SPECIAL REVIEW

Squeezing the most out of existing literature: a systematic re-analysis of published evidence on ecological responses to altered flows

J. ANGUS WEBB*, KIMBERLY A. MILLER†, ELISE L. KING*, SIOBHAN C. DE LITTLE†, MICHAEL J. STEWARDSON†, JULIE K. H. ZIMMERMAN‡ AND N. LEROY POFF§

*Department of Resource Management and Geography, The University of Melbourne, Parkville, Vic., Australia
†Department of Infrastructure Engineering, The University of Melbourne, Parkville, Vic., Australia
‡Bay-Delta Fish and Wildlife Office, U.S. Fish and Wildlife Service, Sacramento, CA, U.S.A.
§Department of Biology, Colorado State University, Fort Collins, CO, U.S.A.

SUMMARY

1. Human-induced changes in river flow regimes are ubiquitous worldwide. Although numerous case studies have identified negative ecological impacts of changes in different aspects of flow regimes (e.g. magnitude, timing), there have been few attempts to systematically review this literature to derive general relationships regarding ecological responses to changes in flow regimes.

2. Systematic literature reviews can inform science and management in ecologically complex systems not amenable to experimentation. However, such analysis of existing literature is often limited by inconsistent study design and data reporting. To attempt to overcome these difficulties, we used the recently developed Eco Evidence method and software to analyse 165 studies of ecological responses to changes in river flow regimes.

3. Eco Evidence provides a rule set and standardised list of terms to assist reviewers to interpret consistently the results of disparate studies. The companion software assists with the synthesis of this information to reach transparent and repeatable conclusions regarding cause–effect hypotheses of ecological responses to environmental drivers.

4. We compared our results to those of a recent, informal systematic review of the same studies, which is proving extremely influential. Stronger conclusions are reached when evidence is weighted, classified and combined according to the rules in Eco Evidence. Compared to the original review, we reached informative conclusions for a larger number of flow–response hypotheses, found that hypotheses for which the most evidence was available returned inconsistent results, addressed hypotheses at levels of conceptual resolution relevant to management and identified where insufficient evidence exists to reach a conclusion.

5. Analyses conducted at several levels of conceptual resolution found strong support for many hypotheses regarding ecological impacts. We found a consistent sensitivity to changes in flow regime for both fish and riparian vegetation across a variety of performance metrics. While macroinvertebrate responses varied among performance metrics (e.g. abundance was negatively affected by increases or decreases in flows, diversity was only negatively affected by flow decreases, and assemblage structure was affected by neither), they were largely consistent within these metrics.

6. We thus conclude that the Eco Evidence approach allowed us to extract more knowledge from the data set than was possible in the original review. Eco Evidence can improve synthesis of the burgeoning ecological literature and improve our general understanding in ecology. Amid widespread

Correspondence: J. Angus Webb, Building 379 (221 Bouverie Street Carlton), The University of Melbourne, Parkville, 3010, Vic., Australia.
E-mail: angus.webb@unimelb.edu.au

calls for 'evidence-based' environmental management, this powerful tool provides managers with a means of using research to help inform complex environmental decision-making.

Introduction

Human alteration of natural flow regimes (flow magnitude, frequency, duration, timing, rate of change; Poff *et al.*, 1997) is ubiquitous, and research on ecological effects of flow regime alteration is being published at an increasing rate globally (Stewardson & Webb, 2010). However, the observational and opportunistic nature of much of this research (Webb, Stewardson & Koster, 2010) makes it inferentially weak compared to manipulative experiments (Johnson, 2002). Thus, while general principles of flow alteration are well accepted, there has been little success in deriving specific predictions about how degraded biota will respond to flow restoration (Souchon *et al.*, 2008). Against this backdrop, Poff & Zimmerman (2010) synthesised the findings of 165 published studies, attempting to derive quantitative flow alteration–response relationships. They employed review methods commonly used in ecology, but also extended these by attempting both qualitative and quantitative systematic syntheses of the literature. However, their approach was informal, in that they did not employ a previously published *systematic review* method (e.g. CEBC, 2010). There were also no formal tests of the underlying hypotheses. They specifically noted the difficulty of synthesising a diverse set of studies that differed in experimental designs, modes of reporting, causes of flow alteration and species studied. Overall, their results supported those of earlier, less comprehensive, narrative literature reviews (Poff *et al.*, 1997; Bunn & Arthington, 2002; Lloyd *et al.*, 2003), but did not extend these earlier findings with quantitative results. In this study, we seek to extend the results of Poff & Zimmerman (2010) by applying a formal systematic review method.

