A statistical distribution for modelling rainfall with promising applications in crop science

Sarah Lennox[1], Peter K. Dunn[2], Brendan Power[3] and Peter DeVoil[4]

[1] Department of Primary Industries and Fisheries, Agency for Food and Fibre Sciences, PO Box 102, Toowoomba Qld, 4350
www.dpi.qld.gov.au Email sarah.lennox@dpi.qld.gov.au
[2] University of Southern Queensland, Department of Mathematics and Computing, Faculty of Sciences, West Street,
Toowoomba Qld, 4350. www.usq.edu.au Email dunn@usq.edu.au
[3] Department of Primary Industries and Fisheries, Agency for Food and Fibre Sciences, PO Box 102, Toowoomba Qld, 4350.
www.dpi.qld.gov.au Email Brendan.Power@dpi.qld.gov.au
[4] Department of Primary Industries and Fisheries, Agency for Food and Fibre Sciences, PO Box 102, Toowoomba Qld, 4350.
www.dpi.qld.gov.au Email peter.devoil@dpi.qld.gov.au

Abstract

The finding of an accurate and reliable rainfall model has been the point of much discussion in previous research and has promising input applications in areas such as crop growth, hydrological systems and simulation studies. In the past it has been necessary to model rainfall as two separate processes: rainfall occurrence (whether the period is dry/wet) and rainfall amounts (rainfall amount observed during a wet period). As the rainfall process involves both discrete (rainfall = 0 mm) and continuous parts (rainfall >0 mm), two separate models have previously been fitted and the information from the two models combined in order to provide a summary of the rainfall model. The Tweedie distribution however is able to combine both aspects to provide one complete rainfall process. This results in a more accurate, reliable and practical model that can then be incorporated into other areas such as crop growth systems.

Media summary

The Tweedie distribution has the potential to allow improved rainfall models to be developed. These improvements could then lead to better crop growth models and simulation studies.

Key Words

Tweedie distribution, Generalized Linear Model (GLM), precipitation, crop model simulation;

Introduction

The modelling of daily rainfall amounts has been the point of much discussion in past research. There are many difficulties that are encountered when modelling rainfall amounts and many suggestions as to how they can be overcome. One of the more prominent difficulties involves the continuity of the data. When a particular day is dry, then a rainfall amount of exactly zero is observed. If it rains then an observed amount greater than zero is recorded. This implies that the data is both discrete (equal to 0) and continuous (greater than 0). For this reason many authors have considered the rainfall process as two separate circumstances. One circumstance includes a model for whether a particular period of time will be dry or wet. The second circumstance examines the amount of rain when the day is wet. There are many methods that have been employed in order to find suitable models for both of these circumstances, many of which use a generalized linear model (GLM) framework and distributions such as the gamma (rainfall greater than 0), Poisson (number of wet days) and binomial distributions (probability of a period of time being wet/dry). The Tweedie distribution is named after M.C.K. Tweedie (Tweedie, 1984) who first introduced the Tweedie distribution in statistics. Recently it has been found that the Tweedie distributions enable the rainfall process to be modelled as a single simplified model and hence has promising applications in numerous fields including crop science.

Once the rainfall process is adequately and appropriately modelled the model can then be used in agricultural planning, may be able to aid in drought and flood predictions, used for impacts of climate change studies, rainfall runoff modelling, crop growth studies and the like. An accurate, reliable and relatively simple rainfall model would in turn suggest possible improvements or advantages in many of the areas mentioned above.

This paper aims to show some of the advantages obtained through the use of the Tweedie distribution. Comparisons are made of the outputs from a crop growth model in which the daily rainfall input is from either observed data or generated from a Tweedie distribution.

Methods

Tweedie distributions belong to a class of distributions called exponential dispersion models (EDMs). Other distributions in this family include the normal, gamma, Poisson and binomial distributions. Distributions in the EDM family form the basis for fitting generalized linear models (GLMs); thus, Tweedie GLMs are a possible tool for modelling.

The Tweedie family of distributions are a three-parameter family (the normal and gamma are two-parameter families, for example), and therefore permit modelling flexibility. The parameter p is called the index parameter and determines the shape of the Tweedie distribution. The normal, Poisson and gamma distributions are special cases of the Tweedie distribution. For various values of p, the following distributions exist:

- p=0: normal distribution;
- p=1: Poisson distribution;
- 1< p< 2: for continuous data with exact zeros;
- p=2: gamma distribution;
- p>2: for positive continuous data.

