

Extreme value analysis for gridded data

Sanabria L.A. and R. P. Cechet

Geoscience Australia, Canberra, Australia

Email: {augusto.sanabria;bob.cechet}@ga.gov.au

Abstract: The Risk & Impact Analysis Group of Geoscience Australia has been developing models to assess the hazard and risk produced by a number of natural phenomena. This paper describes a model to assess severe wind hazard over a region rather than at a recording station. The model integrates three sub-models: a statistical model that calculates return periods for the event using extreme value distributions; a model to extract and process wind speeds from a high-resolution (regional) climate model; and a Monte Carlo simulation model to generate wind gust speeds from mean wind speeds. Large scale high resolution gridded data requires a fast, efficient way to calculate wind hazard. A computer-based algorithm to achieve this aim is presented in this paper. To illustrate the methodology, wind hazard calculation over the Australian island state of Tasmania will be presented.

Keywords: Wind hazard; Natural disasters; extreme value distributions; Monte Carlo simulation; high-resolution climate models.

1. INTRODUCTION

The need to develop a new approach to understand and manage the risk posed by natural hazards in Australia has been acknowledged and emphasised by Australian Commonwealth and state governments. To this effect, the Council of Australian Governments (COAG) commissioned a review of natural disaster relief arrangements in June 2001. A report with the results of the review was published by the Department of Transport and Regional Services in early 2004 [DOTARS, 2004]. The report proposes a fundamental shift in focus beyond relief and recovery towards cost-effective, evidence-based disaster mitigation. Consequently, while disaster response and reaction plans remain important, the move is now towards anticipation and mitigation of natural hazards.

In this context, Geoscience Australia (GA) is developing risk models and innovative approaches to assess the potential losses to Australian communities from a range of sudden impact natural hazards. These models aim to define economic and social impacts of natural hazards in a consistent fashion to allow the direct comparison of risks from different hazards at one location and for a single hazard over different regions. Currently, GA is focused on developing probabilistic risk assessment models for earthquakes, tsunamis, and severe winds [Middelmann, 2007].

Severe wind is one of the major hazards facing Australia. 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 damage to Australian residential buildings [Chen, 2004].

This paper presents an overview of the methodology developed to assess wind hazard in the non-cyclonic regions of Australia. Hazard estimation for the island state of Tasmania is presented.

2. MODEL DESCRIPTION

The severe wind hazard model was initially developed for location-based analysis and later extended to gridded data. The latter allows wind analysts to assess hazard over a region

using climate simulated data. Using simulated data, it is also possible to study the impact of climate change on wind hazard.

The model integrates three sub-models:

- A Statistical Model (ie. data-based model) to quantify wind hazard using extreme value distributions;
- A high-resolution regional climate model (RCM) which produces gridded hourly “maximum time-step mean” wind speed and direction fields; and
- A Monte Carlo method to calculate severe wind hazard produced by gust winds using results from the Statistical Model.

2.1 Statistical model

The core of the statistical wind hazard model is the calculation of return periods by fitting extreme value distributions to the given data. If annual observations are used the return period (RP) is defined as the probability that a geophysical variable’s extreme value (rainfall, wind speed, temperature, etc.) is exceeded on average once a year,

$$RP = 1/(1 - CDF) \quad (1)$$

Where CDF is the cumulative distribution.

More commonly the inverse of (1) is taken and the RP expressed as the average number of years, which takes for the event to be exceeded, ie. if the probability of exceeding the event in a single year is 0.02, we can say that the event can be exceeded, on average, once in 50 years (1/0.02)

If daily maxima are used for the analysis, as in the ‘peaks over threshold’ method, the expression has to be modified to take into account the number of observations per year. The new expression is given below [Gilleland and Katz, 2009],

$$RP = 1/\{(1 - CDF)\}/nopy \quad (2)$$

Where: nopy = number of observations per year

The reason for the popularity of RP in natural phenomena studies is the possibility of extrapolating the RP curve well beyond the range of the observations; this is achieved by using extreme value distributions (EVD). There are two basic methods to fit EVD to a given dataset: a) The block maxima method in which a number of EVD are fitted to the annual maxima (the EVD used in this case are the generalised extreme value distributions or GEV. b) The ‘Peaks over threshold’ method in which the generalised Pareto distribution (GPD) is fitted to values exceeding a given threshold. The latter is the preferred method when maximum daily observations are available [Holmes, 1999].

