Using Bayesian hierarchical models to measure and predict the effectiveness of environmental flows for ecological responses

S.C. de Little[a], J.A. Webb[b], K.A. Miller[a], I.D. Rutherfurd[b] and M.J. Stewardson[a]

[a] *Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Australia*
[b] *Department of Resource Management and Geography, The University of Melbourne, Parkville 3010, Australia*
Email: siobhan.delittle@unimelb.edu.au

Abstract: Quantitative predictions of ecological responses to flow alterations are fundamental to the planning and delivery of environmental water. However, the majority of such predictions are often based on expert opinion, and lack a solid basis in empirical evidence. To derive evidence-based, defensible environmental flow recommendations, new approaches are required that make best use of the available data to predict ecological responses to flow alterations. In this study we use Bayesian hierarchical modelling to explore the impact of changed flow regime on terrestrial vegetation encroachment into river channels. In regulated rivers, encroachment of terrestrial vegetation into the channel is an issue of concern for managers because duration, frequency, and season of inundation are major determinants of plant community development. Environmental flows are often recommended as a way of reducing encroachment, but the assumed response has not been rigorously tested. Neither are there any general quantitative models that describe the predicted benefit (in terms of reduced terrestrial vegetation encroachment) of different flow regimes. In this paper we report a Bayesian approach that identifies the relationship between flow and vegetation encroachment revealed in seven integrated data sets from south eastern Australia. A principal advantage of Bayesian modelling is its flexibility. Thus, one is able to model physical and biological processes as part of a hierarchical statistical analysis. Here, we describe the relationship between terrestrial cover and inundation using a curvilinear function that combines both inundation duration and the number of inundation events. The model also incorporates hydrological data for the 5 years prior to vegetation sampling, weighting the most recent years most heavily. Finally, it accounts for effects of bank slope, season of inundation, and random effects associated with sampling (year of sampling, sampling transect, and uncertainty associated with the survey technique used). The model also improves the precision of estimates by using expert-derived prior probability distributions for model parameters, and by having a hierarchical structure among sites and rivers. Bayesian hierarchical models assume dependency amongst the sampling units, and therefore the model parameter values are assumed to be drawn from a larger common distribution. Relationships from each site 'borrow strength' from other sites, leading to robust influence despite the common problems associated with sample replication in environmental monitoring studies such as this. By combining data across seven different river systems, we are able to quantify relationships between different inundation durations and frequencies and the extent of terrestrial vegetation encroachment. This, in turn, allows us to make predictions of encroachment under different flow regimes. The hierarchical nature of the model allows us to report at the site and river level and also at the state level. These are the scales of interest to local stakeholders and the state funding agencies. Our results highlight the power and flexibility of Bayesian models to make quantitative, evidence-based predictions of ecological responses to changes in flow regimes. Such models will be vital for the future of environmental water management in data-poor situations that are common to environmental monitoring, and produce outcomes that are reportable at different stakeholder and governance levels.

Keywords: Bayesian hierarchical model, environmental flows, terrestrial vegetation encroachment

1. INTRODUCTION

Protecting and restoring the ecological integrity of freshwater systems relies largely on providing a flow regime that mimics the natural regime of that system as closely as possible (Poff et al., 1997). Whilst there is broad international uptake of the policies to provide such 'environmental flows', in reality there has been minimal actual implementation of environmental flow programs (Le Quesne et al., 2010). Environmental flow management requires precise estimates of the relationships between flow delivery (e.g. summer base flow or winter fresh levels) and resultant ecological impacts (e.g. increasing native fish breeding habitat or riparian vegetation community health). Currently, the majority of environmental flow assessments rely heavily on expert opinion, often from informally structured expert panel discussions, to derive flow recommendations (Stewardson and Webb 2010). Although the experts involved are usually experienced and well-respected, the process is not transparent and is potentially open to legal challenge. While there is wide agreement on the general principles regarding the ecological impacts of flow alteration, there has been little success in deriving quantitative, and thus predictive, relationships between specific ecological responses and aspects of the flow regime (Poff and Zimmerman 2010).