Since Tweedie distributions can include exact zero (dry days) and continuous data (rainfall quantity on wet days), this family of distributions is ideal for modelling rainfall. The distributions are also positively skewed, another feature of rainfall patterns. An example of two Tweedie densities that fit this description is displayed in Figure 1.

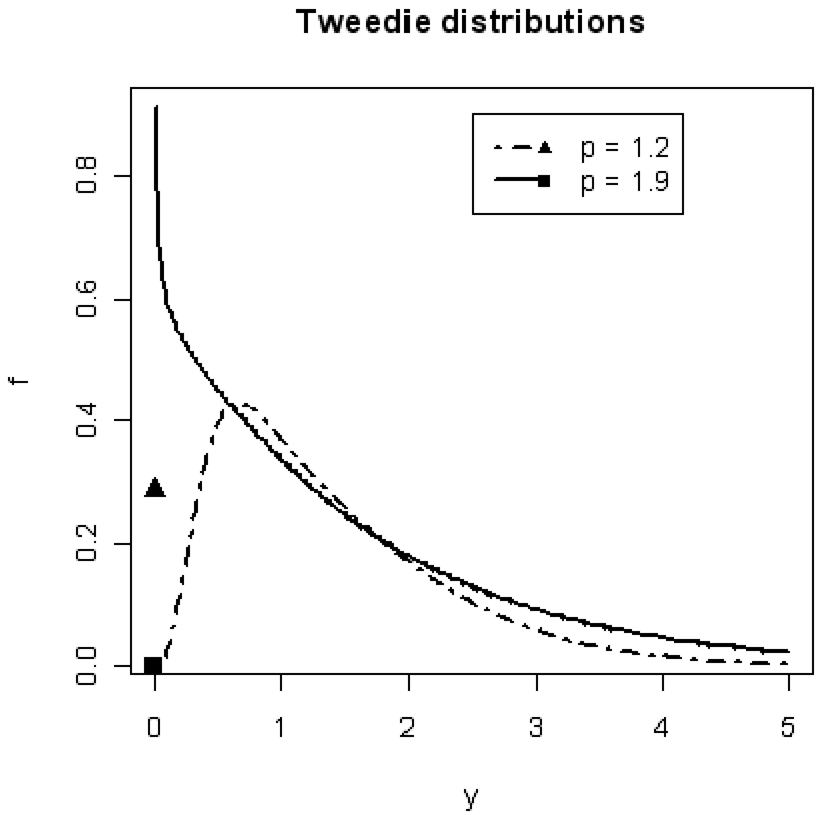

Figure 1. Two Tweedie densities with differing index parameters *p* that allow continuous data with exact zeros to be modelled.

Rainfall at Emerald has been analysed using the Tweedie distribution (Lennox, 2003). Analysis was conducted for Emerald daily rainfall data from the CLIMARC data set (Clarkson, 2002) for 1st January 1900 to 31st December 2002.

One possible application of these rainfall distributions includes the simulation of a number of years of rainfall data for use within a program such as APSIM (Agricultural Production Systems Simulator; Keating et al. 2003). Using the fitted GLM with a Tweedie distribution to simulate future daily rainfall data will allow future rainfall scenarios to be input into crop models to enable risk assessment and decision making tools.

To simulate the daily rainfall data, twelve GLMs (one for each month of the year) were fitted to 102 years of historical data (1900-2002). These models were then used to generate an equivalent 102-year, daily rainfall time series. Using such data as input into agricultural systems simulators such as APSIM will enable assessments of the impact of changing rainfall distributions for future scenarios (e.g. for climate change research).

Results

Emerald daily rainfall data shows a large proportion of dry days. After a GLM was fitted, the quantile residuals were used to test if the choice of the Tweedie distribution was appropriate as proposed by Dunn et al. 1996. If the distribution is appropriate, the normal probability plot of the quantile residuals will be normally distributed. On the normal probability plot in this paper, a thin solid line indicates normality. The normal probability plot of the quantile residuals in Figure 2 shows that the Tweedie distribution is very appropriate for modelling rainfall amounts greater than or equal to zero for the month of May. A GLM was fitted in this paper on a month-by-month basis, however, it is also possible to fit a single model through all of the daily data and this too has shown the Tweedie distribution to be very appropriate (Lennox, 2003).