The major limitation of the GPD in practical work is the selection of the appropriate threshold ‘u’ for the given dataset. Values above the threshold are considered for fitting the GPD. For this reason, the GPD is very sensitive to the threshold selection. To illustrate the problem, consider the curves of wind speed return period using the Sydney airport dataset as shown in Figure 1. This and other wind datasets used in this study were provided by the Australian Bureau of Meteorology (BoM). The dataset’s range of years is shown in the x-axis (1939-2005). The curves were generated by fitting GPD distributions with different thresholds ‘u’. The right-hand side table shows the threshold used, the number of observations in the dataset exceeding the threshold and the shape parameter of the GPD for that threshold.

Selecting the appropriate threshold can be a very difficult task. Although there are methods to help modellers with the selection process they are mostly visual, subjective techniques, prone to producing inaccurate results and inappropriate for large scale applications. So to model wind speeds using GPD distributions it is necessary to develop a technique for

automatic selection of the appropriate threshold for a given dataset. One such technique is briefly described below.

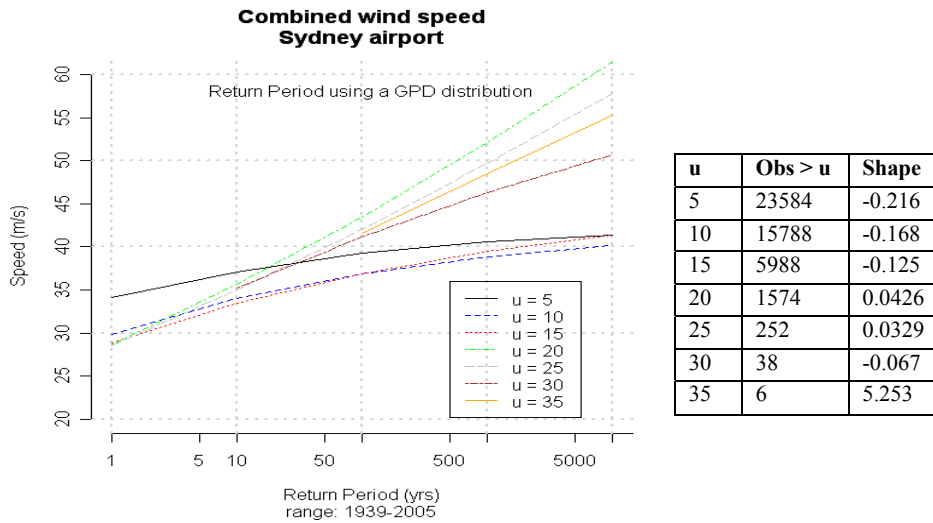


Fig. 1. GPD sensitivity to threshold u.

2.1.1 Algorithm for automatic selection of threshold

Such an algorithm can be developed by observing that there are two different regions in Figure 1: Region 1 (with $u = 20, 25, 35$) is characterised by a shape parameter greater or equal to 0 which produces unbounded curves; these types of curves are inappropriate for modelling wind speed which is a naturally bounded phenomenon [Lechner et al., 1992]. Region 2 ($u = 5, 10, 15, 30$) is characterised by a negative shape parameter which results in bounded curves appropriate for modelling wind speed; for this reason we call this the *feasible* region. Notice however that the curves produced by thresholds 5, 10 and 15, are flat with quick convergence to a very low value and hence are not appropriate for modelling wind speeds. From Figure 1 it is clear that the appropriate threshold from those shown in Figure 1 is $u = 30$ (m/s) which produces a bounded curve (one which asymptotically converges to a limiting value).