Research into the ecological impacts of both flow alteration and restoration is hampered by poor experimental design. Common examples include a lack of replication, spatial and temporal confounding, and the patchy nature of the treatment (i.e. flow delivery is continuous rather than categorical) (Arthington et al., 2006; Stewardson and Webb 2010). These issues make it difficult to use the strongest approaches, such as BACI (before-after control-impact; Underwood, 1997) or related designs to infer the effects of changes in flow. Spatiotemporal gradient studies are the most applicable to the study of flow-ecology relationships, but these data can be difficult to handle statistically. Lack of replication at the correct scale (i.e. sampling from more than one river, rather than from more than one reach), and spatial and temporal confounding greatly impact upon the ability of standard statistical approaches (e.g. ANOVA, GLM) to detect effects. Bayesian statistical approaches, which are inherently flexible, are able to cope with some of these issues (Clark, 2005). Bayesian models are able to be formulated to conform to the requirements of the data, whereas standard approaches force data to comply with the requirements of a relatively small number of model types (McCarthy, 2007). Bayesian methods allow us to model physical and biological processes, rather than simply looking for associations within data sets. Bayesian methods also have the benefit of being able to incorporate existing knowledge of parameter values into analyses through informative prior probability distributions. This reduces the unexplained variation in posterior parameter estimates, and consequently of any quantitative predictions from the model.

Another important feature of Bayesian modelling approaches is that they allow the construction of far more complex models than is possible with traditional statistical approaches, and are particularly suited to dealing with the complexities of spatiotemporal aspects of ecological datasets (Clark, 2005). A Bayesian hierarchical model treats multiple sampling units (e.g. sites or groups of sites) as 'exchangeable' where the parameter values (e.g. mean vegetation cover) that describe each site are expected to be similar (but not necessarily the same). An identical assumption is made in every statistical analysis (frequentist or Bayesian), as replicates within sampling units are assumed to be independent and identically distributed. In Bayesian hierarchical models, exchangeability of parameter values is achieved by assigning a prior distribution, in which the prior parameter values (e.g. site mean) for all sampling units are viewed as a sample from a common distribution. The parameters (e.g. mean across all sites) of this distribution are known as 'hyperparameters'. Therefore, the entire set of observed data can be used to estimate the values of the hyperparameters, even though those values are not observed (Gelman et al., 2004). The practical effects of the Bayesian hierarchical approach is that unexplained variation, both within and among sampling units, is reduced relative to an analysis that does not assume a hierarchical relationship among the sites (Gelman et al., 2004).

Here, we use Bayesian hierarchical modelling to quantify the relationship between different aspects of a managed flow regime (duration, frequency and seasonality of inundation) and the encroachment of terrestrial vegetation into river channels. In regulated rivers, reduced base flows and altered flooding frequencies can lead to growth of terrestrial species within the channel, disrupting normal fluvial processes. In south-eastern Australia, the 'millennium drought' of 2000-2010 saw many river channels heavily colonized by terrestrial species. Terrestrial vegetation encroachment results from a predictable set of conditions, and has globally well-defined consequences for riverine ecosystems (Miller et al., 2013). Environmental flow assessments often recommend flow delivery to remove and/or prevent encroachment in river channels (e.g. VEWH et al., 2011; Konrad et al., 2012), and while the ecological relationships that underpin such flow recommendations are accepted as 'fact', the evidence for and against them often has not been rigorously tested (Sutherland et al., 2004). Our aim was to create a quantitative model of the terrestrial vegetation response to flow regime to by analysing relationships between flow and terrestrial vegetation data from south east Australian rivers.

2. METHODS

The data analyzed in this paper were collected through the Victorian Environmental Flows Monitoring and Assessment Program (Chee et al., 2009; Cottingham et al., 2005). VEFMAP is a major government-funded monitoring program, specifically designed to assess the effectiveness of the state's environmental flows program. Vegetation data were collected from 27 sites across 7 different river systems spread across the entire state (http://tinyurl.com/VEFMAPvegSites). At each site, Braun-Blanquet (Mueller-Dombois and Ellenberg, 1974) cover scores for each species were assessed in 1 x 1 m quadrats. The Braun-Blanquet cover abundance scale consists of 7 classes relating to different amounts of cover: 0 (not present), + (few individuals), 1 (<5%), 2 (5-25%), 3 (35-50%), 4 (50-75%) and 5 (75-100%). The quadrats were spaced across 10 cross-sectional transects that spanned approximately 500 m of the river. This process was repeated up to 3 times (2008, 2010, 2012). Terrestrial vegetation species were classified into functional groups according to Casanova and Brock (2000), and the mid-points of the Braun-Blanquet cover ranges of all the terrestrial species were summed together to give a 'total terrestrial vegetation cover' for each quadrat. Site surveys were used to relate the elevation of each quadrat to hydrological data, and therefore to calculate the number of days of inundation per year and the number of individual inundation events experienced by that quadrat. Overall, the analysis incorporated data from 9464 quadrats.

