Using the Stretched Exponential Distribution to Model Runs of Extremes in a Daily Meteorological Variable

Warwick Grace

Grace Research Network, Adelaide, Australia

(Manuscript received July 2014; accepted June 2015)

The stretched exponential distribution is used to create a statistical model for the frequency and duration of runs of extreme days. Extreme days are regarded here as those days when the meteorological variable concerned exceeds an upper threshold of any chosen ranking percentile or does not reach a lower threshold.

This stretched exponential model represents a generalisation of a simpler modelling framework applied previously to Australian and international data. The model requires only two parameters and these parameters are location-specific and independent of the percentile of the threshold. Using the records of daily maximum and minimum temperatures in Australia and Europe, agreement between model and observations is shown to be good to very good.

Introduction

Extremes and runs of extremes of many meteorological variables have an impact on society and the environment and their frequency of occurrence is of interest. An extreme day is here defined as one where the variable concerned exceeds some chosen upper threshold or remains below a chosen lower threshold; and a run of extreme days is a sequence of consecutive extreme days. The duration of a run is defined here as the maximum number of extreme days in the run.

There appears to be no simple analytical or parametric model of frequency of runs of extremes of a daily variable in the literature other than the Markov model of Grace et al. (2009, hereafter GSH) which has an analytical expression as a geometric or exponential distribution.

Stochastic modelling of runs of extreme days has been related mostly to heatwaves and wet or dry spells. Stochastic time series modelling using first-order autoregressive (AR) models with a knowledge of the monthly mean, standard deviation and autocorrelation provide characteristics of heatwaves that are in good agreement with observations in mid-latitude areas (Mearns et al. 1984, Kysely 2010, Grace 2011). Grace (2013) presented an idealised stochastic autoregressive model requiring only two parameters, which provides fair to good representation of runs of extremes in the form of heatwaves, cold spells, and runs of days with high evaporation, high wind run, low pressure, and low sunshine hours. A disadvantage of the stochastic models is that they have no analytical expression.

The stretched exponential function has been used in physics to model relaxation phenomena such as capacitor discharge, and decay in luminescence and fluorescence (Sornette 2006, Laherrere and Sornette 1998) and in biochemistry to model catalytic activity of enzymes (Flomenbom et al. 2005). The stretched exponential function has been used as (a) a probability density distribution and (b) a complementary cumulative distribution to model a diverse range of phenomena in nature and the economy. For the former, Luevano (2013) cites over 20 papers: for the latter, Laherrere and Sornette modelled distributions of radio and light emissions from galaxies, oilfield sizes, city sizes, currency variations, biological extinction events, earthquakes and journal citations, and showed that the stretched exponential (complementary cumulative) distribu-

Corresponding author address: Warwick Grace, Grace Research Network, 29 Yurilla Drive, Bellevue Heights SA 5050.
Email: wg@graceresearch.com

tion is superior to commonly used power-law distribution models for their examples. In this paper the stretched exponential distribution (SED) is used in the probability density form and will be formally defined later.

The purpose of this paper is to show that the SED model provides a good approximation to the frequency of observed runs of extremes of daily maximum and minimum temperatures. It has a relatively simple two-parameter analytical function.

The testing datasets are described; then the SED model described and developed. Model performance is assessed qualitatively and quantitatively, with discussion and conclusions following.

Data considerations

The data used are the daily maximum and minimum temperature record since 1910 from the Australian Bureau of Meteorology's Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) High Quality Temperature dataset (Bureau of Meteorology 2012, Trewin 2013) and the quality-controlled daily maximum and minimum temperature dataset for sites in Europe, including Russia, and the Mediterranean available from the European Climate Assessment Dataset (Klein Tank et al. 2002 and Klok and Klein Tank 2009). The data were allocated to either calendar (i.e., January to December) or 'Austral summer' years (from July to June) so that the peak period for the extremes is mid-year: this avoids any problems due to incomplete runs occurring at the starts or ends of the records used. A complete, or near-complete, year of record at a station is regarded as one with no more than two missing observations. For the years with one or two missing observations, the missing daily observations were substituted with linearly interpolated values. For example, if the observation for a day is missing, then the missing observation is replaced by the linear interpolation based on the adjacent days. Only complete or near-complete years were used.

For simplicity the confounding effect of global warming was avoided by using only years up to 1970 in the temperature datasets. The period of up to 1970 was chosen as being approximately stationary since it is known that most of the warming trend in maximum temperatures in Australia in the recent past has occurred since 1970 (Commonwealth of Australia, 2012). For the Australian dataset only sites with at least 25 years of complete or near-complete data up to 1970 were used. For the European dataset, only sites with at least 70 years of complete or near-complete data up to 1970 were used. The numbers of sites are shown at Table 1.

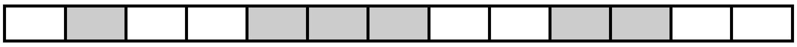
Temperature-percentile relationships were constructed for each site for the whole period of record (up to 1970) and calculated over that entire period regardless of the annual cycle. Thus it is always straightforward to convert runs in terms of percentile thresholds to runs in terms of absolute temperatures. For example, the 90th percentile maximum temperature for Melbourne for the period up to 1970 was 29.4°C, while 30°C corresponds to the 91.2th percentile.

Model description and development

Assumptions and Definitions

An extreme day is one where a daily meteorological variable, T, exceeds (or remains below) a specified upper (or lower) threshold, T_p. The threshold T_p is the pth percentile value calculated over entire record regardless of the annual cycle. Extreme days and non-extreme days are mutually exclusive, and a run is a sequence of extreme days bounded by non-extreme days (see Figure 1 for a schematic illustration of the idea). If T happens to equal T_p the day is taken as non-extreme.

