Using the REML statistical procedure to examine trends due to climate change

Kerry L. Bell, **Sarah Lennox** and Peter DeVoil

Department of Primary Industries and Fisheries, PO Box 102, Toowoomba, Qld 4350.
www.dpi.qld.gov.au
Emails Kerry.Bell@dpi.qld.gov.au; Sarah.Lennox@dpi.qld.gov.au; Peter.DeVoil@dpi.qld.gov.au

Abstract

There is strong evidence that human influences have affected climate change due to continued emissions of carbon dioxide and other greenhouse gases from fossil fuels and other sources. Understanding these changes and projecting future changes is important in order to maintain or increase economic returns and decrease environmental impacts from agricultural production. This requires statistical tools that confidently identify trends in noisy time series. We explore the use of the REML (Restricted Maximum Likelihood) procedure as a statistical tool that accounts for an autoregressive term, cyclic trends and deterministic trends in the form of line and spline terms. The yearly minimum temperature at Emerald in Central Queensland is used as an example.

Media summary

We used Restricted Maximum Likelihood (REML) procedures as a powerful, statistical tool to confidently identify climate change trends in noisy time series.

Key Words

Climate change, time series, REML, Emerald, multi-taper method (MTM)

Introduction

Human activities have already contributed to climate change due to continued emissions of carbon dioxide and other greenhouse gases from fossil fuels and other sources (IPCC 2001). Howden et al. (2003) showed that climate change has already impacted on agricultural production systems in Central Queensland, Australia. This trend is likely to continue and adequate adaptation and mitigation responses have to be developed and implemented to ensure ongoing profitability of the agricultural sector in this region (Potgieter et al. in prep). This requires appropriate statistical tools that can confidently identify trends in noisy time series.

Many statistical techniques can be used to analyse time series, each having their strengths and weaknesses. These include, amongst others, the classical linear regression, which is simple to use, easy to interpret but not adequate for autocorrelated data. The REML procedure (Restricted Maximum Likelihood) (Patterson and Thompson 1971, Genstat 6 Committee 2002) is another way to look at times series, as it can model autoregressive processes. Cycles can be fitted to the model and linear and spline terms can be added and assessed for significance. Although REML has not been designed strictly with time series in mind, it has the ability to model the autocorrelation across residuals, as well as identify trends over time, an issue of particular importance when investigating climate change.

This paper examines the use of the REML procedure to examine trends in climate change data.

Issues concerning time series analysis

There are several aspects of a time series that need to be addressed when modelling the data as follows:

- autocorrelation (ie. interdependence of temporally or spatially adjacent data Figure 1A)
- quasi-cyclical, stationary oscillations in time series of historical climate records (Figure 1B)

- non-stationary trends (either linear or non-linear) in climate records as a footprint of climate change (Figure 1C).

A time series may contain any of these components in addition to white and/or red noise (i.e. long-term irregular behaviour; Maia et al. 2004; Meinke et al. 2004).

The main impact of autocorrelation (A in Figure 1) when modelling trend over time is that traditional regression techniques assume independency of the values when testing for significance of trend. Any autocorrelation violates this assumption, thus invalidating such regression techniques (Stern and Kaufmann 1997).

Cycles in the data (B in figure 1) can be explored through various spectral methods (Ghil et al. 2002). The classical spectral method that is based on a univariate analysis may be inappropriate and lacking power (Stern and Kaufmann 1997). The presence of a unit root (stochastic process), an indication of non-stationarity, also has implications for most spectral analyses. If there is suspected to be a significant change in the climate variable over time (outside of seasonality) then the classical spectral analysis methods are inappropriate.

