gust to mean ratio*.

A MC sampling technique is utilised to provide a range of likely gust wind speed magnitudes which can be used to determine RP wind gust estimates (see Section 5). This approach has been used to create synthetic datasets that attempt to describe the full range of possible outcomes for *peak wind gusts*. In some cases a shorter dataset than the observed *peak wind gusts* was used (cases 2 and 3 below). It is assumed that a shorter record retains the level of variability expected in the longer record. This is true for the majority of mid-latitude observing stations (say for return-periods less than 500 years) where *peak wind gusts* are mainly the result of relatively frequent thunderstorms and intense synoptic weather systems (i.e. not caused by rare tropical cyclone or small-scale tornado activity).

Results from the Monte Carlo simulation are described in Section 6. The synthetic datasets that have been created utilise the *mean* and *gust to mean* observations from the following datasets: **Case 1**: Sydney airport (half-hour sampling) period 1952-2005. **Case 2**: Sydney airport (half-hour sampling) period 1973-2005. **Case 3**: Sydney airport (half-hour sampling) period 1994-2005. **Case 4**: Sydney region (half-hour sampling) period 1973-2005. **Case 5**: (Mixed datasets): Sydney airport, 3-hourly mean and half-hour sampling, period 1973-2005. **Case 6**: (Mixed periods): Sydney airport (half-hour sampling).

In Case 5, the mixed datasets case, the daily *mean* was calculated from the 3-hour dataset while the *gust to mean ratio* was calculated from the half-hour, both in the period 1973-2005. In Case 6, the mixed period case, the *daily mean* was calculated from the whole dataset (1952-2005, see Fig. 2.1c) while the *gust to mean ratio* was calculated from the more reliable part of the dataset, i.e. 1994-2005, as will be explained in Section 6.

The results show that the Monte-Carlo technique is robust and can produce consistent, long-length data records. Examination of subsets of these long records shows a stable process (narrow band of peak wind gust RP with small standard deviations).

The confidence intervals calculated from these longer datasets are narrower than those from the statistical model. In examples shown here, we observe that the results of the statistical model lay within the narrower 95% confidence interval calculated from the MC distributions, adding confidence in our modelling methodology.

2. DATA QUALITY

Most of the wind speed data quality limitations arise as the instruments are calibrated for *mean* wind speeds; the calibration procedure does not determine the transient response of the anemometer as experienced during a short *wind gust* (say of 1-3 second duration). This is a serious limitation for wind gust hazard studies.

Wind speed dataset quality can be examined by considering some of the Sydney Airport wind datasets available from the Australian Bureau of Meteorology (BoM). Figure 2.1a shows the scatter plot of the airport's *maximum daily gust* exceeding 15 m/s. The vertical (dotted) lines indicate the date on which the wind recording instrument was changed as shown in Table 2.1 (BoM, 2006). Figure 2.1a clearly shows three different speed regions; one from the start of the record until 31/12/1973 when the first Synchrotac anemometer was installed; the second period stretches from the installation of the first Synchrotac until the new Synchrotac was installed (01/07/1994); the third region comprises the data after the 1994 upgrade. Each region seems to be drawn from different populations; the first region in particular presents higher speeds than the other two. Regions 1 and 2 show outliers at 42.2 m/s. Studies conducted by BoM show that some stations can be affected by the growth of nearby trees or housing developments, movement of the anemometer, recalibration or changes in the reading frequency (Muirhead et al., 2005).

Table 2.1. Changes in recording instruments.

Date	Change	Instrument
04/01/1939	Installation	Dynes
31/12/1973	Installation	Synchrotac S/N 706
01/07/1994	Installation	Synchotac S/N 706
28/04/1999	Replacement	Synchrotac S/N 706
10/09/1999	Replacement	Synchrotac S/N 897
28/04/2000	Installation	Mast installation
05/09/2003	Installation	Synchrotac cups S/N 732
03/01/2006	Replacement	Synchrotac cups S/N 732

To further examine the dataset, consider the *3-hourly mean speed* from the same observing station, as presented in Figure 2.1b. Notice that the outliers are not present in this dataset. Region 1 still provides the highest speeds in the dataset.