Figure 1 Schematic of a sequence of days with three runs of extreme days. Extreme days are represented as grey. From the left, the three runs are of length 1, 3 and 2 days.



For thresholds at the p^{th} percentile, there is a corresponding expected fraction f of extreme days. For upper and lower thresholds respectively, then

$$f = 1 - 0.01p \tag{1}$$

and

$$f = 0.01p. \tag{2}$$

Earlier models

Grace (2013) showed that for the simplest case, that of a series of a random variable that is stationary and not serially correlated, the runs are anticipated to follow an exponential form, so that

$$R(k,f) = 365\,(1-f)^2 \exp\left[\ln(f)\,k\right] \tag{3}$$

where R is the expected number of runs per annum, and k = 1,2,3,… is the number of days in the run, and f is the fraction of extreme days. The coefficient 365 is a rounding of 365¼ days per annum. The quantity ln(f) is necessarily negative, indicating an exponential decline in the number of runs compared to their duration (from Equation A3 in Grace, 2013).

GSH presented a Markov model for runs of hot days predicated on the concept of regime switching (between continental and maritime air-mass influences). Their expression for the number of runs is

$$R(k,f) = 365 f^{1-M}(1-f^M)^2 \exp\left[M\ln(f)\,k\right] \tag{4}$$

where M is a location-specific coefficient or parameter and k and f are regarded as independent variables (from Equation 12a in GSH). If M =1 then Equation 4 reduces to Equation 3. M is well correlated with the skewness of the January daily maximum temperature (r = 0.91) and M ~0.5 near the coast and ~0.25 inland: the letter M being chosen to indicate the relevance of the maritime influence. This Markov model, referred to as the M model by GSH, performed best for sites on the southern coasts such as Robe and Hobart, and (unpublished) Eucla and Cape Leeuwin. The exponential form of Equation 4 implies that runs against duration tend to plot as straight lines on log-linear axes. Further assessment of this model once the ACORN-SAT dataset became available showed that long runs occur more frequently than the straight-line extrapolation (on the log-linear plot). In other words, in a graphical context a convex curvature was always evident in the plotted data, albeit sometimes only slightly for coastal locations and small f. Examples of the convex curvature are evident in Figures presented later.

Stretched exponential distribution (SED)

The general normalized form of the stretched exponential function is

$$y(t) = c\exp\left[-(\beta\,t)^{\alpha}\right] \tag{5}$$

where t is the independent variable and α and β are parameters with α viewed as a shape parameter and the recipocal of β as a scale parameter (Flomenbom et al 2005). Special cases are α = 2 yielding the probability density function for a normal distribution, and α = 1 the probability density function for an exponential distribution. If t is continuous from t = 0, then

$$c = \frac{\beta}{\Gamma(1+1/\alpha)} \tag{5a}$$

provides the normalisation constraint for Equation 5 to be a probability density function for a random variable constrained to positive values.

Using Equation 5, Equations 3 and 4 are generalised to

$$R(k,f) = 365\;c\exp\left[-(-\beta\ln(f)\,k)^{\alpha}\right] \tag{6}$$

where the additional minus sign is introduced so that $\beta > 0$. The selection of Equation 6 is arbitrary: the only constraints are parsimony and simplicity and that Equations 3 and 4 are recoverable as special cases ($\alpha = 1$ and $\beta = 1$ for Equation 3; and $\alpha = 1$ and $\beta = M$ for Equation 4), as is shown later.

Equation 6 is a two-parameter expression in that the coefficient c is not a free parameter and, as shown below, is itself a function of α, β and f. Values of the duration variable k are positive integers and therefore Equation 5a does not apply. However, the expected number of extreme days per annum is always $365f$ and therefore the coefficient c is constrained by the equation

$$365 f = \underset{k=1}{\overset{\square}{\square}} k\,R(k,f), \tag{7}$$

which leads to

$$c(\alpha,\beta,f) = f\left\{\sum_{k=1}^{\infty} k \exp\left[-\left(-\beta \ln(f)\,k\right)^{\alpha}\right]\right\}^{-1} . \tag{8}$$

Generally this equation has to be approximated numerically, but closed forms exist for some special cases. For the special case $\alpha = 1$ and $\beta = 1$ then

$$c(1,1,f) = f\left\{\sum_{k=1}^{\infty} k \exp\left[\ln(f)\,k\right]\right\}^{-1} \tag{9}$$

which can be recast as

$$c(1,1,f) = f\left\{\sum_{k=1}^{\infty} k\,f^{k}\right\}^{-1} . \tag{10}$$

Using the summation identity (Jolley 1961)

$$\underset{k=1}{\overset{\square}{\square}} k\,f^{k} = \frac{f}{(1-f)^{2}}, \qquad \text{if} \quad f < 1, \tag{11}$$

we have

$$c(1,1,f) = (1-f)^{2}, \tag{12}$$

which is consistent with Equation 3. A similar process shows that for the special case $\alpha = 1$ and $\beta = M$ then

$$c(1,M,f) = f^{1-M}(1-f^{M})^{2} \tag{13}$$

which is consistent with Equation 4. Thus, the Stretched Exponential Distribution model (SED model) for the probability density distribution of runs of extreme days is represented by Equation 6 and the subsidiary Equation 8.

Model fitting – finding *α* and *β*

Assume that observed values of $R(k,f)$ are available. From an initial guess of $\alpha = 1$ and $\beta = 1$, the corresponding model values of $R(k,f)$ are calculated. The general procedure is then to determine values of α and β which minimize the difference between the observed and modelled $R(k,f)$. The particular method used here is a numerical method to minimize the extended likelihood ratio chi-square statistic, χ^{2} (defined below). It became apparent that sometimes a local minimum was found rather than a global minimum and to avoid this, multiple (~20) randomised starting estimates for α and β were used and the final α and β selected from the best of the 20 fits.