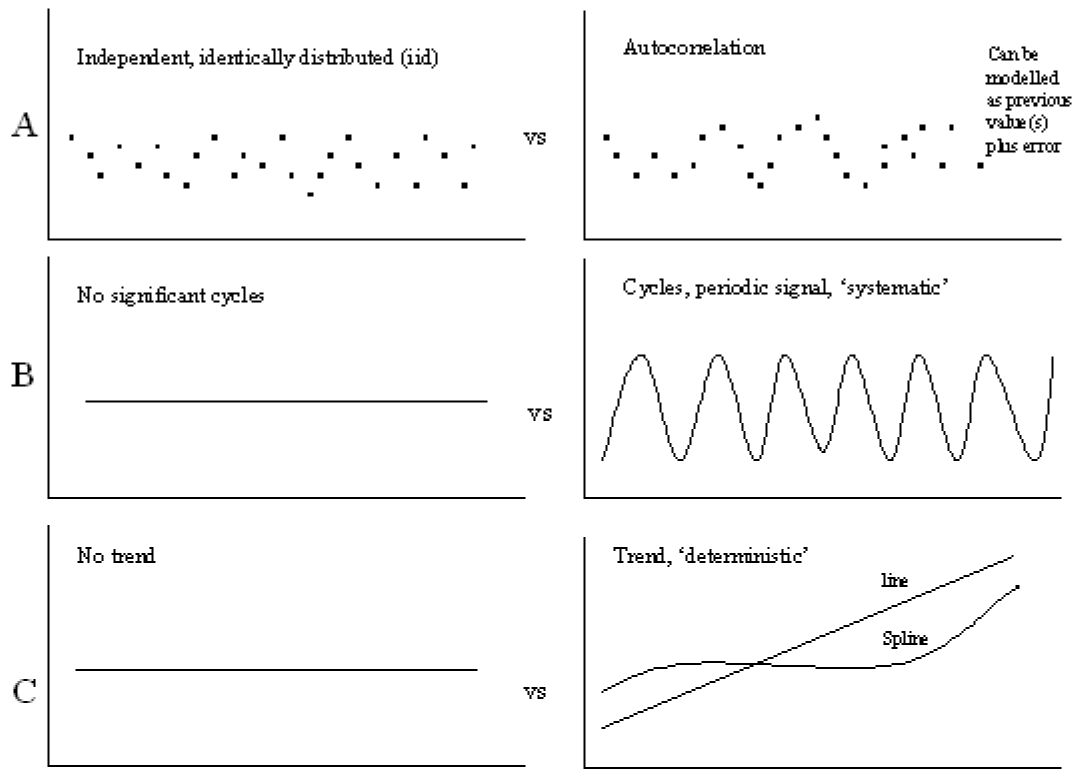

Figure 1. The components of a time series can be made up of a autocorrelation between nearby values (A), a cyclic trend (B) and a trend component (C), or any parts of these. The left hand side shows when these components are not present, while the right hand side shows examples of these.

Methods

In this paper we have:

- examined yearly minimum temperature data for the 93-year period (1910-2002) from Emerald in Central Queensland (this example was chosen because it shows a significant autoregressive term, a strong cyclic term and a non-stationary trend)
- applied the multi-taper spectral analysis method (MTM) (Ghil et al. 2002) to determine periodicities of the cycles in above data and thus identify terms to apply in the REML analysis
- applied the REML analysis through Genstat 6 (Genstat 6 Committee 2002), by initially fitting an autoregressive term (order 1 and 2), terms for the cycles and a line and spline terms for years. The autoregressive and spline term was assessed for significance by successively deleting each term from the model and assessing the change in deviance against a chi-square distribution. The cycles and linear year term were assessed using Wald statistics within REML (compared to a chi-square distribution). Other ways to model the autocorrelation in REML could be through a term for moving average (MA) and a combination of autoregression and moving average (ARMA).

MTM is a non-parametric spectral analysis technique able to detect significant cycles regardless of the existence of autocorrelation and trend. The spectral peaks can be identified with higher resolution and greater confidence yielding a more stable and better spectral estimate than the classical spectral analysis (Ghil et al. 2002; Meinke et al. 2004).

Many time series methods (e.g. ARIMA, Cryer 1986) suggest removing trends from the data sets, usually by differencing, before assessing the model for autocorrelation and for cyclic trends. However, assessing such trends is a key objective in climate change research, hence we do not consider differencing techniques as appropriate approaches for our purposes. Unlike other techniques, structural time series analysis is able to model both cycles and trend amongst the data. It is superior to some techniques, however, it is limited to particular statistical packages and is in some implementations unable to adequately model cycles that have a period greater than one year.

Results and Discussion

Results from the MTM analysis are shown in Figure 2 (see also Meinke et al. 2004, who use monthly minimum temperatures in their analysis). Strong cycles were detected at 1000 year, 93.5 years, 48.8 years and 2.3 years (unmarked). The 1000 year, 93.5 year and 48.8 year are indications of non-stationarity of the data indicative of climate change.

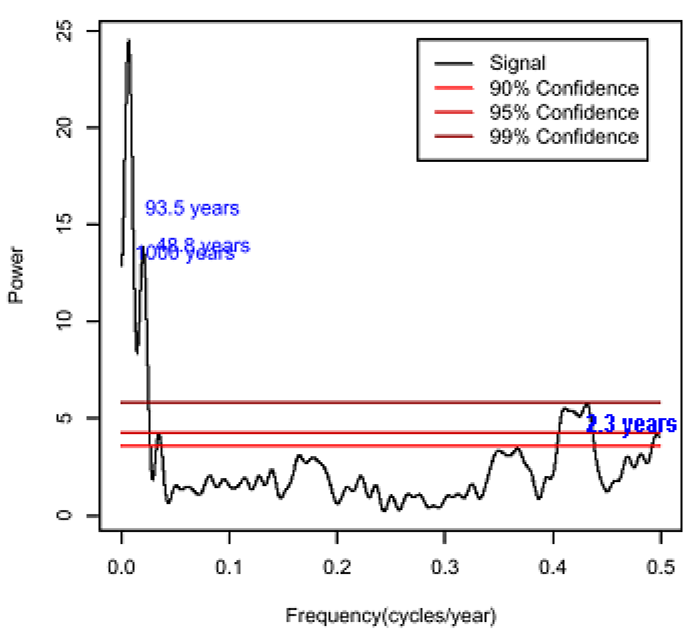

Figure 2. Spectral analysis of yearly minimum temperature for Emerald from 1910 to 2002 using the multitaper method (MTM).