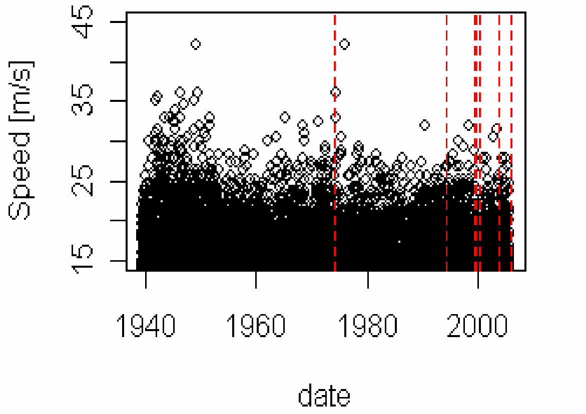


Fig.2.1a. Maximum daily gust speed.

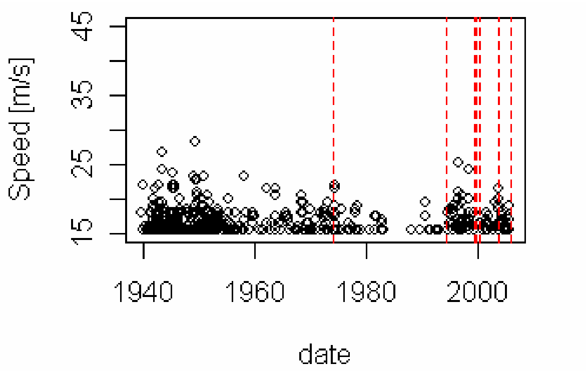


Fig.2.1b. 3-hour mean wind speed

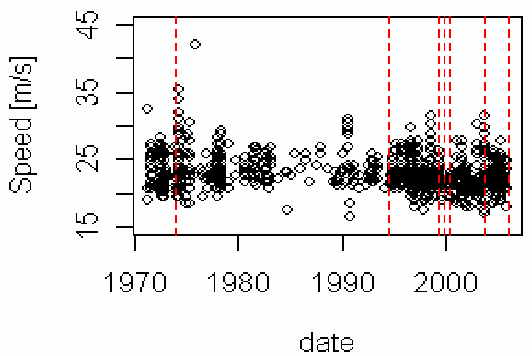


Fig.2.1c. Half-hour maximum gust speed.

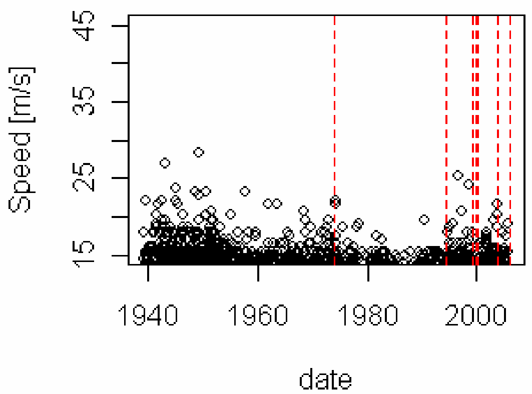


Fig. 2.1d. Daily mean wind speed from 3-hourly observations.

Two more plots of the Sydney Airport wind station are presented in Figures 2.1c and 2.1d, the former was generated from the half hour sampling dataset. This dataset includes half-hour values for *mean* wind speed, i.e. wind speeds averaged over the 10 min previous to observation time; and for *gust* speed. For automatic weather stations (AWS) the *gust* speed recorded is the highest 1 sec value over the last 10 min previous to observation time. These two speeds will be used in Section 3 to calculate a *gust factor*. Figure 2.1c shows the *half hour gust speed* scatter plot. Figure 2.1d shows the *daily mean* wind speed calculated from the 3-hourly dataset. This dataset will be used in case 5. Figure 2.2 shows the relationship between wind *gust* and *mean* speeds as found in the half-hour dataset; the straight lines are the linear regression in periods 1, 2 and 3. The regression lines show a different trend for each time period, with the data for the period 1972-94 showing a long-term trend towards higher values than the other two.