Results of a systematic literature review (Miller et al., 2013) and an expert-based Bayesian network (de Little et al., 2012) indicated that terrestrial vegetation cover within river channels will be highest with little or no inundation, and will drop towards zero with increasing duration of inundation. Breaking the same period of inundation into a number of separate inundation events is expected to reduce the amount by which vegetation cover might be reduced if that inundation period had been uninterrupted (Figure 1).

The expected joint effects of inundation duration and frequency can also be represented mathematically as:

$$c = v_0^{-m\left(\frac{T}{f}\right)} \qquad \text{Equation 1}$$

where c is the % terrestrial vegetation cover expected within the stream channel, v_0 is the maximum cover seen when inundation is zero, m is a parameter that defines how the steepness of the curve from v_0 down to 0 changes with the number of distinct inundation events (f). This explicit mathematical expression of the process by which inundation (T) and event duration (f) interact to reduce terrestrial vegetation cover forms the core of our model.

Figure 1. Diagram showing the relationship between vegetation cover (c) and the duration (T) and frequency (f) of inundation.

We were also interested in the effect of the timing of environmental flow delivery on terrestrial encroachment, so our model also included season of inundation as an additive effect. Season of inundation was quantified as the ratio of days of inundation in winter/spring (May-November) to days of inundation in summer/autumn (December-April). We included hydrological data for 5 years prior to the vegetation sampling. However, inundation in more recent years was assumed to be more important than older inundation events. To capture this effect, the inundation data were weighted by a negative exponential function according to their age. The strength of the negative exponential weighting was determined by a scalar parameter (d) fitted during the analysis. Bank slope differed within transects and also between sites. As steep banks hold less fertile sediments, and are likely to have lower vegetation cover, we also included angle of bank for each quadrat as an additive effect. To account for any pseudo-replication in our sampling design, we included random effects for transect (as cover per quadrat was assessed along 10-15 transects at each site) and sample year (sites were sampled in multiple years: 2008, 2010 or 2012) in the model. Finally, we square-root transformed the cover data prior to analysis in order to achieve a better spread of points around the fitted function, and to allow us to assume Gaussian-distributed residuals.

The final model structure is thus:

$$\sqrt{c_{(SR)}} \sim N\left(\mu_{(SR)}, \sigma^2_{[c]}\right) \qquad \text{Equation 2}$$

$$\mu_{(SR)} = v_{0_{(SR)}}^{-m_{(SR)}\left(\frac{T_{(SR)}}{f_{(SR)}}\right)} + \beta Se_{(SR)} + \gamma_{(S)} Sl_{(SR)} + \delta_{(xs)} + \varepsilon_{(year)} \qquad \text{Equation 3}$$
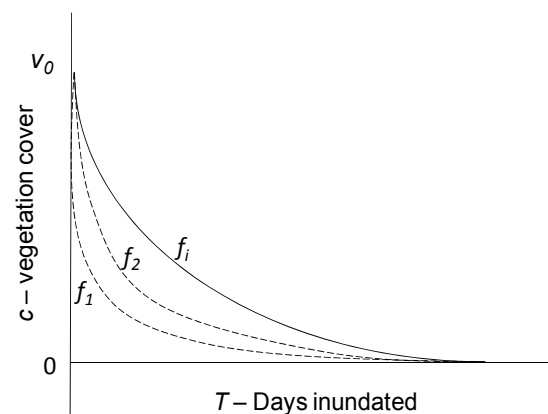
where $c_{(SR)}$ is vegetation cover data within each site (S) within each river (R). N refers to a normal distribution, $\mu_{(SR)}$ is the mean of the modelled (square root transformed) cover, and $\sigma^2{}_{[c]}$ is the variance of this distribution; the bracketed subscript for a variance parameter indicates the parameter or data for which variance is estimated. $Se_{(SR)}$ and $Sl_{(SR)}$ are the season of inundation and bank slope for each quadrat, respectively. The parameters β and $\gamma_{(S)}$ are coefficients for the season and slope covariates (the subscript S denotes that this parameter was estimated at the site-scale only, not at the river-scale). $\delta_{(xs)}$ is the random effect of transect (xs) and $\varepsilon_{(year)}$ is the random effect of sampling year.