A more direct method is to recast Equation 6 as

$$\ln\left[\ln(365c)-\ln(R)\right]-\alpha\ln\left[-\ln(f)\right]=\alpha\ln(k)+\alpha\ln(\beta), \tag{14}$$

and then utilise the fact that Equation 14 is in the slope-intercept form of a linear equation ($y = mx+b$) where y corresponds to the left hand side of Equation 14, x corresponds to ln(k), the slope m corresponds to α and the intercept b corresponds to α ln(β). Starting with initial values of α and β, Equation 14 is solved to give updated values α and β and the process reiterated to an acceptable level of convergence. This ordinary least squares fitting of a double logarithm quantity will generally be inferior to the minimization method above. A similar, and simpler, method is also applicable to the M model.

Cowan (1998, see Ch 6) provides a detailed description of the extended likelihood ratio chi-square statistic, χ^2 for Extended Maximum Likelihood Estimation (EMLE) defined as

$$\chi^2=2\sum_{k=1}^{K}\left[\left(E_k-O_k\right)+O_k\ln\left(\frac{O_k}{E_k}\right)\right], \tag{15}$$

where O_k and E_k refer to the observed counts and model expected counts respectively in the kth bin of K bins. For models normalized on the total number of O then the total E must equal total O, and the extended likelihood ratio chi-square statistic reduces to the usual Maximum Likelihood Estimation (MLE) likelihood ratio chi-square statistic given by

$$\chi^2=2\sum_{k=1}^{K}\left[O_k\ln\left(\frac{O_k}{E_k}\right)\right]. \tag{16}$$

However in our case the normalizing is by the total number of extreme days which by definition is invariant at 365f per annum. In other words, the model necessarily produces the same total number of extreme days as observed, but not necessarily the same number of runs. There are two ramifications. Firstly, the EMLE statistic is appropriate. Secondly, the number of degrees of freedom is equal to the number of non-zero bins minus the number of parameters estimated in fitting the model, which here is two (being one each for α and β). In comparison, for the MLE statistic there is further reduction of the degrees of freedom by one for each row.

Cowan (1998) notes that the binned MLE or EMLE techniques above do “not encounter any difficulties if some of the bins have few or no entries.” In the expression for the extended χ^2 statistic, a row of K bins corresponds to a specified value of f. We can increase the number of rows (designated r) to three to accommodate f = [0.01, 0.05, and 0.1], thus

$$\chi^2=2\sum_{r=1}^{3}\sum_{k=1}^{K}\left[\left(E_{r,k}-O_{r,k}\right)+O_{r,k}\ln\left(\frac{O_{r,k}}{E_{r,k}}\right)\right]. \tag{17}$$

The use of the extended χ^2 statistic at Equation 17 provided fast and reliable convergence.

Model Performance

Qualitative assessment

Some graphical examples of runs of extreme days provide qualitative assessment of the model: heatwaves can in some respects be regarded as a run of days with extremely high maximum temperatures, and cold spells as a run of days with extremely low minimum temperatures. In Figures 2 to 5, observed runs rates (shown as per century rather than per annum for convenience) are shown as dots; the full line curve is the SED model with the *α,β* parameters tuned so as to provide the best fit. 95% confidence intervals for the observed number of runs are shown by vertical bars: if no runs of a duration have been recorded then no confidence interval is plotted. It is reasonable to expect that the runs of any given duration occur randomly among the years of record (although not within the years). For rare events (corresponding to runs of longer duration and/or small *f*) randomly distributed across time it is appropriate to calculate confidence intervals for runs per annum using the Poisson distribution. For larger numbers, one would usually use the normal distribution but the Poisson distribution still applies.

The two Australian stations of Figures 2 and 3 are chosen because they are geographically and climatically different. The two European stations of Figures 4 and 5 are selected on the grounds that their data records are the longest in the database (~200 years). These plots and others in the European dataset covering over 100 years of record show that the observations agree reasonably with the model in the tail area out to runs lengths of 15 or more days. The SED models shown at these Figures are qualitatively similar in the quality of agreement for all other stations in the datasets. This comment applies to other percentile thresholds above 90% and below 10%, not just 90, 95 and 99% and 1, 5 and 10%.

Quantitative assessment

Quantitative assessment of the SED model's accuracy and reliability is performed for each of the temperature datasets for heatwaves and cold spells, for $f = [0.01, 0.05, \text{ and } 0.1]$. Performance measures used were the χ^2 test and normed chi-square, *C*. *C* is defined as

$$C = \frac{\chi^2}{\nu} \tag{18}$$

where ν is the number of degrees of freedom, with $\chi 2$ and ν being calculated using Equation 17. The χ^2 test is whether to reject or accept (strictly, to reject or not reject) the null hypothesis that the observed and model distributions of runs frequencies are equivalent at a 5% level of significance. There is good agreement between model and observations if the null hypothesis is not rejected. To apply this test, both the $\chi 2$ statistic and ν are required. The *C* value serves as a useful single number guide to goodness-of-fit and this is illustrated by the contour plot of the acceptance level of significance with respect to *C* and ν at Figure 6. For the region below the 5% contour, there is good agreement between model and observations. For example, if $C< 1$ the χ^2 test is passed regardless of ν. With $\nu \sim 20$ to 50, then the χ^2 test is passed if *C* is less than about 1.3.

