

Bootstrap Confidence Intervals for Predicted Rainfall Quantiles

Peter K. Dunn

Department of Mathematics and Computing

University of Southern Queensland, Toowoomba

email: dunn@usq.edu.au

Telephone: +61 7 4631 2772

Fax: +61 7 4631 1775

February 25, 2003

Abstract

Rainfall probability charts have been used to quantify the effect of the Southern Oscillation Index on rainfall for many years. To better understand the effect of the SOI phases, we discuss forming confidence

intervals on the predicted rainfall quantiles using percentile bootstrap methods.

KEYWORDS: bootstrap; southern oscillation; confidence intervals; rainfall.

1 Introduction

The relationship between the Southern Oscillation Index (SOI) and rainfall has been well known for many years (for example, Troup (1965), Quinn and Burt (1972) and others). Using the SOI, Stone and Auliciems (1992) constructed five phases of the SOI which can then be used to study the effect of the SOI. The use of phases has proved to be far more constructive than using just the raw SOI itself. The phases have since been used to study the effects on cropping (for example, Meinke *et al.*, 1996), frost (for example, Willcocks and Stone, 2000 and Stone *et al.*, 1996a), assessing planting opportunities (Stone and McKeon, 1993), wheat crop management (Hammer *et al.* 1996), the effects of sorghum ergot on sorghum production (Meinke and Ryley 1997) and other phenomena. In addition, the SOI phases have been used in the APSIM computer simulation software (McCown *et al.* (1995)) to model the

growing of crops.

Despite the successful use of the method, one shortcoming, however, is that no confidence bands have been placed on the quantiles used in such applications. Indeed, the determination of confidence limits for quantiles on rainfall is a difficult, if not impossible, problem analytically. Confidence bands would also serve to indicate the accuracy of the results reported in the rainfall probability charts. Using the bootstrap method, however, such confidence intervals can be generated using computer simulation.

In this paper, we add bootstrap confidence intervals to the rainfall probability charts (introduced by Stone and Auliciems (1992)) commonly used to show the effects of the SOI phases on rainfall quantiles. In the past, Kruskal-Wallis or Kolmogorov-Smirnoff non-parametric tests (see, for example, Conover (1999)) on medians have been used to quantify the effects of the SOI (for example, Stone and Auliciems 1992). Confidence intervals provide more information, and enable the results to potentially be carried into calculation based on the SOI phases.

In the next section, we briefly discuss the SOI phases. We then discuss the bootstrap method in Section 3, and different bootstrap methods for confidence intervals in Section 4. In Section 5, we examine rainfall proba-

bility charts produced using the bootstrap confidence intervals. Finally, in Section 6, we conclude and make further comments.

2 The SOI Phases

Many attempts have been made to quantify the relationship between the SOI and rainfall. The novel approach taken by Stone and Auliciems (1992) is based on a principal components analysis and cluster analysis of the SOI. This results in the monthly SOI and the *changes* in the monthly SOI being used to classify the SOI trend into five different “phases”. Broadly speaking, the five phases correspond to the SOI falling rapidly, staying consistently negative, staying consistently near zero, staying consistently positive, and rising rapidly. Based on historical records, the rainfall (or any other phenomenon) at any location with sufficient records can be gathered for each of the five different phases, and the distributions of rainfall for these phases can be displayed using boxplots or rainfall probability charts. These charts usually plot the proportion of years for which the rainfall exceeds given levels. For a more complete discussion of the SOI phases, see Stone and Auliciems (1992). Stone *et al.* (1996b) recently applied the phases to global rainfall prediction,

and to the effect of the SOI on rainfall in the whole of Australia (Stone *et al.* (1999)).

In seasonal rainfall prediction, the SOI phases are usually used to predict the total rainfall in the three months following the month in which the average monthly SOI is calculated. We then define the p th rainfall quantile, r_p , as that value of rainfall below which a proportion p of years have received greater rainfall. That is, the p th rainfall quantile of rainfall R is defined as $\Pr(R > r_p) = p$. Note that this is different to the usual statistical definition of quantiles, where we would have $\Pr(R < r_p) = p$.