The algorithm generates GPD curves in the *feasible* region in steps of 0.25 (m/s) and selects the appropriate threshold for modelling the given dataset within this region. It is generally the threshold producing the highest return-period curve. Based on return periods generated for a number of BoM datasets, a set of rules for selection of the appropriate threshold were compiled and coded to produce the automatic algorithm [Sanabria and Cechet, 2007]. Figure 2 shows the selection algorithm at work using the same dataset as before; RP curves in steps of 0.25 have been generated in the feasible region (the corresponding threshold is shown in the left hand side box). In this case the algorithm returns 29.0 (m/s) as the appropriate threshold to fit the GPD to the given dataset.

In wind hazard applications it is important to separate the components of the wind dataset such as thunderstorm, tornadoes, synoptic winds, etc. because they pose a different hazard to people and the built environment. The Statistical Model can also separate these components using the weather description dataset provided also by BoM. Figure 3 illustrates the case: the Sydney Airport wind speed dataset has been separated into thunderstorm and synoptic winds and RP curves using the GPD distribution have been calculated for each component. The original completed dataset, called ‘combined winds’, is also shown [Sanabria and Cechet, 2007].

2.1.2 Confidence interval

Calculation of return periods should be considered incomplete if a confidence interval for the results is not presented. A confidence interval shows the range of values in which the true value of the RP exceedance level lies for a given probability. In this work we are interested in finding confidence intervals with 95% probability, in other words, we want to find the RP exceedance level of wind speed with the interval in which the true value of the exceedance level can be found in 95% of cases.

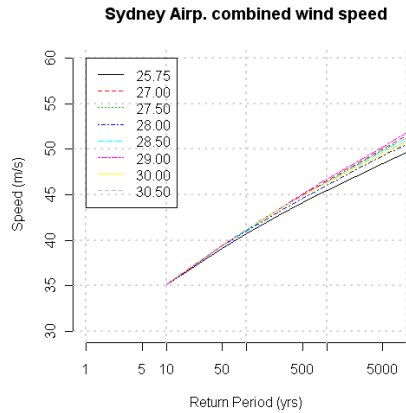


Fig. 2. Threshold selection algorithm.

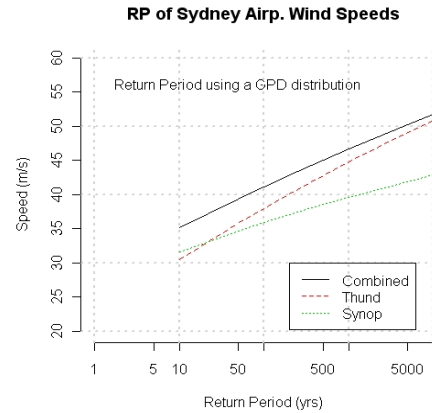


Fig. 3. Separation of wind components

The confidence interval depends on the amount and structure of the dataset, particularly the variance-covariance matrix which measures the spread of the samples around their mean.

There are two basic algorithms for calculation of the confidence intervals (CI) of results produced by extreme value distributions: The ‘Delta’ method and the ‘Profile Likelihood’ method. Both methods have been implemented in the R environment by Gilleland and Katz [2009] based on Coles [2001]. Applying the methods to temperature data, they found out that the Profile-likelihood method gives better results because it considers the asymmetry of the data [Gilleland and Katz, 2005]. Figure 4 shows the RP of the Sydney Airport combined winds with 95% Confidence Interval calculated using the Profile-likelihood method (since winds are highly asymmetric). The exceedance levels for the 500-year RP with its 95% confidence interval is shown in red, this is the prescribed wind speeds for design of structures in Australia [AS/NZS 1170.2:2002].