The parameters $v_{0(SR)}$, $m_{(SR)}$, and $\gamma_{(S)}$ were modeled hierarchically. The prior for each site-level parameter was assumed to be drawn from river-level distributions of parameter values with their own mean and variance, and these river-level parameters were again assumed to be drawn from a state-level distribution of parameter values. For example:

$$v_{0(SR)} \sim N\big(r.v_{0(R)}, \sigma^2{}_{[v_0]}\big) \qquad \text{Equation 4}$$
$$r.v_{0(R)} \sim N\big(g.v_0, \sigma^2{}_{[r.v_0]}\big) \qquad \text{Equation 5}$$

Where $r.v_{0(R)}$ indicates the river-level mean of the distribution of $v_{0(SR)}$ values and $\sigma^2{}_{[v_0]}$ dictate the deviation of individual $v_{0(SR)}$ values from $r.v_{0(R)}$. This structure is the same for $m_{(SR)}$, and $\gamma_{(S)}$, although $\gamma_{(S)}$ does not have a river-level hyperparameter (see Fig 2). Parameters that are not subscripted with S or R ($g.v_0, g.m, g.\gamma, \beta, \delta_{(xs)}, \varepsilon_{(year)}$) are calculated at the highest levels of the hierarchy – the state. The hierarchical structure of the model is illustrated in Figure 2.

All parameters in a Bayesian model require a prior probability distribution, which is then updated by data. It is common practice to use minimally-informative prior distributions when little is known about the parameter. We used minimally-informative priors for parameters relating to pseudo-replication and confounding background variation ($\gamma_{(S)}$, $\delta_{(xs)}$, $\varepsilon_{(year)}$, and d). However, for the three parameters where we were particularly interested in quantifying their relationship with terrestrial vegetation cover ($g.v_0$, $g.m$, and β), we used the expert-derived Bayesian belief network (BBN) from de Little et al. (2012) to produce informative prior distributions. The expert-quantified BBN was used to

Figure 2. Hierarchical structure for the vegetation model. Refer to Methods section for definitions of the parameters.

create new data sets that corresponded to different states of parent variables in the BBN, and their consequent distributions of vegetation covers. These BBN-derived data were then run through the hierarchical Bayesian model described here with minimally-informative priors for all parameters, and the posterior distributions generated for $g.v_0$, $g.m$, and β were used as informative prior distributions for the final hierarchical model.
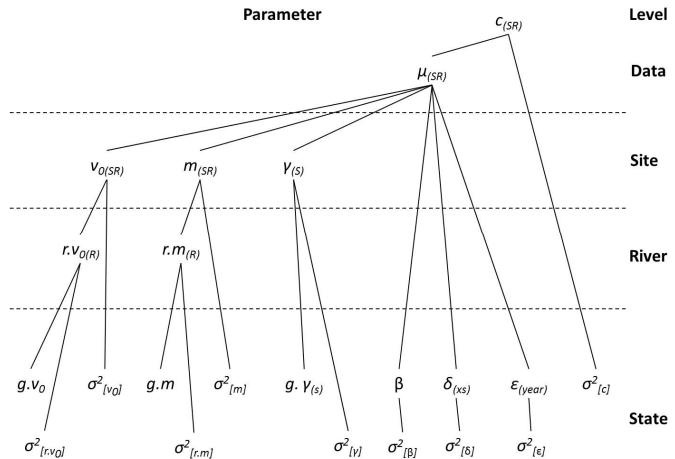
To test the effect of using a hierarchical approach, we also ran non-hierarchical versions of the models at the river level. This was done by removing the higher-level hyperparameters from the model, and instead assigning the expert-derived prior distributions to the parameters at the river level. All model fitting was done using the OpenBUGS software package (Lunn et al., 2009). We used 2 parallel Markov chains for model estimation. Each chain was burned in for 10000 iterations, with a further 20000 iterations for parameter estimation. The Brooks Gelman Rubin diagnostic (Brooks and Gelman 1998) was used to confirm convergence of the chains by the end of the burn-in. These operations resulted in a total sample size for parameter estimation of 40000.