Of 264 χ^2 tests of the SED model (see Table 1 and Figure 7 for details), the median *C* was 1.05 and 83% passed the χ^2 test. For the separate classes of heatwaves and cold spells for Australian and European stations, the results were similar or better other than that the model performed worse for the class of cold spells at European stations. From these results of the measures of *C* and the χ^2 tests, it is concluded that the SED model is typically good to very good in its estimation of the frequency of runs and duration of runs.

The performances of the stochastic model of Grace (2013) and the *M* model of GSH were re-assessed for comparison using the same dataset and methodology as for the SED model (Table 1). It is clear that the SED model greatly out-performs the other models: compared to an overall 83% acceptance for the χ^2 tests for the SED model, the stochastic model and *M* model rated 29 and 19% respectively. The stochastic model is superior to the *M* model although not uniformly so with the *M* model having marginally better acceptance rate (36% compared to 34%) for heatwaves at Australian stations.

Figure 2 Runs frequency versus runs duration for daily maximum temperature (*Tx*) exceeding the 90th (top), 95th (middle) and 99th (bottom) for Melbourne (left) and Alice Springs (right). The observed numbers of runs are shown as dots, and their associated 95% confidence intervals as vertical lines. The fitted SED model is shown as a thick line, with the fitted parameters α and β being given on each plot: note that these are always independent of the threshold percentile p. For Alice Springs there is a significant number of years discarded due to insufficient data – see earlier comments regarding data considerations.

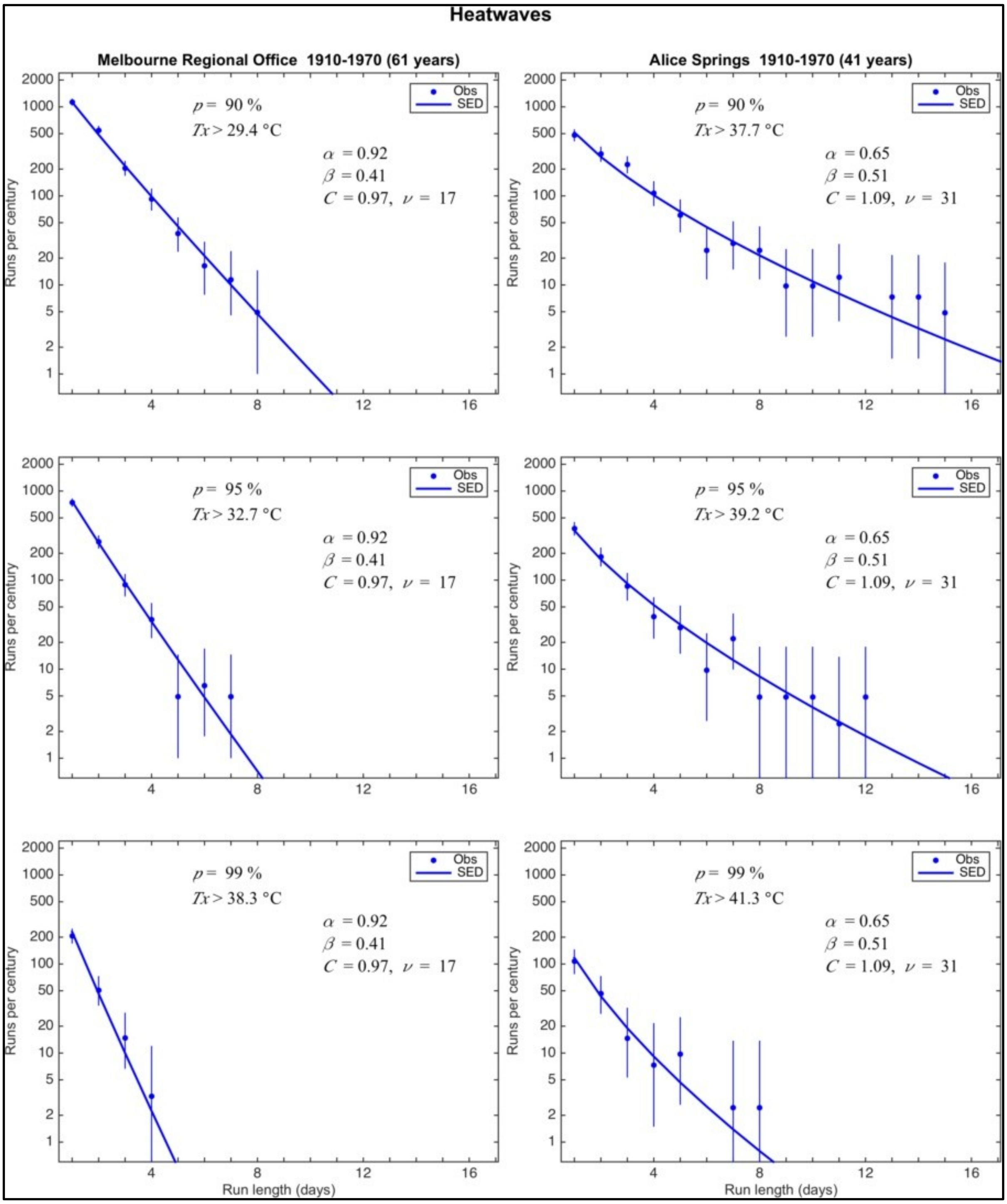


Figure 3 As per Figure 2, but for the 10th (top), 5th (middle) and 1st (Bottom) percentiles of minimum daily temperature (*Tn*) at Melbourne (left) and Alice Springs (right). For Alice Springs, 14 years were discarded due to insufficient data.