2.1.3 Bias correction

The bias introduced by the fact that CCAM calculates area-average values whilst the recording stations give location-based winds has to be corrected otherwise biased results could be obtained. To correct this bias consider the observed and CCAM-modelled RP of mean wind speeds as shown in Figure 5. A linear regression expression (LR) between observed and CCAM-modelled mean wind speeds can be found to correct the CCAM-modelled RP and the same expression can be used to correct the bias in the curves of gust speed RP for both station and gridded results as shown below. For gridded results we combined the data from the Launceston and Wynyard sites to undertake the correction because they are located in the same region (north of the state). The LR correction at 500-yr RP for these stations is 4.2 m/s. The correction for Hobart calculated from Figure 5 is 7.2 m/s. The correction for Tasmania was then taken as the average of the two, ie. 5.7 m/s.

2.2 Statistical model using gridded data

Having a fast algorithm for calculation of RP using a GPD allows us to obtain RP for a given region rather than a recording station as show above. In climate simulation models a region is represented as a consisting of a number of cells, the higher the grid resolution the higher the number of cells. To illustrate the application of the statistical model to a grid,

consider the island state of Tasmania, south of the Australian main land. Wind speeds produced by the simulations of a high-resolution regional climate model can be extracted and used to calculate wind hazard over this area. It is important to point out that the discussion below uses only synoptic winds, ie. the green curve in Figure 3. Work is underway to also include thunderstorm downburst gust winds in order to calculate a combined RP curve for gust wind speeds (black line in Figure 3).

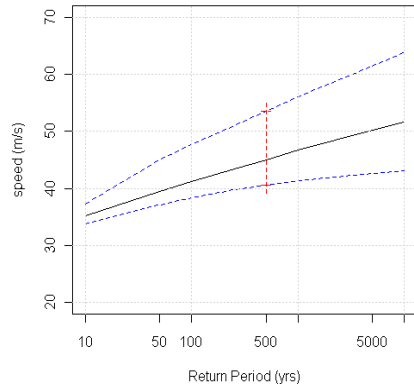


Fig. 4. RP of combined winds with 95% CI.

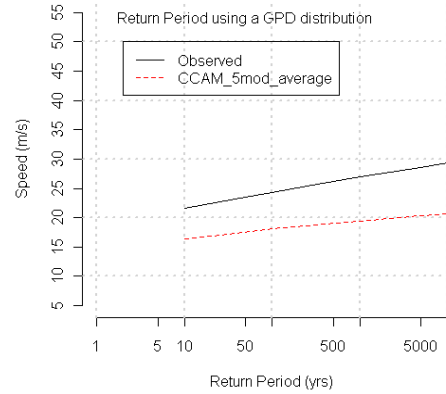


Fig. 5. RP curves of mean winds

The climate simulation data used for this project was obtained from CSIRO's Conformal-Cubic Atmospheric Model (CCAM). Simulations focusing on Tasmania, using IPCC scenario A2, for the period 1960 to 2100, were used. Five coupled general circulation models (CGCM) were used to drive these simulations (dynamic downscaling) as explained by Corney et al. [2010]: CSIRO mark 3.5; ECHAM 5; GFDL_CM 2.1; MIROC 3.2 and UK Hadley_CM3. We extracted maximum time-step hourly mean wind speeds (simulated for 10 metre height) from each of the five simulations using NCO tools [NCO, 2010]. Figure 6 shows the Tasmanian contour map with the grid. The grid has a latitude-longitude resolution of 0.1×0.1 degrees, which gives a total of 2856 cells. Then the maximum mean hourly speeds extracted in each cell were transformed to maximum daily mean speeds using the R package 'zoo' [Zeileis and Grothendieck, 2005]. The final dataset utilised is the average of the 5 simulations considered. Figure 7 shows the exceedance levels of mean wind speeds for a 500-yr RP.