3. RESULTS

We present here example results for several sites and scenarios (no inundation, 50 days continuous inundation in winter, 50 days continuous inundation in summer, 50 days inundation in summer divided into three episodes) that are representative of the findings for every site included in the analysis (27 sites in total).

The results show that moderate levels of inundation (50 days) will be effective at preventing vegetation encroachment for most sites (Figure 3a&d). Under zero inundation, all sites showed moderate to high cover of terrestrial vegetation within the channel. These predicted cover values decreased significantly, at some sites to almost zero cover, with 50 days of continuous inundation. Inundation in summer leads to slightly lower cover than inundation in winter at sites where initial cover is high (Figure 3c). Breaking the inundation into several distinct wetting events or episodes reduces its effectiveness (Figure 3a&b).

The use of expert-derived priors, rather than minimally-informed priors, allowed us to incorporate expert opinion, as well as field data, into our model, and thus reduced the level of uncertainty around the parameter estimates (e.g. $g.m$; Figure 4a) and model predictions (Figure 4c). Parameter values for each river differed somewhat between the hierarchical and

Figure 3 Predicted median percent cover of terrestrial vegetation in river channels under different inundation scenarios **(a) & (b)** show the difference between the predicted effect inundation delivered continuously or over 3 wetting episodes, **(c)** shows the predicted effect of 50 days of inundation delivered in summer versus winter, and **(d)** shows the predicted effect of 50 days of inundation on sites on four different rivers. The median values are plotted, along with the 95% credible interval for the parameter distribution.

independent sites analyses, with the median values being more consistent across rivers for the hierarchical model (e.g. $r.m$; Figure 4b). The uncertainty (95% credible intervals) of parameter values was substantially higher for all rivers in the independent rivers model, as was the uncertainty around predicted terrestrial cover values. This was particularly evident for the Broken and Macalister rivers (Figure 4b&d).
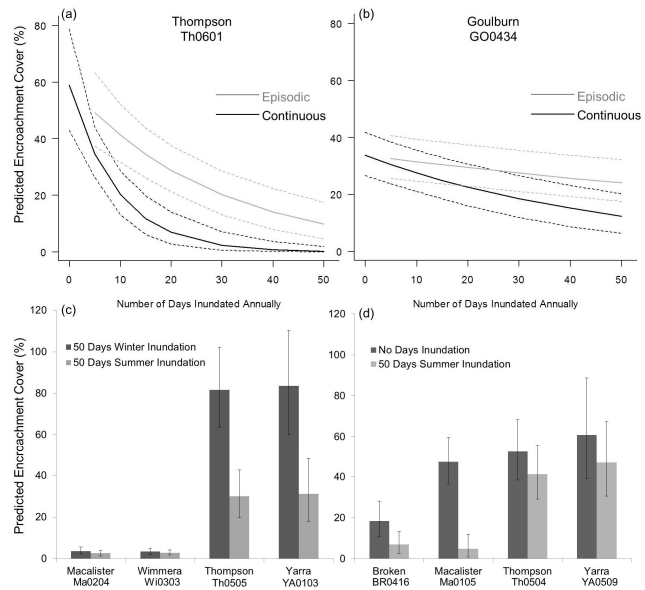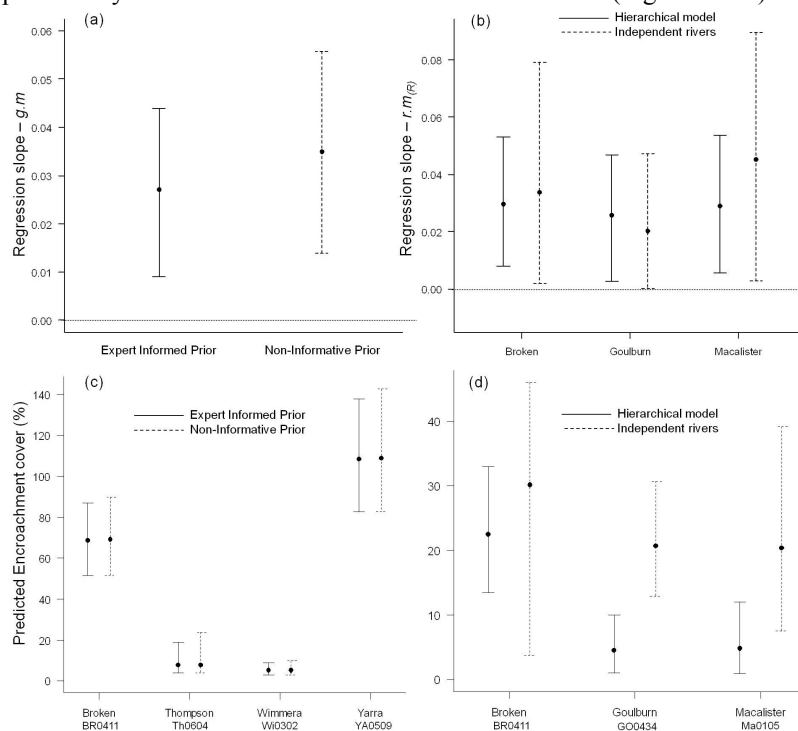