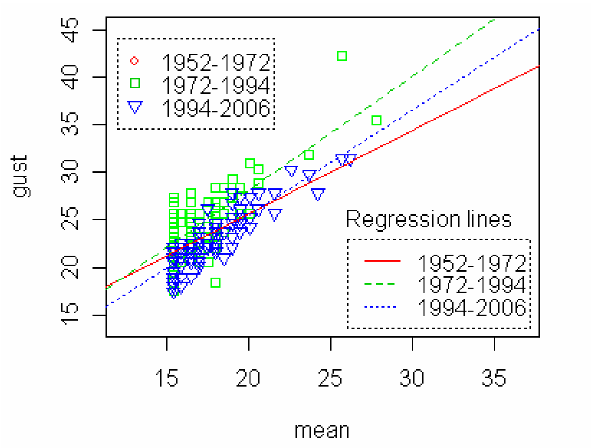


Fig. 2.2. Gust to mean ratio (half-hourly).

3. STATISTICAL MODEL OF PEAK WIND GUST SPEED

To facilitate comparison of the statistical model results with the results produced by the MC model, Figure 3.1 (taken from Sanabria & Cechet, 2007) shows RP of *gust* speed using the Sydney Airport *max. daily gust speed*. This dataset covers a range of 66 years from 1939 to 2005. The circles are the actual return periods; the solid line shows the calculated RP using an Extreme Value Distribution, in this case the Generalised Pareto Distribution (GPD). Notice that the RP can be extended from 66 up to 10000 years by using the GPD. The dotted lines show the 95% confidence interval. The confidence interval at 500-year RP is indicated by the vertical line. The 500-year RP speed is 44.9 m/s, this value is between the confidence limits of 36.8 and 52.8, theoretically there is a 95% probability that a value in the range is correct. In practice the confidence interval may be too wide and hence its usefulness will be dependent on the level of risk/confidence required for the application.

Accurate results for wind RP are important as they are used in the energy generation and the building construction industries. For example the 500-year RP wind gust is utilised in the Australian building codes for residential and commercial constructions. These considerations provided the impetus for the Monte Carlo simulation discussed in this paper.

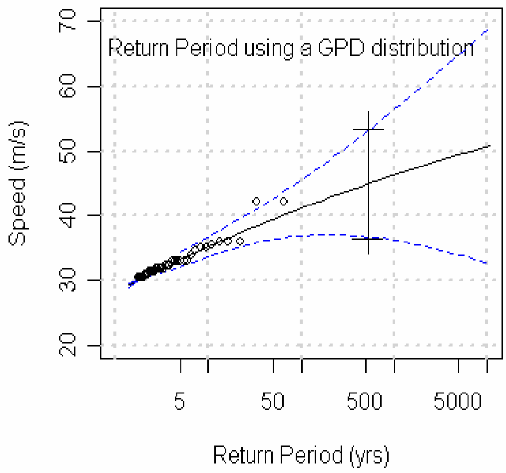


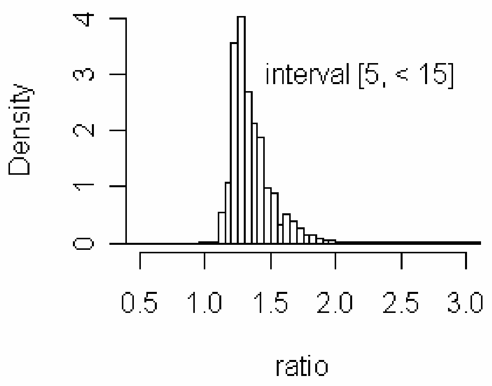
Fig. 3.1. Return period of gust wind speed

