Journal of Hydrology 519 (2014) 720–731

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Optimal development of calibration equations for paired catchment projects

Leon J. Bren *, Patrick N.J. Lane

Department of Forest & Ecosystem Science, The University of Melbourne, Parkville, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 15 July 2014
Accepted 28 July 2014
Available online 6 August 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Paired-catchments
Calibration
Verification
Nash–Sutcliffe
Experiments

SUMMARY

Data from two Australian paired-catchment projects were used to assess the effect of length of the calibration period on the quality of calibration achieved. The data were divided into a "calibration" and a "verification" set. The Nash–Sutcliffe (N–S) coefficient of efficiency was used to assess the quality of prediction of the verification set as a function of the length of the calibration set. The results showed a rapid initial increase in quality of calibration with increasing calibration length. This then "plateaued". With simple linear regression models, reasonable calibration (N–S > 0.7) was achieved in 60 days and good calibration (N–S > 0.8) in 100 days. More complex models achieved good calibration after 300 days of data. In general, there were no increases in the N–S value achieved after 3 years – the main advantage of longer calibrations appeared to be lower mean errors. Similar results were obtained with daily, monthly, quarterly, or annual subdivisions of flows. All residuals suffered from autocorrelation and non-normality; the former was removed by an autoregressive technique, but the latter appears implicit in the technique. Simulation of the use of an *n*-fold data examination technique to monitor the development of calibration as data flowed in substantially reproduced this result. This appears to be a good strategy for hydrologists monitoring development of calibration in a continuing project. Paired-catchment experimentation is a robust experimental technique but would benefit from application of a set of protocols prescribing techniques.

1. Introduction

The genesis for this paper came from three sources; the first was a conference discussion in which it was argued that one paired catchment project was better than another because it had a longer calibration period. The second was an issue in refereeing of a paper using paired catchment data about whether the presence of autocorrelation in a residual sequence was of importance. The third has been a continuing demand in Australia for information on the water use of regrowth native forests (relative to old-growth) as a function of forest age. In attempting to resolve these issues it was realised that although paired catchment projects have been a major source of information on the hydrology of forested catchments, there is actually little information on the optimal analytical approaches which might or should be used on these.

This paper uses measured, high-quality data sequences from the pre-treatment phase of two Australian paired catchment projects. The questions examined are:

1. What is the gain in information over time as the calibration period extends? Do we find "diminishing returns" after long periods of calibration? How long a calibration period do we need to get stable results? Is there an "optimal length" of the calibration period?
2. What is the relative gain or loss of information in going from daily to monthly to seasonal to annual data? Does autocorrelation of residuals in calibration models materially reduce the value of shorter time subdivisions?
3. Do calibration models formed using shorter sequences of data give similar models to those using longer sequences of data?
4. How might the practising hydrologist monitor their calibration development in real time, to obtain the most efficient calibration?

Of necessity, these questions also forced an examination of the rationale of data analysis in modern paired catchment projects. The work has used a "case study" approach with the aim of seeing whether there are clear trends in at least these two cases. It is hoped that this can be repeated with a wider, international, data set in the coming year.

* Corresponding author.

http://dx.doi.org/10.1016/j.jhydrol.2014.07.059
0022-1694/

2. About paired catchment forest hydrology projects

Paired catchment studies were developed in the forest hydrology world to provide information on the relative water use of trees versus other land uses. The technique was devised about a century ago (McCulloch and Robinson, 1993). In various forms, it has become a common method in water studies around the world for a wide range of comparative studies of water quality and quantity (e.g. Brown et al., 2005); most commonly the method is used to estimate changes of catchment outflow associated with change in vegetation. The method uses two similar catchments with the same initial land use. Data are collected from the two catchments for some years. During this time an appropriate "calibration model" is formed allowing the outflow of one catchment ("control") as a real time predictor of the outflow of the second "to-be-treated" catchment. The second catchment then undergoes a land-use change ("treatment"). The difference between the "treated" catchment outflow and the estimated flow based on the control catchment outflow is a measure of the land use effect (plus error) relative to the original land-use. The use of the control catchment avoids many issues of co-variation over time associated with the vagaries of rainfall and climate; indeed, a common finding with control catchments is that additional variables such as rainfall add little predictive power because they are implicitly included in the control catchment data. The modern view is that the treatment imposes a time-variant signal of changed water use as a function of tree age and annual rainfall; the role of the paired catchment project is to elucidate this.

Hewlett and Pienaar (1973) and Hewlett (1971) provide a good account of the advantages and disadvantages of the technique. Andreassian (2004) provides a more modern view of the method and notes their effectiveness in providing information. Hewlett and Pienaar (1973) argued that the method had, to that date, provided most of the world's information on long term water use of vegetation. We believe that this statement is still true today, although plot-based methods of measurement have sometimes provided alternative and supplementary techniques.