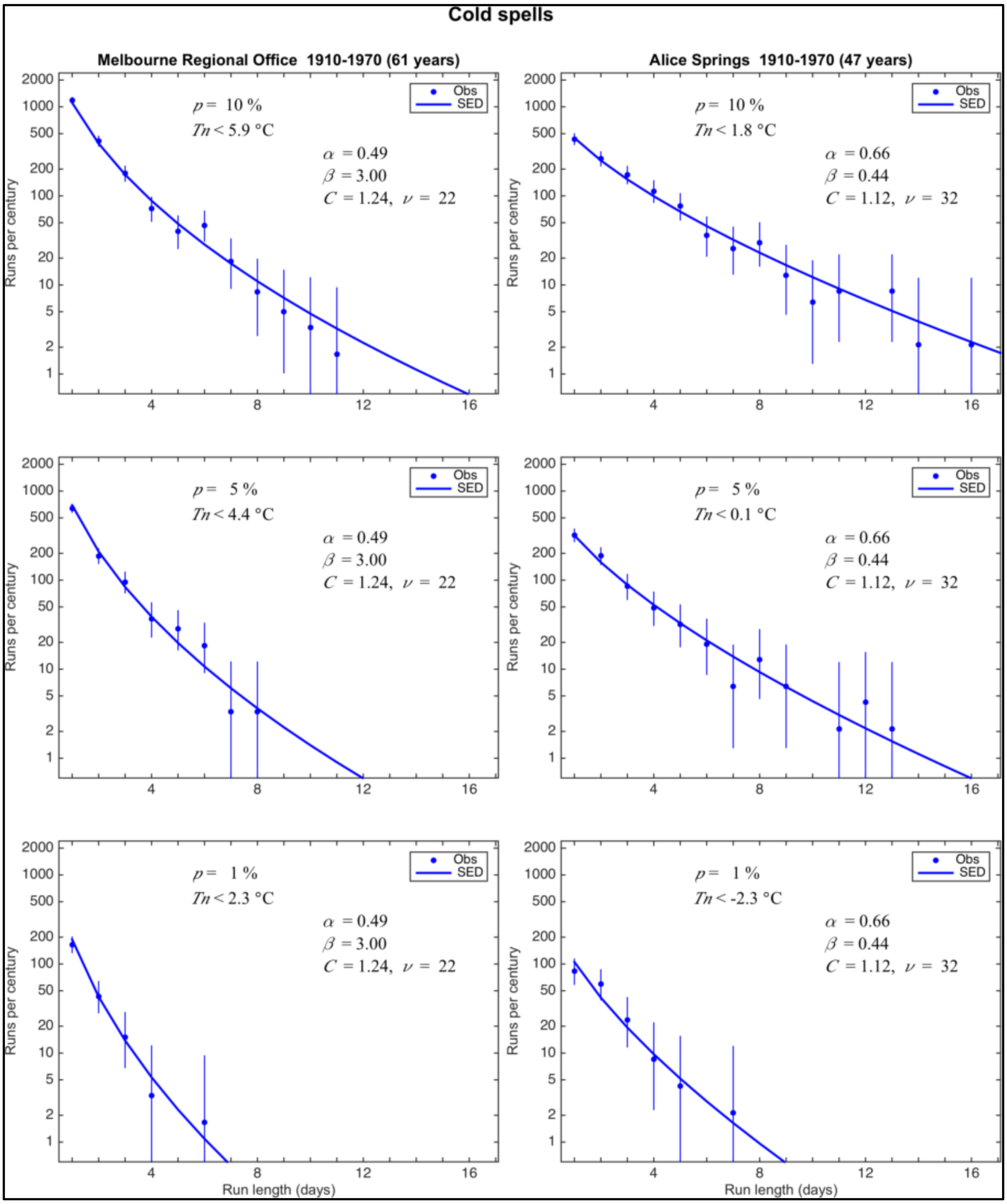


Figure 4 As for Figure 2, but for Milan (left) and Prague (right).