4. GUST TO MEAN RATIO

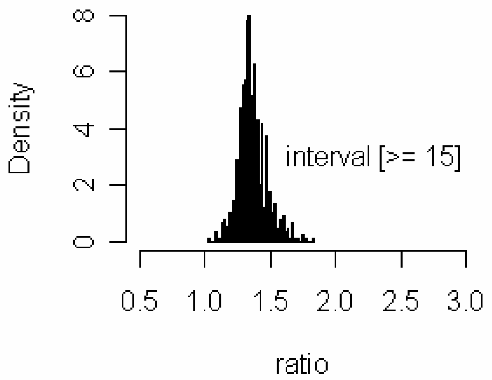
In most of the datasets used to develop the algorithm it was necessary to split the *mean* wind speed axis into two intervals to better capture the *gust to mean ratio* (gust factor) variability across the range of *mean* wind speed observed. The *mean* speed intervals selected were [5, < 15] and [≥ 15] m/s. The corresponding values of *gust speed* found in these *mean* wind speed intervals were selected for calculation of the *gust to mean ratio* in each interval. A minimum mean wind speed of 5 m/s was chosen as gust factors associated with light winds were of no concern in this study.

The Sydney Airport *half-hour* dataset was used to test the algorithm (Case 2 of Section 1). The dataset covers the range 1952-2005 (54 years of data) but only the length of record after the installation of the Synchrotac anemometer on 31/12/1973 was used. This part of the data record is considered more reliable (Muirhead *et al.*, 2005).

Figure 4.1(a) and (b) shows the histograms of *gust to mean ratio* in the two intervals. Table 4.1 summarises the characteristics of the *gust to mean ratio* in each interval.



(a)



(b)

Fig. 4.1. Histograms of *gust to mean ratio* for intervals (a) [5, < 15m/s] and (b) [≥15m/s]

Table 4.1. Characteristics of the *gust to mean ratio*

	[5, < 15m/s]	[≥ 15m/s]
Minimum	1.00	1.03
Maximum	3.32	1.84
Mean	1.36	1.35
Std. deviation	0.17	0.11

5. DESCRIPTION OF THE MC SIMULATION

The MC simulation used in this project relies on the fact that the *mean* wind records are more reliable than the *gust* records. *Gust* and *mean* speeds were extracted from the BoM half-hour dataset. The *gust factor* (*gust to mean ratio*) was calculated as the ratio of the vectors of *gust* and *mean* speed.

Using the ratios of Figure 4.1 the empirical cumulative distribution function (CDF) of the *gust to mean ratio* was determined; this is the probabilistic function used for the sampling process in the Monte Carlo simulation.

In the simulation a numerical convolution of maximum *daily mean* wind speed and a distribution function of *gust to mean ratio* was carried out. The maximum *daily mean* wind speed was found from the half-hour *mean* speeds using the R-package “zoo” (Zeileis & Grothendieck, 2005). The numerical convolution was carried out by taking a sample from the max. *daily mean* wind speed dataset in an ordered fashion, then a sample from the *gust to mean ratio* 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

The maximum *daily mean* wind speed vector has 4893 observations, and hence the max *daily gust* speed, calculated as explained in Section 5, has 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 898,985 wind speed elements corresponding to 2512 years of consistent *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 (see top part of Figure 6.1). 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 must be removed.

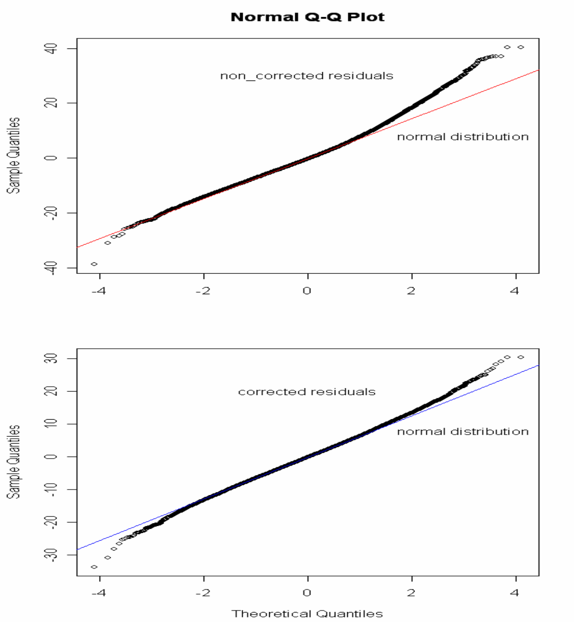


Fig. 6.1.Residuals with and without improved tail.

