

A Statistical Model of Severe Winds

GEOSCIENCE AUSTRALIA
RECORD 2007/12

by

L.A. Sanabria¹ and R.P. Cechet¹



Australian Government
Geoscience Australia

1. Risk and Impact Analysis Group, Geoscience Australia, GPO Box 378, Canberra ACT 2601

Department of Industry, Tourism & Resources

Minister for Industry, Tourism & Resources: The Hon. Ian Macfarlane, MP

Parliamentary Secretary: The Hon. Bob Baldwin, MP

Secretary: Mark Paterson

Geoscience Australia

Chief Executive Officer: Dr Neil Williams

© Commonwealth of Australia 2007

This work is copyright. Apart from any fair dealings for the purpose of study, research, criticism or review, as permitted under the Copyright Act 1968, no part may be reproduced by any process without written permission. Copyright is the responsibility of the Chief Executive Officer, Geoscience Australia. Requests and enquiries should be directed to the **Chief Executive Officer, Geoscience Australia, GPO Box 378 Canberra ACT 2601.**

Geoscience Australia has tried to make the information in this product as accurate as possible. However, it does not guarantee that the information is totally accurate or complete. Therefore, you should not solely rely on this information when making a commercial decision.

ISSN: 1448-2177

ISBN: 978 1 921236 43 3

GeoCat No. 65052

Bibliographic reference: Sanabria, L.A., Cechet, R.P., 2007. A Statistical Model of Severe Winds. Geoscience Australia Record 2007/12, 60p.

Table of Contents

INDEX OF FIGURES AND TABLES.....	v
FIGURES	v
TABLES	v
EXECUTIVE SUMMARY	vii
1. INTRODUCTION.....	1
2. DESCRIPTION OF THE MODEL	1
3. EXTREME VALUE DISTRIBUTIONS	3
3.1 GENERALISED EXTREME VALUE DISTRIBUTION	5
3.2 GENERALISED PARETO DISTRIBUTION	7
3.2.1 <i>Selection of the appropriate threshold ‘u’</i>	8
4. SYDNEY REGION WIND DATABASES	8
5. SEPARATION OF WIND TYPES.....	10
5.1 WEATHER CLASSIFICATION.....	10
5.2 MAXIMUM DAILY WIND GUST	12
6. RETURN PERIODS FOR SELECTED BOM WIND STATIONS	15
6.1 RETURN PERIODS USING A GPD.....	15
6.2 TESTS FOR THRESHOLD SELECTION.....	18
6.3 ALGORITHM FOR AUTOMATIC SELECTION OF APPROPRIATE ‘U’	21
6.4 RETURN PERIODS FOR THUNDERSTORM WINDS	23
6.5 RETURN PERIODS FOR OTHER WIND STATIONS	28
7. RETURN PERIODS FOR THE SYDNEY REGION.....	32
8. MODEL VALIDATION	38
9. CONFIDENCE INTERVAL	42
10. RELATIONSHIP BETWEEN WIND MEANS AND GUSTS	46
11. A NOTE ON UNCERTAINTY MODELLING	47
12. CONCLUSIONS.....	48
13. ACKNOWLEDGEMENTS.....	49
14. REFERENCES	49
APPENDIX A	51
RETURN PERIODS FOR BANKSTOWN, RICHMOND, WILLIAMSTOWN AND NOWRA WITH CORRESPONDING AUTOMATIC THRESHOLD SELECTION.	51
APPENDIX B.....	55
B1. THE R-STATISTICS PACKAGE	55
B2. R-SOFTWARE DEVELOPED FOR THIS PROJECT.....	56

APPENDIX C.	57
SUMMARY OF RULES FOR AUTOMATIC SELECTION OF THRESHOLD 'U'	57
APPENDIX D.	58
TABLE D1. PRESENT WEATHER CLASSIFICATION (W.M.O. PRESENT WEATHER CODE).	58
TABLE D2. PAST WEATHER CLASSIFICATION (W.M.O. PAST WEATHER CODE).	60

INDEX OF FIGURES AND TABLES

Figures

Figure 2.1.	Return period of wind speed with confidence interval.	2
Figure 3.1.	Fitting normal and GEV distributions to a histogram of random samples.	4
Figure 3.2a.	GEV distribution with three different shape parameters.	6
Figure 3.2b.	Extreme value analysis using Sydney airport data with.	7
Figure 6.1.	Sensitivity of the GPD to threshold ‘u’	16
Figure 6.2.	Sensitivity of the GPD to data range (years).	17
Figure 6.3.	Mean Residual Life plot.	18
Figure 6.4.	Model-based plot.	19
Figure 6.5a.	Probability plot with Sydney data	20
Figure 6.5b.	Quantile plot with confidence interval for Sydney data.	20
Figure 6.7.	Return period for thunderstorms (Sydney Airport).	24
Figure 6.8.	Return period of the three categories of gust wind speeds (Sydney Airport) ...	27
Figure 6.9a & 6.9b.	Return periods for (a) <i>combined</i> wind speeds and b) <i>thunderstorms</i> ...	31
Figure 7.1a.	Sydney Region <i>combined</i> wind threshold sensitivity.	32
Figure 7.1b.	Sydney Region combined wind with threshold selection.	33
Figure 7.2a.	Sydney Region <i>thunderstorm</i> threshold sensitivity.	34
Figure 7.2b.	Sydney Region <i>thunderstorm</i> with threshold selection.	35
Figure 7.3a.	Sydney Region <i>synoptic</i> wind threshold sensitivity.	40
Figure 7.3b.	Sydney Region <i>synoptic</i> wind with threshold selection.	37
Figure 8.1.	Sydney Region Wind Gusts	38
Figure 8.2.	Combined wind speeds for the Sydney region and its components	40
Figure 8.3.	Sydney region with full range of record (142 years)	41
Figure 8.4.	Comparison of full and reduced range curves.	41
Figure 9.1.	Return period (years) for Sydney airport <i>combined</i> wind.	43
Figure 9.2.	Sydney airport <i>thunderstorm</i> speed with confidence interval.	48
Figure 9.3a.	Sydney region combined wind speed with 95% confidence interval.	45
Figure 9.3b.	Sydney region <i>thunderstorm</i> wind speed with 95% confidence interval.	45
Figure 9.3c.	Sydney region <i>synoptic</i> wind speed with 95% confidence interval.	46
Figure 10.1.	Correlation between gust and mean wind speed	47
Figure A1.	Return periods for Bankstown <i>combined</i> gust wind speeds.	51
Figure A2.	Return periods for Richmond <i>combined</i> gust wind speeds	52
Figure A3.	Return periods of <i>combined</i> wind speeds for Williamtown.	53
Figure A4.	Return periods of <i>combined</i> wind speeds for Nowra	54

Tables

Table 4.1.	Wind datasets acquired from BoM.	8
Table 4.2.	Wind observing stations	9
Table 5.1a.	Present weather description.	11
Table 5.1b.	Past weather description.	11
Table 5.2.	Approximation of max daily dates to three-hourly dates	12
Table 5.3.	WS-WC for Sydney airport	13
Table 5.4.	Highest (combined) wind gust speeds for Sydney airport.	14

Table 6.1.	Parameters for Figure 6.1	15
Table 6.2a.	GPD parameters (Sydney airport dataset)	22
Table 6.2b.	GPD parameters (Sydney airport; thunderstorms)	24
Table 6.3.	Highest 25 thunderstorm wind speeds for Sydney airport	25
Table 6.4.	Highest <i>combined</i> wind speeds for Bankstown airport	28
Table 6.5.	Highest (<i>combined</i>) wind speeds for old Richmond RAAF observing site.....	29
Table 6.6.	Temporal range of max daily wind gust speed datasets	29
Table 6.7.	Williamstown and Nowra highest gust wind speeds.	30
Table 6.8.	Return periods of <i>combined</i> Wind Gust Speeds in m/s.	30
Table 6.9.	Return periods of <i>thunderstorm</i> Wind Gust Speeds in m/s	30

Executive Summary

Severe wind is one of the major natural hazards facing the Australian continent. While cyclonic winds are a northern Australian phenomenon, winds driven by thunderstorms and tornadoes also inflict serious damage and sometimes cause loss of life in southern states. Severe winds are responsible for about 40% of damage to Australian residential buildings. Impact on Australian houses is significantly higher than for other natural hazards such as floods (22%), bushfires (19%), and earthquakes (6%) (Chen, 2004).

This report presents a recently developed statistical model to investigate severe wind hazard. The model is part of the mathematical models and corresponding software tools for hazard analysis and assessment developed by the Risk and Impact Analysis Group (RIAG) at Geoscience Australia. The group's objective is to develop state of the art tools to study the risk posed by natural hazards in the Australian region. This work is part of the Disaster Mitigation Australia Package coordinated by the Department of Transport and Regional Services (DOTARS). DOTARS has commissioned Geoscience Australia to develop national, systematic and rigorous risk assessments not only for severe winds but also for a variety of sudden impact natural hazards.

Wind hazard is assessed by calculating return periods of maximum wind gust (generally considered as 1-3 second duration gusts) from observational data. The return periods for these wind gust speeds were obtained using the application of statistical extreme value distributions. Parameters to fit these distributions were calculated from data provided by the Australian Bureau of Meteorology. Initially wind station datasets in the Sydney region are used as a case study; the work will be extended to other regions in the near future.

Validation of the model was carried out by comparing these results with those from the same region published by Holmes (2002) and by the Australian/NZ standards for wind loading of structures (AS/NZS 1170.2, 2002). To further validate the results we are developing a Monte Carlo simulation technique to generate return periods using synthetic gust speed datasets based on observed mean wind speeds and also the short duration gust to mean wind ratio from the later part of the observed record.

The methodology for wind hazard assessment and results presented in this report will be used by others in the RIAG to estimate the risk that severe winds pose to commercial and residential buildings and to critical infrastructure. The risk depends on the wind hazard weighted by a number of multipliers acting on a building stock of known exposure and vulnerability (Cechet et al., 2007). This information will also be used by emergency authorities to build preparedness, take measures to mitigate future impact, educate and make communities more resilient, and develop effective disaster recovery procedures.

1. Introduction

The aim of this report is to present a statistical model for the study of severe wind hazard by assessing the return period of maximum wind gust from observational data

The statistical model described in this report is part of the Risk and Impact Analysis Group's (RIAG) mathematical models and corresponding software tools for hazard analysis and assessment. The aim of the RIAG is to develop state of the art tools to study the risk posed by natural hazards in the Australian region. Although statistical models have limitations, they are useful for assessment of potential severe winds at discrete points in a region and are fundamental for the calculation of hazard from records of observational gust and mean wind speed datasets. Statistical methods are exclusively utilised in the current Australian wind loadings standard (AS/NZS 1170.2, 2002).

The focus of this report is on the calculation of return periods for wind gust speeds using statistical extreme value distributions. The distributions are calculated from data obtained from a number of observational wind stations. Datasets of wind records were acquired from the Australian Bureau of Meteorology (BoM) for these studies. Initially wind station records from the Sydney region are used as a case study; the work will be extended to other regions in the near future, and will be used to support the review of the current wind loadings standard.

The software for these studies was developed within the R-statistical package on a computer running the LINUX operating system. A general description of the software is presented in Appendix B.

The core of the model is the development of effective algorithms for separation of wind components and for selection of the appropriate threshold to fit Generalised Pareto Distributions (GPD) to given data. To assess the uncertainty inherent to model results confidence intervals for most return periods have been calculated and are presented in Section 9.

Validation of the model was carried out by comparing our results with results published by Holmes (2002) and by the Australian/NZ standard for wind loading of structures (AS/NZS 1170.2, 2002). We have concentrated mainly on three meteorological observing stations in the 'Sydney region' as defined in Holmes (2002).

2. Description of the model

The model consists of fitting extreme value distributions to meteorological data (wind measurements) so that return periods well beyond the range of the available data can be calculated. A return period 'RP' is a method to assess the maximum wind speed that could be exceeded on average once every return period (in years). Mathematically it is the inverse of the complementary cumulative distribution function

$$RP = 1/(1 - CDF) \quad (2.1)$$

where: CDF = Cumulative distribution function
RP = Return Period

Thus a 50-year return period wind speed has a probability of exceedance of "one in fifty" or 2% in any one year.

Expression (2.1) is valid for datasets of annual maxima which have one observation per year, these are the typical datasets used in extreme value distributions. If daily maxima are used, as in the ‘peaks over threshold’ method (Coles, 2001), the expression has to be modified to take into account the number of observations per year as explained in Section 6.1. The new expression is given by Equation (2.2) (Gillelland and Katz, 2005a),

$$RP = 1/\{(1 - CDF)/nopy\} \quad (2.2)$$

where: $nopy$ = number of observations per year

Utilising extreme value distributions, it is possible to calculate the probability of exceedance of maximum wind speeds in the next 50, 100 or 1000 years and beyond, based on records spanning only 20 or 30 years. Confidence limits for the results can be calculated. To illustrate the concept of return periods consider Figure 2.1. The circles show the return period (in a logarithmic x-axis) of the Sydney airport maximum daily wind speed calculated using equation (2.2). The parameter “ $nopy$ ” were calculated by noticing that there are 24107 non-missing records in the dataset covering a range of 66 years (1939-2005), hence

$$nopy = 24107/66 = 365.25 \text{ obs/year}$$

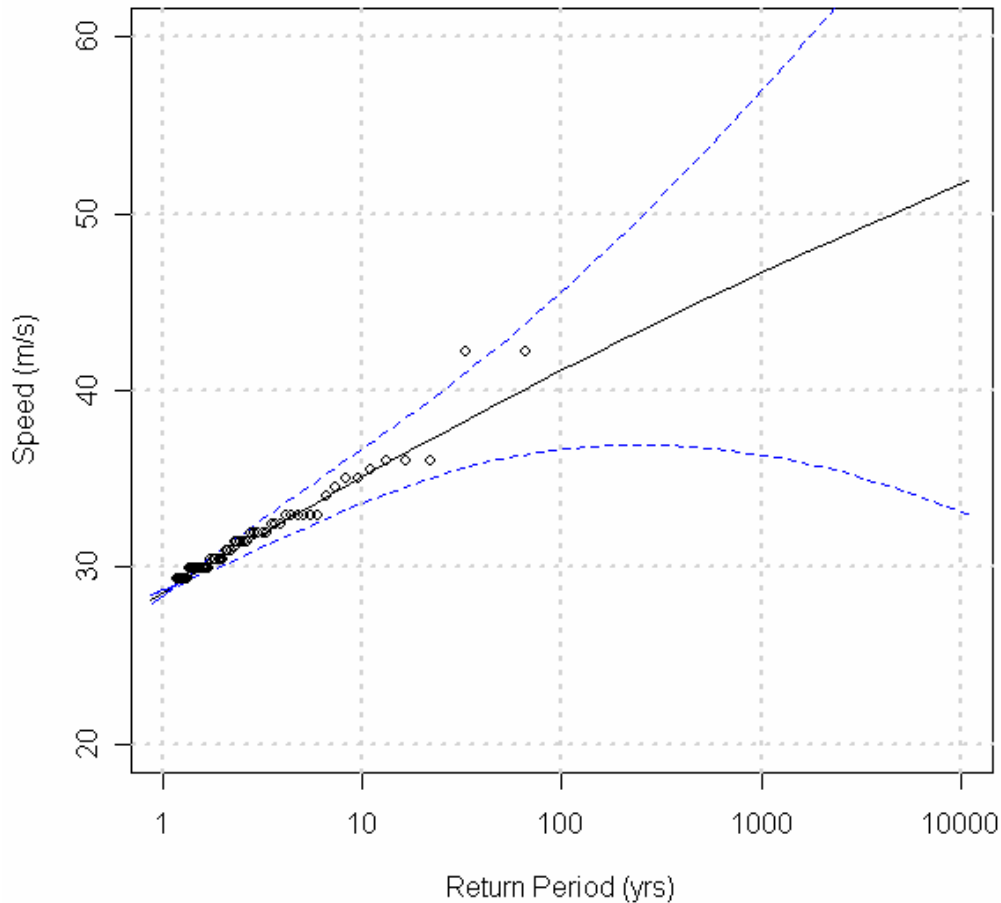


Figure 2.1. Return period of wind speed with confidence interval.

The solid curve shows the corresponding fit using an extreme value distribution. The dotted curves (in blue) show the 95% confidence interval. The extreme value curve fits the observations well except for the two outliers at 42.2 m/s. Notice also that the curve has been extended up to 10000 years using only 66 years of data. This report details the development of an effective methodology to produce curves similar to Figure 2.1 for wind gust speeds from a number of wind stations in South East Australia.

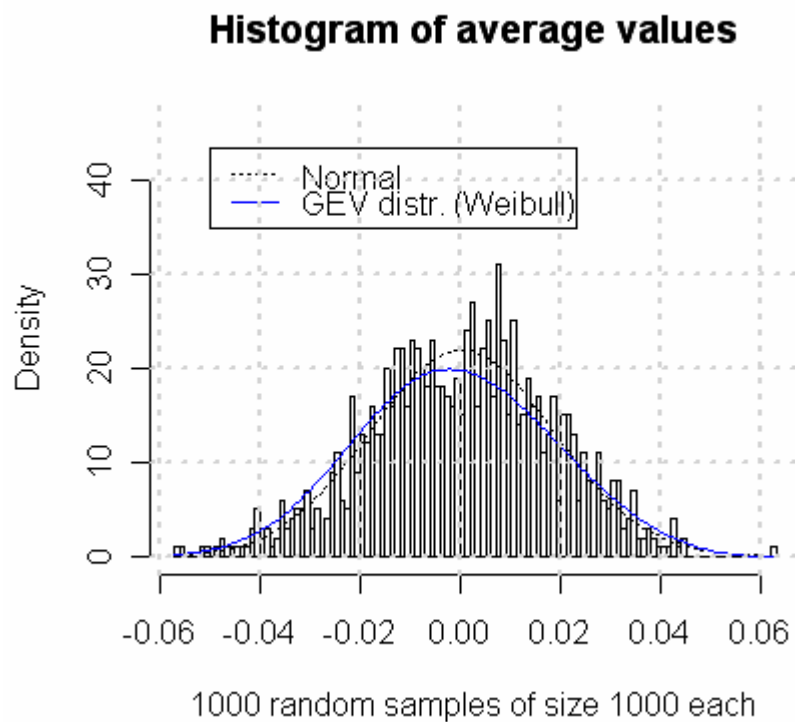
As in other mathematical models it is important to be aware of the model limitations otherwise incorrect conclusions can be drawn. The main limitations of any model based on extreme value distributions are:

- The results are valid in the limit, an idealised mathematical space of infinite observations, so there is a degree of uncertainty when used with finite samples,
- The short length of records may not be representative of either a location or a region. North Queensland, for instance was struck by two cyclones of category 4-5 in recent years; most observational datasets of wind speed in the region will not include severe gusts associated with these types of events (i.e. the return period for intense cyclonic activity at a location is much greater than the length of the current observational record),
- The model itself is based on observational data that may have large but unknown errors (i.e. extreme gusts are not calibrated; instruments are calibrated for mean response), and
- The records may not be based on independent observations.

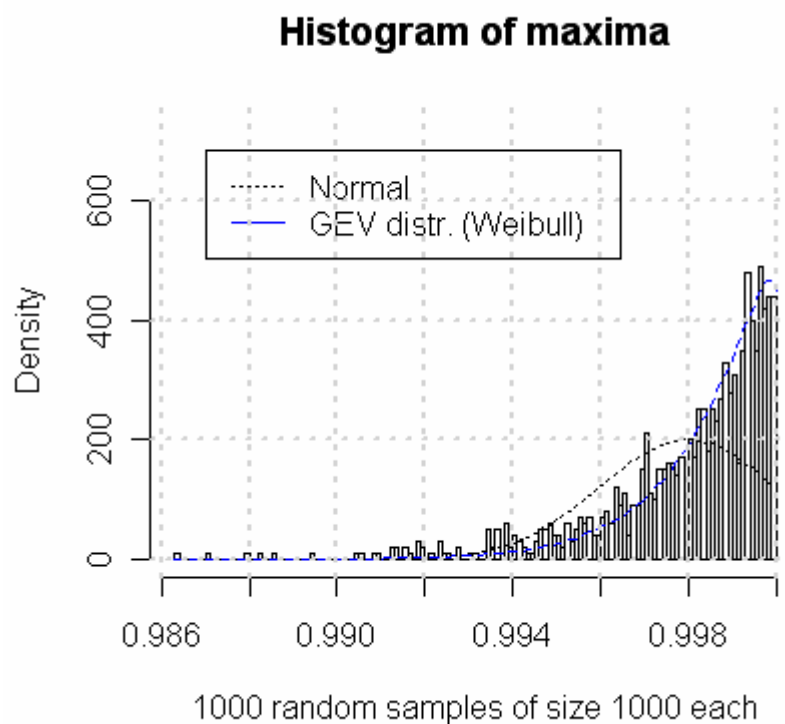
In spite of these limitations, the model discussed here can be useful to assist engineering, planning and emergency authorities with their decision-making regarding hazardous winds.

3. Extreme Value Distributions

Extreme value distributions are asymptotic functions that allow modellers to extrapolate limited data samples to possible maximum values by fitting these functions to the data. The theory behind extreme value distributions is similar to the Central Limit Theorem (CLT); both infer the limiting distribution of independent, identically distributed (iid) random variables. According to the CLT, the mean value of a sample of iid random variables converges to a standard normal distribution. Similarly if the maxima of a large number of iid random variables converge to a distribution, this distribution has to be a member of the Extreme Value Distributions (Jagger and Elsner, 2006). Figure 3.1 illustrates the case (idea borrowed from Gilleland and Katz, 2005b). In Figure 3.1a the histogram of mean values of 1000 vectors, each containing 1000 samples generated at random between -1 and 1, is shown. A normal (solid line) and a Generalised Extreme Value distribution (dotted line) are used to fit the histogram. It is clear that the normal distribution is better at fitting the histogram of mean values than the Generalised Extreme Value distribution (GEV). Figure 3.1b shows the histogram of maxima, this time the GEV is better at fitting the histogram of peak values of each vector than the normal distribution. Notice that the GEV distribution appropriate to fit the given data is a GEV distribution type III or Weibull distribution.



(a)



(b)

Figure 3.1. Fitting normal and GEV distributions to a histogram of random samples.

Although the extreme value distributions have been derived considering a set of infinite data samples, they are regarded as good approximations to the behaviour of limited data. There are two basic types of functions to fit extreme values: Generalised Extreme Value distributions (GEV) and the Generalised Pareto Distribution (GPD).

3.1 Generalised Extreme Value distribution

A unifying expression to represent the family of extreme value distributions was introduced by Jenkinson in 1955; it is known as the Generalised Extreme Value (GEV) distribution (Coles, 2001) and is given by,

$$G(z) = \exp\{-[1 + \xi (z - \mu)/\sigma]^{-1/\xi}\} \quad (3.1)$$

the expression is defined on the set $\{z : 1 + \xi (z - \mu)/\sigma > 0\}$

where: ξ = shape parameter
 μ = location parameter
 σ = scale parameter

These parameters must satisfy:

$$-\infty < \mu < \infty, \quad \sigma > 0, \quad \text{and} \quad -\infty < \xi < \infty,$$

The value of the shape parameter defines which specific distribution is being fitted to the given data,

- If $\xi < 0$, a GEV type III or **Weibull** distribution is used. This distribution has a convex (bounded) curve.
- If $\xi > 0$, a GEV type II or **Frechet** distribution is used. This distribution has a concave (unbounded) curve.
- If $\xi = 0$, a GEV type I or **Gumbel** distribution which is a straight line, is used.