For model development and validation three BoM recording stations were selected: Hobart, Launceston and Wynyard Airports (shown by the circles in Figure 7). These sites were selected because their weather stations and anemometer measurements are located at airports, avoiding the problem of houses or trees affecting the instruments, and also due to these sites having wind gust records.

2.2.1 Wind hazard for future climate

One of the advantages of using climate-simulated data is that it allows wind analysts to study the wind conditions in a given region under future climate forcing. These types of studies are important in the development of adaptation strategies. To examine future wind conditions in Tasmania, wind data from a 20-year window around 2070 was extracted from the simulations as explained before.

2.3 Monte Carlo simulation (MC)

The wind map presented above shows only mean wind speed conditions. Severe wind hazard requires however RP of *gust* wind speeds. A MC-based technique to generate wind gust from mean wind speeds has been developed [Sanabria and Cechet, 2010].

The MC works by simulating the physics of wind generation for synoptic wind conditions. 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 separately 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. The process consists of the numerical convolution of mean wind and the gust factor to produce wind gust speeds.

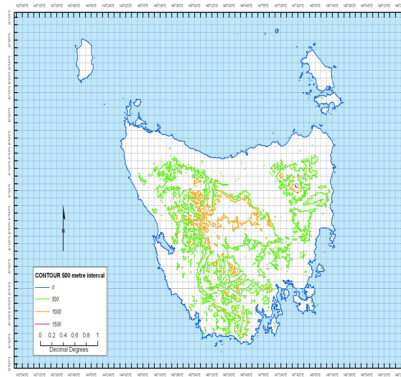


Fig. 6. Tasmanian contour map.

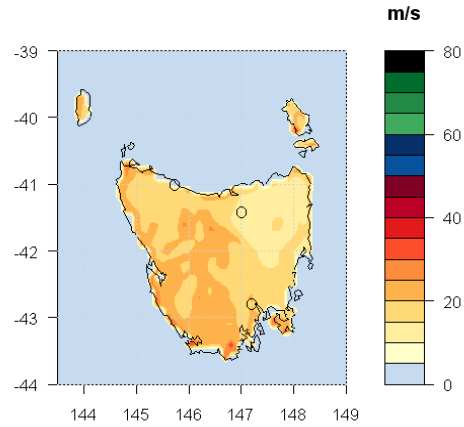


Fig. 7. Mean speed exceedance for 500-yr RP.

2.3.1 Gust factors (GF)

The gust factor (GF) is defined as the ratio of maximum wind speed (*gust*) and *mean* wind speed for the same time period. In this project the GF was calculated from the *half hourly* speed datasets provided by BoM and used for the sampling process in the Monte Carlo simulation. In practice the *x*-axis is split into a number of intervals and their corresponding GF distribution function is calculated for each interval in order to capture the characteristics of the distribution's long tail [Sanabria and Cechet, 2010]. As an illustration consider the GF calculated for Sydney Airport shown in Figure 8.

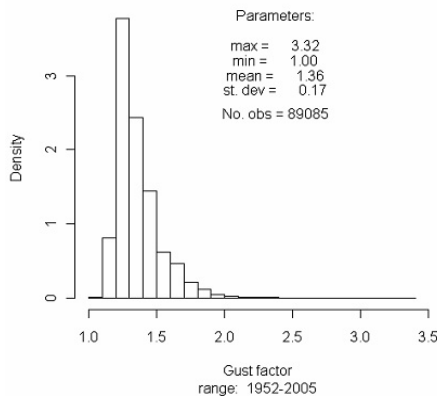


Fig. 8. Sydney Airport Gust Factor.

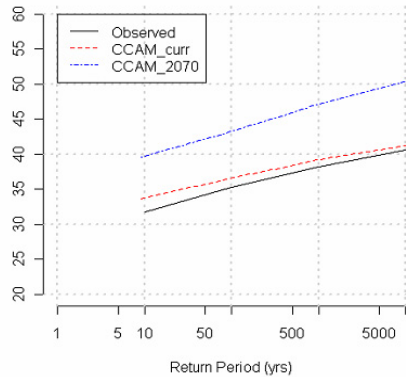


Fig. 9. RP of corrected gust speeds.