To remove the *tail* bias of the distributions produced by the simulation it was necessary to split the axis of the observed maximum *daily gust* into 5 m/s 'bins' and calculate the percentage of values falling within each bin. 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 6.1 (bottom) compares the residuals against the normal distribution (straight line) after the tail bias was removed; notice the closer agreement. The completed simulations (200 synthetic datasets) were combined to form the resulting distribution of the MC simulation. Figure 6.2 shows the CDF of both the observed *gust* speed and the MC simulation. The density functions are also shown (right y-axis).

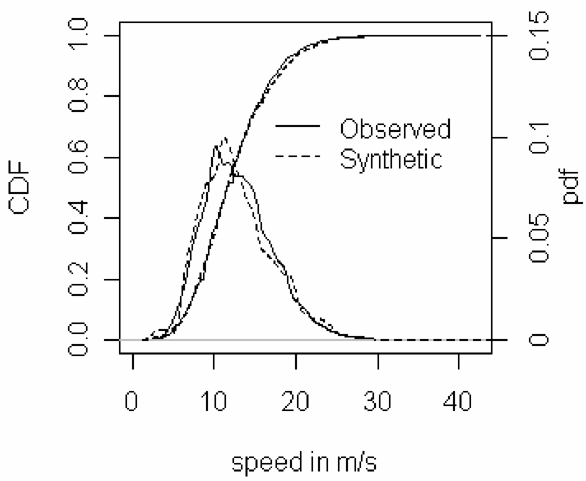


Fig. 6.2. Observed and synthetic prob. functions.

The GPD was used to calculate the RP for the synthetic *gust* speed dataset generated by the MC process. Figure 6.3 shows the results. Since the synthetic dataset is equivalent to 2512 years of data, the confidence interval has narrowed reflecting a greater reliability in the calculations. Comparison of Figures 3.1 and 6.3 shows close agreement between the *mean* of the synthetic and the observed wind speed return periods; this is expected since both have a similar distribution function as shown in Figure 6.2.

The MC process explained above was run 999 times. Each time one large vector of *gust* speeds made up of the 200 synthetic datasets was produced, with data equivalent to 2512 years. Return periods, without confidence intervals, for each of the 999 runs were calculated and plotted as shown in Figure 6.4. A summary of the results is presented in Table 6.1.

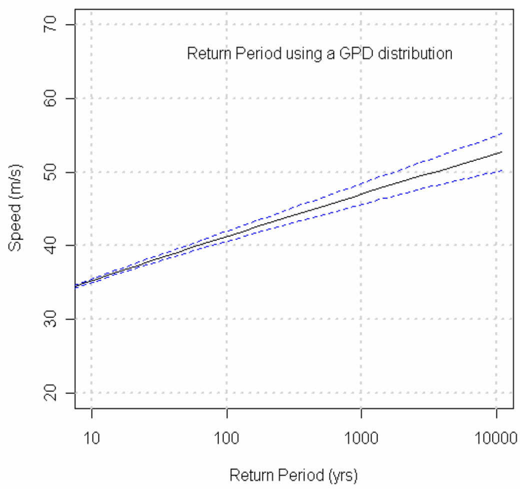


Fig. 6.3. Return periods of a synthetic dataset.

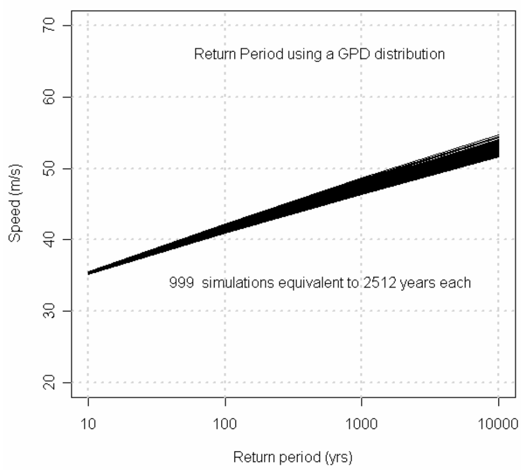


Fig. 6.4. Return periods of 999 simulations

Table 6.1. Summary of Monte Carlo simulation wind speeds (Sydney airport 1973-2005).