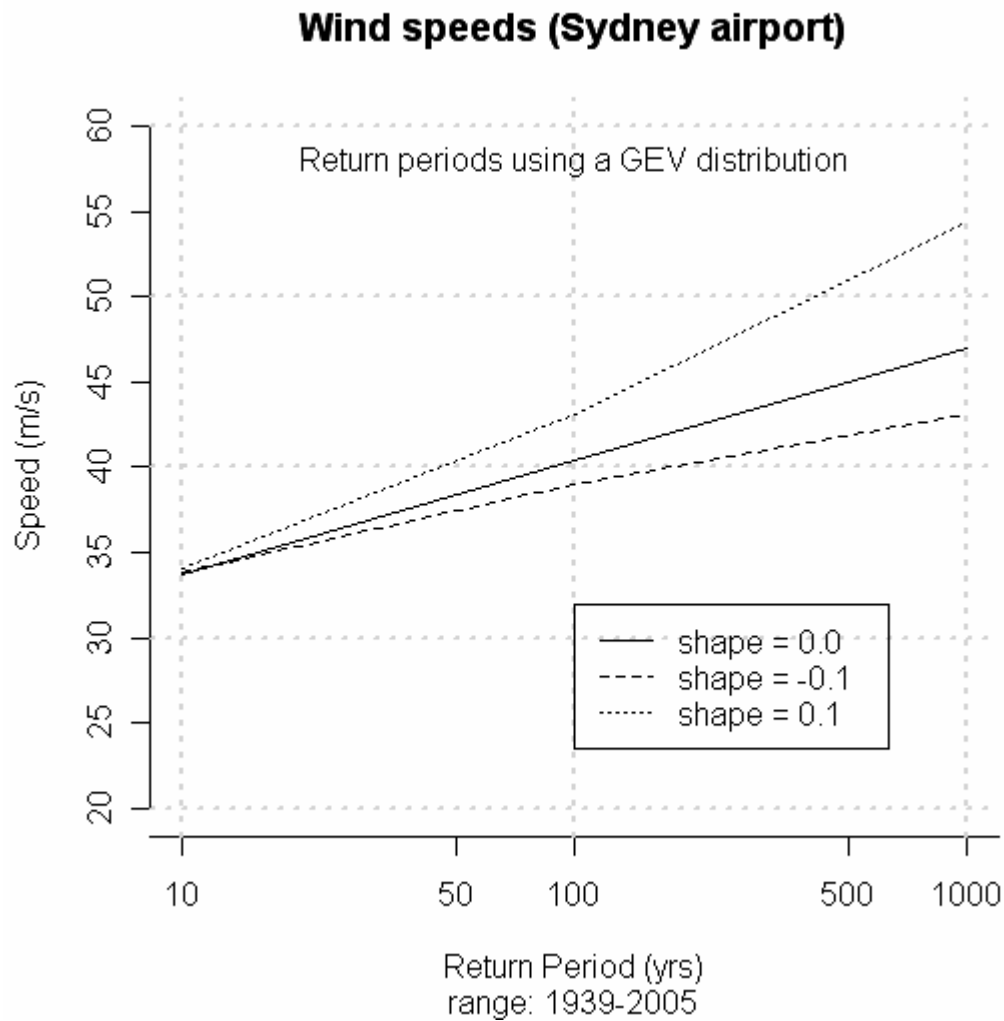


Figure 3.2a. GEV distribution with three different shape parameters.

Figure 3.2a shows the three types of curves obtained using the arbitrary shape parameter ' ξ ' indicated in the figure, that is: $\xi = 0.0$ (straight line), $\xi = -0.1$ (convex curve) and $\xi = 0.1$ (concave curve). The figure was generated using data from Sydney airport maximum yearly wind speed. Range refers to the number of years of available data used for fitting the GEV distribution.

A number of techniques to fit these distributions have been developed; the most effective are the maximum likelihood (ML) and the probability weighted moments (PWM). Figure 3.2b shows the Sydney airport data fitted using these techniques; the shape parameter was calculated from the data. Although in most cases they give similar results other examples (not presented here) show that the most effective technique to fit the available wind datasets is the ML method. Palutikoff et al., (1999) and Seguro and Lambert (2000) report similar results; for this reason the ML technique will be used throughout this work.

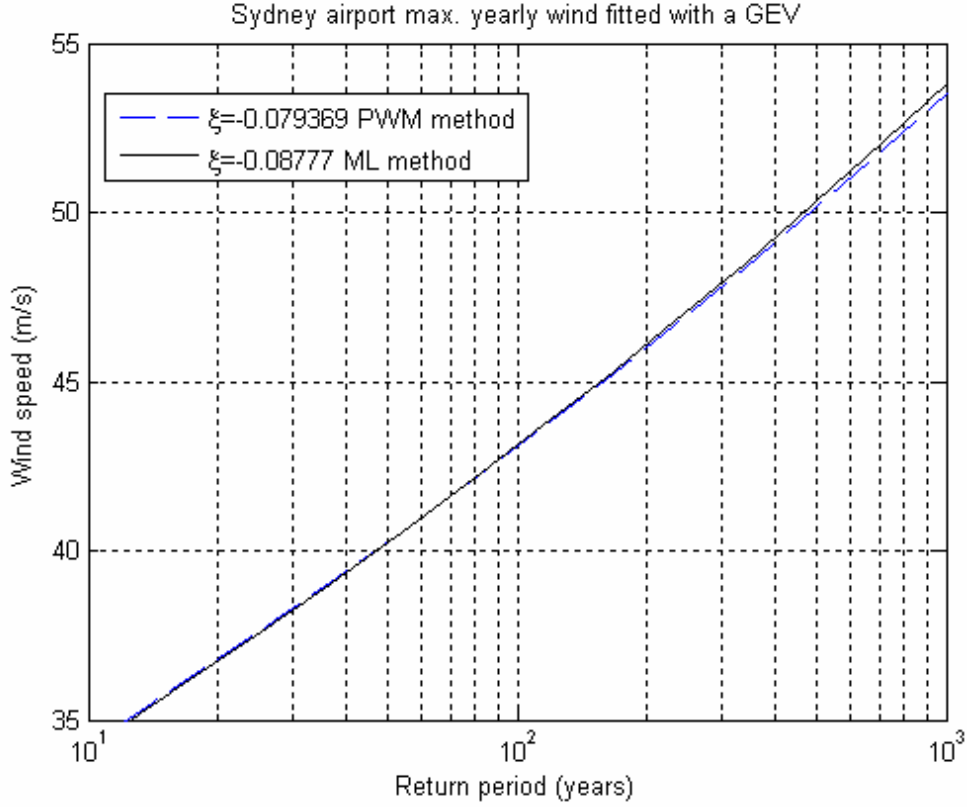


Figure 3.2b. Extreme value analysis using Sydney airport data with two different fitting techniques.

One of the problems using GEV distributions in modelling work is that only a few samples of the records available are used. A typical result uses yearly maximum wind speeds, that is, one observation per year; the other observations are not used. This is a problem because it is possible that the highest values below the maximum in a single year exceed the maximum of another year and yet they are not included in the model. For this reason a new asymptotic distribution has been introduced for extreme value analysis, known as the Generalised Pareto Distribution (GPD).

3.2 Generalised Pareto distribution

The GPD belongs to a family of threshold models because it uses all values of a dataset exceeding a given threshold. It has certain advantages; first it uses a lot more data than the GEV, secondly by setting the threshold high enough, the data will be better distributed in time, so it is likely that the data samples are independent from each other, one of the conditions of extreme value distributions.

The GPD is defined by the expression,

$$H(y) = 1 - (1 + \xi y/\tilde{s})^{-1/\xi} \quad (3.2)$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\tilde{s}) > 0\}$

where: $\tilde{s} = \sigma + \xi(u - \mu)$ and $u = \text{threshold value}$

“ σ ” and “ μ ” are the same parameters of the GEV defined above, i.e. if the original data can be fitted with a GEV distribution, values above the threshold can be fitted with a GPD, so the parameters of (3.2) can be calculated from those of (3.1).

A limited number of data samples were processed with both GEV and GPD distributions during the course of this project. The results produced by the GEV distributions were considered too poor for the type of application reported here so this report focuses only on results produced by the GPD as explained in Section 6.

3.2.1 Selection of the appropriate threshold ‘u’.

One of the problems found in fitting a GPD to given data samples is the selection of the appropriate threshold value ‘u’. High threshold values result in the selection of only a few data points, most likely not enough for a good fitting of the distribution. Low values result in too many samples which are most likely not independent from each other. On the other hand return period calculation using GPD distributions are very sensitive to the threshold selection. Although there are methods to help modellers select the appropriate threshold for a given dataset they are mostly visual, subjective techniques, prone to producing inaccurate results and inappropriate for large scale applications. 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. An algorithm to do this is presented in Section 6.

4. Sydney region wind databases

The Wind Risk Activity (WRA) at Geoscience Australia acquired a series of datasets from Bureau of Meteorology (BoM) containing records of wind speeds and directions from a number of wind stations located in southern New South Wales including a few stations in the Sydney area.

Some of the problems of these datasets are that they include wind speed records for different ranges of years, and even more problematic, the interval at which these speeds were recorded is not consistent. The interval varies from one minute to one day. Table 4.1 presents the type of wind datasets available for these studies. To save paper only the datasets corresponding to Sydney airport, wind station 66037, are shown. The dataset titled ‘Max. yearly’ (bottom of table), was calculated during the course of this project. ‘Range’ refers to the years of data records in the dataset. ‘Dir’ is the directory where the dataset is located in the LINUX machine.

Table 4.1. Wind datasets acquired from BoM

SYDNEY AIRP	Range	Dir
1min mean	1998-2005	/mnt/store/Winddata/Sydney/1min
½ hour mean	1952-2005	/mnt/store/Winddata/Sydney/half_hr
3 hour mean	1939-2005	/mnt/store/Winddata/Sydney/3hour
Max. daily	1939-2005	/mnt/store/Winddata/Sydney/MxDai_gust
1min gust	1998-2005	/mnt/store/Winddata/Sydney/1min
10min gust	1952-2005	/mnt/store/Winddata/Sydney/half_hr
3 hour wind type	1939-2006	/mnt/store/Winddata/Sydney/w_char
Max. yearly	1939-2005	/mnt/store/Winddata/Sydney/MxYrly

Similar datasets for 23 wind stations were also acquired from BoM, and some of them were used in this project. Table 4.2 lists all 23 wind stations whose data was acquired for this project; the number in the second column is the identification number given by BoM to the wind station.

Table 4.2. Wind observing stations

1	66037	SYDNEY AIRPORT AMO
2	66062	SYDNEY (OBSERVATORY HILL)
3	66137	BANKSTOWN AIRPORT AWS
4	67033	RICHMOND RAAF
5	67105	RICHMOND RAAF
6	63039	KATOOMBA (MURRI ST)
7	61087	GOSFORD (NARARA RESEARCH STATION) AWS
8	68192	CAMDEN AIRPORT AWS
9	68188	WOLLONGONG UNIVERSITY
10	61078	WILLIAMTOWN RAAF
11	68076	NOWRA RAN AIR STATION
12	68072	NOWRA RAN AIR STATION AWS
13	63224	LITHGOW (BIRDWOOD ST)
14	63291	BATHURST AIRPORT AWS
15	68102	BOWRAL (PARRY DRIVE)
16	68239	MOSS VALE AWS
17	70330	GOULBURN AIRPORT AWS
18	70014	CANBERRA AIRPORT
19	62101	MUDGEE AIRPORT AWS
20	61275	SINGLETON ARMY
21	61397	SINGLETON STP
22	61388	MAITLAND VISITORS CENTRE
23	61054	NELSON BAY (NELSON HEAD)

Representative wind speeds for Sydney were obtained from the Sydney airport station (number 66037) not from the Sydney Observatory Hill (66062). Wind speeds measured at Observatory Hill are not representative of the Sydney area. On the one hand they were observed on an elevated region compared to the surroundings, where winds are subjected to topographic acceleration. On the other hand the station has been affected by the construction of tall buildings in recent years. To assess the differences, return periods of both stations will be presented in Tables 6.9 & 6.10.

Another important dataset acquired from BoM is a dataset giving three-hour present and past weather classification according to the conventions of the World Meteorological Observation Codes (W.M.O.). This dataset includes the date and time of the recording, a description of the ‘present weather’ defined as the weather observed at the station during the past hour; and a description of the ‘past weather’ defined as the weather during the past three hours but not including the weather during the past hour. This information will be used to develop an algorithm to separate winds into three classes as explained in the next Section.

It is important to point out that all these datasets were processed as they were received from BoM. Except for correction of obvious errors in wind speed values, quality analysis of these datasets was not attempted.

5. Separation of wind types

A number of researchers have recognised the need for a more detailed analysis and modelling of the different components of the wind data sets. In particular they have shown more realistic results when each wind component is modelled independently (Gomes & Vickery, 1978) (Holmes, 2002).

The model discussed here follows the technique proposed by Holmes (1999) which allows the extraction of *thunderstorm downbursts* or *mesoscale* winds from the dataset. The remaining dataset is termed *synoptic* winds and the complete dataset is called the *combined* wind dataset. Regions also experiencing wind gusts from tropical cyclones and/or tornadoes could have this component dealt with separately.

It is important to separate the wind gusts into these components or classes because wind multipliers as detailed in the Australian/NZ Wind Loadings Standard (AS/NZS 1170:2, 2002) affect these classes differently. In particular, the topographic multiplier for thunderstorm downbursts is significantly reduced compared to synoptic winds. Synoptic winds travel chiefly in a horizontal direction towards a target location whereas thunderstorm downbursts have a significant vertical component and are affected to a lesser extent by topographic acceleration.

Discussion of how the wind direction and topographic characteristics affect the wind hazard is not included here; however interested readers can refer to a separate Geoscience Australia publication which deals with this problem (Nadimpalli et al., 2007).

Holmes (2002) used a visual inspection of the anemometer charts for extraction of the *thunderstorms* from the original dataset. The method is appropriate for modelling of a limited number of cases, typically for Australian major cities, but is not very efficient for the larger task of analysing and modelling Australia-wide wind hazard. For this a more efficient technique must be developed. The following section presents such a technique

5.1 Weather classification

Australia's BoM publishes a three-hour interval classification of weather conditions according to the W.M.O. code (BoM, 1982). The 'present weather' is described by 100 symbols abridged from the W.M.O. code indicating the type of weather conditions during the past hour. These symbols are numbered from 0 to 99; see Table D1 in Appendix D. From this classification, we identified 12 weather classes that can be considered as thunderstorms. They are classes 17, 27, 29 and classes 91 to 99 as presented in Table 5.1a. Classes 9, 33 and 35 refer to short-period dust storms or sand storms but they have also been included because they impede visibility and hence are related to nearby mesoscale phenomena because of the sudden change in wind conditions.

Our algorithm is based on the merging by date of the wind speed dataset with the present & past weather information. Once the 'present weather' and 'past weather' classifications are included in the wind dataset, it is possible to extract the wind speeds corresponding to thunderstorms, synoptic scale events, and also tornados and tropical cyclones if they are included in the observational record.

Ten symbols for the 'past weather' are also given to represent the most significant weather conditions within the past three hours of observation but not during the most recent hour. BoM numbers these symbols from 0 to 9; see Table 5.1b and Table D2 of Appendix D. In our algorithm we select records with the present weather numbers shown in Table 5.1a and past weather with numbers starting with 9 or 3 (Table 5.1b).

Table 5.1a. Present weather description.

Present weather number	Description
17	Thunder heard but no precipitation at the station.
27	Showers of hail, or of hail and rain during past hour, but no at time of obs.
29	Thunderstorm (with or without precip) during past hour, but not at time of obs.
91	Slight rain at time of obs; thunderstorm during past hour not at time of obs.
92	Moderate or heavy rain at time of obs; TS during past hour not at time of obs.
93	Slight snow and/or rain/hail at time of obs; TS during past hour not at time of obs.
94	Moderate or heavy snow and/or rain/hail at time of obs; TS past hour not at obs time.
95	Slight or moderate thunderstorm without hail but with rain and/or snow at obs time.
96	Slight or moderate thunderstorm with hail at time of obs.
97	Heavy thunderstorm without hail but with rain and/or snow at time of obs.
98	Thunderstorm combined with duststorm or sandstorm at time of obs.
99	Heavy thunderstorm with hail at time of obs.
9	Duststorm or sandstorm within sight of station or at station during past hour
33	Severe duststorm or sandstorm, has decreased during past hour
35	Severe duststorm or sandstorm, has increased during past hour.

Table 5.1b. Past weather description.

Past weather number	Description
0	Clear, few clouds at observing station.
1	Partly cloudy, scattered or variable sky at obs station.
2	Cloudy (broken) or overcast at observing station.
3	Sandstorm or duststorm during past 3 hours.
4	Fog or smoke or thick dust haze during past 3 hours.
5	Drizzle.
6	Rain.
7	Snow, or rain and snow.
8	Shower(s).
9	Thunderstorm, with or without precipitation.

5.2 *Maximum daily wind gust*

One of the problems with the algorithm presented in the previous section is that return periods for wind hazard studies have to be calculated using maximum daily gust speeds (max daily). Unfortunately the BoM weather classification is based on three-hour periods. In this case it is necessary to fit the max daily dates into the existing three-hour periods. To do this the dates in the maximum wind speed dataset are approximated to the next three-hour period. If the date is less than the mid point of the three-hour period, it is approximated down to the previous three-hour period otherwise it is approximated up to the next three-hour period, Table 5.2 illustrates the process.

Table 5.2. Approximation of max daily dates to three-hourly dates

Max. daily wind speed		3-hour weather dataset		
Date (y,m,d,h)	Wind speed	Date(y,m,d,h)	Present	Past
		1986,02,09,1200	3	2
		1986,02,09,1500	2	80
		1986,02,09,1800	95	95
1986,02,09,1900	19.6			
		1986,02,09,2100	5	5
		1986,02,09,0000	3	2

The date 1986,02,09,1900 (Feb-9-1986:07:00 PM) of the wind dataset is transformed into 1986,02,09,1800 because 7:00 PM is less than 7:30 PM the mid point of the period 6:00-9:00 PM. The present weather classification dataset indicates that a weather class 95 ('slight or moderate thunderstorms without hail at time of obs; TS past hour not at obs time') was present on 1986,02,09,1800. The number 95 in the 'past weather' column is a combination of the digits 9 and 5 meaning that 'thunderstorm, with or without precipitation' has been present in addition to 'drizzle', during the past three hours. When thunderstorms are extracted from the merged dataset, the 19.6 m/s wind speed will be included in the thunderstorm dataset with the date 1986,02,09,1800 and the classification 95 for present weather (9 & 5 for past weather).

It is important to point out that the algorithm also shifts the 'past weather' column one interval upwards so that wind speeds corresponding to thunderstorms in the previous interval are also selected, i.e. the wind speed corresponding to Feb-9-1986:03:00 PM, if included in the max. wind speed dataset, would also be selected in to the thunderstorms dataset. In our case the max daily for Feb-9-1986 occurs at 7:00 PM so the 3:00 PM interval is not included in the merged dataset and hence cannot be selected.

The algorithm produces a dataset of max daily wind speeds with dates moved to the three-hour records and columns giving the corresponding weather classification. However there are a lot of missing values in the three-hour weather classification dataset. The arduous task of manually inspecting BoM anemometer paper records maybe required to complete this dataset. BoM has commenced a process to systematically scan anemometer records to produce digital records for each observing station. The R-programming environment marks missing values with 'NA'. The R-package includes a function which allows the user to 'clean' a dataset of missing values and this complements our algorithm.

Table 5.3. WS-WC for Sydney airport

Date	wind speed	weather	
		present	past
2002-04-06:09:00	10.3	2	2
2002-04-07:15:00	8.8	5	5
2002-04-08:15:00	8.2	3	3
2002-04-09:18:00	11.3	2	2
2002-04-10:15:00	10.3	3	3
2002-04-11:15:00	10.8	5	5
2002-04-12:15:00	12.9	5	5
2002-04-13:21:00	13.4	3	3
2002-04-14:15:00	16.5	3	3
2002-04-15:12:00	9.8	3	3
2002-04-16:09:00	7.7	5	5
2002-04-17:12:00	9.8	16	81
2002-04-18:12:00	14.9	25	15
2002-04-19:15:00	16.5	2	15
2002-04-20:09:00	8.8	2	3
2002-04-22:12:00	23.2	50	50
2002-04-23:09:00	10.3	2	14
2002-04-24:15:00	7.2	3	3
2002-04-25:18:00	9.8	2	2
2002-04-26:12:00	11.8	2	3
2002-04-27:15:00	11.8	1	2
2002-04-28:09:00	11.3	15	15
2002-04-29:03:00	8.8	1	3
2002-04-30:15:00	5.7	14	2
2002-05-01:12:00	6.7	5	3
2002-05-02:12:00	9.8	2	3
2002-05-03:12:00	6.2	5	16
2002-05-04:15:00	5.1	15	1
2002-05-05:15:00	7.2	1	2
2002-05-06:09:00	8.8	5	5
2002-05-07:12:00	6.2	5	5
2002-05-08:15:00	8.2	5	5
2002-05-09:09:00	7.2	5	5
2002-05-10:09:00	13.4	1	1
2002-05-11:12:00	14.4	15	15
2002-05-12:09:00	7.2	2	2
2002-05-13:15:00	12.4	2	3
2002-05-14:09:00	13.9	80	80

The final maximum daily wind speed-weather classification dataset (WS-WC) is much smaller than the original max daily wind speed dataset after the missing records have been eliminated. Table 5.3 shows a few values of this dataset for Sydney airport daily max (station number 066037) with wind speeds in m/s.

Table 5.4 presents the 50 highest wind speeds of the Sydney airport combined dataset with the corresponding weather classification (lines with a '#' are collected in the thunderstorm dataset). Wind speeds are given in m/s.