3 The Bootstrap Method

The bootstrap method is a technique used for determining, among other things, the accuracy of statistics. It relies heavily on computer simulations. Traditionally, standard errors have been calculated using well known formulae often based on assumptions that are not satisfied or only approximately satisfied, or perhaps worse, where it is not known if the assumptions hold or not. The use of bootstrap methods overcomes these problems. For an excellent, easy to understand introduction to the subject, the reader is re-

ferred to Efron and Tibshirani (1993) or DiCiccio and Efron (1996). More mathematical details can be found, for example, in Hall (1992). In essence, the bootstrap method relies on resampling with replacement from the given sample and calculating the required statistic from these repeated samples. The values of the statistic from the repeated sampling can then be used to generate standard errors and confidence intervals for the statistic. It may be clear to some readers that the bootstrap method shares some common elements with other computational methods, such as the jackknife and cross-validation. An illustrative, easy to follow example can be found in Efron and Tibshirani (1993), Chapter 1.

One of the assumptions made, even in the bootstrap, is that the data are independent. See Zwiers (1990), Remark 2.2 of Singh (1981), for example; Wilks (1997), Künsch (1989) and Liu and Singh (1992) discuss blocking bootstrap as a solution. The SOI series and the rainfall series themselves naturally exhibit serial correlation. However, the current application does not use the straight SOI or rainfall series, but selections from these that satisfy certain criteria of SOI phase and month. For example, if we are interested in the rainfall distribution for the three months starting in July based on a June with an SOI Phase 5, our entire data set would comprise

the three month rainfall totals from the 28 years 1890, 1891, 1894, 1895, 1898, 1899, 1907, 1908, 1913 and so on, and serial correlation is not an issue. Indeed, the autocorrelation function of the rainfall for the above years shows no significant autocorrelations.

4 Applying the Bootstrap Method

More formally, we suppose a random sample of size n , $\mathbf{x} = (x_1, x_2, \dots, x_n)$ from an unknown population distribution F . We let the parameter of interest be $\theta = t(F)$, estimated from the sample \mathbf{x} . We wish to find an estimate of θ (say $\hat{\theta}$) using the information in the sample \mathbf{x} ; we let this statistic of interest, then, be $\hat{\theta} = s(\mathbf{x})$. We then define a *bootstrap sample* as \mathbf{x}^* as a random sample of size n drawn *with replacement* from the given sample \mathbf{x} . From this bootstrap sample, the statistic of interest can again be determined; this bootstrap replication of $\hat{\theta}$ determined from \mathbf{x}^* is then denoted by $\hat{\theta}^*$. The bootstrap estimate of the standard error of θ is then the standard error of $\hat{\theta}$ for the datasets of size n randomly drawn with replacement from \hat{F} .

In the context of confidence intervals, there are many bootstrap methods available. The *percentile-t method* uses the sample itself to generate

the equivalent of t -scores in the usual normal theory confidence intervals. These intervals are transformation respecting (Efron and Tibshirani (1993), page 175). If an estimate is transformation respecting, the confidence interval of the transformation of the parameter of interest, $m = T(\theta)$, is just $[m_{\text{low}}, m_{\text{upp}}] = [T(\hat{\theta}_{\text{low}}), T(\hat{\theta}_{\text{upp}})]$. In practice, we can see how the confidence intervals of rainfall quantiles would be used in further dependent studies (in cropping simulations for example), and the transformation respecting property may be of some importance.

The *bias-corrected and accelerated* version, usually abbreviated to BC_a , allows for bias correction, and also introduces a skewness adjustment. A major shortcoming of the BC_a method is that of undercoverage with small samples and large values of α (Hall, 1992, §3.10.5). It is quite possible that small samples will arise in this application (for example, when a particular SOI phase has not historically occurred very often in a particular month). The BC_a method is also transformation respecting and second order accurate (see Efron and Tibshirani, 1993, §14.3 for further details). The percentile method is only first order accurate.

A further method is the *approximate bootstrap* confidence intervals, that attempt to reduce the computation involved. This method, however, uses a

Taylor series expansion and hence requires that $\hat{\theta} = s(\mathbf{x})$ be smooth, which is not the case for quantiles.

There are numerous methods for confidence intervals; we have just previewed some of those. For the present application, we must be aware the quantiles are not smooth, and in particular, finding confidence for quantiles close to 0 and 100 must be treated carefully. For example, Johns (1988) uses the concept of importance resampling to reduce the number of bootstrap samples that need to be taken to produce desirable properties.

We decide that the percentile- t method is the best compromise to use in the current application.