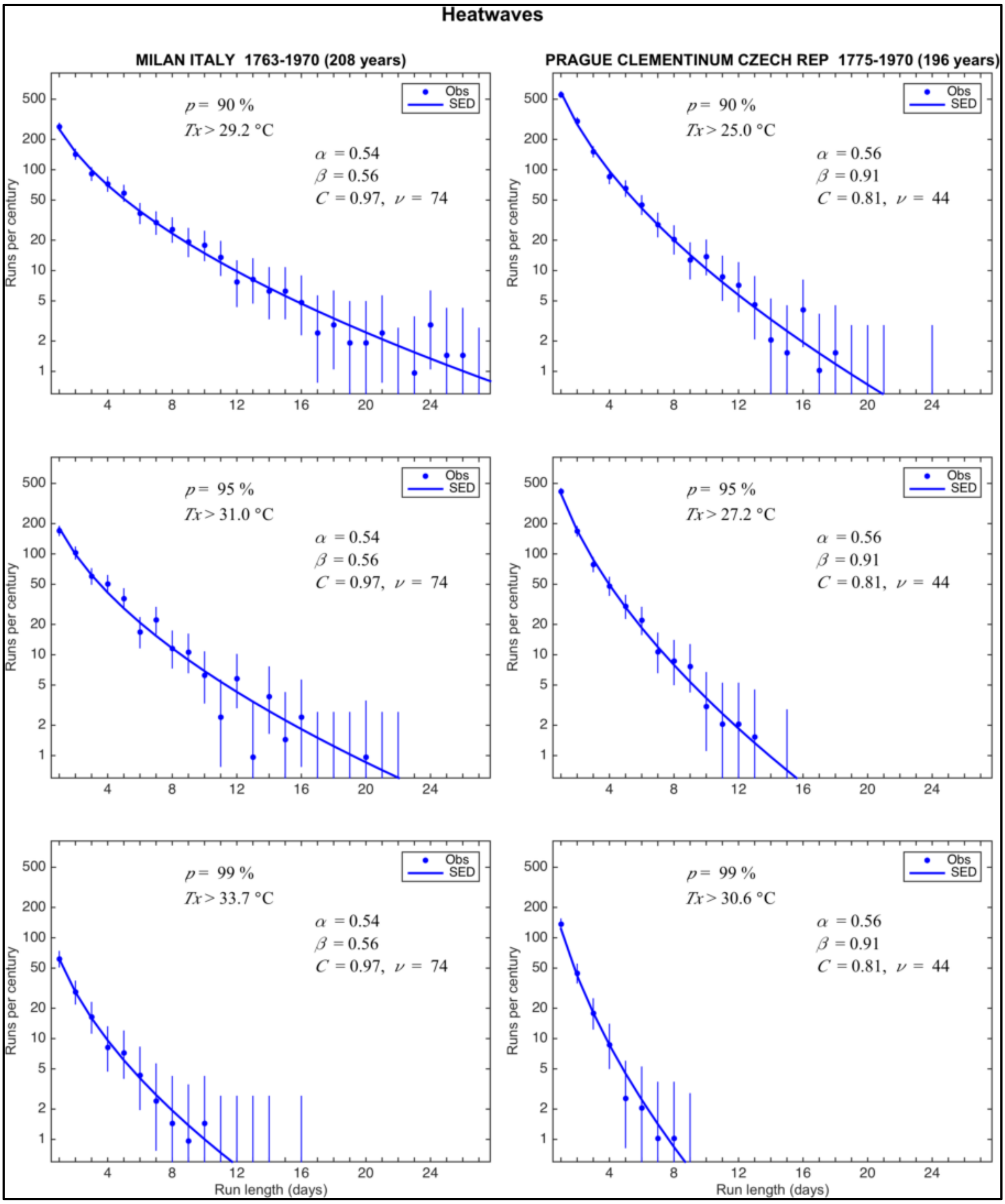


Figure 5 As for Figure 3, but for Milan (left) and Prague (right).

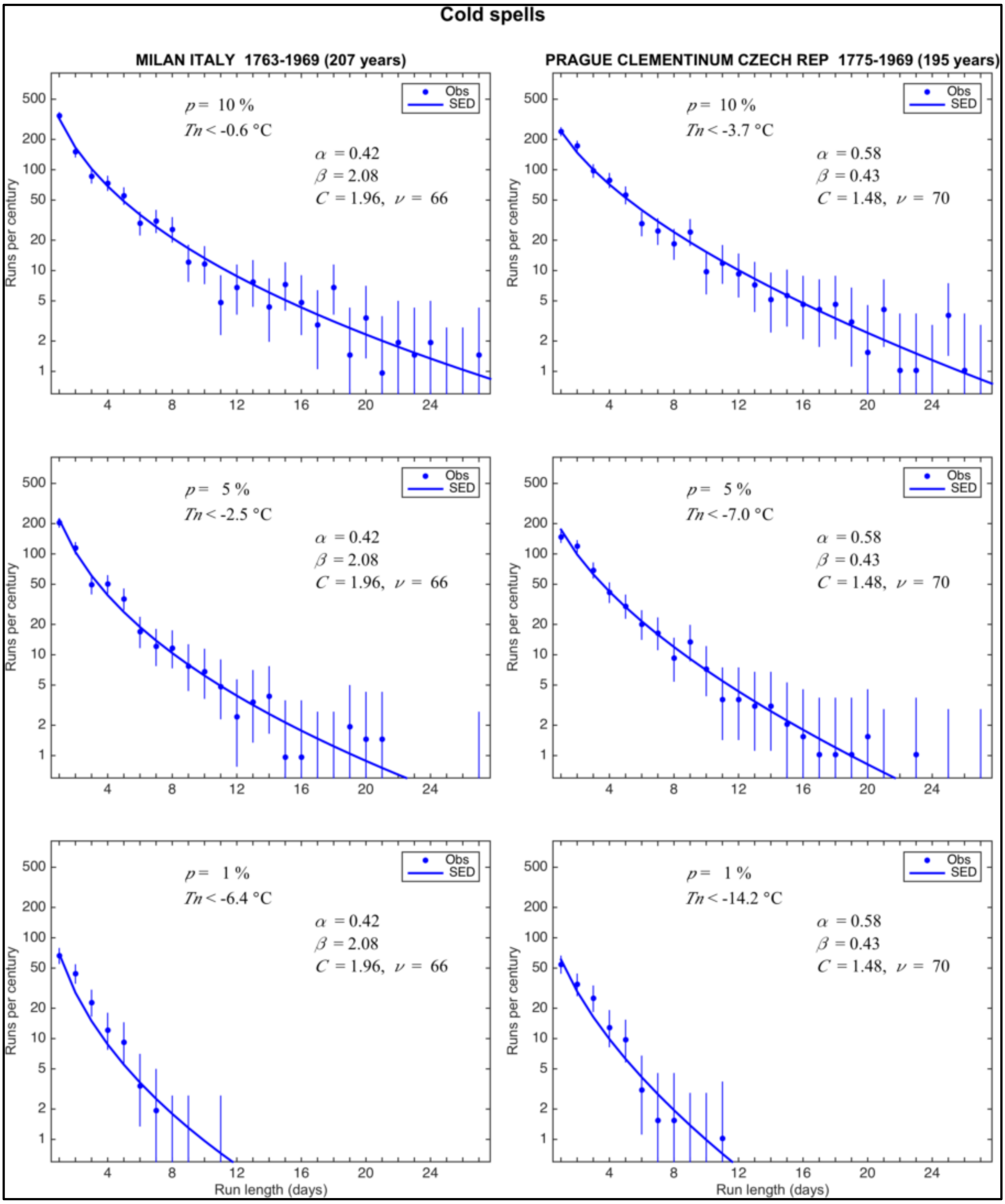
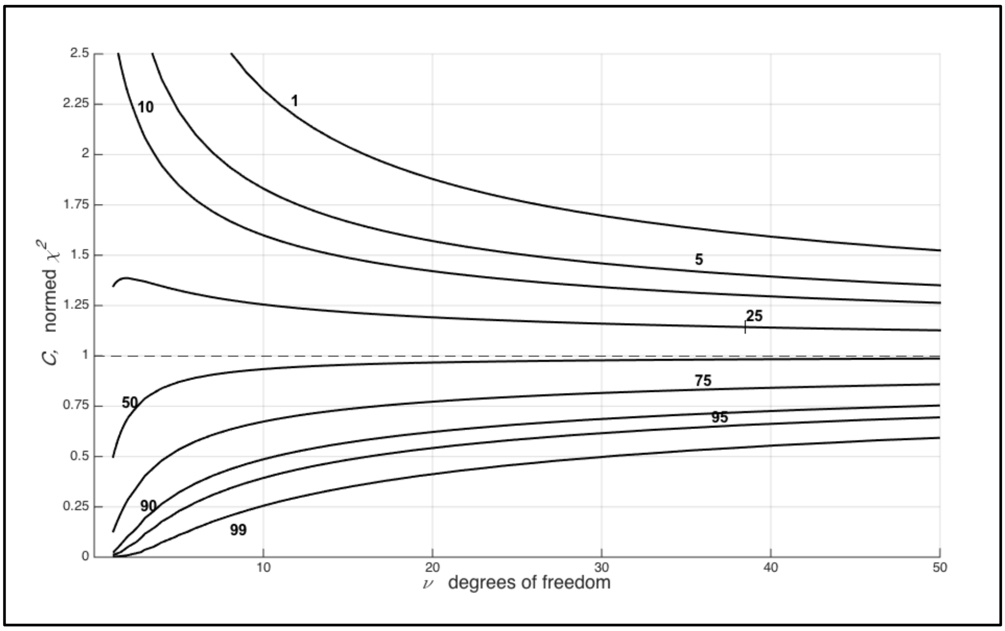