Table 5.4 Highest (combined) wind gust speeds for Sydney airport

Date	wind speed	weather	
		present	past
1949-01-15:15:00	42.2	NA	92 #
1975-11-23:18:00	42.2	4	95 #
1946-05-27:12:00	36.0	NA	NA
1949-06-04:21:00	36.0	NA	61
1974-05-26:00:00	36.0	61	62
1941-12-20:15:00	35.0	NA	NA
1948-10-28:12:00	35.0	NA	NA
1946-12-14:12:00	34.5	NA	NA
1943-05-15:18:00	32.9	61	61
1946-05-19:15:00	32.9	5	5
1951-10-24:18:00	32.9	95 #	61
1965-05-16:18:00	32.9	15	80
1974-05-25:18:00	32.9	60	13
1944-12-21:15:00	32.4	NA	NA
1971-05-04:18:00	32.4	60	1
1949-11-27:18:00	31.9	NA	NA
1951-08-09:12:00	31.9	NA	NA
1968-05-15:09:00	31.9	65	81
1990-08-03:09:00	31.9	61	80
1998-08-07:15:00	31.9	61	62
1945-04-07:21:00	31.4	NA	61
1948-12-18:18:00	31.4	NA	NA
1948-12-22:15:00	31.4	NA	NA
2003-08-24:18:00	31.4	1	2
1950-11-04:12:00	30.9	NA	NA
1968-09-30:21:00	30.9	32	1
1942-10-04:15:00	30.4	NA	NA
1947-07-28:15:00	30.4	NA	NA
1951-03-06:18:00	30.4	NA	NA
1963-08-30:03:00	30.4	63	63
1975-06-12:15:00	30.4	61	62
2003-01-08:18:00	30.4	80	80
1941-10-26:12:00	29.9	17 #	NA
1943-06-22:18:00	29.9	NA	NA
1944-10-17:18:00	29.9	NA	NA
1946-11-13:15:00	29.9	NA	NA
1949-11-03:09:00	29.9	NA	NA
1962-02-18:06:00	29.9	60	61
1968-11-11:12:00	29.9	15	95 #
1996-08-31:03:00	29.9	65	62
1941-05-24:12:00	29.3	NA	NA
1945-11-09:18:00	29.3	61	61
1948-10-22:06:00	29.3	NA	NA
1950-03-04:15:00	29.3	95 #	95 #
1972-06-23:12:00	29.3	60	61
1978-06-02:03:00	29.3	80	80
1997-05-11:06:00	29.3	61	62
1942-09-23:15:00	28.8	NA	NA

6. Return periods for selected BoM wind stations

In recent years most risk researchers have changed to utilising GPD distributions rather than GEV distributions for extreme value analysis, due to its superior performance; see for instance Hecker et al. (1998), Holmes & Moriarty (1999) and Brabson & Palutikof (2000). For this reason this report focuses only on results produced by the GPD.

6.1 Return periods using a GPD

As mentioned earlier the central problem of modelling wind speed data using a GPD is the selection of the appropriate threshold 'u'. The GPD is very sensitive to the choice of the threshold. The Sydney airport max daily dataset can be used to illustrate the case. Figure 6.1 shows return periods (in years) for different thresholds. The range (in years) of the data available is shown under the x-axis label, in this case is 1939-2005. It is important to point out that if actual return periods from this dataset are calculated, Equation (2.2) should be used. In this case the parameter 'nopy', is calculated as the total number of wind speed records with non-missing observations divided by the range of years in the dataset. The 'nopy' parameter is used in the R 'GPD' function to scale the wind speed (y-axis) (Stephenson, 2004).

In Figure 6.1 the full or combined dataset was used (i.e. thunderstorm winds were not extracted from the original dataset). The total number of observations in the dataset is 24140. The maximum speed is 42.2 m/s. The threshold was varied between $u=5$ and $u=35$ m/s in increments of 5. Table 6.1 shows the number of observations exceeding the threshold and the corresponding shape parameter ' ξ '.

Notice the variation in sign of the shape parameter. As explained in Section 3.2, positive or zero shape parameters make the function unbounded which is physically incorrect since natural wind speeds have a finite maximum (Lechner et al., 1992). Negative values make the function converge to a maximum. Note: Authors such as Holmes (2002) and Palutikof (1999) use a different sign for the shape parameter in the GPD expression of Equation (3.2) and so obtain the opposite result: positive shapes make the function converge to a maximum. In this report we follow the convention used in Coles (2001) and the R software developed by Stephenson based on Coles' book (Stephenson, 2004). It is important to point out that Stephenson's package fits GPD distributions using the Maximum Likelihood (ML) technique.

Table 6.1. Parameters for Figure 6.1

Threshold 'u'	No. of observations exceeding 'u'	Shape parameter
5	23584	-0.216
10	15788	-0.168
15	5988	-0.125
20	1574	0.0426
25	252	0.0329
30	38	-0.067
35	6	5.253

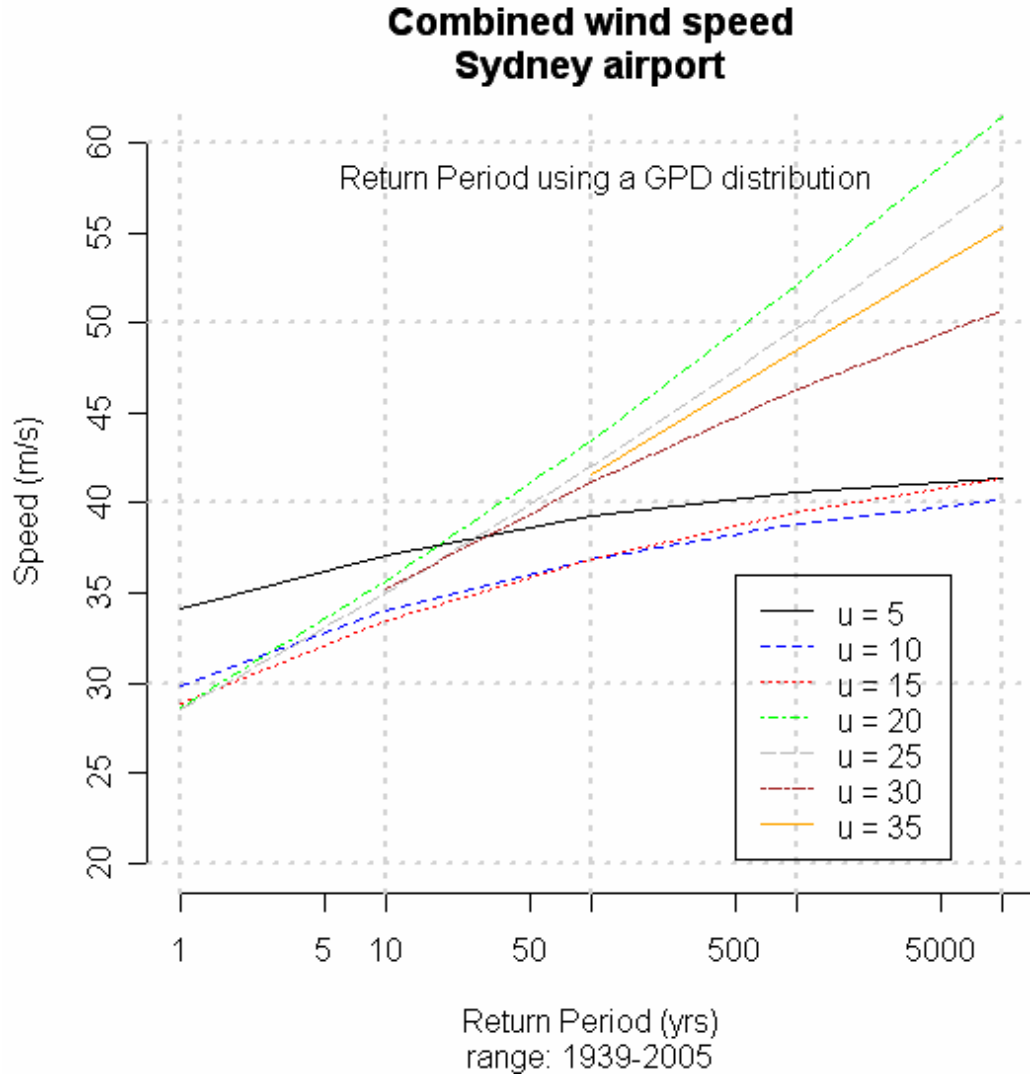


Figure 6.1. Sensitivity of the GPD to threshold ‘u’

Figure 6.1 shows that the GPD is very sensitive to the threshold selection. For values of $u = 20, 25$ and 35 the function is unbounded. The function for values between $u = 5$ and 15 is bounded but flat with a quick convergence to its limit. It is observed that the best threshold of the ones considered would be $u = 30$, which determines a convex function (because $\xi < 0$), and hence guarantees convergence to a limiting value. The GPD is also very sensitive to the range of values available for calculation of parameters. Smaller ranges mean fewer values over the threshold and hence have a direct impact on the shape of the distribution. Figure 6.2 shows return periods for different range of years, starting with 1939-2005, for the same dataset of Figure 6.1. All return periods were generated using the same threshold of 20 m/s.

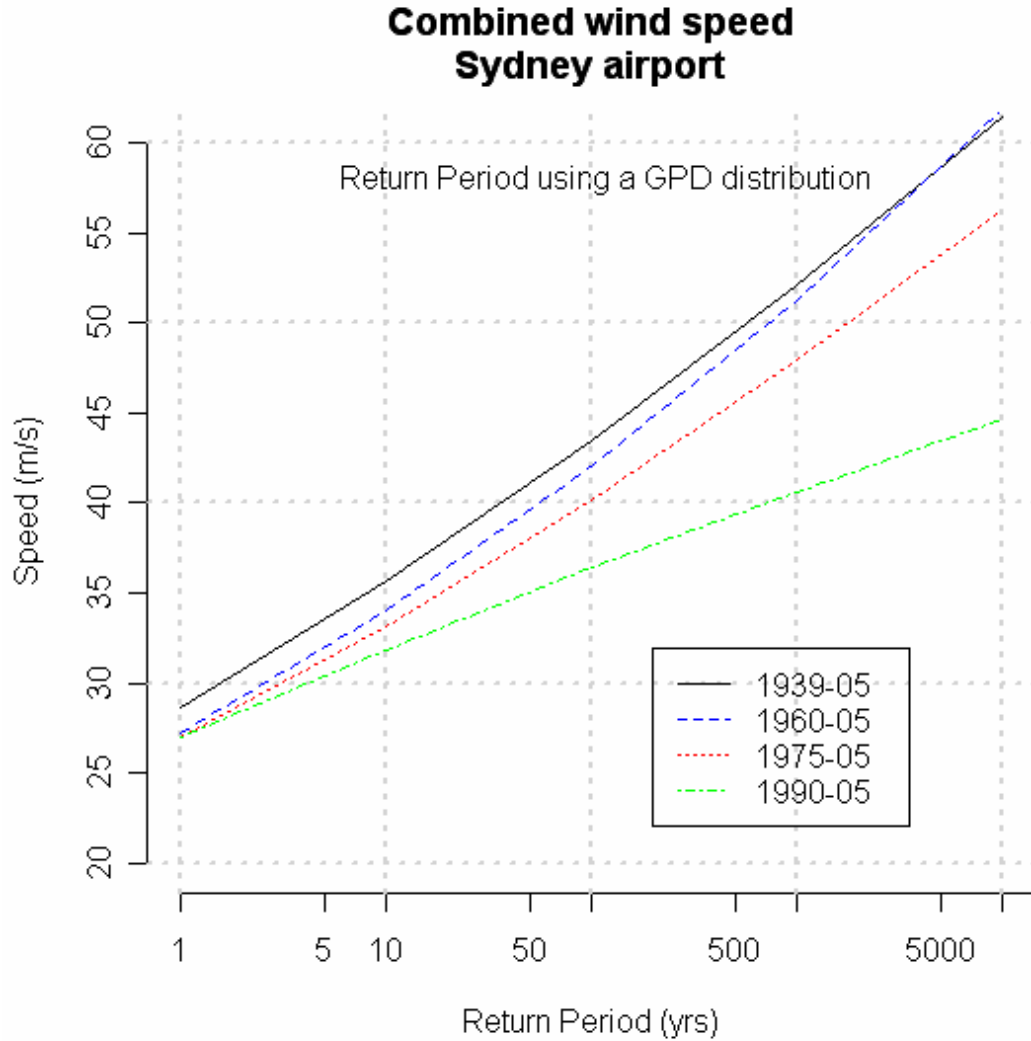


Figure 6.2. Sensitivity of the GPD to data range (years)

The sensitivity results confirm the importance of the threshold selection in the calculation of return periods using the GPD. A too low threshold means too many points for fitting the GPD distribution; in this case independence of the random variables of sample points may be violated. The end result is a low, flat curve with quick convergence to the limit, as shown in the curves with threshold from 5 to 15. Too high thresholds mean fewer points for fitting the GPD distribution, in this case the distribution may not converge at all as shown by the large, positive shape parameter of the curve with $u = 35$ ($\xi = 5.253$). Values between $u = 20$ and $u = 25$ result in positive shape parameters close to 0, in this case the GPD becomes the Gumbel distribution (GEV distribution type I) which is linear and unbounded and not appropriate for modelling phenomena with finite bounds like wind speed. Figure 6.1 indicates that the best threshold from the ones considered is $u = 30$ m/s, which is the highest, bounded curve. However, as mentioned before, there are visual methods to help analysts select the appropriate threshold to fit the GPD; they are presented and discussed in the next section.

6.2 Tests for threshold selection

The high sensitivity of the GPD to the threshold selection has led some researchers to find techniques to help in the selection of the most appropriate value. Two visual techniques have been developed in recent years to solve this problem. The first one is called The Mean Residual Life plot (MRL) also known as the Conditional Mean Exceedance (CME) method (Lechner et. al., 1992); the second one is the Model Based Check (MBC). Both of these techniques are explained in Coles (2001).

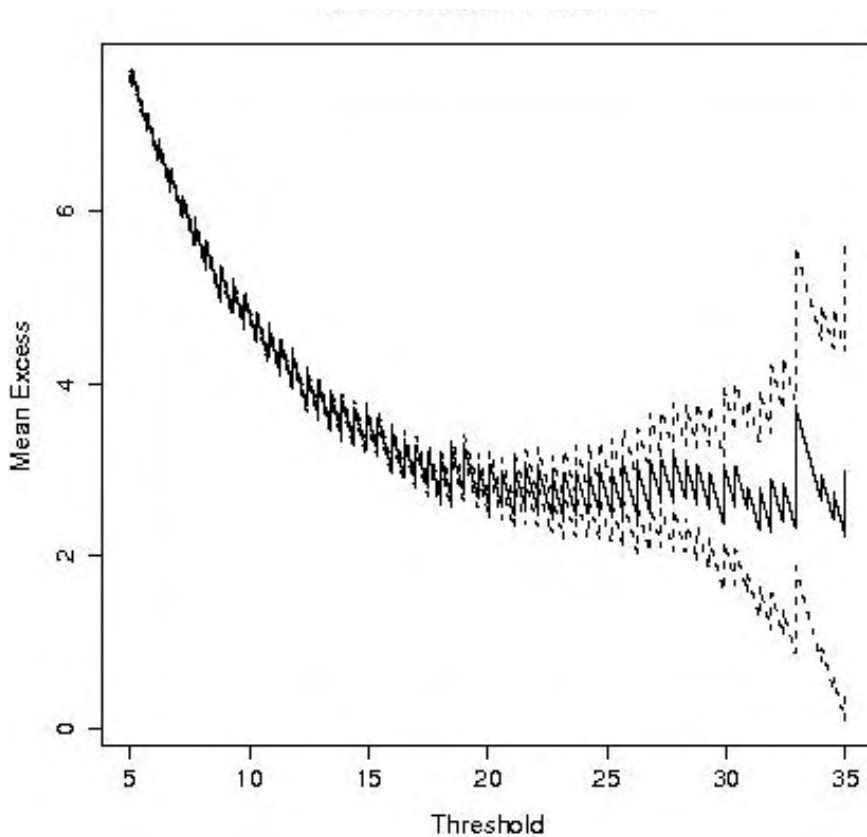


Figure 6.3. Mean Residual Life plot.

Figure 6.3 shows the MRL plot for the Sydney airport data where the dotted lines show the 95% confidence interval. This plot should be approximately linear for valid values of 'u'. The plot clearly shows a curve for values less than about 22 and a linear form for thresholds higher than this value. This suggests that 'u' should be higher than 22. For values exceeding 33, the curve presents a jump showing stability problems due to lack of data. There are only 10 values exceeding 33. Using this method the threshold should be selected between 22 and 33.

Figure 6.4 shows the MBC plot for the same data with the vertical lines showing the 95% confidence intervals. The plot should be approximately constant for valid values of 'u'. The plot is approximately constant for values exceeding 27 except for 29, 31 and 33 which represent small perturbations probably due to sampling errors. This plot supports the selection of 30 as the most appropriate value for the threshold.

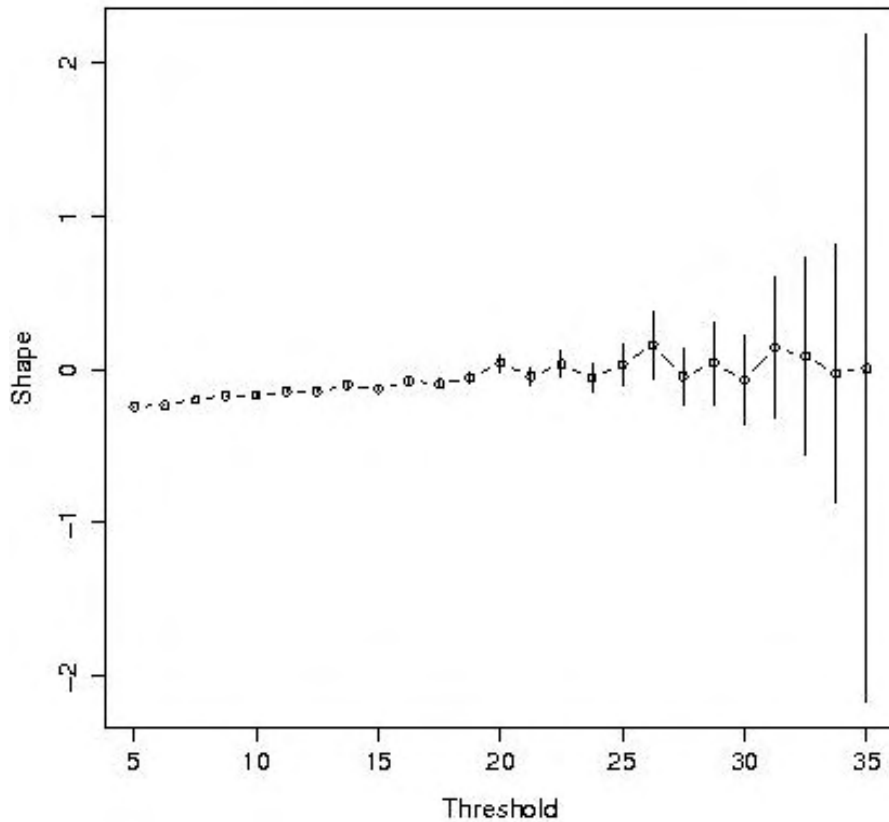


Figure 6.4. Model-based plot.

Other diagnostics to judge the GPD modelling capabilities have been developed and are implemented in the R-package. Two of them are known as ‘Probability’ plot and ‘Quantile’ plot (Coles, 2001). Figure 6.5a and 6.5b show these plots for the Sydney airport dataset. The dotted lines are the 95% confidence intervals. The first one shows the plot of empirical wind speeds against the modelled values using the GPD with $u = 30$. If the model is appropriate the plot should be approximately linear (a line has been added to the plot to facilitate checking). The second plot shows empirical quantile values against the modelled ones using the GPD. Similarly, if the model is correct, the plot should follow a linear relationship. In both cases the plot follows the line. For higher wind speeds the ‘quantile’ plot has problems fitting values exceeding 38, there are only two of these values, both are 42.2 m/s. These values can be seen as outliers. So in general it seems that the GPD with threshold 30 m/s does a reasonable job fitting the Sydney airport data. Notice the variation of the confidence interval of the quantile with the number of years considered: the larger the number of years the higher the uncertainty and hence the wider the confidence interval. This topic will be discussed in Section 9.

Although the techniques for threshold selection presented in this Section can help the analyst to check whether their selection is appropriate, these techniques are too subjective to give accurate results. Therefore a robust, automatic technique to select the appropriate threshold for a given dataset is needed in GPD applications. Such a technique is presented in the next Section.

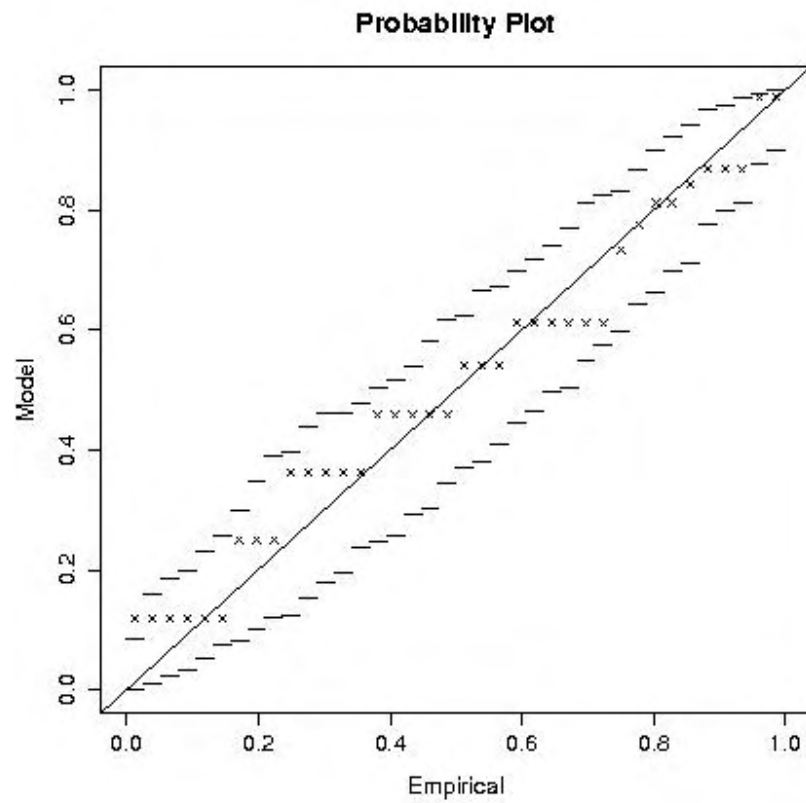


Figure 6.5a. Probability plot with Sydney data

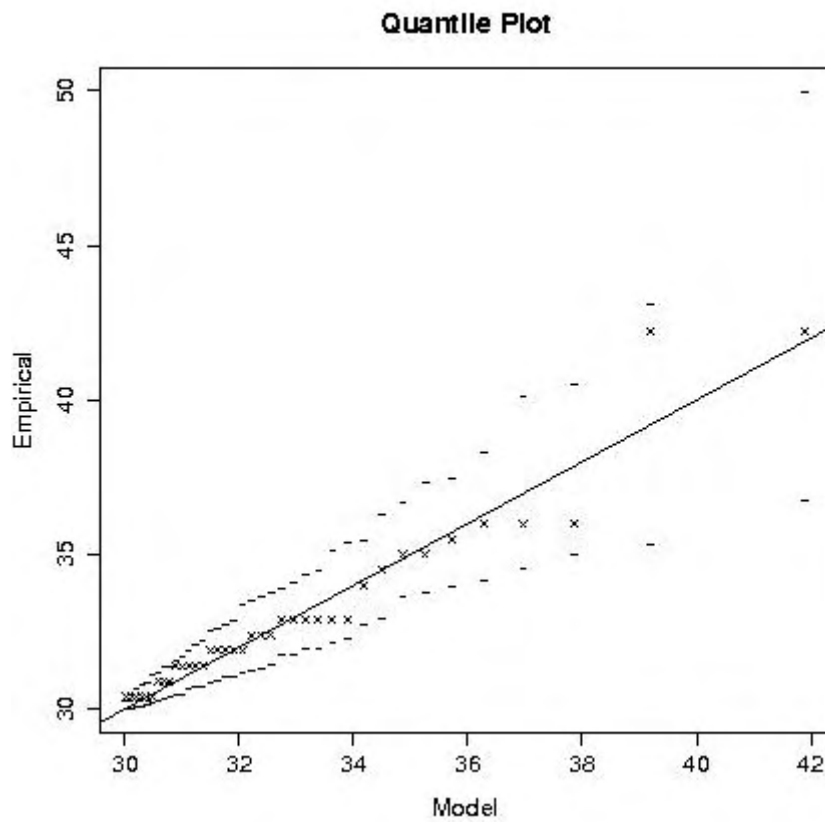


Figure 6.5b. Quantile plot with confidence interval for Sydney data.