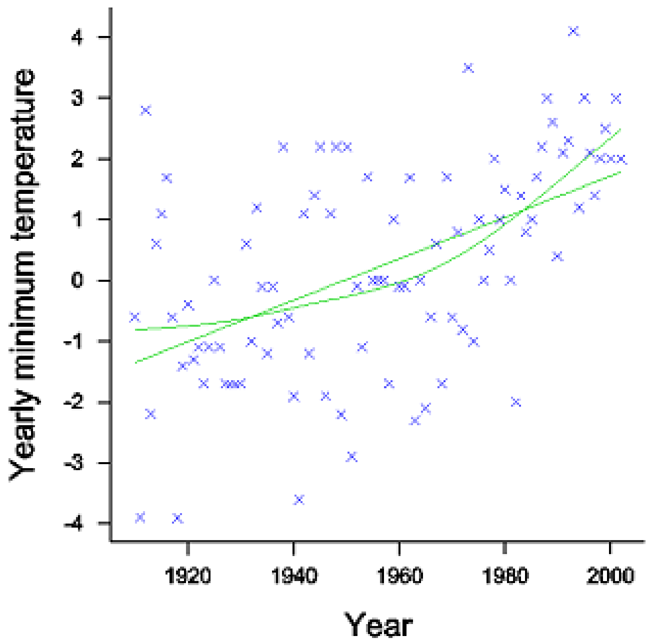

Figure 3. Plot of the yearly minimum temperature for Emerald, the fit as a line and the fit with a spline.

The 2.3 year cycle identified via MTM analysis was approximated as a set of cubic polynomials in the fixed part of the REML model. This periodic term was found to be non-significant and dropped from the model, as it didn't provide useful additional information in explaining the variability of minimum temperature. It is interesting that the cycles found in the MTM analysis were not significant in the final REML model, this could be either due to how they were fitted or to lack of strength of these cycles in the presence of other terms.

The resulting REML model included a significant autoregressive term (order 1) and a significant spline term over years. Figure 3 shows the original data, with a linear regression fit and a fit with the spline term. The spline shows an increase in the yearly minimum temperature since approximately 1970. This is consistent with other analyses of the same data, and our understanding of climate change dynamics.

Fitting this data using REML had the advantage of being able to model the correlation across time and also fit and display the trend over time as a spline.

Caution should always be taken in extrapolating the fitted trends over time. An example of an incorrect extrapolation could be assuming that a trend may continue at the same rate as the last 30 years, where in fact it could be flattening out or accelerating. An increase in variables such as temperature may be part of a larger cycle that is greater than the length of our data set, and we could possibly be only measuring a rising part of such a cycle.

Conclusion

For climate change research, it is essential to model the residual structure of time series data correctly so that trend components can be tested and assessed appropriately. The REML procedure is one method that allows the fitting of, for instance, an autoregressive term (or any other REML compatible term) as well as linear and spline terms in the one model.

Known disadvantages of REML are lack of convergence for certain models and the need for the data to be normally distributed.

REML is another way to look at the trends of climate change, which was particularly useful for this application as we can describe climate trend while modelling the autocorrelation across time. Understanding climate change should aid in managing future agricultural production systems, in decisions such as crop variety selection and in optimal planting times.

References

Cryer JD (1986) Time Series Analysis. Duxbury Press, Boston.

Genstat 6 Committee (2002) The Guide to GenStat Release 6.1, Part 2: Statistics. VSN International, Oxford, UK.

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Roberston AW, Saunders A, Tian Y, Varadi F and Yiou P (2002) Advanced spectral methods for climatic time series. Reviews of Geophysics 40, pp 1-1 - 1-41.

Howden SM, Meinke H, McKeon GM and Power B (2003) Risk management of wheat in a non-stationary climate: frost in Central Queensland. Modsim'03, International Congress on Modelling and Simulation Proceedings.

IPCC. Climate Change (2001) The scientific basis. Houghton JT, Ding, Griggs DJ, Noguer M, van der Linden JJ and Xiaosu D (eds). Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, UK. Pp. 944.

Maia de H N A, Meinke H, Lennox S, (2004) Mapping probabilistic forecast skill using P-values. 4th Int. Crop Science Congress, Brisbane, September 2004, these proceedings.

Meinke H, Donald L, deVoil P, Power B, Baethgen W, Howden M, Allan R and Bates B (2004) How predictable is the climate and how can we use it in managing cropping risks? Invited Symposium Paper, 4th Int. Crop Science Congress, Brisbane, September 2004, these proceedings.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

Potgieter AB, Hammer GL and DeVoil P (in prep.). A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia. Submitted to Agricultural. Forest Meteorology.

Stern DI and Kaufmann RK (1997) Time series properties of global climate variables: detection and attribution of climate change. Working Papers in Ecological Economics Number 9702, Centre for Resource and Environmental Studies, Australian National University, Canberra.