Table 1 Performance measures of *F* and median *C* for each dataset for runs of days with maximum temperatures above the 90th, 95th and 99th percentiles and minimum temperatures below the 10th, 5th and 1st percentiles for Australian and European datasets for years up to 1970. Australian (European) datasets were restricted to sites with at least 25 (70) years of continuous record. *C* is normed χ^2, *F* is percentage of sites for which null hypothesis is accepted, under the χ^2 test at 0.05 significance level. "Stoch" refers to the stochastic model of Grace (2013); "*M*" refers to the *M* model of GSH. On all measures SED model is superior to the other models for all datasets examined.

Dataset	*Number of Sites*	*Performance Measure*					
		F %			*Median C*		
Models →		SED	Stoch	M	SED	Stoch	M
Heatwaves							
- Australian	67	85	34	36	1.10	1.65	1.72
- European	65	89	28	3	0.99	1.56	3.38
Cold spells							
Australian	70	95	46	36	0.96	1.57	1.67
European	62	63	12	1	1.23	1.91	3.53
Combined or grand median	264	83	29	19	1.05	1.72	2.49

Figure 6 Normed χ^2, *C*, and *v*, degrees of freedom, with contours of the corresponding levels of significance (expressed as percentage). Generally for a level of significance of 5%, the region below the 5% contour indicates acceptance of the null hypothesis that the model and observations have the same distribution.



Discussion and Conclusions

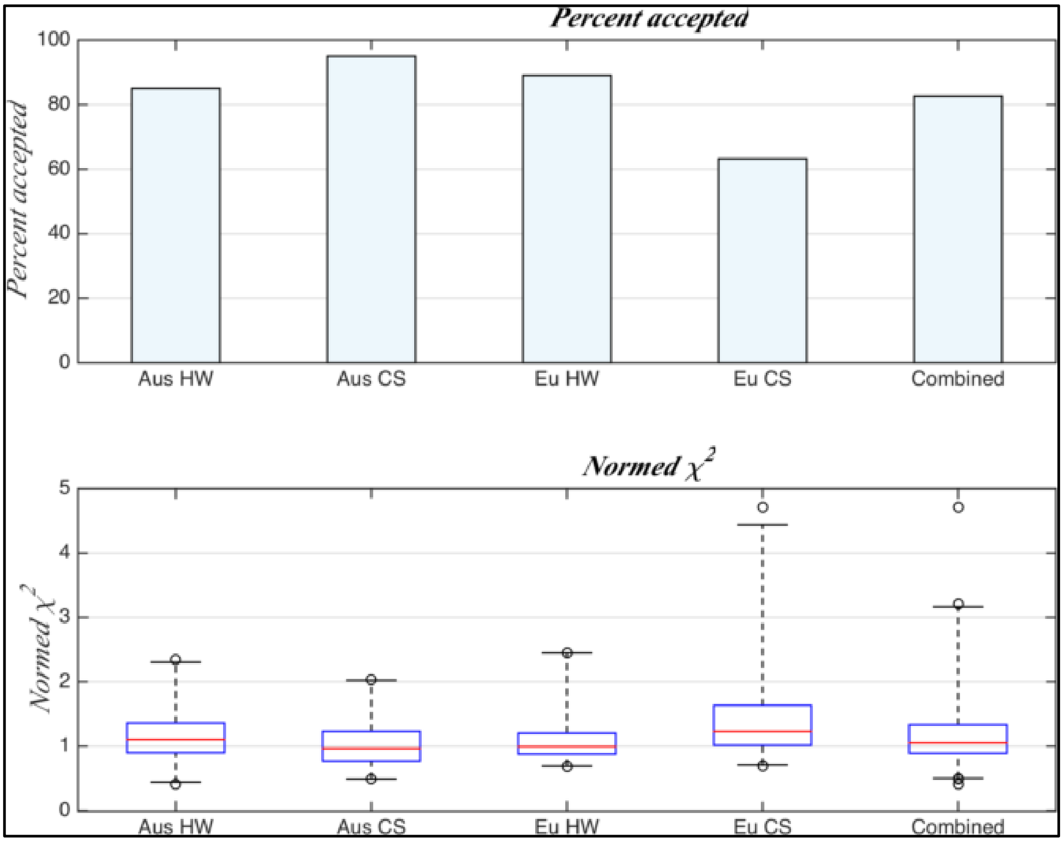
The theoretical grounds for choosing the stretched exponential distribution (SED) model are limited. Firstly, it was formed as an arbitrary generalisation of an exponential expression applicable to runs of independent random variables (Equation 3, in particular). On the other hand, a wide range of other natural phenomena is well described using the stretched exponential distribution. However, the empirical support as presented here is strong although not comprehensive.