Systematic reviews concisely summarise and synthesise the literature on a research question, providing insights beyond those possible with single studies. In contrast to the *narrative* reviews more common in ecology and environmental science, systematic reviews explicitly treat the literature as data and conduct analyses to test hypotheses (Khan *et al.*, 2003). Systematic reviews are common in several research fields that must deal with complex, multivariate cause–effect relationships, most notably medical research, where they are a key component of 'evidence-based medicine' (Pullin &

Knight, 2009). Recent calls to embrace evidence-based methods in environmental management (Sutherland *et al.*, 2004; Pullin, Knight & Watkinson, 2009) argue that systematic reviews should be used more widely and have led to the establishment of a foundation to promote their use (Pullin & Knight, 2009). Systematic reviews that supply environmental managers with the synthesised findings of research in an easy to understand form may be able to improve the input of science into policy development (Skinner *et al.*, 2012) and thus help managers to fulfil legislative requirements to use 'best available science' in policy development (Ryder *et al.*, 2010). However, they have yet to be widely adopted in environmental management (Pullin & Stewart, 2006).

The inferential strength of individual studies in environmental science is limited by an inability to randomise treatments, confounding environmental variables, and little or no replication. Systematic reviews of such studies need to take these limitations into account. Such problems of weak inference are also faced in epidemiological research. In response, epidemiologists developed 'causal criteria analysis', a formal method for combining multiple individually weak pieces of evidence to reach strong conclusions about cause–effect relationships (Hill, 1965; Susser, 1991; Tugwell & Haynes, 2006). The recently published Eco Evidence method for systematic review (Norris *et al.*, 2012) is modelled on causal criteria analysis and uses the literature, an underexploited source of evidence, for assessing questions of causality. The method provides a rule set that assists reviewers to interpret the results of individual studies, weighting them by the strength of their experimental design. This promotes consistency and repeatability when large numbers of studies are being reviewed. Evidence from multiple studies is then synthesised using a standardised framework to test individual cause–effect hypotheses and overall research questions. Hypotheses may be retained or may be falsified. The method is supported by freely available software (Webb *et al.*, 2011, 2012a), which automates the synthesis of evidence across studies, banks the evidence in a reusable online database and provides a standard report to maximise transparency of the process. To date, Eco Evidence has been successfully used in a number of topic-specific systematic reviews of river, wetland and floodplain environments (Harrison, 2010; Greet, Webb & Cousens, 2011; Grove

et al., 2012; Webb, Wallis & Stewardson, 2012b). Here, we use the Eco Evidence method and software to synthesise literature on ecological responses to human-altered river flow regimes, demonstrating its ability to synthesise the literature on a management issue of global significance (Dudgeon *et al.*, 2006; Arthington *et al.*, 2010). More specifically, we compare the Eco Evidence results to those of Poff & Zimmerman (2010; hereafter PZ2010) demonstrating the particular benefits and promise of this standardised approach compared to a more traditional literature review.

We used the Eco Evidence method for weighting and combining evidence to reanalyse the same 165 studies as PZ2010. We conducted four analyses that tested whether different features of Eco Evidence led to demonstrably stronger conclusions than the original review. While supporting the general findings of PZ2010, our analyses reached more definitive conclusions about the ecological effects of flow alteration. We were able to test hypotheses at scales directly relevant to management (i.e. effects of directional changes in flow components on specific taxa) and to identify where evidence was consistent, conflicting, or insufficient to reach a conclusion. We use this reanalysis not only as a synthesis of current knowledge in freshwater ecology, but as a case study to demonstrate the utility of the Eco Evidence approach to inform a range of complex environmental management issues.

Methods

The Eco Evidence framework

The Eco Evidence framework and supporting software have been described fully elsewhere (Nichols *et al.*, 2011; Webb *et al.*, 2011, 2012a; Norris *et al.*, 2012). Briefly, the framework (Norris *et al.*, 2012) consists of eight steps (Fig. 1), which may be grouped as:

Problem formulation (Steps 1–4, Step 6). The reviewer defines the overall question they are seeking to address (e.g. Do introduced fish species reduce native fish diversity?), the context within which the question is being asked (e.g. temperate lowland rivers of North America), and develops a conceptual model of the hypothesised cause–effect relationships within the overall question. The conceptual model and cause–effect hypotheses can be revised during the process.

Literature review and evidence extraction (Step 5). The hypothesised cause–effect relationships are used to guide a systematic literature search, and the reviewer extracts evidence from the relevant literature. Studies must report primary data; studies that cite other research do not contribute evidence. This avoids issues of double counting of some data (where one might include the primary study plus another study that cites it) and also guards against perpetuating any misinterpretation of an original study by a citing author. An 'evidence item' from a single study consists of a classification of hypothesised cause and effect, their trajectories (increase, decrease, change, no change), the nature of the association between them and a classification of the experimental design and replication.