3. RESULTS

Gust wind exceedance levels at 500-year RP for Tasmania under current climate conditions and around 2070 was generated from CCAM-simulated mean wind speeds using the MC model and is presented in Figures 10 and 11. As expected the gust wind speeds follow

closely the texture of the mean values (Figure 7), in particular note the increase of synoptic wind speeds due to the effect of the mountain slope in both the west and the north-east parts of Tasmania. This is consistent with detailed studies of the effect of mountain slopes in wind speeds [Holmes, 2007].

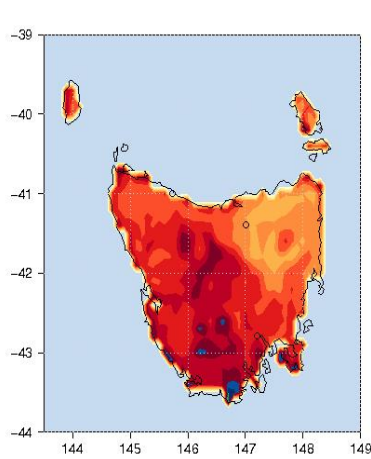


Fig. 10. Wind gust exceedance for 500-yr RP (current climate).

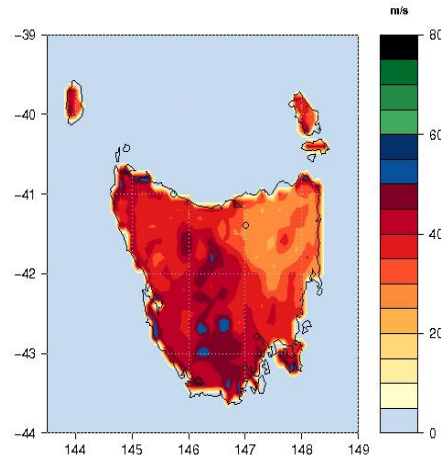


Fig. 11. Wind gust exceedance for 500-yr RP around 2070.

Comparing Figure 10 with Figure 11, it is possible to visually observe an increase in gust wind speed hazard by 2070 compared with current climate. The increase is more noticeable in the north-east coastal areas of the state and the mountainous region of the north-west. The results are highly dependent on the quality of the mean speeds provided by the climate-simulation models, and also the quality and representativeness of the observations (ie. is the record length representative of a long-term climatology for extreme wind gusts?).

3.1 Future climate gust wind hazard at the Tasmanian stations.

Plotting both the future-simulated and observed RP exceedance levels at the meteorological stations allows the determination of the trend in gust wind speed for a changing climate. Figure 9 (in previous page) shows the observed and corrected CCAM-modelled RP of gust wind for the Hobart Airport. Similar results were obtained for the other two stations but are not presented here because of space limitations; however Table 1 presents the corrected exceedance levels of gust wind speeds generated by the model for the selected stations. The station-location results confirm the trend observed in the RP maps: there is increase in wind gust hazard by 2070 with respect to current climate conditions. These results agree with similar studies carried out for Victoria [McInnes et al., 2005].

Table 1. Corrected gust wind speed exceedance levels (500-yr RP) at stations

	Hobart	Launceston	Wynyard
Observed	37.5	37.1	36.3
CCAM current	39.1	36.1	32.7
CCAM 2070	47.0	38.9	38.4

4. CONCLUSIONS

A general overview of the methodology to assess severe wind hazard has been presented in this paper. The methodology quantifies wind hazard by fitting an extreme value distribution to gust wind speed. The methodology has been used to assess wind hazard at a recording station as well as over a region using gridded model data. To assess regional wind hazard,

gridded data obtained from a high-resolution (regional) climate model is used. However climate-simulated wind speeds provide only mean wind speeds whilst in severe wind hazard the value of interests is gust wind speed. A Monte Carlo simulation model was developed to calculate gust wind from mean wind speed via sampling from the corresponding gust factor.