6.3 Algorithm for automatic selection of appropriate ‘u’

Based on the return periods presented in earlier versions of this report a set of rules for selection of the appropriate threshold were compiled and used to develop an automatic algorithm. The return periods of this and the next sections were recalculated using the automatic algorithm. To save paper both the original results and the results produced by the automatic algorithm will be presented only for a few cases.

Figures 6.1 and 6.2 show that there are three distinctive *regions* in a family of return periods calculated using the GPD. The *first region* corresponds to curves that have a positive shape parameter and hence are concave and unbounded which can be modelled by Type II distributions. As explained in Section 6.1 these types of distributions are not appropriate in wind modelling and hence thresholds which produce these types of distributions are not acceptable. The *second region* is characterised by very small or zero shape parameters which produce straight lines (Type I distributions). These types of distributions are not appropriate to model wind speed either, because it is a naturally bounded phenomenon, hence thresholds which give rise to return periods in *region 2* should also be rejected. The *third region* is the region of interest for this type of study. This *region* is characterised by a negative shape parameter which gives rise to convex curves which tend to a limiting asymptotic point. This asymptotic value ‘a’ is given by the expression (Coles, 2001);

$$a = (\mu - \sigma)/\xi \quad (6.1)$$

where the parameters were explained in Sections 3.1 and 3.2.

Curves falling in the *third region* are appropriate for wind modelling for this reason we call it the *feasible region*.

The algorithm iteratively calculates the shape parameter for a number of possible thresholds, only thresholds which produce GPD curves in the *feasible region* are considered for wind speed modelling. A good starting point for the algorithm is 1/2 of the maximum speed in the dataset, i.e. the first candidate threshold is given by,

$$u_1 = (1/2) \max(w_s) \quad (6.2)$$

where w_s is the vector of wind speeds in m/s. The next thresholds are generated by increasing u_1 in steps of 0.25 m/s until the number of values exceeding the threshold is less than or equal to 20. Figure 6.6 shows the results of applying the algorithm to the data presented in Figure 6.1. Notice that the algorithm plots only curves in the *feasible region*. The table of results produced by the algorithm is presented as Table 6.3. To save paper, the starting threshold ‘ u_1 ’ in Fig 6.6 was set to a higher value than that calculated by Equation 6.2. The valid thresholds are shown on the left-hand side of Figure 6.6.

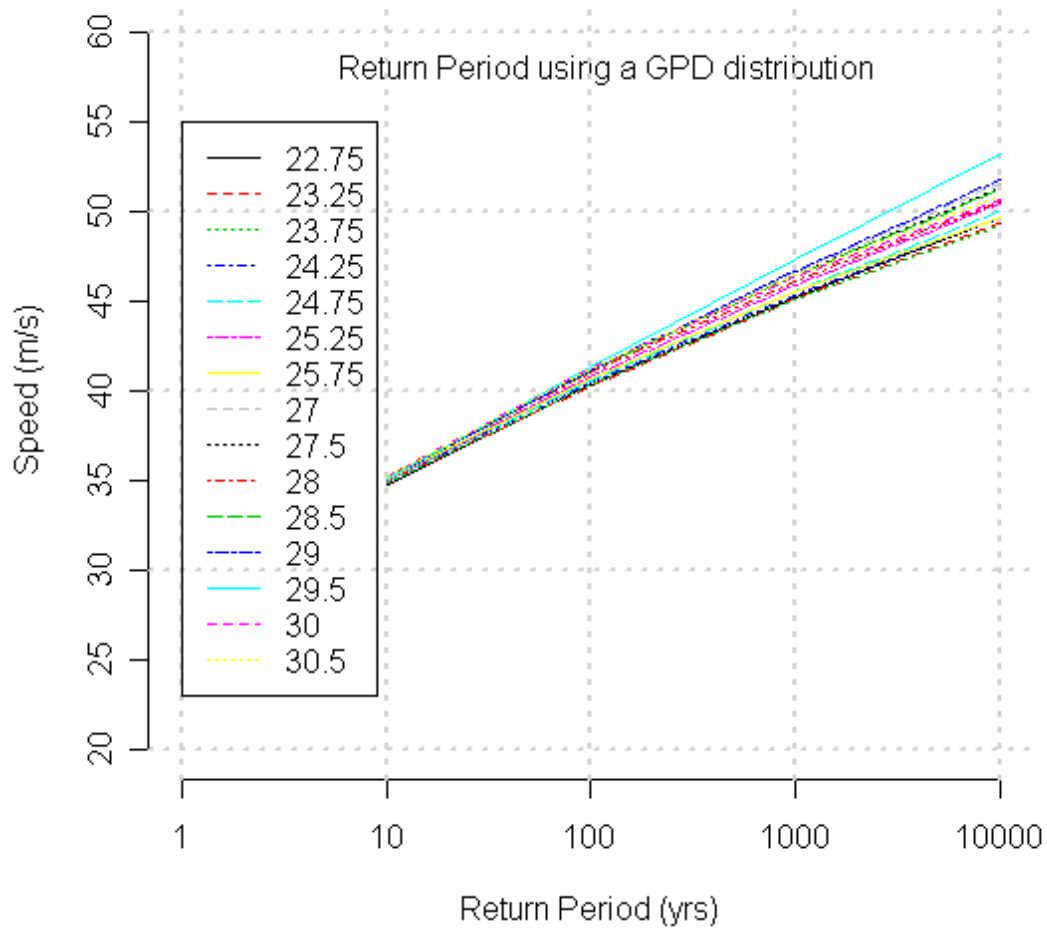


Figure 6.6. Automatic selection of threshold (Sydney airport dataset).

Table 6.2a. GPD parameters (Sydney airport dataset)

u	No>u	shape	mx_pnt	q1000	q10000	qdiff
22.75	536.0	-0.04	183.62	45.23	49.67	4.44
23.25	438.0	-0.05	152.82	45.06	49.29	4.24
23.75	362.0	-0.06	142.18	45.08	49.26	4.19
24.25	302.0	-0.05	147.14	45.31	49.60	4.29
24.75	252.0	-0.05	155.13	45.58	50.03	4.45
25.25	210.0	-0.05	163.16	45.83	50.41	4.58
25.75	171.0	-0.06	122.66	45.50	49.61	4.12
27.00	112.0	-0.04	191.97	46.53	51.49	4.97
27.50	94.0	-0.05	174.11	46.45	51.27	4.82
28.00	78.0	-0.06	127.94	46.07	50.52	4.45
28.50	67.0	-0.05	161.33	46.40	51.21	4.82
29.00	57.0	-0.04	201.82	46.66	51.78	5.12
29.50	49.0	-0.01	618.27	47.33	53.17	5.84
30.00	38.0	-0.07	113.46	46.30	50.70	4.39
30.50	32.0	-0.07	118.18	46.43	50.91	4.48

```

max ws = 42.2
acceptable shape < -0.01
Av1000  approp u  Av10000  approp u
51      29.5      56      29.5

```

Column 1 of Table 6.3 shows the possible thresholds, the second column shows the number of values exceeding the threshold. Column 3 shows the shape parameter for given threshold. Notice that only shape parameters with values less than -0.01 are accepted, if they are too close to 0 they determine a Type II (Gumbel) distribution (straight line). The column marked as 'mx_pnt' is the asymptotic value 'a' given by equation (6.1). The next two columns show the wind speed calculated by the GPD at quantiles 1000 and 10000 (years). The last column shows the difference between these quantiles, that is, $(q_{10000} - q_{1000})$. A big difference between these quantiles is a sign of numerical stability problems in calculations involving large return periods; hence the maximum difference is limited to 12% of q_{10000} .

The final step in the automatic selection of the threshold is to calculate the average value of both q_{1000} and q_{10000} , in other words the average of columns 5 and 6. The appropriate threshold is the one which produces the closest curve to the average plus 15% at q_{10000} .

$$u_{app} = \min(\text{curve}_i - 1.15 * \text{average}) \quad (6.3)$$

where 'i' refers to the curves of Figure 6.6 (or entries of Table 6.3).

The average plus 15% of columns 5 and 6 are shown at the bottom of Table 6.3 as 'Av1000' and 'Av10000'. The result of calculating equation (6.3), i.e. the appropriate threshold, is shown next to the averages as 'approp u'. For the Sydney airport dataset, the algorithm suggests that the appropriate threshold is 29.5 which gives a higher return period curve than that of the threshold selected by inspection in Section 6.1 (i.e. 30 m/s). Table 6.3a suggests that the asymptotic value (mx_pnt) is approaching a Type II (Gumbel) and that the "concave" curve is an outlier compared to adjacent threshold values (see page 31 for further discussion). In some cases the threshold suggested by Av1000 could be different from the value suggested by Av10000, as a general rule the analyst should select a threshold which produces a return period of *combined* winds which envelops *thunderstorm* and *synoptic* wind curves as shown in Figure 6.8 (Holmes, 2006b).

The rules for automatic selection were developed by trial and error using about half of the datasets shown in Table 4.2. A summary of these rules is presented in Appendix C.

6.4 Return periods for thunderstorm winds

The Sydney airport dataset was separated into *thunderstorm* and *synoptic* winds using the algorithm presented in Section 5 and return periods for each type of wind gust were plotted. Figure 6.7 shows return periods for *thunderstorms* using a GPD. The corresponding automatic selection procedure is presented in Table 6.3b. The algorithm suggests a threshold of 17.75 or 18.75 m/s for this dataset, both produce curves very close to each other, but as indicated before, the actual value should be selected by comparing *combined* and *thunderstorm* winds as shown in Figure 6.8. The *combined* speed datasets has higher values than the *thunderstorm* dataset hence if the selected threshold is correct the return periods of *combined* wind should be higher than those of *thunderstorms*. The highest speeds of the *thunderstorm* dataset are shown in Table 6.3.

The relative location of each one of the return period curves is as shown in Figure 6.8 if there are no errors in the records. Unfortunately this is not the case in most wind datasets. In most cases there are a lot of missing observations in the merged dataset as shown in Tables 5.4 and 6.3. The user should analyse the datasets and correct obvious mistakes. A common error is when there is no weather observation in the max wind speed. In this case the algorithm does not pick up this observation as corresponding to a thunderstorm and hence the observation is incorrectly located in the synoptic wind dataset. The end result is a RP curve of synoptic wind higher than the corresponding thunderstorm curve.

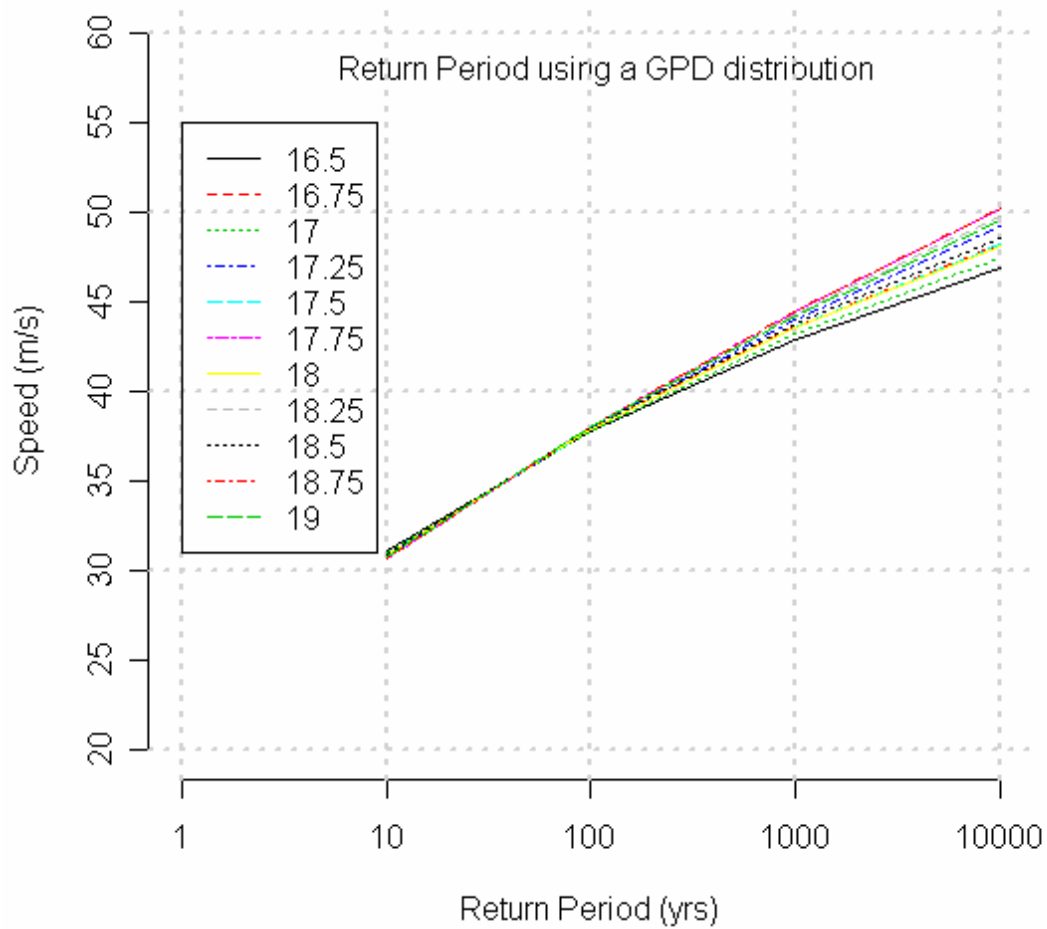


Figure 6.7. Return period for thunderstorms (Sydney Airport)

Table 6.2b. GPD parameters (Sydney airport; thunderstorms)

u	No>u	shape	mx_pnt	q1000	q10000	qdiff
16.50	236.0	-0.11	98.36	42.90	46.85	3.94
16.75	236.0	-0.09	127.55	43.53	48.20	4.67
17.00	217.0	-0.10	110.50	43.22	47.49	4.27
17.25	217.0	-0.07	154.57	44.01	49.16	5.15
17.50	198.0	-0.09	125.42	43.57	48.22	4.65
17.75	198.0	-0.06	193.56	44.49	50.20	5.71
18.00	174.0	-0.09	120.29	43.58	48.14	4.56
18.25	174.0	-0.06	176.60	44.32	49.80	5.48
18.50	156.0	-0.09	128.09	43.78	48.51	4.74
18.75	156.0	-0.06	194.79	44.48	50.24	5.76
19.00	143.0	-0.07	161.19	44.22	49.49	5.27

max ws = 42.2
acceptable shape < -0.01
Av1000 approp u Av10000 approp u
50 17.75 55 18.75

Table 6.3. Highest 25 thunderstorm wind speeds for Sydney airport (Geoscience Australia) and the highest thunderstorm speeds in Holmes data for the same wind station (2006a).

Geoscience Australia (1939-2005); Holmes, 2006a (1969-1992)

ws = wind speed (gust in m/s) wpr = present weather wpa = past weather

Geoscience Aust. Date	ws	wpr	wpa	Holmes (2006) Date	ws
1949-01-15:15:00	42.2	NA	92	23/11/1975	42.2
1975-11-23:18:00	42.2	4	95	04/05/1971	32.4
1951-10-24:18:00	32.9	95	61	29/03/1975	28.8
1941-10-26:12:00	29.9	17	NA	25/01/1975	28.3
1968-11-11:12:00	29.9	15	95	05/09/1974	27.8
1950-03-04:15:00	29.3	95	95	02/12/1969	27.3
1957-12-02:18:00	28.8	34	34	20/01/1990	27.3
1975-03-29:15:00	28.8	3	97	22/12/1969	26.7
1975-01-25:15:00	28.3	4	95	30/09/1982	26.7
1958-01-19:15:00	27.8	NA	95	20/11/1973	26.7
1963-09-11:15:00	27.3	9	4	22/04/1972	25.7
1969-12-02:15:00	27.3	9	9	21/01/1977	25.7
1990-01-20:18:00	27.3	95	95	08/01/1988	25.7
1969-12-22:18:00	26.8	4	95	14/12/1971	24.7
1972-04-22:12:00	26.8	3	96	13/12/1983	24.2
1973-11-20:18:00	26.8	91	95	26/01/1991	24.2
1996-02-08:18:00	26.8	4	95	10/02/1971	23.7
1951-12-09:21:00	26.3	61	95	23/12/1985	23.7
1953-05-07:00:00	26.3	97	95	28/03/1975	23.1
1967-10-28:15:00	26.3	6	95	16/01/1978	23.1
1951-09-25:18:00	25.7	91	91	28/08/1992	23.1
1962-02-14:15:00	25.7	5	95	11/02/1979	22.6
1964-03-01:15:00	25.7	2	95	24/04/1981	22.6
1965-06-23:00:00	25.7	63	95	21/08/1981	22.6
1977-01-21:15:00	25.7	4	97	14/12/1970	22.6

It is interesting to compare the data shown in Table 6.4, which relates to the *thunderstorm* information extracted from the Geoscience Australia dataset (1939-2005) and the *thunderstorm* information extracted from Holmes (2006a). Both tables are similar for the higher wind speed values. It is therefore expected that the Geoscience Australia curve of return periods will look similar (for high return periods) to the curve calculated by Holmes. However some points in Holmes thunderstorm dataset are not present in our dataset, for instance, the second point, a wind speed of 32.4 m/s which occurred on 04/05/1971, is not in our dataset. To investigate this difference, consider Table 6.2 where the original (combined) wind speeds are listed. An entry of 32.4 m/s can be seen on the date 1971-05-04:18:00 however the Table clearly indicates that the weather conditions at the time of the observation were 60: "Intermittent rain (not freezing), slight at time of obs". The conditions in the previous 3 hours (not including the previous hour) are listed as 1: "Partly cloudy (scattered) or variable sky". Since staff writing down these conditions at the Sydney airport have access to radar information it is unlikely that they would miss a thunderstorm. The error could be in the visual inspection of the anemometer chart. These kinds of differences may also result in differences in the return period curves. To increase the confidence in our results it would be necessary to locate these differences and manually check the correct entries by looking at the

anemometer charts. It is expected that this activity can be realised in collaboration from BoM in the future.

Presented in Figure 6.8 is a plot of return periods for all wind types using a threshold of 30 m/s for the *combined* winds, 18.75 m/s for the *thunderstorm* dataset and 21.5 m/s for *synoptic* winds. The value 30 rather than 29.5 for the threshold of *combined* winds was selected because it produces a combined curve closer to the *thunderstorm* and *synoptic* curves as the *combined* curve should closely envelop the other two curves (Holmes, 2006b). The figure shows that return periods of *combined* winds are higher than the corresponding *thunderstorm* winds but the latter tends asymptotically to the *combined* winds. *Synoptic* wind speeds are consistently lower than *combined* wind speeds. They are also lower than *thunderstorm* speeds for medium and high return periods but higher for low return periods (less than 30 years). The figure shows that wind databases for some locations (but not all) are dominated by *thunderstorm* winds and that separation of the three types of winds is important in order to obtain a more realistic modelling of wind return periods.

Figure 6.8 also includes a curve using the Holmes' *thunderstorm* data presented in Table 6.4. Although Holmes' data is taken from a reduced range of years, both curves show similar characteristics. The Holmes (2006a) data has only 39 points compared with the 508 data points for our dataset. Figure 6.8 shows that Holmes' *thunderstorm* curve is higher than our corresponding *thunderstorm* curve for low values but lower for high values. This is because of the mix of values in our curve: we have a lot more low to medium samples hence our curve is pushed down by these values but we have also more high values than Holmes and hence our curve is higher for high return periods. The relativities of our plot agree with Holmes', the highest curve is the *combined* wind speed followed by *thunderstorms*; *synoptic* winds are higher than *thunderstorms* for low values but lower for high return periods. The three-wind return period curve can also be used to fine-tune the threshold selection procedure. If the location of the curves with respect to each other does not appear correct based on the analyst's experience or literature review, it is always possible to go back to Figures 7.3a and 7.3b to choose a more appropriate threshold.

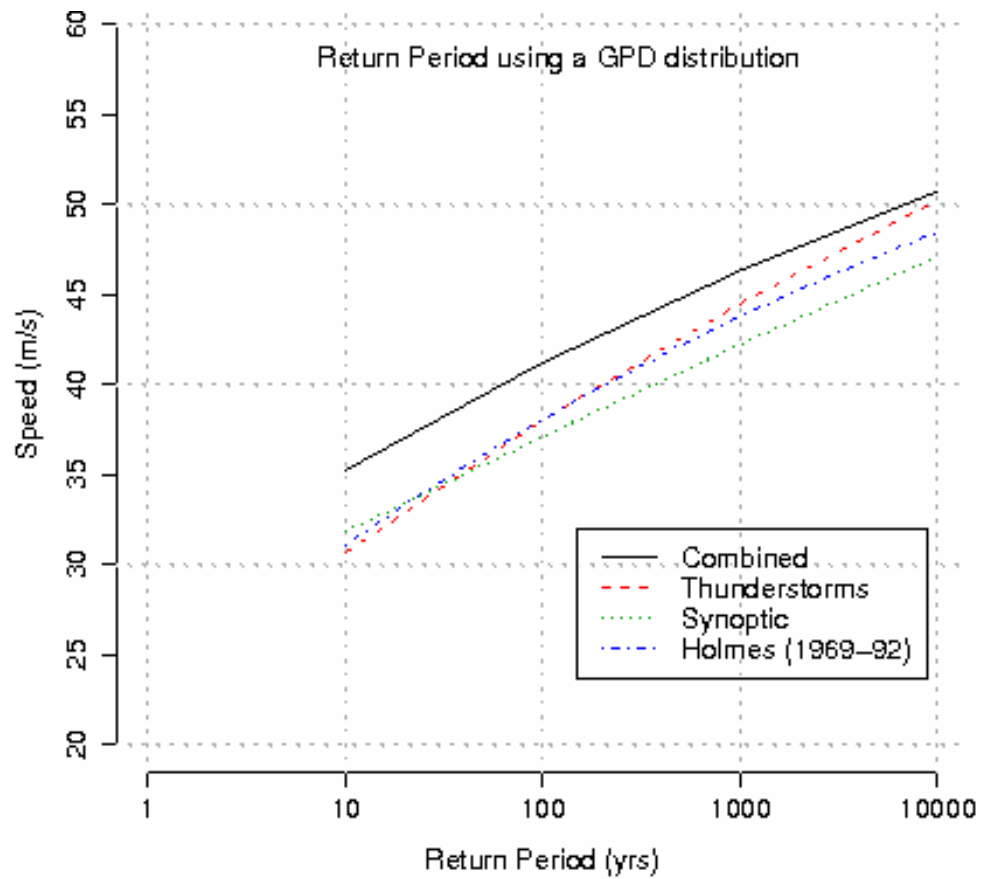


Figure 6.8. Return period of the three categories of gust wind speeds (Sydney Airport)

6.5 Return periods for other wind stations

Return periods for the Sydney region wind observing stations of Bankstown and Richmond (both airports with near ideal exposure to undisturbed wind flow) were also calculated and plotted. They are presented in Appendix A. These stations are important to calculate return periods for the ‘Sydney region’ as defined in Holmes (2002), as comparison of our results with those reported in Holmes (2002) is an important part of this study.