Figure 4 (a) shows an example of the parameter $g.m$ estimated with expert-informed and minimally informed priors, **(b)** $m_{(SR)}$ values for different rivers for hierarchical and non-hierarchical models, predicted terrestrial encroachment cover after 50 days continuous winter inundation **(c)** for expert-informed versus minimally informed prior models, and **(d)** hierarchical versus non-hierarchical models. The median values are plotted, along with the 95% credible interval for the parameter distribution.

4.   DISCUSSION AND CONCLUSIONS

In the example presented here, the findings of previous environmental flow assessments – that flow management can be used to reduce terrestrial vegetation encroachment – have been supported by the results of our Bayesian hierarchical analysis. However, modelling with a Bayesian context has allowed us to mathematically represent the physical processes inherent in this relationship: we are able to incorporate the negative exponential effect of inundation duration on terrestrial vegetation, the rate at which the frequency of inundation changes the 'concavity' of that relationship (Figure 1), and weight the effect of different years of inundation data according to their age. Studies that focus on BACI or similar 'flow delivery as treatment' based designs are not able to represent the flow-ecology relationships in this way. Rather, they simply test for associations within the data, as they lack the sophistication to accurately represent the complexity of the underlying physical processes. The flexibility of Bayesian modelling allows us to formulate physically-realistic models of response to flow, and these models can be updated as new knowledge and data become available via an iterative cycle of model development.

Modelling within a Bayesian framework also allowed us to incorporate expert-derived information, via the prior distribution in Bayes theorem, to improve the accuracy and precision of our estimates (Figure 4a&c). Restoration studies have notoriously 'patchy' datasets, and the ability to combine different sources of information can allow for greater inference in field data poor situations. Also in this vein, the hierarchical approach allowed us to incorporate information from a large number of sites (27) from 7 different river systems, which greatly improved the precision and accuracy of our predictions (Figure 4b). This concept of exchangeability means that conclusions will be greatly strengthened in the data-poor situations that are common in environmental flows studies.

The modelling process outlined in this paper can be used to make specific predictions (together with the uncertainties of those predictions) of ecological responses to different flow regimes. Particular to the hierarchical model structure, this approach allows us to report results at multiple scales (results not shown). While land owners may be interested in the relationship at the site scale, catchment management authorities are looking for quantified relationships to deliver flow components at a river scale, and finally the environmental flow funding body or water holder (e.g. The Victorian Department of Environment and Primary Industries) is interested in generalising such relationships to the entire state of Victoria. By quantifying the major model parameters (e.g. the effect of duration and frequency of inundation) at these three scales within a hierarchical model, we can make predictions of vegetation cover at all these scales. The other advantage of the Bayesian hierarchical modelling approach is that it provides a mathematically valid framework for making predictions at sites that have not been monitored. The river-level distributions of parameters can be sampled from to provide predictions for a new hypothetical site on that river, or the state-level distributions can be used to similarly make predictions for another river entirely. Such predictions use only the fitted high-level parameters (e.g. river level, state level) to make predictions, and so are less precise than those produced for sites where monitoring data have been collected. However, the predictions are non-biased, and provide a solution to the well-known problem in ecology of scaling up small-scale results to larger scales.