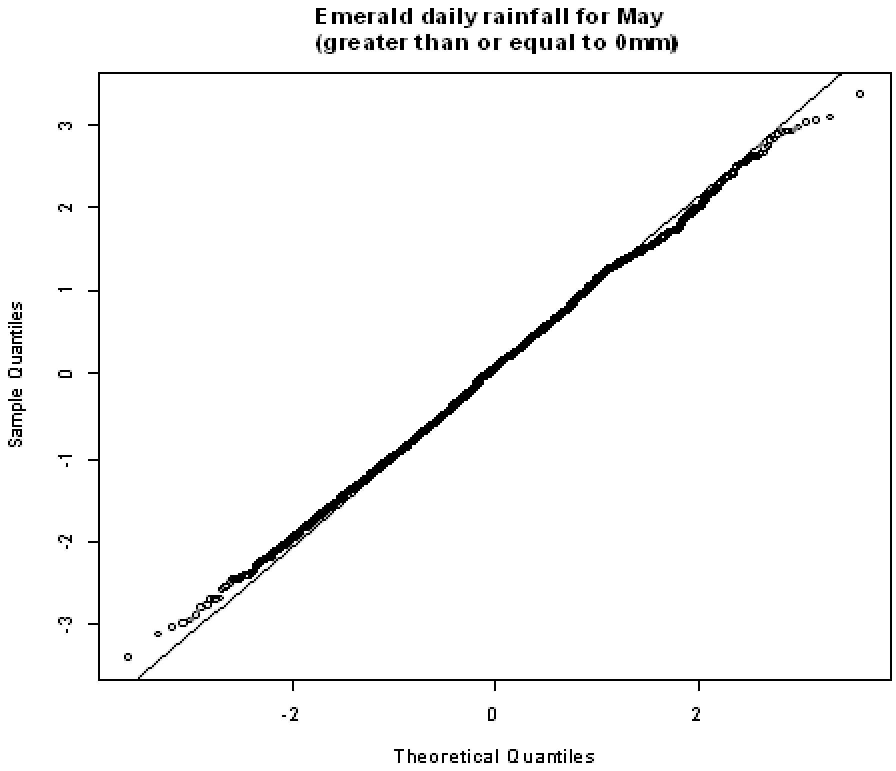

Figure 2: Normal probability plot for Daily Emerald data (1889-2002). Ideally the points should lie as close as possible to the thin solid line.

Figure 3 shows probability of exceeding yield (sorghum) plots derived from APSIM using either the observed historical daily rainfall records, or one hundred generated realisations based on the Tweedie distributions as input. Preliminary analysis for sorghum indicates that the yields derived from observed rainfall records are not significantly different from those obtained by generated (synthetic) rainfall for the same period. The maximum difference that can be obtained from the Emerald yield distribution (red line) in Figure 3(a) before the difference becomes statistically significant is displayed by the dotted lines and indicates that all our simulated yield outputs are not significantly different from the Emerald yield distribution: based on a Kolmolgorov-Smirnov(KS) test. Figure 3(b) also shows that the resulting distribution from the combination of 100 simulations results in a smoothed distribution very similar to the observed rainfall distribution. The KS test again indicates that there is no significant difference between these two distributions. This confirms that the modelling and simulating of daily rainfall data using Tweedie distributions can provide accurate and relevant results. Tweedie distributions therefore offer a sound and efficient way to explore differences in crop yield that are attributed to changing rainfall distributions.

The process has also been applied to wheat yields, however preliminary analysis has not proved as successful and could be an indication of a greater need for auto-correlation in the simulated rainfall models particularly for winter crops. Further analysis of summer versus winter crops will lead to a better understanding of how the Tweedie distributions should be implemented for various cropping systems.

A large scale analysis of these distributions and the various models from numerous locations would assist in developing better climate-related risk management and decision making capabilities and could lead to a better understanding of the variations in crop yields due to variations in rainfall distributions.