The highest *combined* wind speeds for Bankstown and Richmond are shown in Tables 6.5 and 6.6 respectively (records with ‘#’ are selected for the thunderstorm dataset).

Table 6.4. Highest *combined* wind speeds for Bankstown airport

Date	wind speed	weather	
		present	past
1977-01-21:15:00	34.5	2	97 #
1979-11-26:06:00	31.4	15	99 #
1989-11-17:15:00	30.4	14	95 #
1990-03-18:15:00	30.4	80	96 #
1975-06-21:09:00	29.9	64	64
1981-08-21:12:00	29.9	2	2
1971-10-02:15:00	28.3	2	2
1978-01-24:15:00	28.3	80	97 #
1983-10-19:15:00	28.3	80	95 #
1981-09-27:12:00	26.8	2	NA
1978-03-19:09:00	26.3	80	81
1980-09-15:09:00	26.3	60	80
1991-03-11:15:00	26.3	17 #	96 #
1971-10-03:15:00	25.7	1	2
1972-08-11:12:00	25.7	2	2
1980-08-31:12:00	25.7	6	6
1968-11-23:09:00	25.2	7	4
1975-06-12:15:00	25.2	21	80
1976-09-10:12:00	25.2	1	2
1977-11-10:15:00	25.2	2	2
1971-07-25:12:00	24.7	2	1
1974-05-01:15:00	24.7	2	2
1975-03-29:15:00	24.7	2	92 #
1979-11-24:12:00	24.7	29 #	NA

Table 6.6 presents the highest combined wind gust for Richmond old RAAF observing site (station 67033).

One of the problems with the comparison of our results with those of Holmes (2002) is that the available range of values in our datasets is greater than that of Holmes. Our datasets, acquired from BoM in 2005 include observations up to 2005. Table 6.7 shows the difference in ranges for the datasets of the stations referred to above. As explained in Section 6.2, the GPD is very sensitive to the range of values, so to be able to do a proper comparison we have reduced our range of values to the same range reported in Holmes (2002). The results are presented in the next section.

Table 6.5. Highest (*combined*) wind speeds for old Richmond RAAF observing site.

Date	wind speed	weather	
		present	past
1945-12-08:12:00	37.6	NA	96 #
1964-07-12:12:00	35	18	18
1944-12-10:12:00	34	31	32
1968-03-25:12:00	34	15	99 #
1957-10-10:15:00	33.5	7	NA
1945-09-04:09:00	32.4	NA	NA
1961-02-23:18:00	31.9	17 #	95 #
1983-10-19:15:00	31.9	97 #	95 #
1963-09-16:06:00	31.4	10	5
1944-12-21:12:00	30.9	7	7
1962-08-28:12:00	30.9	5	5
1961-02-08:18:00	30.4	15	95 #
1964-02-15:18:00	30.4	NA	NA
1965-11-16:12:00	30.4	7	1
1962-02-10:15:00	29.9	NA	NA
1944-11-13:15:00	29.3	7	7
1945-06-23:12:00	29.3	NA	NA
1991-01-12:15:00	29.3	2	95 #
1994-01-03:15:00	29.3	2	NA
1945-11-03:15:00	28.8	7	NA
1983-12-13:18:00	28.8	15	98 #
1944-10-17:15:00	27.8	NA	80
1945-11-09:18:00	27.8	7	7
1968-09-30:21:00	27.8	9	60
1983-11-08:21:00	27.3	13	2

Table 6.6. Temporal range of max daily wind gust speed datasets

	Holmes	this study
Sydney airport	1969-1992	1939-2005
Bankstown	1970-1991	1968-1992
Richmond	1970-1992	1942-1994

The algorithm was applied to two other wind stations Williamtown, station 61078, and Nowra, station 68076. A new automatic weather station was installed in Nowra in 2001 (station number 68072) but it was not used in these studies because the data range available is too small. Figures A3 and A4 in Appendix A show the return periods for *combined* wind gusts of Williamtown and Nowra and the corresponding automatic threshold selection table.

Return speeds for the Williamtown wind station are higher than for the Richmond station; this is important for an appropriate definition of a ‘Sydney region’ as discussed in the next section. Table 6.8 shows the top 20 combined wind speeds of Williamtown and Nowra in m/s. Return period wind speeds for 10, 100, 1000 and 10000 years and for the following observing stations: Bankstown, Richmond, Nowra, Williamstown, Sydney airport and Sydney Observatory Hill are presented in Tables 6.9 and 6.10. The wind speeds considered correspond to the *combined* and *thunderstorm* categories respectively. Return periods for the new Richmond station and the old one (used in this report) are presented to allow comparison between both stations. Return periods for Sydney airport using the full range of data and using the reduced range 1969-92 have been calculated to allow comparison of our results with the Holmes results. All return periods in the tables have been rounded up to the next integer, i.e. 34.50 becomes 35.0.

Table 6.7. Williamstown and Nowra highest gust wind speeds.

Williamstown	Nowra
38.1	40.7
38.1	36.5
37.6	36
36.	35
35.5	34.5
34	33.5
34	33.5
32.9	32.9
32.4	32.9
32.4	32.4
31.4	32.4
30.9	32.4
30.9	32.4
30.0	32.4
30.9	32.4
30.4	32.4
30.4	31.9
30.4	31.9
30.4	31.9
29.9	31.9

Table 6.8. Return periods of *combined* Wind Gust Speeds in m/s.

Return period (years)	10	100	1000	10000
Bankstown (1968-1992)	32	38	44	49
Richmond (1994-2005)	27	29	31	32
Richmond (1942-1994)	32	39	44	50
Nowra (1955-1997)	35	39	44	48
Williamstown (1942-2005)	33	39	45	50
Sydney (this report) (1939-2005)	35	41	46	51
Sydney (reduced range) (1969-1992)	34	40	46	51
Observatory Hill (1955-1992)	35	41	46	51

Table 6.9. Return periods of *thunderstorm* Wind Gust Speeds in m/s

Return period (years)	10	100	1000	10000
Bankstown (1968-1992)	30	36	39	41
Richmond (1994-2005)	24	26	27	28
Richmond (1942-1994)	30	38	44	50
Nowra (1955-1997)	29	37	43	48
Williamstown (1942-2005)	30	37	42	47
Sydney (this report) (1939-2005)	31	38	44	50
Sydney (reduced range) (1969-1992)	31	39	45	51
Observatory Hill (1955-1992)	35	40	44	48
Sydney (Holmes)	31	38	44	48

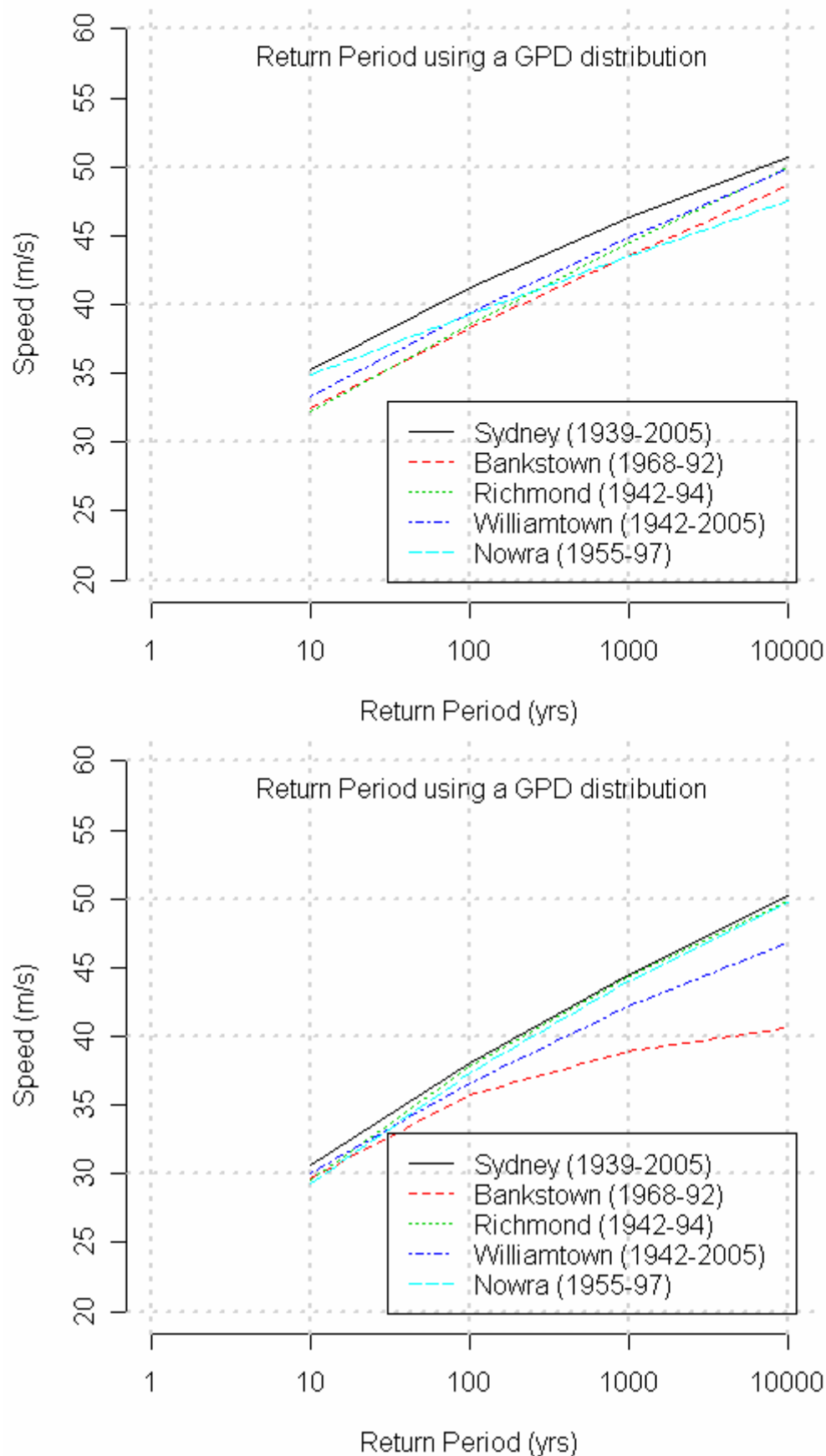


Figure 6.9a & 6.9b. Return periods for (a) *combined* wind speeds and (b) *thunderstorm* wind for five sites in the Sydney region.

7. Return periods for the Sydney region

In Holmes (2002) the Sydney region is made up of three wind stations: Sydney airport, Bankstown airport and Richmond RAAF Airforce Base. Following this convention our 'Sydney region' wind dataset was made up by joining the wind datasets for Sydney airport, Bankstown and Richmond. Separation of thunderstorm and synoptic winds from these datasets was carried out using the algorithm described in Section 6.

Making up the Sydney region using only Sydney, Bankstown and Richmond airports can be questioned because the last two sites are both located inland of the coastal interface. Coastal sites with likely greater wind speeds such as Nowra and Williamstown (or a range of observing stations with short records; automatic weather stations available since the mid-1980's) may be more representative of the wind hazards affecting the Sydney region and may need to be included (i.e. winds for well exposed sites, that are representative of a large area, could be aggregated into one long record).

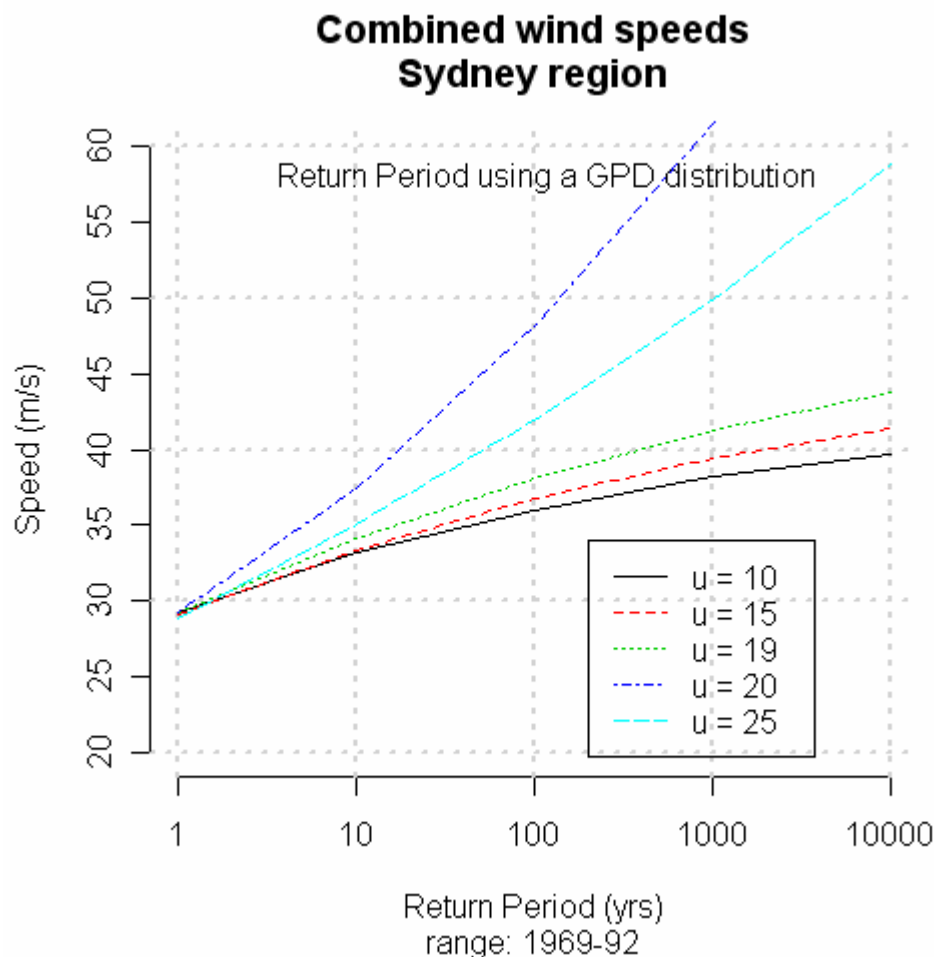


Figure 7.1a. Sydney Region *combined* wind threshold sensitivity (Holmes(1999) study period of record).

Figure 7.1a shows the sensitivity of return periods to threshold selection for the Sydney region *combined* winds. It is clear that $u = 10$ gives a very low curve. Thresholds of 20 and 25 make the curves unbounded, so the appropriate threshold for this dataset seems to be either $u = 15$ or $u = 19$. The 19 m/s threshold gives a higher curve so the threshold $u = 19$ is the recommended one. Notice the high sensitivity of the regional wind dataset, curves produced using $u = 20$ are concave and unbounded whilst curves using $u = 19$ are convex and appropriate for modelling wind speeds.

The same dataset was used to produce Figure 7.1b, plotted using the algorithm for automatic threshold selection explained in Section 6.3. Notice that the algorithm selects $u = 24.25$ as the best threshold for given data which gives a much higher curve than the ones shown in the *feasible* region of Figure 7.1a.

Similarly Figure 7.2a shows the sensitivity of the GPD Sydney region *thunderstorm* to the selection of 'u'. A value of $u = 10$ results in a low curve, values of 20 and 25 produce a concave curve which violates the requirement of convergence to a limiting speed, so the appropriate threshold seems to be either $u = 15$ or $u = 19$. Figure 7.2b generated by the automatic threshold selection algorithm shows that the most appropriate threshold for this dataset is actually $u = 24.25$ m/s (not shown in Figure 7.2a)

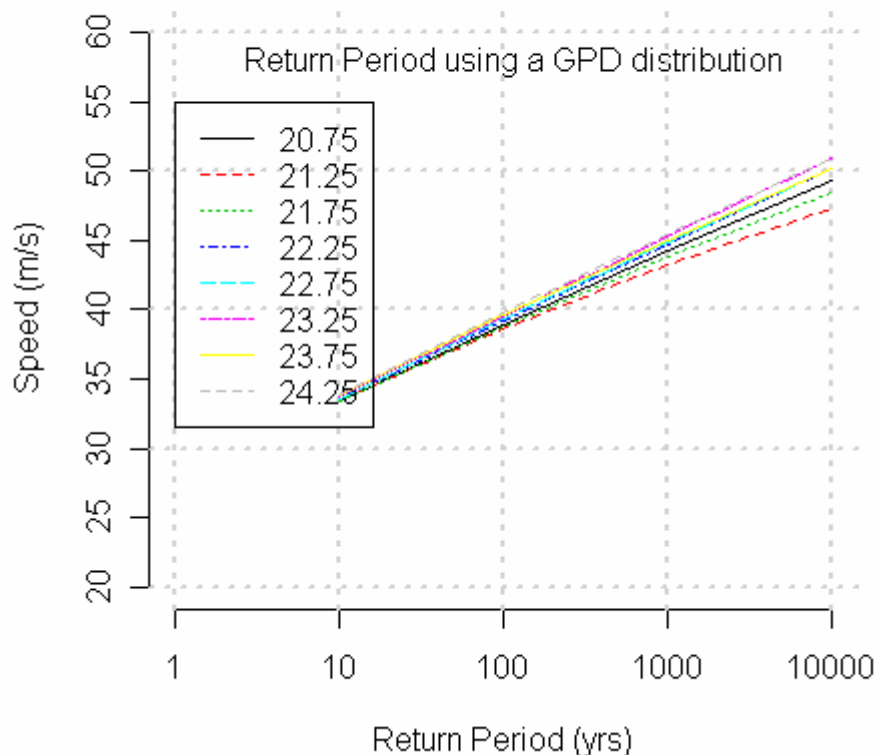


Figure 7.1b. Sydney Region combined wind with threshold selection. (Holmes(1999) study period of record).

Sydney region combined winds. GPD parameters						
u	No>u	shape	mx_pnt	q1000	q10000	qdiff
20.75	304.0	-0.021	338.06	44.23	49.29	5.06
21.25	240.0	-0.047	154.28	43.22	47.36	4.14
21.75	204.0	-0.034	211.66	43.80	48.40	4.61
22.25	174.0	-0.014	516.94	44.71	50.20	5.49
22.75	139.0	-0.017	422.29	44.88	50.25	5.37
23.25	113.0	-0.013	537.34	45.27	50.87	5.60
23.75	89.0	-0.028	253.95	45.05	50.20	5.15
24.25	74.0	-0.02	359.25	45.05	50.90	5.41

max ws = 42.2
acceptable shape < -0.01
Av1000 approp u Av10000 approp u
52 24.25 58 24.25

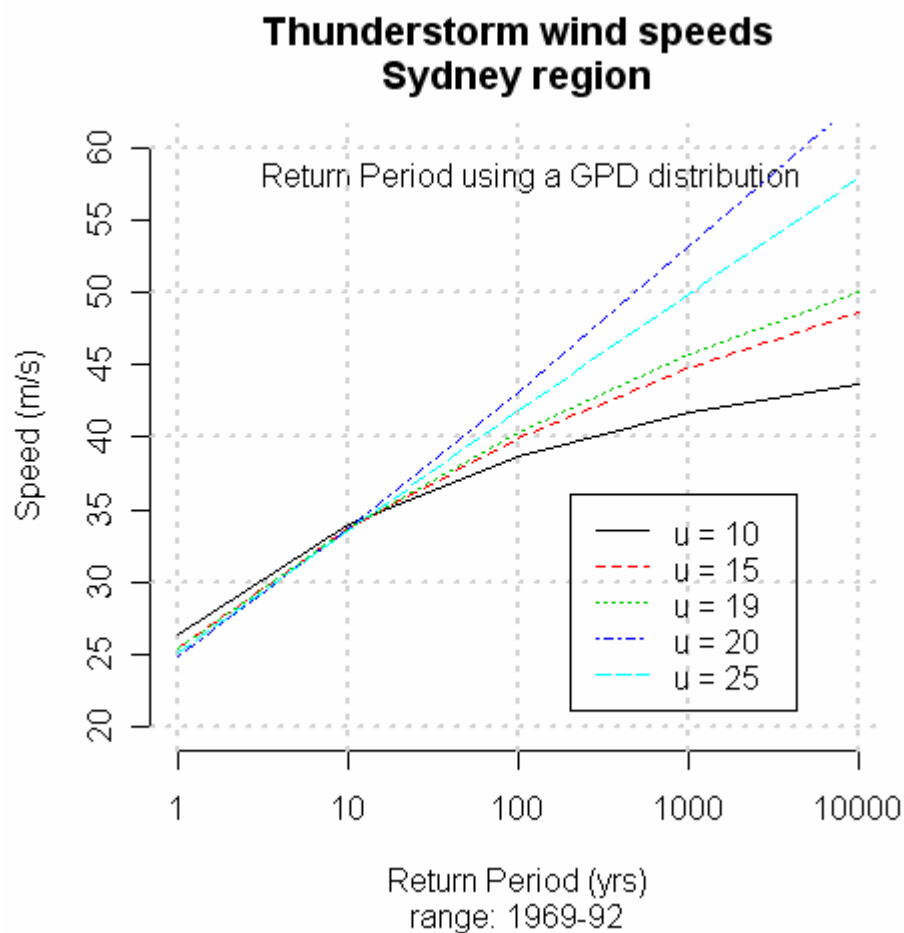


Figure 7.2a. Sydney Region *thunderstorm* threshold sensitivity

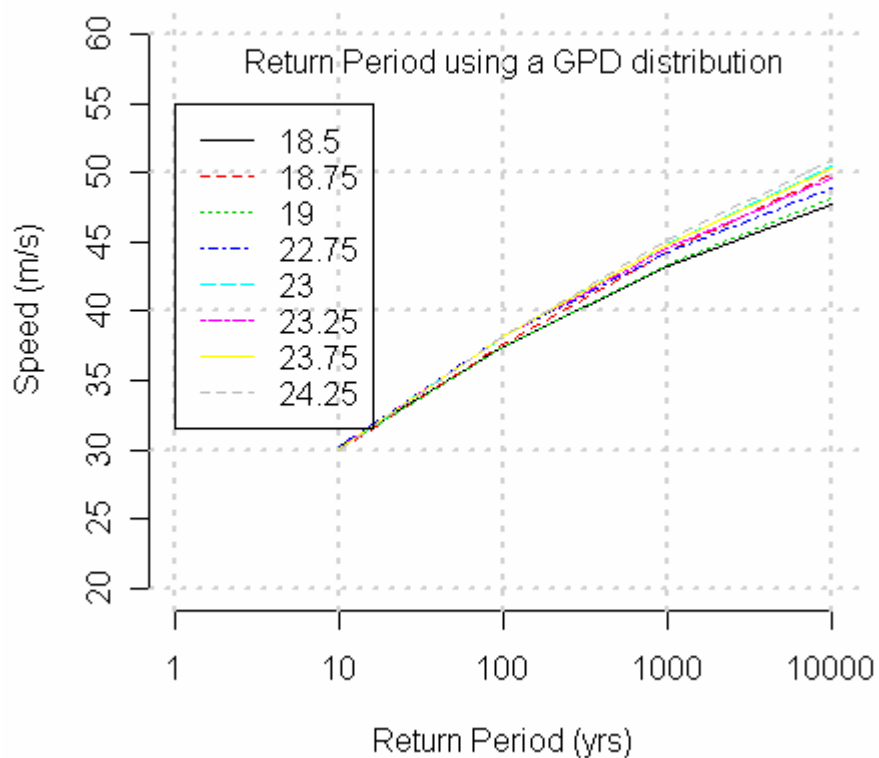


Figure 7.2b. Sydney Region *thunderstorm* with threshold selection [Holmes(1999) study period of record].