A confidence interval can be calculated to assess the uncertainty of the model results.

The model may also been used to assess the impact of climate change in wind hazard from future projections. The model has been used successfully to assess wind hazard in the Australian state of Tasmania under current climate and by the end of this century.

ACKNOWLEDGEMENTS

The first author wishes to thank the Climate Futures Tasmania project staff: Prof. N. L. Bindoff (project leader) for providing access to their computer resources and for the financial support to attend this conference; Mr. Leigh Gordon (IT support); and Dr. C. J. White (analyst) provided many fruitful discussions regarding extreme value analysis.

REFERENCES

- Australian/New Zealand Standard. Structural design actions. Part 2: Wind actions. AS/NZS 1170:2:2002.
- Australian Bureau of Meteorology (BoM). <http://www.bom.gov.au/>
- Brasseur O., Development & Application of a Physical Approach to Estimating Wind Gusts. *Monthly Weather Review*. Vol. 129, 5-25, 2001.
- Chen K., Relative Risks Ratings for Local Government Areas. *Risk Frontiers quarterly newsletter*, Macquarie University, Australia. Vol. 3 issue 3, March 2004.
- Coles S., *An Introduction to Statistical Modeling of Extreme Values*. Springer series in statistics. London 2001.
- Corney, S., J.J. Katzfey, J.L. McGregor, M. Grose, J. Bennett, C.J. White, G. Holz and N.L. Bindoff, Climate Futures for Tasmania technical report: methods and results on climate modelling, Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Tasmania, in press (due 2010).
- DOTARS. Natural Disasters in Australia – Review of mitigation, relief and recovery arrangements. A report to the Council of Australian Governments (COAG). 2004.
- Gillelland, E., and R.W. Katz, Analyzing Seasonal to Interannual Extreme Weather and Climate Variability with the Extremes Toolkit. National Center for Atmospheric Research (NCAR). Boulder CO, USA. 2005.
- Gillelland E., R. Katz and G. Young, “extremes”: Extreme value toolkit. R package version 1.60. 2009. <http://CRAN.R-project.org/package=extRemes>.
- Holmes J.D., *Wind Loading of Structures*. Taylor & Francis 2nd Ed. 2007.
- Holmes J.D., and W.W. Moriarty, Application of the Generalised Pareto Distribution to extreme value analysis in wind engineering. *Journal of Wind Eng. and Industrial Aerodynamics* 83 (1999) 1-10.
- Lechner J.A., Leigh S.D. and Simiu E. Recent Approaches to Extreme Value Estimation with Application to Wind Speeds. Part I: the Pickands Method. *Journal of Wind Eng. and Industrial Aerodynamics*, 41-44 (1992) 509-519.
- Middelmann M. Natural Hazards in Australia. Identifying Risk and Analysis Requirements. Geoscience Australia. 2007.
- McInnes, K.L., Abbs D.J. and J. A. 2005 Bathols. Climate Change in Eastern Victoria. Stage 1 Report: The effect of climate change on coastal wind and weather patterns. Report to Gippsland Coastal Board.
- NetCDF Operator (NCO Tools). <http://nco.sourceforge.net>
- Sanabria L.A., and R.P. Cechet, *A Statistical Model of Severe Winds*. Geoscience Australia, Record 2007/12.
- Sanabria L.A., and R.P. Cechet, Severe Wind Hazard Assessment using Monte Carlo Simulation. *Journal of Environmental Modeling & Assessment*. Volume 15, Issue 2, (147-154). 2010. Springer, Netherlands.
- Zeileis A., and G. Grothendieck, “zoo”: Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14(6), 1-27. 2005.