The idea of percentile intervals is reasonably simple. Suppose we generate n_B bootstrap samples from our data, and for each sample we determine the statistic of interest, $\hat{\theta}^* = s(\mathbf{x}^*)$. (The statistics of interest in our application are the rainfall quantiles.) If we let the distribution function of $\hat{\theta}^*$ be $\hat{F}(\hat{\theta}^*)$, we can hence determine the $1 - 2\alpha$ confidence interval as

$$[\hat{F}^{-1}(\alpha), \hat{F}^{-1}(1 - \alpha)].$$

We can then define $\hat{F}^{-1}(\alpha) = \hat{\theta}^{*(\alpha)}$, and write the confidence interval as

$$[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}].$$

Essentially, we determine the statistic for each bootstrap replication, and the $100(1 - \alpha)$ confidence interval is the α th and $(1 - \alpha)$ th empirical percentiles.

5 Application of the Bootstrap to Rainfall Quantiles

The percentile bootstrap method discussed in the previous section can now be applied to the forming of rainfall quantiles based on phases of the SOI, and displayed using rainfall probability charts. The calculations themselves are relatively straight forward. The calculations were done in MATLAB (The MathWorks Inc (1997)) but could also have be done in C++, FORTRAN, S-PLUS (MathSoft (1997)) or something else. MATLAB was chosen since the code is more transparent.

For the purpose of an example, we use the rainfall records at Charleville to demonstrate the technique. In the Figures, we plot the rainfall probability

Figure 1: The 95% Confidence Intervals on the Rainfall Percentiles for Predicted Rainfall from August to October Based on a Phase 2 SOI in the Previous July.

Figure 2: The 95% Confidence Intervals on the Rainfall Percentiles for Predicted Rainfall from September to November Based on a Phase 2 SOI in the Previous June.

charts with the 95% confidence intervals supplied in dotted lines. We select a few cases to show the types of plots that result. In practice, we found that at least 1000 bootstrap samples were needed for reasonably stable bootstrap confidence intervals, in line with the discussion in Efron and Tibshirani (1993) §19.3.

It is clear from the figures that there is some measure of confidence that we can place on the predicted rainfall quantiles. For example, using Figure 1, the probability of obtaining at least 100mm of rain from August to October based on a Phase 2 SOI in July is about 45%. We can also add some measure of confidence to the prediction, stating that the 95% confidence interval is between approximately 30% and 60%. These confidence intervals could potentially be useful in modelling situations and computer simulations based on

Figure 3: The 95% Confidence Intervals on the Rainfall Percentiles for Predicted Rainfall from March to April Based on a Phase 3 SOI in the Previous February.

Figure 4: The 95% Confidence Intervals on the Rainfall Percentiles for Predicted Rainfall from November to January Based on a Phase 2 SOI in the Previous October.

Figure 5: The 95% Confidence Intervals on the Rainfall Percentiles for Predicted Rainfall from January to March Based on a Phase 4 SOI in the Previous December.

the SOI, as discussed earlier, to introduce a confidence interval into forecasts.

6 Conclusion and Discussion

In this paper, we have discussed using SOI phases to predict rainfall patterns. We have introduced the notion of adding confidence limits on predicted rainfall quantiles using the percentile form of the bootstrap. The results provide useful bounds on the predicted rainfall quantiles based on SOI phases, which serve to indicate the accuracy of the rainfall probability charts used by Stone and Auliciems (1992). The methods could also be used to provide bounds on other quantities that are based on SOI phases.

The concept of importance resampling (see Lange (1999), Johns (1988) or Efron and Tibshirani (1993)) could also be used to increase the efficiency of the procedure when finding quantiles with small or large α .

7 Acknowledgements

The author would like to thank the two reviewers, whose comments and recommendations substantially improved this paper. We also thank the Queensland Department of Primary Industries for the data.