The advantages of the SED model are that it has a straightforward analytical expression (Equation 6); it is potentially applicable to a wide range of meteorological variables; it has only two free parameters, which are location-specific; and in arriving at those parameters information from all of the thresholds at percentiles of 90, 91, …, 99% may be used, not just from any single threshold. Therefore, with some caution the model may be used in data sparse areas of very high or very low thresholds (low *f*) and/or long duration (large *k*).

The disadvantages are that the parameters of the model have no clear physical meaning other than being scaling or shape parameters; and that determination of the parameters requires the counting and compilation of observed runs and durations from records with no gaps, or at least, very few.

For a daily meteorological variable, an empirical two-parameter stretched exponential function model of the frequency of runs of extremes in relation to duration (in days) and intensity (as measured by ranking percentiles of the variable) was presented and shown to provide a good to very good approximation to the frequency of observed runs of extremes of daily maximum and minimum temperatures in Australia and internationally. It is possible that the SED model could be applied to a variable exhibiting a long term trend by transforming the variable to a departure from trend.

Figure 7 Bar plots of *F* (top) and *C* (bottom) for heatwaves (HW) and cold spells (CS) for Australian (Aus) and European (Eu) stations, and combined. Median shown by red line, quartiles by box-ends, and whiskers are 1st and 99th percentiles. For *C*, the normed χ^2, values below about 1.5 indicate an acceptable model.



Acknowledgements

Reviewer and editorial suggestions and earlier comments from Bruce Brooks, formerly of Bureau of Meteorology, lead to greater rigour and readability. This paper was developed from previous research funded by the Grape and Wine Research and Development Corporation.

References

Bureau of Meteorology, 2012. Data available at http://www.bom.gov.au/climate/change/acorn-sat/. Accessed 1 May 2012.

Commonwealth of Australia, 2012. State of the Climate 2012. A joint publication of Bureau of Meteorology and CSIRO. 12 pp.

Cowan, G., 1998. Statistical Data Analysis. Oxford University Press. Pp 200.

Flomenbom, O., K. Velonia, D. Loos, S. Masuo, M. Cotlet, Y. Engelborghs, J. Hofkens, A.E. Rowan, R.J.M. Nolte, M. Van der Auweraer, F.C. De Schryver and J. Klafter. 2005. Stretched exponential decay and correlations in the catalytic activity of fluctuating lipase molecules. Proc. Natl. Acad. Sci. USA , 7, 2368-2372. doi: 10.1073/pnas.0409039102.

Grace, W.J., Sadras, V.O. and Hayman, P.T., 2009. Modelling heatwaves on viticultural regions of southeastern Australia. Aust. Met. Oceanogr. J., 58, 249-262.

Grace, W.J., 2011. Modelling heatwaves: Connecting an Empirical Markov Model with an Autoregressive Model. Aust. Met. Oceanogr. J., 61, 43-52.

Grace, W.J., 2013. A stochastic model for runs of extreme days for a daily meteorological variable. Aust. Met. Oceanogr. J., 63, 473-486.

Jolley, L.B.W., 1961. Summation of Series. Dover Publications. 2nd Ed. pp 251.

Klein Tank, A.M.G., J.B. Wijngaard, G.P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, and P. Petrovic, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. Climatol.. 22, 1441-1453. Data and metadata available at http://eca.knmi.nl. Accessed 30 Dec 2011.

Klok, E.J, and Klein Tank, A.M.G. 2009. Updated and extended European dataset of daily climate observations. Int. J. Climatol.. 29, 1182-1191.

Laherrere, J., and Sornette, D., 1998. Stretched Exponential Distributions in Nature and Economy: 'Fat Tails' with characteristic scales. European Physical Journal B, 2, 525-539.

Luevano, J.R., 2013. Statistical features of the Stretched Exponential Densities. Journal of Physics: Conference Series, 475. 1-8. doi:10.1088/1742-6596/475/1/012008.

Sornette, D., 2006. Critical Phenomena in Natural Sciences. Springer 2nd Ed. 528 pp.

Trewin, B., 2013. A daily homogenized temperature data set for Australia. Int. J. Climatol., 33, 1510–1529. doi: 10.1002/joc.3530.