Sydney region thunderstorm wind speeds. GPD parameters						
u	No>u	shape	mx_pnt	q1000	q10000	qdiff
18.50	110.0	-0.10	88.37	43.20	47.72	4.52
18.75	110.0	-0.067	138.06	44.26	49.96	5.70
19.00	99.0	-0.097	94.54	43.44	48.14	4.70
22.75	36.0	-0.12	77.29	44.27	48.94	4.66
23.00	36.0	-0.08	113.88	44.86	50.51	5.64
23.25	33.0	-0.10	90.32	44.55	49.63	5.08
23.75	30.0	-0.085	107.91	44.84	50.33	5.49
24.25	27.0	-0.071	129.31	45.12	51.0	5.88
max ws = 42.2						
acceptable shape <		-0.01				
Av1000	approp u	Av10000	approp u			
51	24.25	57	24.25			

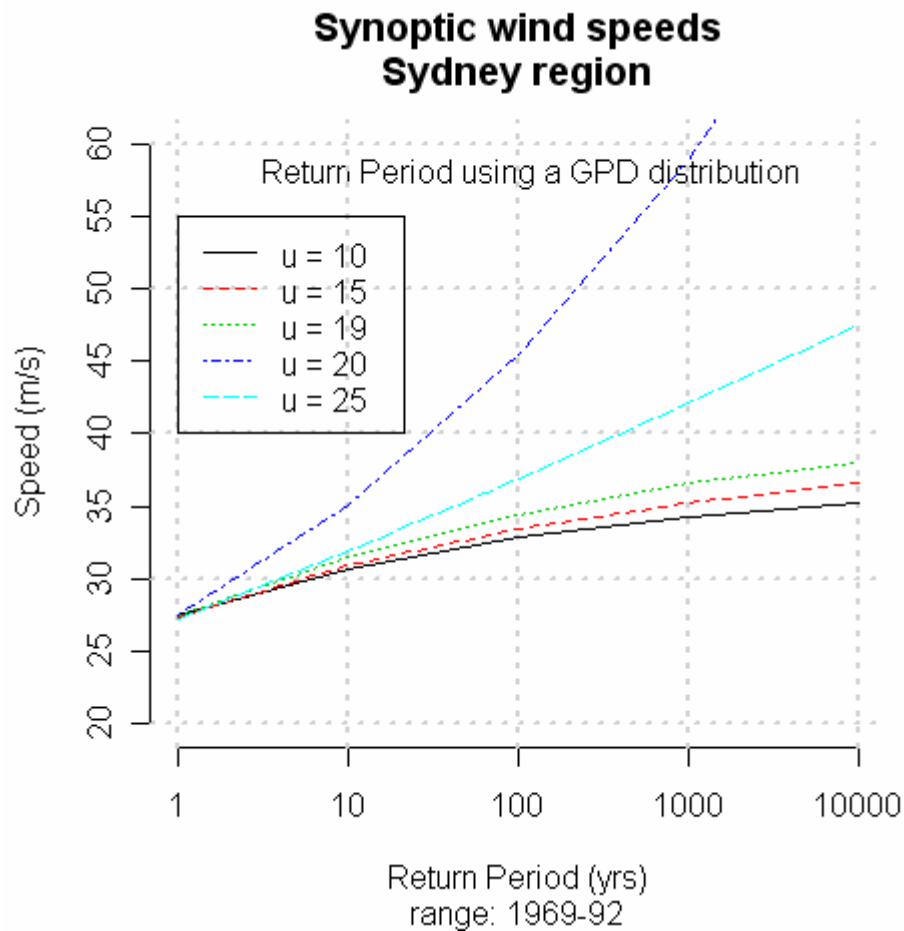


Figure 7.3a. Sydney Region *synoptic* wind threshold sensitivity

Figure 7.3a shows return periods for the *synoptic* winds (the wind gusts remaining following the extraction of the *thunderstorm* events from the *combined* wind dataset). The appropriate threshold for this dataset appears to be $u = 19$ since $u = 20$ and $u = 25$ produce unbounded curves. The automatic threshold selection algorithms selects $u = 27$ m/s which gives a curve much higher than $u = 19$ as shown in Figure 7.3b.

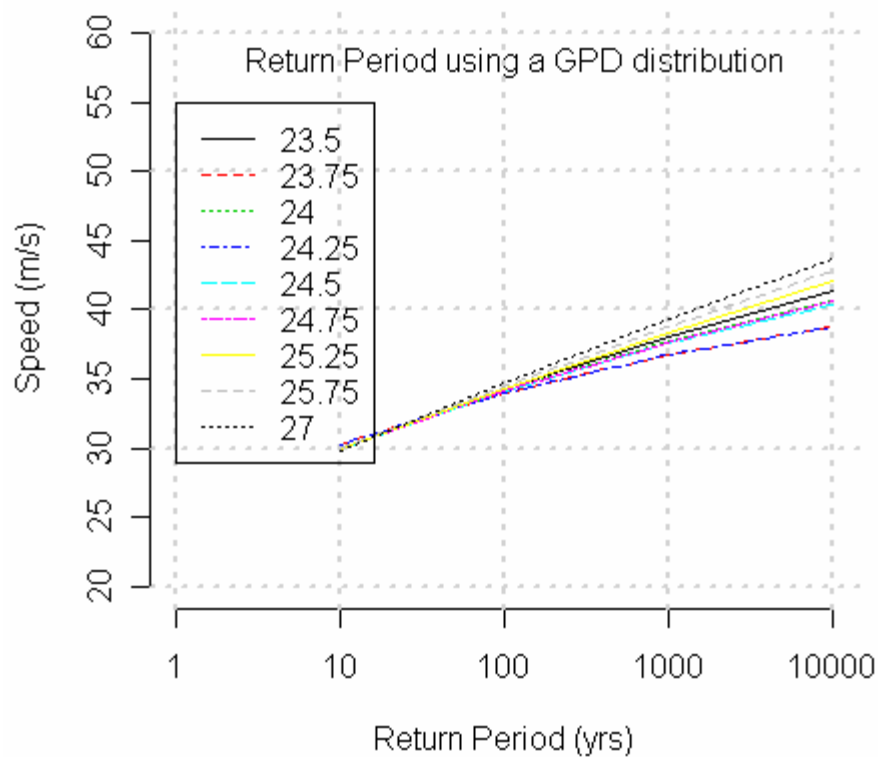


Figure 7.3b. Sydney Region *synoptic* wind with threshold selection (Holmes(1999) study period of record).

Sydney region synoptic winds. GPD parameters						
u	No>u	shape	mx_pnt	q1000	q10000	qdiff
23.50	126.0	-0.058	109.06	38.07	41.39	3.32
23.75	101.0	-0.14	44.69	36.77	38.75	1.99
24.00	101.0	-0.074	84.85	37.71	40.66	2.95
24.25	82.0	-0.15	43.30	36.81	38.77	1.96
24.50	82.0	-0.081	78.06	37.57	40.42	2.85
24.75	74.0	-0.075	84.21	37.68	40.65	2.96
25.25	60.0	-0.039	160.05	38.38	42.10	3.72
25.75	46.0	-0.029	215.81	38.76	42.77	4.01
27.00	23.0	-0.030	207.08	39.30	43.65	4.35
max ws = 36						
acceptable shape < -0.01						
Av1000	approp u	Av10000	approp u			
44	27	48	27			

As an evaluation exercise for the automatic technique, in the next section we next undertake a comparison of the three classes of winds (synoptic, thunderstorm and combined) with the Holmes (2002) results.

8. Model validation

Validation of the statistical model of wind speeds reported here was carried out by comparing our results with the results published in the Australian/NZ standards for design of structures subjected to severe winds (AS/NZS 1170, 2002).

Figure 8.1 presents a comparison of return periods for the three types of winds considered in this report over the reduced range 1969-1992. The initial threshold for each wind type, found in the previous Section, is as follows: *combined* winds: 24.5 m/s, *thunderstorms*: 24.25 m/s and *synoptic* winds: 27 m/s. However the final selection of threshold was done by plotting Figure 8.1 in an iterative fashion as explained in Section 6.3. It was found that the threshold for *thunderstorms* was too high and that the appropriate threshold was 22.75 rather than 24.25 m/s. Return periods for *combined* winds as presented in AS/NZS Table 3.1 'Regional Wind Speeds' for region A (1 to 7) has been added to the plot to facilitate comparison. The parameter 'nopy' of Equation (2.2) was calculated considering that the Sydney region dataset covers a range of years made up of the individual stations, that is, Sydney airport = 23 years, Bankstown = 21 years and Richmond = 22 years, hence the Sydney region range = 66 years.

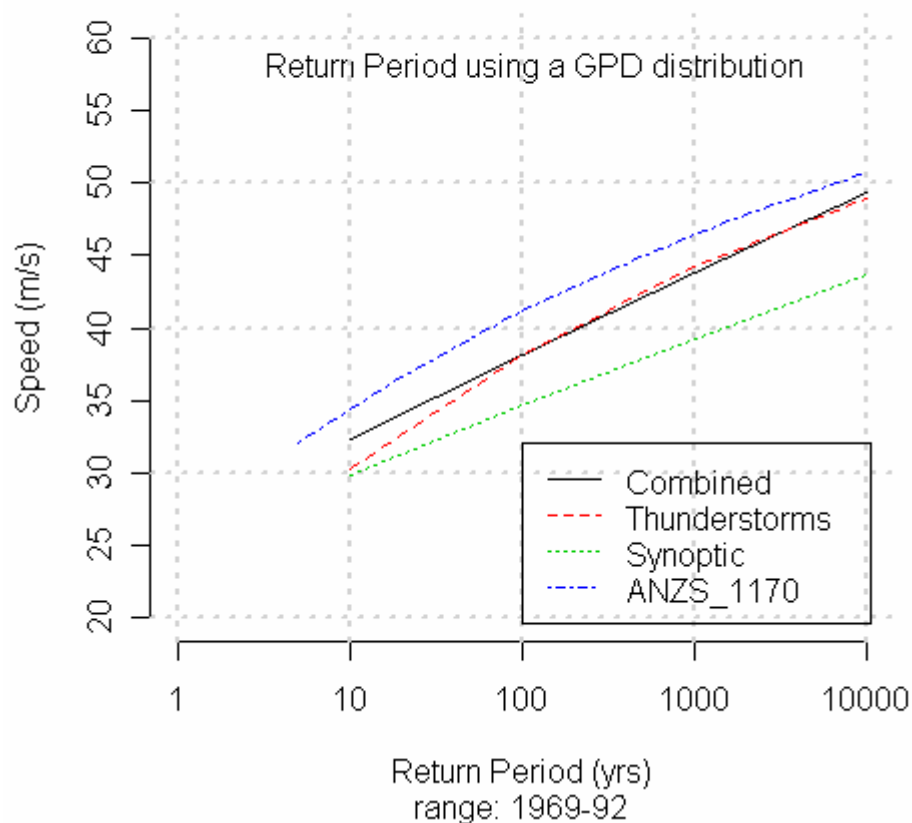


Figure 8.1. Sydney Region Wind Gusts

Figure 8.1 shows that return periods for *combined* and *thunderstorm* winds are very close to each other, particularly for mid to high values and hence the plot is dominated by *thunderstorms*. The *synoptic* wind dataset is consistently lower than combined winds for all return periods. It is also lower than thunderstorm winds except for low return periods with the crossover between synoptic winds and thunderstorms occurring at a return period of ten years. Notice that our *combined* dataset plot is close to the AS/NZS combined plot for high and low values, although it is consistently lower for mid values. This is expected since we are using different datasets, as discussed in Section 6.4.

The relativities of the *combined*, *thunderstorm* and *synoptic* wind datasets agree with previous results. The *combined* wind dataset is dominated by *thunderstorm* winds and hence *combined* and *thunderstorm* return periods are very close to each other. Once *thunderstorm* winds are removed, the remaining, synoptic winds produce a low speed curve. The figure also shows that our results are slightly lower than those presented in the AS/NZS standard (different dataset but approximately the same period of record). A larger range of years for our return periods will be considered below. Table 8.1 summarises the return period results for each of the stations which make up the Sydney region over the range indicated in the Table. As explained in Section 6.5 all return periods in the tables have been rounded up to the next integer.

These return periods are presented in graphical form in Figure 8.2. The Figure shows that *combined* wind speeds for the Sydney region are dominated by the Sydney airport wind speeds and that wind speeds from the Bankstown station present similar characteristics to the Sydney airport particularly for high return periods. The lowest wind speeds over the reduced range occur in the Richmond station.

Table 8.1. Return periods for *combined* winds – Sydney region (reduced period).

Return period (years)	10	100	1000	10000
Bankstown (1970-1991)	33	39	45	51
Richmond (1970-1992)	29	34	39	44
Sydney (reduced range) (1969-92)	34	40	46	51
Sydney region (1969-92)	32	38	44	49
AS/NZS_1170.2	34	41	46	51

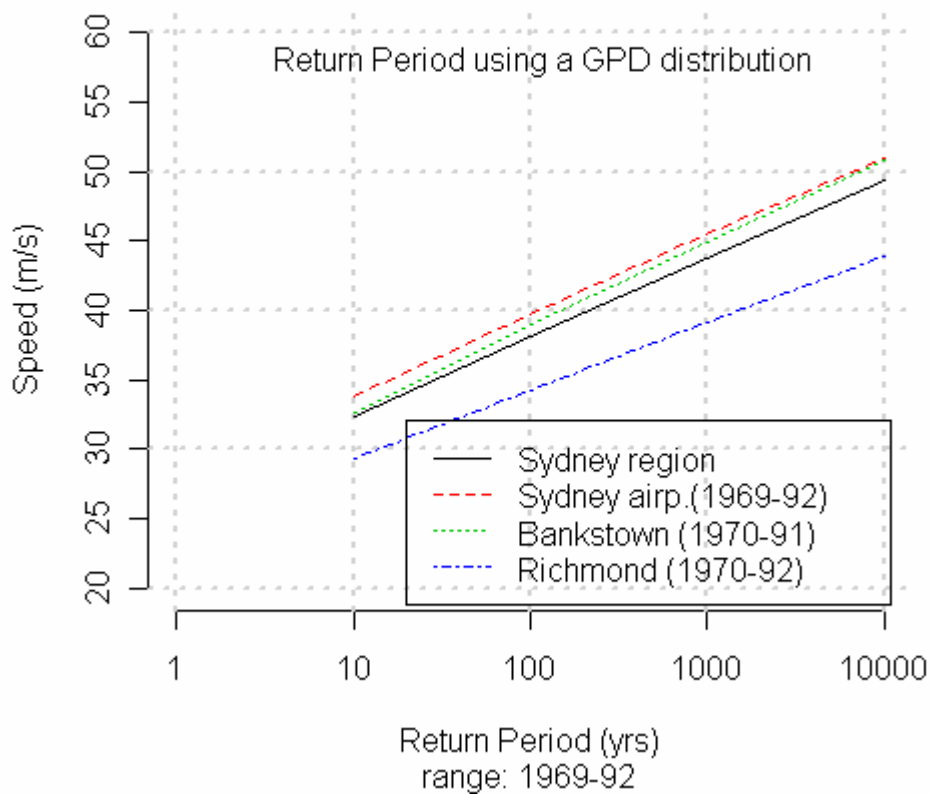


Figure 8.2. Combined wind speeds for the Sydney region and its components (Sydney, Richmond and Bankstown).

To further examine our results we calculated return periods for the Sydney region using all available data. The range of years for the Sydney region using the full range of the individual stations was 142 years (Sydney airport = 66, Bankstown = 24, and Richmond = 56 years). Figure 8.3 shows the results. The thresholds used were: *combined* wind: 26, *thunderstorm* winds: 18.75; and *synoptic* winds: 22 m/s.

Return periods for the Sydney region using the full data confirm our previous results; *thunderstorms* dominate the plot and tend to join the *combined* wind speeds curve for high return periods, while *synoptic* winds show lower speeds than *thunderstorms* for mid to high return periods but higher speeds for all other return periods, the crossover occurs at a return period of 12 years. The relative location of these and the *synoptic* wind curve follow the same tendency shown by the previous curves.

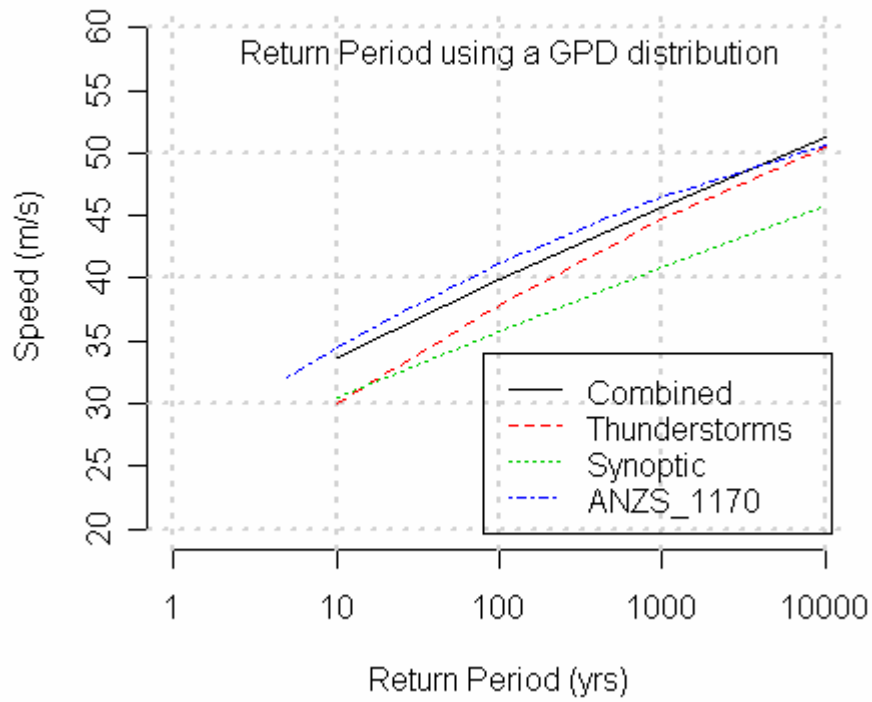


Figure 8.3. Sydney region with full range of record (142 years)

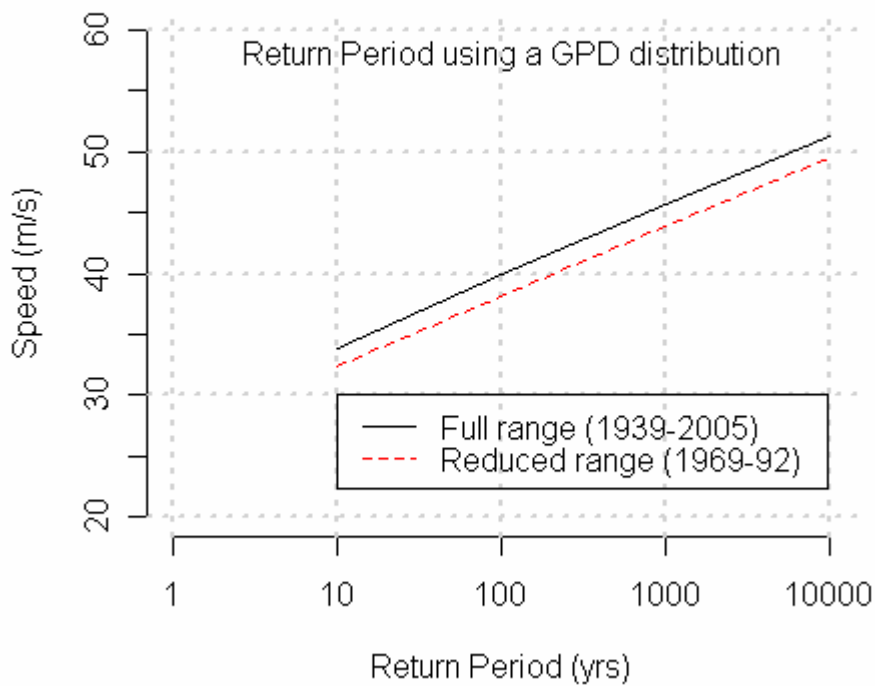


Figure 8.4. Comparison of full and reduced range curves

The curve presented in Figure 8.4 compares the return periods for combined winds using the Holmes (2002) reduced range of years with the full range of values used in our studies. The curves are almost parallel with the return periods using the full range of values higher than for the corresponding reduced range. This behaviour is consistent with the results presented previously and adds confidence to our model.

Table 8.2 compares our return periods of *combined* wind speeds for the Sydney region with results published by AS/NZS and Holmes (2002). It shows that our results using the full range of available data is very close to the results for region A published by the standards (AS/NZS 1170, 2002).

Table 8.2. Return period for *combined* winds – Sydney region.

Return period (years)	10	100	1000	10000
Sydney region (reduced range)	32	38	44	49
Sydney region (full range)	34	40	46	51
ANZS_1170	34	41	46	51
Holmes (2002)	33	39	46	53

9. Confidence Interval

A statistical model to calculate return periods of wind speed 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 return period 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 return period of wind speeds with the interval in which the true value of the return period can be found in 95% of cases.

The confidence interval depends on the amount and structure of the data samples, particularly the variance-covariance matrix which measures the spread of the samples around their mean. There are two basic techniques for calculation of confidence intervals for GPD-based return periods: the Delta and the Profile-likelihood method. The Delta method assumes normality of the samples and calculates the confidence interval using maximum likelihood. The Profile-likelihood method assumes a χ^2 distribution for the samples and maximises the log-likelihood function to calculate the appropriate confidence interval. Both methods have been implemented in the R environment by Gilleland and Katz (2005a) based on Coles (2001). Applying the methods to temperature data, Gilleland and Katz found out that the Profile-likelihood method gives better results because it considers the asymmetry of the data (2005b). For this reason the Profile-likelihood method as implemented in the R ‘extRemes’ package by Gilleland and Katz (2005a) is used in this work.

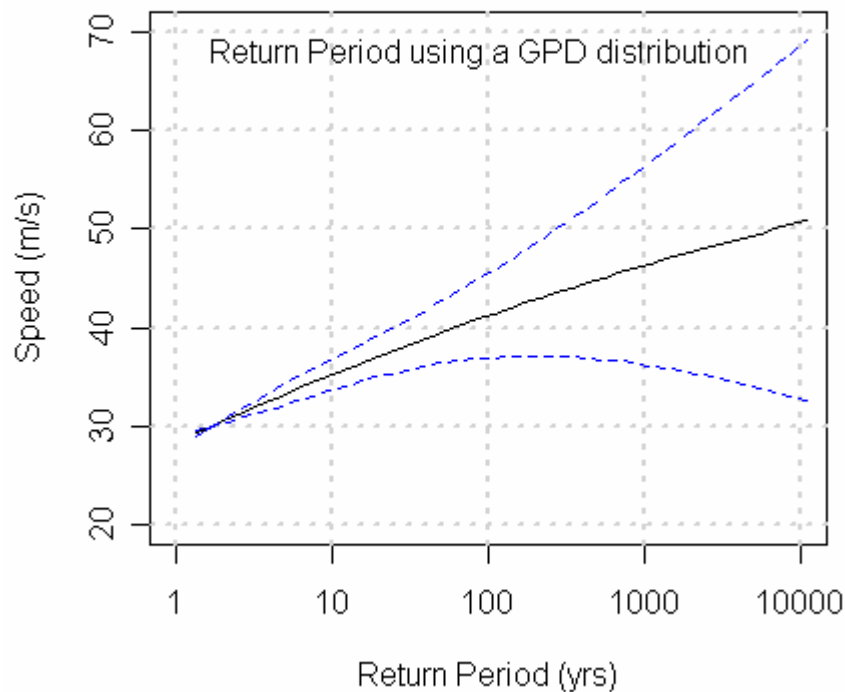


Figure 9.1. Return period (years) for Sydney airport *combined* wind gust speeds with 95% confidence interval.