Paired catchment projects are expensive to install and can involve substantial areas of valuable forest. Bren and McGuire (2012) in a survey of the technique in Australia concluded that once the initial capital of installation has been overcome they are relatively inexpensive to maintain. This has been an important factor in the robustness of the technique during periods of volatile forestry organisation change in Australia. Secondly, the duration of the experiment is effectively the length of the calibration + length of the treatment. This can be many years and poses organisational problems of staffing and maintenance of enthusiasm. For instance, Wicht, in his South African forest hydrology work (e.g. Wicht (1967), ultimately proposed a 32 year calibration period (Kruger and Bennett, 2013). Such long periods of time are a major disadvantage of paired catchment experiments. Our view is that anything that shortens the calibration adds to the viability of the technique. We also believe a major reason for achieving a shorter calibration is that, once the treatment is administered, the project is "embedded in the managing organisation". It is then more likely to be brought to a successful conclusion (i.e. refereed papers documenting the results).

Although the methodology has undoubtedly been effective, the technique occupies an interesting place in scientific methodology. Statistical inference is often limited by the difficulty that true replication is probably impossible plus some consistent refusals of hydrologic data to conform to the normal distributions beloved of statisticians. The experiments are certainly quantitative and involve years of measurements, with many complications associated with missing data, droughts, and floods. Allocation of treatments between the control and the catchment to be treated usually involves "practicality" factors and is rarely random. The methods could possibly be made to meet the experimental criteria of clinical trials (e.g. Kaptchuk, 1998) in which the data analyst is separated from the experimentalist and that "blinds" are used so that the analyst has no idea which catchment was treated or even when the treatment took place. However the authors can find no examples of where this strictness has been applied. We believe that development of more rigorous analytical techniques which apply strictures of experimentation will be a useful and possibly fruitful field for catchment experimentalists.

The role of statistics in paired catchment projects has been somewhat chequered. Hewlett and Pienaar (1973) noted that the major difficulty with regression analysis is the possible lack of normality in the errors about the regression. Their advocated method was regression using dummy variables to help distinguish a "before and after" treatment effect. From our point of view, this does not allow a clear separation of the time-varying effects associated with the forest treatment. Hewlett and Pienaar (1973) concurred with Wicht's (1967) conclusion that, in many cases, the treatment effect is so clear that statistical confirmation is, to the forest practitioner at least, superfluous. Hydrologists appear divided into those who feel that the statistical analysis is important and those who feel that the results "speak for themselves" and that statistical analysis is, at best, a nice but non-essential embellishment of such experiments. This has been partly due to the absence of clear statistical methodology other than analysis of covariance on annual data. Hewlett and Pienaar (1973) commented on the then "total lack of progress in the theoretical aspects of catchment experimental design" and this does not appear to have much changed. Other than a few papers looking at aspects of paired catchment analysis (e.g. Watson et al., 2001) there is not a recent, coherent body of work giving guidance on the statistical techniques which might be applied.

An early development of statistical techniques for paired catchment experiments was the work of Wilm (1944, 1949). These asked the fundamental question of "how long should experimental watersheds be calibrated?" The methodology used at the time was analysis of covariance using annual data. The difficulty with this approach was that it assumed a constant treatment effect over time for at least some years after treatment. Selection of annual data possibly reflected the numerical load of computation in those days. Kendall (1946), as quoted by Elashoff (1969) noted that "analysis of covariance is not a mill which will grind out results automatically without care or forethought. It is a rather delicate instrument but requires skill as well as enthusiasm." This aspect is evident in applying analysis of covariance techniques to paired catchment data. Interestingly, Elashoff (1969) also quotes Kendall (1946) as saying that "the reader who roves among the literature of the subject will sometimes find elaborate analysis applied to data in order to prove something which was almost obvious from careful inspection right from the start." Wilm (1944, 1949), amongst others, noted that the underlying assumptions of normality and no autocorrelation of annual data are probably only approximated, but did not consider the matter further.

Kovner and Evans (1954) extended and simplified the analysis of Wilm (1949) to give graphical solutions of how long the calibration period would need to be for a given treatment length and mean difference attributable to the treatment. The analysis uses years of data as the minimum time sub-division. As an example quoted (using data from Coweeta), it was concluded that to detect a 5% change in the mean yield at a significance level of $p = 0.05$, the experiment (and presumably the treatment effect) would have to last for 12 years calibration and another 12 years post-treatment. From the point of view of research hydrologists, the method