In conclusion, while processed-based Bayesian hierarchical models are a significantly more complex to build and understand than standard statistical analyses, the richness of information that comes from them makes the investment of funds into such effort worthwhile. This is exactly the type of general quantitative ecological response model that environmental flows science has been lacking. Such quantitative relationships have the potential to help move environmental flows into a new era of 'evidence-based' environmental management.

de Little *et al.,* Using Bayesian hierarchical models to measure the effectiveness of environmental flows

REFERENCES

Arthington, A.H., Naiman, R.J., McClain, M.E., Nilsson, C. (2010) Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities. *Freshwater Biology*, 55, 1-16.

Brooks, S.P., Gelman A. (1998) General methods for monitoring convergence of iterative solutions. *Journal of Computational and Graphical Statistics*, 7, 434-455.

Casanova, M.T., Brock, M.A. (2000) How do depth, duration and frequency of flooding influence the establishment of wetland plant communities? *Plant* Ecology, 147, 237-250.

Chee, Y., Webb, J.A., Stewardson, M.J., Cottingham, P. (2009) Victorian Environmental Flows Monitoring and Assessment Program: Monitoring and assessing environmental flow releases in the Thompson River. eWater Cooperative Research Centre, Available: http://tinyurl.com/VEFMAP-Thomson

Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, 8, 2–14.

Cottingham, P., Stewardson, M.J., Webb, J.A., (2005) Victorian Environmental Flows Monitoring and Assessment Program. Stage 1: Statewide Framework. Melbourne Water Corporation, Available: http://tinyurl.com/2feys3

de Little, S.C., Webb, J.A., Patulny, L., Miller, K.A., Stewardson, M.J. (2012) Novel methodology for detecting ecological responses to environmental flow regimes: using causal criteria analysis and expert elicitation to examine the effects of different flow regimes on terrestrial vegetation. 9th International Symposium on Ecohydraulics 2012 Proceedings. Available: http://www.ise2012.boku.ac.at/papers/13896_2.pdf

Gelman, A., Rubin, J.B., Stern, H.S., Rubin, D.B. (2004) Bayesian Data Analysis. Chapman & Hall ∕CRC, Boca Raton, FL.

Konrad, C.P., Warner, A., Higgins, J.V. (2012) Evaluating dam re-operation for freshwater conservation in the sustainable rivers project. *River Research and Applications*, 28, 777-792.

Le Quesne, T., Kendy, E., Weston, D. (2010) The Implementation Challenge: taking stock of government policies to protect and restore environmental flows, The Nature Conservancy & WWF.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. (2009) The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28, 3049-82.

McCarthy, M.A. (2007) Bayesian Methods for Ecology. Cambridge University Press, Cambridge, New York.

Meuller-Dombois, D., Ellenberg, H., (1974) Aims and Methods in Vegetation Ecology. John Wiley and Sons, New York.

Miller, K.A., Webb, J.A., de Little, S.C., Stewardson, M.J. (2013) Environmental flows can reduce the encroachment of terrestrial vegetation into river channels: a systematic literature review. *Environmental Management*, 1-11, 10.1007/s00267-013-0147-0

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C. (1997) The natural flow regime. *BioScience*, 47, 769-84.

Poff, N.L., Zimmerman, J.K.H. (2010) Ecological responses to altered flow regimes: a literature review to inform the science and management of regulated rivers. *Freshwater Biology*, 55, 194-205.

Stewardson, M.J., Webb, J.A. (2010) Modelling ecological responses to flow alteration: making the most of existing data and knowledge. Ecosystem Response Modelling in the Murray-Darling Basin, CSIRO Publishing, 37-49.

Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M. (2004) The need for evidence-based conservation. *Trends in Ecology and Evolution*, 19, 305-308.

Underwood, A.J. (1997) Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance. Cambridge University Press, New York.

VEWH, Authority G-HCM, Authority WCM (2011) Seasonal watering plan 2011-12, schedule 7: Wimmera-glenelg system. Victorian Environmental Water Holder, Victoria, Australia. Available at http://www.vewh.vic.gov.au/__data/assets/pdf_file/0018/121365/Schedule-7-Wimmera-Glenelg-system.pdf