Figure 9.1 shows the return periods of the Sydney airport combined wind speeds (using the full range of available data as presented in Figures 6.6 and 6.8) with its 95% confidence interval (dotted lines). Notice that the confidence interval substantially increases as the return periods increase indicating a higher degree of uncertainty when making inferences far beyond the range of the data (66 years). The confidence intervals suggest that return periods calculated beyond 1000 years are too unreliable for use in practical applications.

Figure 9.2 shows the return periods of *thunderstorm* wind speeds with their 95% confidence interval for the Sydney Airport site. To interpret these results consider the return period of thunderstorm wind speed at 1000 years as shown in Figure 9.2. It shows that the expected value is within [36, 50] 95% of the time. The maximum likelihood value returned by the R function [44 m/s] is within the range and hence is likely to be correct (i.e. in the correct range) with 95% probability based on the data considered.

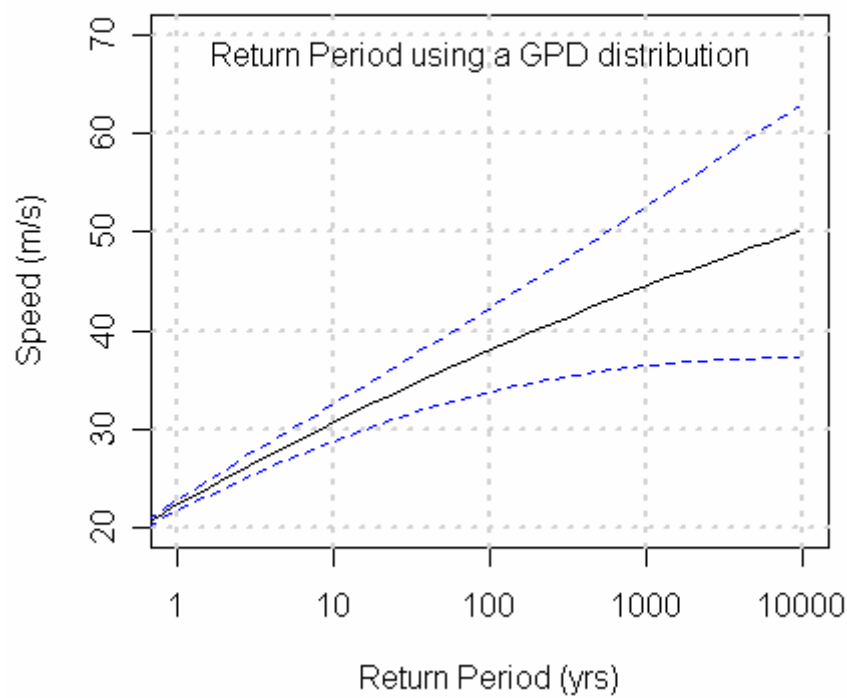


Figure 9.2. Sydney airport *thunderstorm* speed with confidence interval.

To complete this analysis, Figure 8.3 will be repeated showing the 95% confidence intervals for the three wind classes; see Figures 9.3a, 9.3b and 9.3c. As before the 95% confidence interval becomes wider as the return period increases. However the interval's width is different for each wind type with the widest interval occurring for *combined* winds. This indicates a cumulative degree of uncertainty for this type of wind which is made up of the other two wind classes.

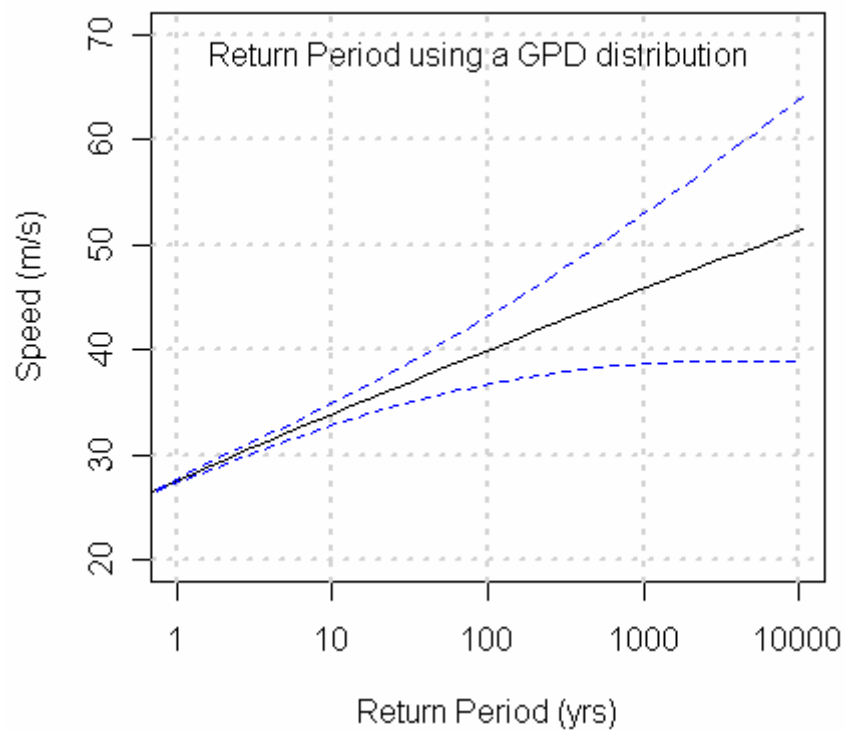


Figure 9.3a. Sydney region combined wind speed with 95% confidence interval. (Range: 1939-2005).

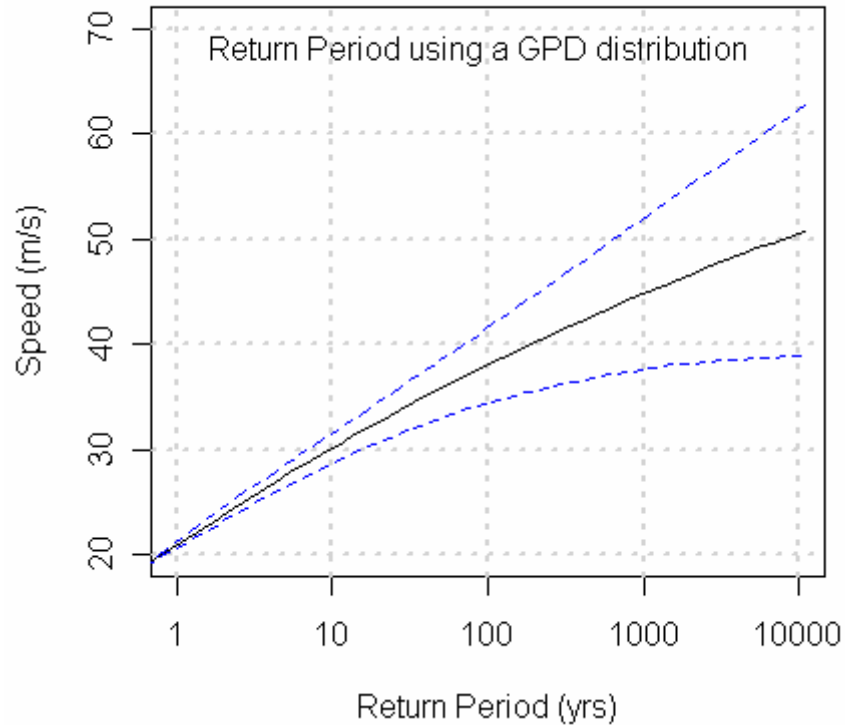


Figure 9.3b. Sydney region *thunderstorm* wind speed with 95% confidence interval. (Range: 1939-2005).

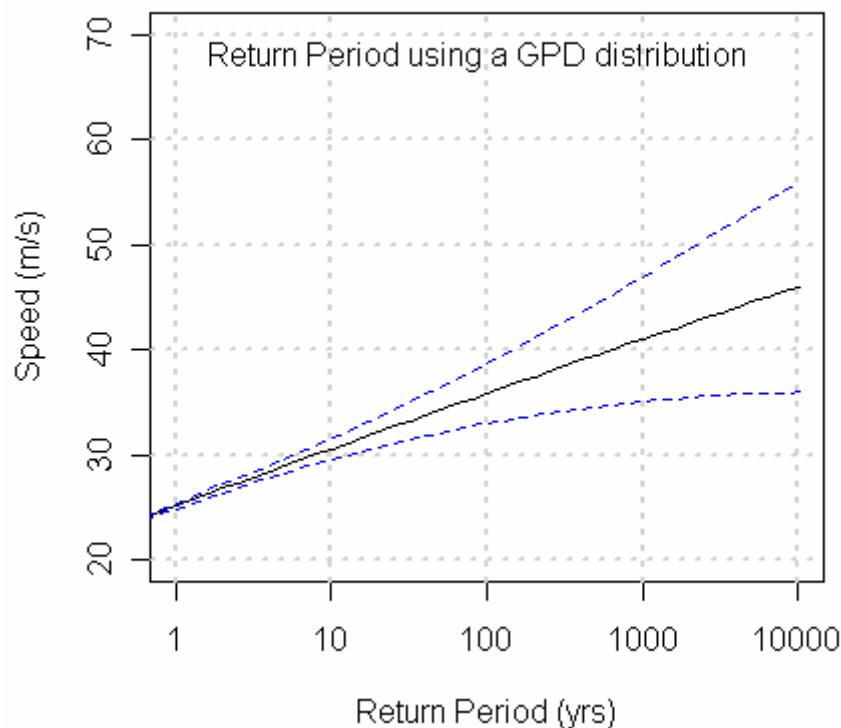


Figure 9.3c. Sydney region *synoptic* wind speed with 95% confidence interval. (Range: 1939-2005).

10. Relationship between wind means and gusts

As mentioned earlier, the calibration of anemometers involves calibrating the mean response of the instrument (i.e. record the response of the instrument under a constant wind forcing in a wind tunnel). This calibration procedure does not determine the transient response of the anemometer as experienced during a short wind gust (say of 1-3 second duration). Different types of anemometers may have different transient responses, and for older installations (where paper chart recorders were utilised) it is the instrument and recording system response that is important.

In Figure 10.1, the Sydney airport dataset has been split into three periods to determine whether the relationship between wind means and gusts remains constant throughout the whole dataset. Significant differences were found between the earlier and latter two periods. This phenomenon requires further investigation, however we can say that there is more confidence in the mean measurements of wind speed than in the gust measurements, and that there is more confidence in the measurements from the synchrotac anemometer (introduced during the 1980s) than in the older Dynes anemometers (Muirhead et al., 2005).

We are developing a methodology that employs Monte Carlo simulation to create a synthetic daily maximum wind gust dataset based on the daily maximum wind mean dataset and the relationship between mean wind and gust calculated from the more recent part of the dataset (from the 1980s in most cases).

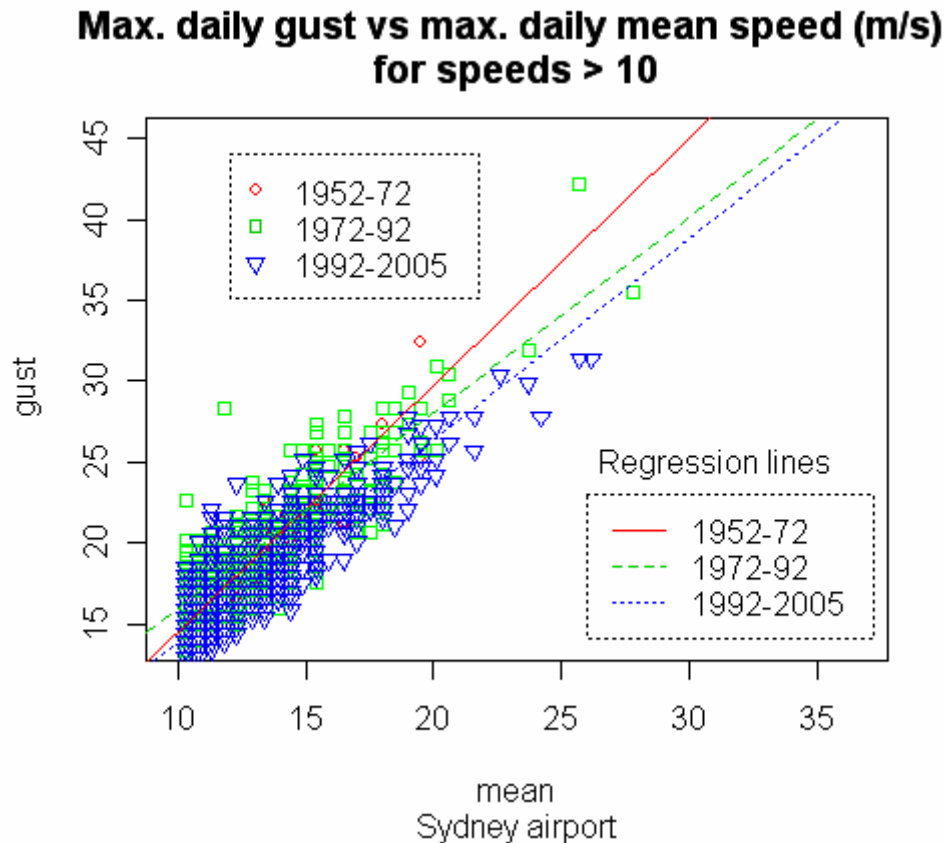


Figure 10.1. Correlation between gust and mean wind speed

11. A note on uncertainty modelling

The Risk Research Group (RRG) has produced guidelines for the incorporation of uncertainty into the risk models developed by the group. The aim of these guidelines is to make sure that all our models are consistent. This is essential for a proper comparison of the risk produced by different hazards (Sanabria & Dhu, 2005). The guidelines state that it is important to distinguish and represent two kinds of uncertainty: aleatory and epistemic. Aleatory uncertainty includes the natural variability and inherent randomness of complex natural phenomena; it can be estimated but cannot be reduced. This type of uncertainty can be modelled by using random variables. In this work aleatory uncertainty of wind speed gusts is modelled by using extreme value distributions.

Epistemic uncertainty, on the other hand, is the result of inadequate data and incomplete model development due to limitations in knowledge of the phenomenon's physics. This kind of uncertainty can be reduced with better data collection, advances in knowledge of the phenomenon physics and refinement of models to represent it. A good illustration of the problem can be seen in Table 8.2. The return period of wind speed for the Sydney region calculated by our method gives a value of 49 m/s at 10000 years while Holmes method gives 53 m/s for the same range of years. A good way to deal with epistemic uncertainty is to calculate this return period using both methods; the final result could be a weighted average of the individual results. The weight could be determined by the analyst after discussion with colleagues, literature review, etc. (Sanabria & Dhu, 2005).

Another source of epistemic uncertainty in this work is the quality of the data, as discussed in the previous section. As indicated before we are developing a Monte Carlo simulation method to generate synthetic wind gust speeds in order to assess the size of the epistemic error in our results.

Table 12.1. Summary of results

Return period (years)	10	100	1000	10000
AS/NZS_1170.2	34	41	46	51
Holmes (2002)	33	39	46	53
Sydney region (reduced dataset)	32	38	44	49
Sydney region (available data to 2005)	34	40	46	51
Sydney airport (1939-2005)	35	41	46	51
Bankstown (1968-1992)	32	38	44	49
Richmond (1994-2005)	27	29	31	32
Richmond (1942-1994)	32	39	44	50
Nowra (1955-1997)	35	39	44	48
Williamtown (1942-2005)	33	39	45	50
Observatory Hill (1955-1992)	35	41	46	51

12. Conclusions

A statistical model of severe winds has been developed based on a family of Extreme Value distributions known as the Generalised Pareto Distribution. The model allows users to calculate and plot return periods for maximum gust wind speeds based on datasets from observing stations provided by meteorological agencies such as the Australian Bureau of Meteorology. In spite of the fact that the given datasets cover only a few decades of data, using these distributions it is possible to calculate return periods from 100 to 5000 years and beyond. 95% confidence intervals for the return periods were calculated showing that the uncertainty in the results increases with the magnitude of the return period.

The algorithm for separation of *thunderstorm* and *synoptic* wind components presented in Section 6 shows results similar to those reported by Holmes (2002) and also to the wind gust speeds published in the wind loadings standard (AS/NZS_1170.2, 2002). Comparison of the Sydney region return period wind gust speeds and also Sydney region observing station return period wind gust speeds is shown in Table 12.1. The advantage of our algorithm is that it can be easily computationally implemented using observing station datasets available from the BoM. Hence our algorithm is more efficient than visual inspection of anemometer charts and can be used for large scale investigation of wind hazards in Australia.

Our results also show that the Generalised Pareto Distribution (GPD) can produce acceptable results over a large range of return periods. The main limitation of this distribution is the selection of the appropriate threshold for a given dataset. As discussed in Sections 7 and 8 the results are very sensitive to the threshold selection and hence an automatic procedure to solve this problem is necessary. Such a procedure, developed during the course of this work, was presented in Section 6.3.

13. Acknowledgements

We want to express our gratitude to Dr. J. Holmes for his generous help during the preparation of this report. A few points were clarified in a useful discussion with him in Melbourne and he gave us copy of the processed wind speed data utilised for the derivation of the AS/NZS wind loading standard. His assistance facilitated the validation of the algorithm. He also reviewed the final version of this report. We also want to express our appreciation to Doerte Jacob and Ian Muirhead from the BoM who reviewed this document. Comments from all reviewers greatly improved the final document.

We thank our colleagues Ole Nielsen who provided support with the installation of R packages and system management of the LINUX machine; and Jane Sexton for her review of the final version of this document.

14. References

AS/NZS 1170.2 (2002) Structural design actions, Part 2: Wind actions, *Australian/New Zealand Standard*, 2002.

AS/NZS 1170.2 Supp 1 (2002) Structural design actions — Wind actions— Commentary, *Australian/New Zealand Standard*, 2002.

Bureau of Meteorology (1982). Department of Science and Technology (1982). Present Weather Plotting Diagram (Description abridge from W.M.O. Code). Jan. 1982.

Brabson B.B. and Palutikof J.P. (2000). Test of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *Journal of Applied Meteorology*, Vol. 39, 1627-1640.

Cechet P.R., Nadimpalli K., Edwards M. and Sanabria L.A. (2007). Severe Wind Gust Risk for Australian Capital Cities – A National Risk Assessment approach. Accepted for presentation to MODSIM-07. Christchurch, NZ. Dec. 10-13, 2007.

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

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

Gillelland E. and Katz R.W., (2005a). *Extremes Toolkit (extremes): Weather and Climate Applications of Extreme Value Statistics*. National Center for Atmospheric Research (NCAR). Boulder CO, USA.

Gillelland E. and Katz R.W., (2005b). *Analysing Seasonal to Interannual Extreme Weather and Climate Variability with the Extremes Toolkit*. National Center for Atmospheric Research (NCAR). Boulder CO, USA.

Gomes L. and Vickery B.J. (1977). On The Prediction of Extreme Wind Speeds from the Parent Distribution. *J. Ind. Aerodyn.*, 2:21-36.

- Heckert N.A., Simiu E. and Whalen T. (1998). Estimates of Hurricane Wind Speeds by 'Peaks over Threshold' method. *Journal of Structural Eng.* April 1998. 445-449.
- Holmes J.D. (2002). A Re-analysis of Recorded Extreme Wind Speeds in Region A. *Australian Journal of Structural Engineering*. Vol. 4 No. 1.
- Holmes J.D. and Moriarty W.W. (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Eng. and Industrial Aerodynamics* 83 (1999) 1-10.
- Holmes J.D. (2006a). Sydney airport downburst wind speeds 1969-1992. Personal communication. April 2006.
- Holmes J.D. (2006b). Comments on 'A Statistical Model of Severe Winds'. Personal communication. November 2006.
- Jagger T.H. and Elsner J.B. (2006). Climatology Models for Extreme Hurricane Winds near the United States. *Journal of Climate*. Vol. 19, 3220-3236.
- Nadimpalli K, Cechet R.P. and Edwards M. (2007). First Steps Towards a National Assessment of Australian Wind Risk. Accepted for presentation to the 12th Int. conference on Wind Eng. Cairns, Australia. 1-6 July, 2007.
- Lechner J.A., Leigh S.D. and Simiu E. (1992). 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.
- Maindonald, J. H. (2005). The R system – and introduction and overview. Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne.
- 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. November 2005.
- Palutikof J.P., Brabson B.B., Lister D. H. and Adcock S.T. (1999). A Review of Methods to Calculate Extreme Wind Speeds. *Meteorol. Appl.* 6, 119-132.
- Seguro J.V. and Lambert T.W. (2000). Modern Estimation of the Parameters of the Weibull Wind Speed Distribution for Wind Energy Analysis. *Journal of Wind Eng. and Industrial Aerodynamics* 85 (2000) 75-84.
- Sanabria L.A. and Dhu T. (2005). A Methodology for Consistent Modelling of Natural Hazards. Proceedings Int. Congress on Modelling and Simulation MODSIM05. University of Melbourne, Australia. December 2005.
- Stephenson A (2004). A User's Guide to the 'EVD' Package (Version 2.1). Department of Statistics. Macquarie University. Australia.

Appendix A

Return periods for Bankstown, Richmond, Williamstown and Nowra with corresponding automatic threshold selection.