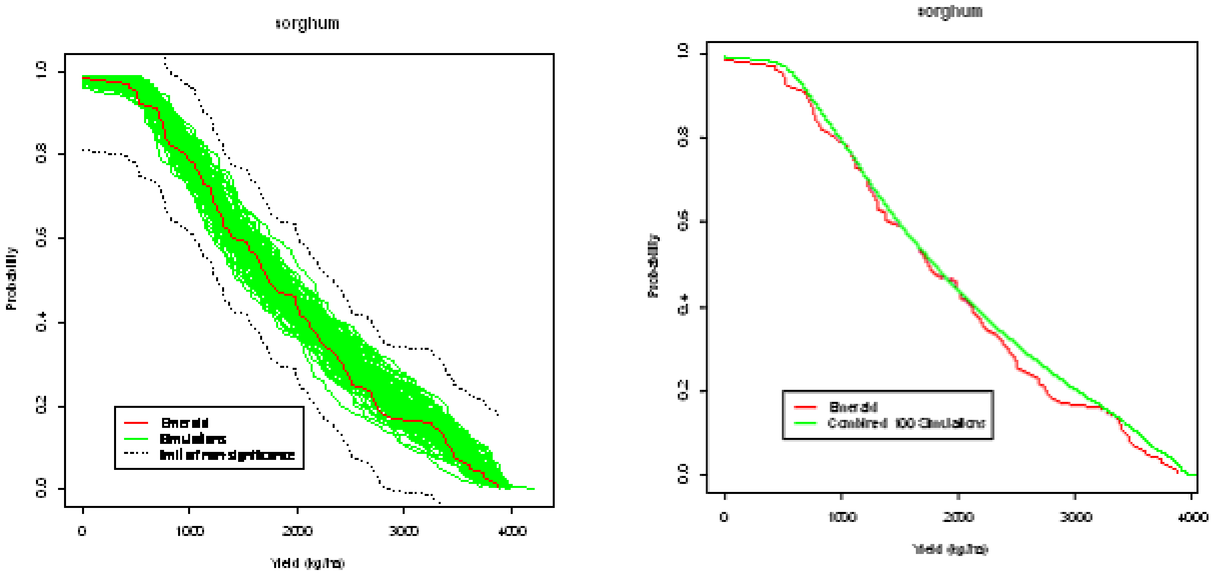

Figure 3: Probability of exceeding yield plots for various simulated rainfall distributions for sorghum crops (a) showing output from 100 simulations and observed rainfall data and (b) showing output from the combined 100 simulations. APSIM output is based on observed Emerald rainfall data and is compared with 100 replications (simulated values over the period 1901-2000) of rainfall data.

Conclusion

In research conducted thus far, it appears that there is great promise and many advantages in using Tweedie distributions for rainfall modelling. Being able to find one combined model for the rainfall process instead of two separate processes results in both an improvement and simplification over previously researched methods of analysis.

If reliable, accurate and simplified models such as those shown in this paper can be produced using these Tweedie distributions then there are many applications in crop science and beyond that could greatly benefit from this work. The input of various rainfall distribution simulations into applications such as APSIM and other crop models could then also lead to better risk management and future decision-making capabilities.

References

Clarkson, N.M. (2002). CLIMARC: a project to extend the Australian computerised CLIMate ARChive.Final Report to Land & Water Australia, Project No. QPI 43.

Coe R, Stern RD (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology* **2**, 1024-1031.

Coe R, Stern RD (1984). A model fitting analysis of daily rainfall data. *Journal of Royal Statistical Society* 147(1), 1-34.

Dunn PK, Smyth GK (1996). Randomised quantile residuals. *Journal of Computational and Graphical Statistics* **5(3)**, 236-244.

Keating BA, PS Carberry, et al. (2002). An overview of APSIM, a model designed for farming systems simulation. European Journal of Agronomy 18, 267-288.

Lennox SM (2003). A statistical approach to rainfall modelling. Bsc (Hons) thesis, University of Southern Queensland, Australia.

Tweedie MCK (1984). An index which distinguishes between some important exponential families. In ‘Statistics Applications and New Directions’, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. (Ed. JK Ghosh and J Roy) pp. 579-604. (Indian Statistical Institute: Calcutta).