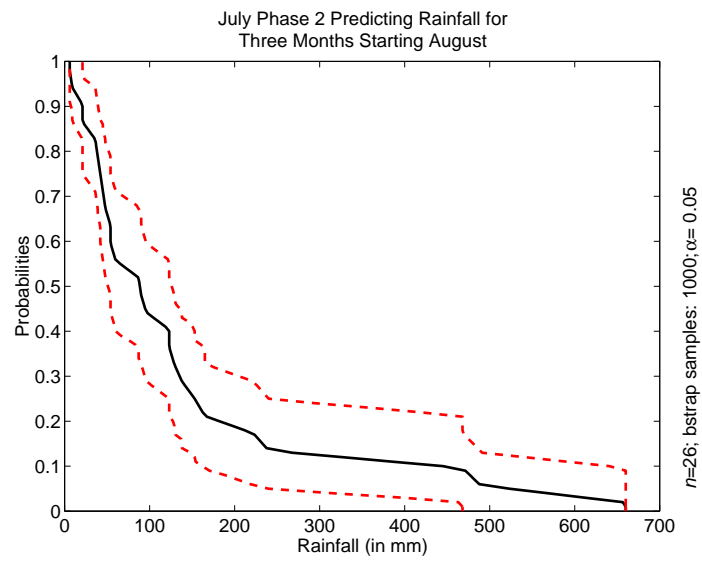
References

1. Conover WJ. 1999. *Practical Nonparametric Statistics*, 3rd edition. John Wiley and Sons: New York.
2. DiCiccio TJ, Efron B. 1996. Bootstrap confidence intervals. *Statistical Science*. **11**(3): 189–228.
3. Efron B, Tibshirani RJ. 1993. *An Introduction to the Bootstrap*. Chapman and Hall: New York.
4. Hall P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag: New York.
5. Hammer GL, Holzworth DP, Stone RC. 1996. The value of skill in seasonal climate forecasting to wheat crop management in a region with

- high climatic variability. *Australian Journal of Agricultural Research*. **47**: 717–737.
6. Johns MV. 1988. Importance sampling for bootstrap confidence intervals. *Journal of the American Statistical Association*. **83**(403): 709–714.
 7. Künsch HR. 1989. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*. **17**: 1217–1241.
 8. Lange K. 1999. *Numerical Analysis for Statisticians*. Springer-Verlag: New York.
 9. Liu RY, Singh, K. 1992. Moving blocks bootstrap captures weak dependence, in Lepage R, Billard L (eds) *Exploring the Limits of the Bootstrap*. John Wiley: New York, 225–248.
 10. MathSoft (pub). 1997. *S-Plus 4 Guide to Statistics*. Data Analysis Products Division, MathSoft: Seattle.
 11. The MathWorks Inc (pub). 1997. *The MATLAB Reference Guide version 5*, MA:Author.
 12. McCown RL, Hammer GL, Hargreaves JNG, Holzworth D, Huth NI.

1995. APSIM: an agricultural production system simulation model for operational research. *Mathematics and Computers in Simulation*. **39**: 225–231.
13. Meinke H, Stone RC, Hammer GL. 1996. SOI phases and climatic risk to peanut production: A case study for northern Australia. *International Journal of Climatology*. **16**: 783–789.
 14. Meinke H, Ryley M. 1997. Effect of sorghum ergot on grain sorghum production: a preliminary analysis. *Australian Journal of Agricultural Research*. **48**:1241–1247.
 15. Quinn WH, Burt WV. 1972. Use of southern oscillation in weather prediction. *Journal of Applied Meteorology*. **11**: 616–628.
 16. Stone RC, Auliciems A. 1992. SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology*. **12**: 625–636.
 17. Stone RC, McKeon GM. 1993. Prospects for using weather prediction to reduce pasture establishment risk. *Tropical Grasslands*. **27**: 406–413.
 18. Stone RC, Nicholls N, Hammer G. 1996a. Frost in northeast Australia:

- Trends and influences of phases of the southern oscillation. *Journal of Climate*. **9**(8): 1896–1909.
19. Stone RC, Hammer GL, Marcussen T. 1996b. Prediction of global rainfall probabilities using phases of the southern oscillation index. *Nature*. **384**: 252–255.
 20. Stone, R, Marcussen, T, Partridge, I. 1999. *Seasonal Rainfall in Australia based on Phases of the SOI*. DPI Information Series CON00997, Queensland Department of Primary Industries: Toowoomba
 21. Troup AJ. 1965. The “Southern Oscillation”. *Quarterly Journal of the Royal Meteorological Society*. **91**: 490–506.
 22. Wilks, DS. 1997. Resampling Hypothesis Tests for Autocorrelated Fields. *Journal of Climate*. **10**: 65–82.
 23. Willcocks, JR, Stone, RC. 2000. *Frost Risk in Eastern Australia and the Influence of the Southern Oscillation*. DPI Information Series QI00001, Queensland Department of Primary Industries: Toowoomba.
 24. Zwiers FW. 1990. The Effect of Serial Correlation on Statistical Inferences Made with Resampling Procedures. *Journal of Climate*. **3**:



1452–1461.