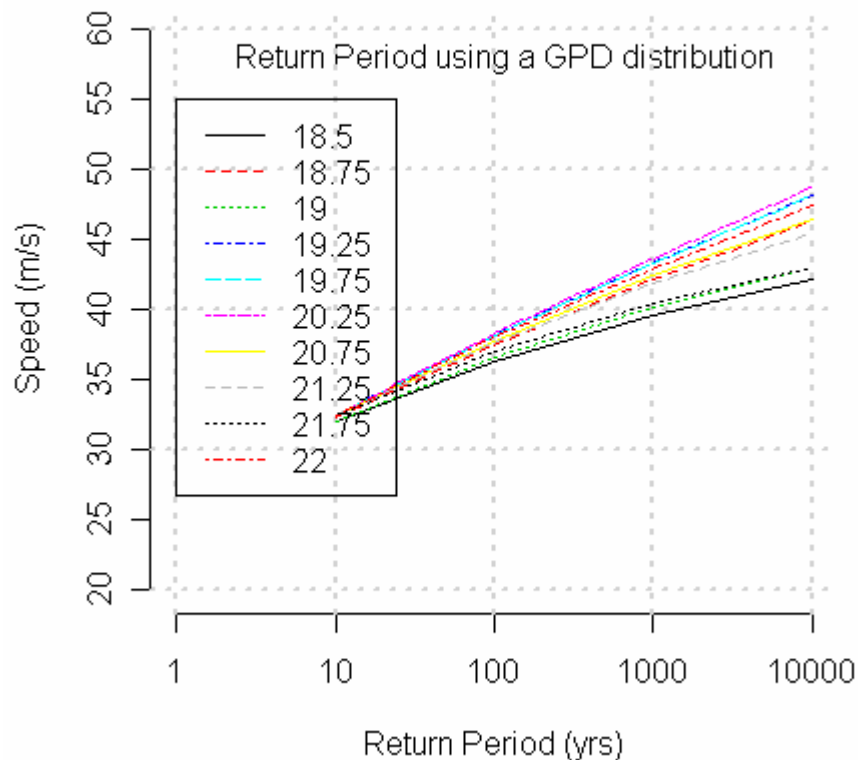


Figure A1. Return periods for Bankstown *combined* gust wind speeds and corresponding table for automatic selection of threshold (below).

u	Bankstown total wind speed. GPD parameters					
	No>u	shape	mx_pnt	q1000	q10000	qdiff
18.50	289.00	-0.12	55.32	39.57	42.08	2.51
18.75	289.00	-0.058	110.37	42.14	46.25	4.12
19.00	252.00	-0.10	62.65	40.18	42.98	2.81
19.25	252.00	-0.037	173.31	43.34	48.18	4.84
19.75	211.00	-0.039	162.82	43.33	48.13	4.81
20.25	178.00	-0.037	175.03	43.55	48.63	5.08
20.75	146.00	-0.062	102.35	42.36	46.43	4.07
21.25	122.00	-0.076	84.64	41.91	45.48	3.57
21.75	99.00	-0.12	52.58	40.41	43.01	2.60
22.00	99.00	-0.051	125.82	42.91	47.45	4.541

```

max ws = 37.1
acceptable shape < -0.01
Av1000  approp u    Av10000  approp u
49      20.25       53       20.25

```

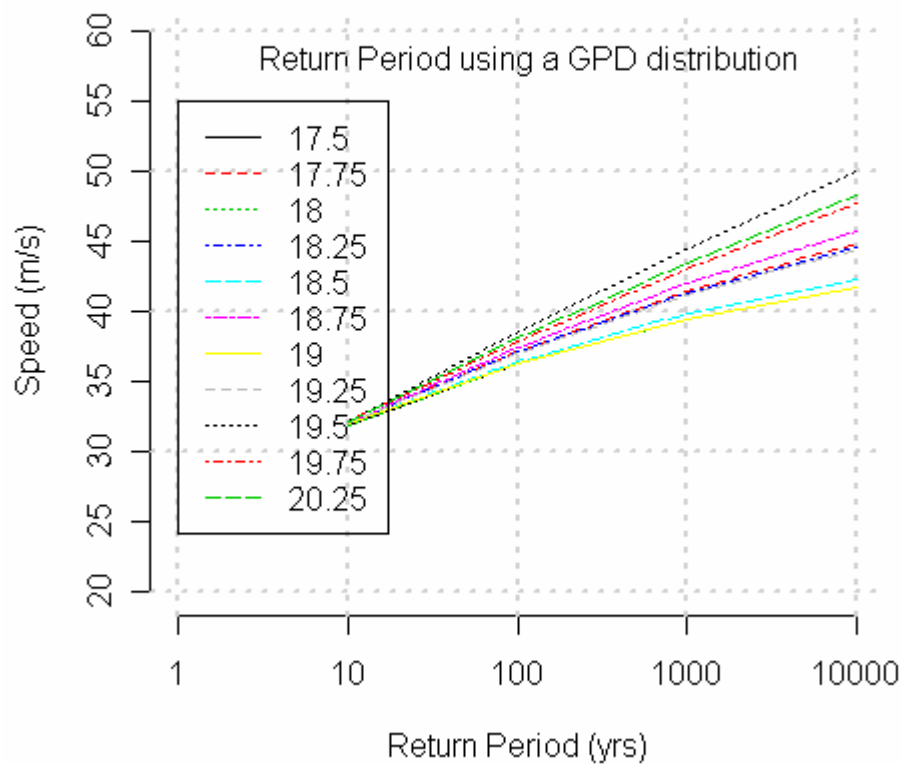


Figure A2. Return periods for Richmond *combined* gust wind speeds and corresponding table for automatic selection of threshold (below).

u	Richmond total wind speed. No>u	GPD parameters shape	mx_pnt	q1000	q10000	qdiff
17.50	549.0	-0.14	36.02	39.42	41.73	2.31
17.75	549.0	-0.089	55.90	41.39	44.81	3.43
18.00	478.0	-0.14	35.26	39.41	41.68	2.27
18.25	478.0	-0.092	53.69	41.21	44.57	3.36
18.50	424.0	-0.13	38.21	39.84	42.32	2.48
18.75	424.0	-0.077	64.66	41.94	45.72	3.78
19.00	361.0	-0.14	34.32	39.50	41.74	2.24
19.25	361.0	-0.095	52.02	41.15	44.42	3.27
19.50	361.0	-0.028	178.77	44.46	49.97	5.51
19.75	326.0	-0.054	91.30	42.10	47.64	4.64
20.25	280.0	-0.046	107.21	43.47	48.30	4.82

```

max ws = 37.6
acceptable shape < -0.01
Av1000 approp u Av10000 approp u
48 19.5 52 19.5

```

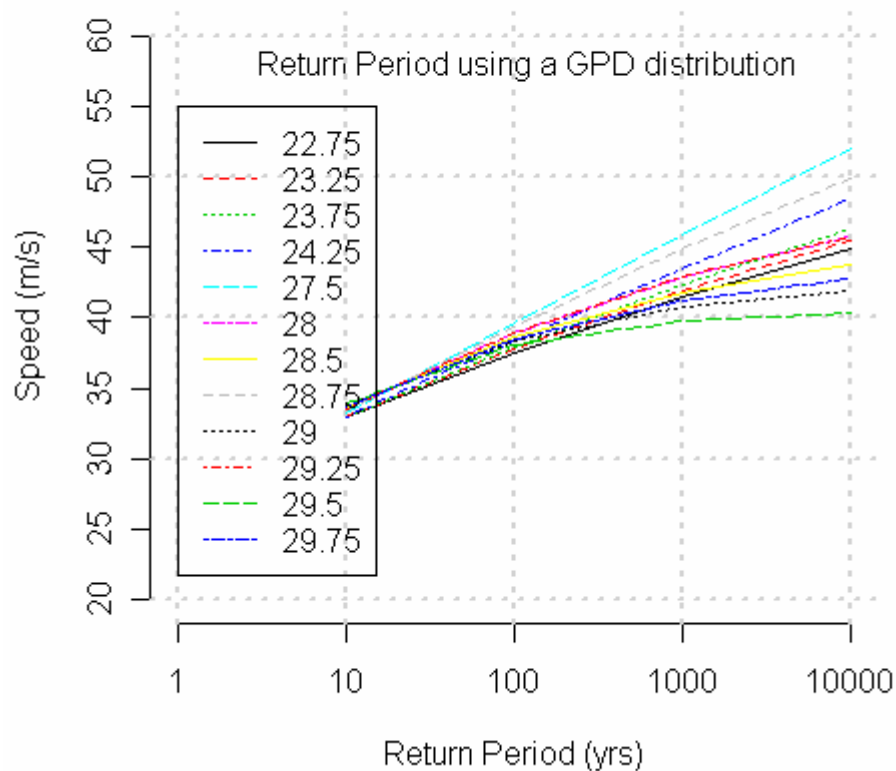


Figure A3. Return periods of *combined* wind speeds for Williamstown and corresponding table for automatic selection of threshold (below).

Williamstown total wind speed. GPD parameters						
u	No>u	shape	mx_pnt	q1000	q10000	qdiff
22.75	367.0	-0.069	105.95	41.50	44.87	3.37
23.25	303.0	-0.062	116.77	41.86	45.40	3.54
23.75	251.0	-0.05	144.58	42.39	46.33	3.94
24.25	211.0	-0.021	341.43	43.51	48.49	4.98
27.50	47.0	-0.014	528.60	45.95	52.04	6.09
28.00	37.0	-0.14	52.30	42.86	45.71	2.85
28.50	31.0	-0.20	35.92	41.81	43.82	2.00
28.75	31.0	-0.048	153.20	44.87	49.82	4.95
29.00	26.0	-0.29	24.83	40.73	41.92	1.19
29.25	26.0	-0.14	53.02	42.85	45.72	2.87
29.50	22.0	-0.42	17.53	39.72	40.33	0.61
29.75	22.0	-0.25	29.71	41.24	42.80	1.55

```

max ws = 38.1
acceptable shape < -0.01
Av1000 approp u Av10000 approp u
49 27.5 53 27.5

```

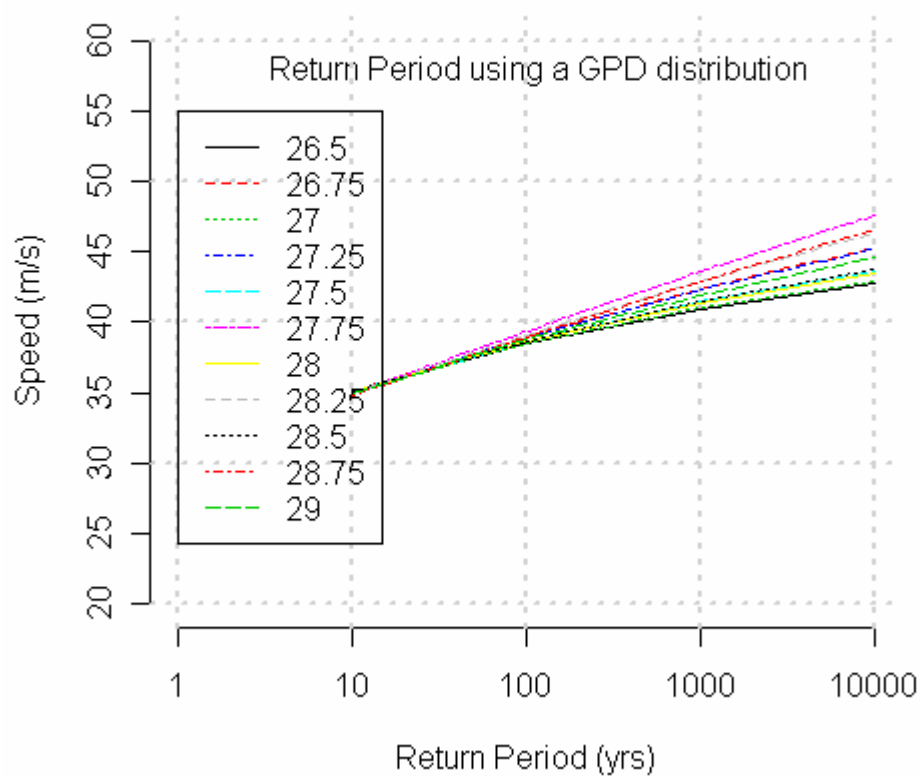


Figure A4. Return periods of *combined* wind speeds for Nowra and corresponding table for automatic selection of threshold (below).

u	Nowra total wind speed. GPD parameters					
	No>u	shape	mx_pnt	q1000	q10000	qdiff
26.50	213.0	-0.13	48.54	40.96	42.78	1.82
26.75	213.0	-0.067	95.73	42.39	45.30	2.92
27.00	178.0	-0.13	49.83	41.03	42.90	1.86
27.25	178.0	-0.063	101.22	42.35	45.24	2.89
27.50	153.0	-0.10	62.07	41.44	43.65	2.21
27.75	153.0	-0.018	354.35	43.57	47.60	4.04
28.00	123.0	-0.11	58.51	41.38	43.51	2.13
28.25	123.0	-0.035	184.27	42.84	46.35	3.50
28.50	101.0	-0.10	62.90	41.51	43.76	2.25
28.75	101.0	-0.026	247.83	42.89	46.57	3.69
29.00	85.0	-0.072	89.60	41.96	44.67	2.71

max ws = 40.7

acceptable shape < -0.01

Av1000	approp u	Av10000	approp u
49	27.75	52	27.75

Appendix B.

B1. The R-statistics package

The R-statistic package is an open source programming language and corresponding environment developed by volunteers from industrial and research institutions for statistical applications. One of the most interesting characteristics of the R-environment is the periodic installation of new packages, as they become available from their developers. This means that the R-environment is constantly updated and complete packages for specialised statistical work are available for all types of applications.

Most of the software developed for this project uses the ‘EVD’ package developed by A. Stephenson (2004) from Macquarie University, based on Coles’ book *An Introduction to Statistical Modelling of Extreme Values* (Coles, 2001). This package contains a collection of functions to fit extreme value distributions to data.

The R-language itself is rich in features and has been developed using state-of-the-art software techniques like object oriented and vector oriented data modelling. Being open source software and supported by a very enthusiastic community of programmers and developers has contributed to the growing popularity of this programming environment (Maindonald, 2005).

R-package versions for Windows, LINUX and Mac machines have been developed. The work reported in this document was conducted in a LINUX environment; this environment was selected because it offers two main advantages: memory capacity and processing speed. These characteristics are particularly important when working with big datasets like most of the Australian Bureau of Meteorology (BoM) wind datasets.

User interaction with the R-environment is done using the Command Line Interface (CLI), it allows the user to type commands directly in to the shell for processing. For this reason it is customary to develop the software by cutting and pasting statements directly into the shell, which facilitates development, testing and checking of the software. By saving these statements into files it is possible to develop a large collection of software tools for different tasks. Some of these tools are functions that can be called from programs to add flexibility to the software.

In this work the files of R-software were written in plain text (Unicode text) and were given the extension ‘*.r’ to distinguish them from other documents. Lines starting with the sign ‘#’ are comments. To facilitate use and maintenance of this software a brief description of the program, its main purpose and required parameters is included in the first comment line.

Appendix B2 contains a comprehensive list of all R-programs associated with this project.

B2. R-software developed for this project

Program name	Description
WStn_select.r	This program reads an external BoM dataset and loads it into the R-shell as a dataframe. Most programs listed here require this dataframe to work. The program is interactive, ie. the user has to cut and paste some lines of the code into the shell as indicated. Requires functions: 'pick_file.r' and 'sel_file.r'
Act_retp.r	Calculates actual return periods from given data. It also calculates and plots histogram of data, pdf and CDF. Requires functions: 'pick_file.r' and 'calc_aretp.r'
retp_full.r	This program merges weather description dataset with max. daily wind speed and extracts thunderstorm and synoptic winds into vectors. It works with the full dataset. Returns a 4-column matrix. Requires: 'snatch.r'
retp_cut.r	The same as previous program but this one cuts the range of available data in to a range defined by the user.
EV_tests.r	This program implements 2 tests to check threshold selection in fitting a GPD to given data.
GPD_distr.r	Calculate return periods for given data using a GPD (fitted using ML). Returns 2-column matrix of return periods (retp, ws)
GEV_distr.r	Calculate return periods for given data using a GEV (fitted using ML). Returns 2-column matrix of return periods (retp, ws)
my_Rplot.r	Plots multiple return periods in a single plot (Figure 7.1)
sel_approp_u.r	This is the algorithm for automatic selection of the appropriate threshold to fit the GPD to given dataset (the rules presented in Appendix C have been coded here). It writes into file 'GPD_table.txt' a summary of procedure. Returns suggested 'u' values. Requires: 'GPD_distr.r', 'my_Rplot.r', 'calc_qsh.r', 'closest_u.r'
Holm_retp.r	Generates return periods for Sydney airport thunderstorms calculated by Dr. Holmes. Returns a 2-D vector (retp,ws).
retp_ANZS.r	Returns a 2-col matrix of return periods as presented in ANZS_1170 (retp,ws)
Calc_aretp.r	Calculates actual return period from given dataset. Returns a 2-D vector (ws,retp)
cnfInt.r	Function to calculate and plot return periods (using a GPD) with 95% confidence interval. Requires: 'my_rlplot.r'.

Appendix C.

Summary of rules for automatic selection of threshold 'u'

Rule 1: Calculate entries for 'GPD_table' and plot acceptable GPD curves. Start with an initial threshold of $\text{int}[(1/2)*\max(ws) + 0.5]$. Where ws is the vector of wind speeds and 'int' means 'round up to next integer'.

Rule 2: Increment initial threshold in steps of 0.25 m/s.

Rule 3: Eliminate entries with shape parameter greater than or equal to -0.01.

Rule 4: Eliminate entries with $q1000 = \text{NA}$ or $q10000 = \text{NA}$ (in other words when the algorithm cannot calculate the quantiles corresponding to 'years' = 1000 ($q1000$) or 'years' = 10000 ($q10000$) because of numerical problems caused by the large quantile).

Rule 5: Eliminate entries with $(q10000 - q1000) > 0.12*q10000$ (if rule is not met it is a sign of numerical stability problems calculating the large quantile 'years' = 10000).

Rule 6: Calculate average of quantile columns. Used them to evaluate expression

$$\text{Av}q10000 = 1.15*\text{average}(\text{column } q10000)$$

Select GPD curve closest to $\text{Av}q10000$ and return corresponding threshold 'u'.

Repeat previous calculation for $\text{Av}q1000$. The selected threshold 'u' should be the same as in previous rule (in cases in which two or more curves are too close to each other, the result of $\text{Av}q1000$ and $\text{Av}q10000$ could be different, in these cases the user should select the higher value).

Rule 7. In some cases the maximum likelihood (ML) method used by the R GPD function cannot find a global maximum and generates a message warning the user. These warning messages can be transformed into error messages to eliminate entries which present these kinds of problems.

Appendix D.

Table D1. Present weather classification (W.M.O. present weather code).

Present Weather (descriptions abridged from W.M.O. Code)






















	0	1	2	3	4	5	6	7	8	9
00	 Cloud development not observed or not observable during past hour	 Clouds generally dissolving or becoming less developed during past hour	 State of sky on the whole unchanged during past hour	 Clouds generally forming or developing during past hour	 Visibility reduced by smoke	 Haze	 Widespread dust in suspension in the air, not raised by wind, at time of obs	 Dust or sand raised by wind, at time of obs	 Well developed dust devil(s) within past hour	 Duststorm or sandstorm within sight of station or at station during past hour
10	 Light fog	 Patches of shallow fog at station not deeper than 6 feet on land	 More or less continuous shallow fog at station not deeper than 6 feet on land	 Lightning visible, no thunder heard	 Precipitation within sight, but not reaching the ground	 Precipitation within sight, reaching ground, but distant from station	 Precipitation within sight, reaching the ground, near to but not at station	 Thunder heard but no precipitation at the station	 Squall(s) within sight during past hour	 Funnel cloud(s) within sight during past hour
20	 Drizzle (not freezing, not showers) during past hour, not at time of obs	 Rain (not freezing, not showers) during past hour, not at time of obs	 Snow (not falling as showers) during past hour, not at time of obs	 Rain and snow (not falling as showers) during past hour, not at time of obs	 Freezing drizzle or rain (not showers) during past hour, not at time of obs	 Showers of rain during past hour, but not at time of obs	 Showers of snow, or of rain and snow during past hour, but not at time of obs	 Showers of hail, or of hail and rain during past hour, but not at time of obs	 Fog during past hour, but not at time of obs	 Thunderstorm (with or without precip) during past hour, but not at time of obs
30	 Slight or moderate duststorm or sandstorm, has decreased during past hour	 Slight or moderate duststorm or sandstorm, no appreciable change during past hour	 Slight or moderate duststorm or sandstorm, has increased during past hour	 Severe duststorm or sandstorm, has decreased during past hour	 Severe duststorm or sandstorm, no appreciable change during past hour	 Severe duststorm or sandstorm, has increased during past hour	 Slight or moderate drifting snow, generally low	 Heavy drifting snow, generally low	 Slight or moderate drifting snow, generally high	 Heavy drifting snow, generally high
40	 Fog at distance at time of obs but not at station during past hour	 Fog in patches	 Fog, sky discernable, has become thinner during past hour	 Fog, sky not discernable, has become thinner during past hour	 Fog, sky discernable, no appreciable change during past hour	 Fog, sky not discernable, no appreciable change during past hour	 Fog, sky discernable, has begun or become thicker during past hour	 Fog, sky not discernable, has begun or become thicker during past hour	 Fog, depositing rime, sky discernable	 Fog, depositing rime, sky not discernable
50	 Intermittent drizzle (not freezing), slight at time of obs	 Continuous drizzle (not freezing), slight at time of obs	 Intermittent drizzle (not freezing), moderate at time of obs	 Continuous drizzle (not freezing), moderate at time of obs	 Intermittent drizzle (not freezing), thick at time of obs	 Continuous drizzle (not freezing), thick at time of obs	 Slight freezing drizzle	 Moderate or thick freezing drizzle	 Drizzle and rain, slight	 Drizzle and rain, moderate or heavy
60	 Intermittent rain (not freezing), slight at time of obs	 Continuous rain (not freezing), slight at time of obs	 Intermittent rain (not freezing), moderate at time of obs	 Continuous rain (not freezing), moderate at time of obs	 Intermittent rain (not freezing), heavy at time of obs	 Continuous rain (not freezing), heavy at time of obs	 Slight freezing rain	 Moderate or heavy freezing rain	 Rain or drizzle and snow, slight	 Rain or drizzle and snow, moderate or heavy
70	 Intermittent fall of snowflakes, slight at time of obs	 Continuous fall of snowflakes, slight at time of obs	 Intermittent fall of snowflakes, moderate at time of obs	 Continuous fall of snowflakes, moderate at time of obs	 Intermittent fall of snowflakes, heavy at time of obs	 Continuous fall of snowflakes, heavy at time of obs	 Ice needles (with or without fog)	 Granular snow (with or without fog)	 Isolated starlike snow crystals (with or without fog)	 Ice pellets (select. U.S. definition)
80	 Slight rain shower(s)	 Moderate or heavy rain shower(s)	 Violent rain shower(s)	 Slight shower(s) of rain and snow mixed	 Moderate or heavy shower(s) of rain and snow mixed	 Slight snow shower(s)	 Moderate or heavy snow shower(s)	 Slight shower(s) of soft or small hail, with or without rain, and/or snow	 Moderate or heavy shower(s) of soft or small hail, with or without rain, and/or snow	 Slight shower(s) of hail, with or without rain and/or snow, not assoc with thunder
90	 Moderate or heavy shower(s) of hail and/or rain/snow, not associated with thunder	 Slight rain at time of obs; TS during past hour not at time of obs	 Moderate or heavy rain at time of obs; TS during past hour not at time of obs	 Slight snow and/or rain/hail at time of obs; TS during past hour not at time of obs	 Moderate or heavy snow and/or rain/hail at time of obs; TS during past hour not at time of obs	 Slight or moderate thunderstorm without hail but with rain and/or snow at time of obs	 Slight or moderate thunderstorm with hail at time of obs	 Heavy thunderstorm without hail but with rain and/or snow at time of obs	 Thunderstorm combined with duststorm or sandstorm at time of obs	 Heavy thunderstorm with hail at time of obs

Table D2. Past weather classification (W.M.O. past weather code).

Past weather symbols - These symbols represent the most significant weather within the past three hours of the observation but not during the most recent hour.

Code	0	1	2	3	4	5	6	7	8	9
Symbol	N/A	N/A	N/A							
Description	Clear or Few Clouds (not plotted)	Partly cloudy (scattered) or variable sky (not plotted)	Cloudy (broken) or overcast (not plotted)	Sandstorm or dust storm, or drifting or blowing snow	Fog, or smoke, or thick dust haze	Drizzle	Rain	Snow, or rain and snow mixed, or ice pellets (sleet)	Shower(s)	Thunderstorm, with or without precipitation