Return period (years)	Mean	Standard deviation	Min	Max
10000	52.9	0.41	51.6	54.7
1000	47.2	0.28	46.4	48.6
500	45.4	0.26	44.8	46.7
100	41.4	0.19	41.0	42.3
10	35.3	0.07	35.2	35.6

The results show a narrow band of RP wind speeds with a small standard deviation, indicating a very stable process. The synthetic datasets contain a significantly 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. This produces a smaller range for confidence limits whilst assuming that the full range of possible outcomes is encompassed by the dataset utilised (i.e. both the distributions of means and gust to mean ratio). The Sydney airport wind speeds at RP of 10, 100, 1000 and 10000 years produced by the statistical model (Fig. 3.1) compare well with the Monte Carlo results of Figure 6.4 as shown in Table 6.2. The same characteristic was observed for the 'Sydney region' as shown in Case 4, bottom of Table 6.2.

The input datasets discussed in Section 1 were run to further test the simulation algorithm. In cases 1, 2 and 3 different ranges of the half-hour dataset were used for the calculation of a *daily mean* and for *gust to mean* sampling. These simulations produce identical results for each range as shown in Table 6.2. Case 6 is a mixed period case, it also uses the half-hour dataset but the range is different for *mean* calculation (1952-2005) and for *gust to mean* sampling (1994-2005). Notice that the simulation in Case 5, the dataset mixed case (*mean* from the 3-hour while *gust ratio* from the half-hour) and Case 6 produce similar results to the other cases. The results also show that the MC simulation produces similar results to the Statistical Model. In practical terms the MC methodology provides upper and lower bounds for the statistical results. This characteristic can be used in the validation of the statistical model.

Mixing input datasets for the MC simulation is an important feature as it allows the calculation of wind hazard in locations where wind stations only record *mean* values. In addition, for non-metered sites (i.e. no observations) *gust ratio* distributions for either nearby/neighbouring stations or similar landscapes could be utilised in conjunction with 3D modelled *mean* wind calculations to determine the *gust* hazard using the MC simulation.

7. CONCLUSIONS

A MC simulation method to calculate wind hazard has been developed at Geoscience Australia. The model has been used to validate the results of a recently developed Statistical Model of wind hazard. A large number of MC simulations were run which produced RP winds within a narrow band providing upper and lower bounds for the statistical results. Taking advantage of the large number of equivalent years of data generated by the method, better defined RP winds (narrower confidence intervals) were generated. In these cases, the results of the statistical method were found within this interval adding confidence to the quality of the data used in the wind hazard studies.

Table 6.2. RP of wind speed in m/s. Comparison of MC (mean) and statistical results.

Station \ RP	10	100	1000	10000
Sydney airport:				
Statistical model	35	41	46	51
Case 1 (1952-2005, half-hour dataset)	35	41	47	53
Case 2 (1973-2005, half-hour dataset)	35	41	47	53
Case 3 (1994-2005, half-hour dataset)	35	41	47	53
Case 5 (1973-2005, 3-hour dataset)	35	41	47	53
Case 6 (mixed, half-hour dataset)	35	41	47	52
Sydney region:				
Statistical model	34	40	46	52
Case 4 (1952-2005, half-hour dataset)	35	41	47	53

8. REFERENCES

Bureau of Meteorology (2006). Basic Climatological Station Metadata – Current Status. Metadata compiled: 29-Sept. 2006.

Chen K. (2004). Relative Risks Ratings for Local Government Areas. Risk Frontiers quarterly newsletter, Macquarie University. Vol. 3, issued 3, March 2004.

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

Muirhead I.J., Grant I.F., Jacob D. and Glowacki T.J. (2005). Developments in Climate Data for a Sustainable Future. Proceedings of the ANZSES Solar 2005 conference. Dunedin NZ. Nov 2005.

Sanabria L.A. and Cechet R.P. (2007). A Statistical Model of Severe Winds. Geoscience Australia, Geocat No. 65052, 2007 (in press).

Zeileis A. and Grothendieck G. (2005). The "zoo" package V1.0-3.Dec. 2005.