Weighting evidence and judging causation (Steps 7–8). Each evidence item receives an evidence weight based on its experimental design and level of replication, with higher weights assigned to studies with greater replication and/or study designs that better control for confounding variables (Table 1). These weights are summed for all evidence in favour of the hypothesis and for all evidence that refutes the hypothesis. The two sums are compared to a threshold value (default value 20 points), resulting in one of four conclusions for that hypothesis. The default weights and threshold value were derived from an expert consultation process (Norris *et al.*, 2012) and can be changed by the user if there is justification for doing so (e.g. Grove *et al.*, 2012). 'Support for Hypothesis' (≥20 points in favour, <20 points against) implies that the causal hypothesis is corroborated by the evidence, but is not considered proved. 'Support for Alternate Hypothesis' (<20 in favour, ≥20 against) implies that the hypothesis has been falsified by the evidence and that a new hypothesis should be tested. 'Inconsistent Evidence' (≥20 in favour, ≥20 against) is another form of falsification that occurs when there is ample evidence both for and against the hypothesis. It may indicate that the hypothesised relationship exists only for a subset of the conditions or organisms assessed in the review. 'Insufficient Evidence' (<20 in favour, <20 against) implies that we cannot say anything about the validity of the hypothesis and may indicate a knowledge gap in the literature (Norris *et al.*, 2012). Lastly, the reviewer collectively considers the results of the individual cause–effect hypotheses to assess the level of support for the overall question developed at Step 1.

The approach is flexible, in that not all eight steps have to be followed rigidly. In this study, because we were seeking partly to compare the results of an Eco Evidence analysis to those of an existing review, we did not employ the full framework. The question and context (Steps 1–2), the collection of literature (part of Step 5) and to some extent the cause–effect hypotheses to be tested (Steps 3–4) had been previously defined by
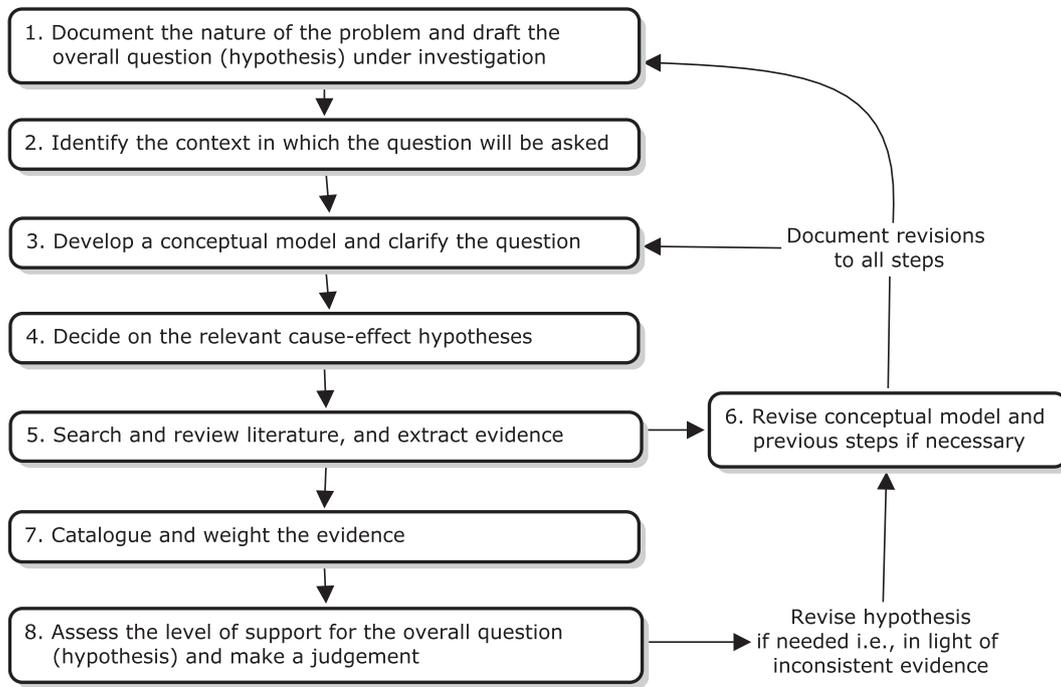
Fig. 1 The Eco Evidence framework. Reproduced from Norris *et al.* (2012).

Table 1 Default weights applied to study types and the number of control/reference and impact/treatment sampling units. Reprinted from Norris *et al.* (2012)

Study design component	Weight
Study design type	
After impact only	1
Reference/Control versus impact no before	2
Before versus after no reference/control	2
Gradient response model	3
BACI or BARI MBACI or Beyond MBACI	4
Replication of factorial designs	
Number of reference/control sampling units	
0	0
1	2
>1	3
Number of impact/treatment sampling units	
1	0
2	2
>2	3
Replication of gradient response models	
<4	0
4	2
5	4
>5	6

B, before; A, after; C, control; R, reference; I, impact; M, multiple. A study's overall evidence weight is the sum of design weight and replication weight/s.

PZ2010. This study focuses on the evidence extraction component of Step 5 and the weighting and synthesis of evidence to judge causation (Steps 7 and 8).

The Eco Evidence software

The Eco Evidence software was developed to facilitate systematic review using the framework. It automates the synthesis of evidence across studies, but not the extraction of evidence from individual studies, which must still be done by the reviewer. The software is freely available at www.toolkit.net.au/tools/eco-evidence. It consists of two components: the online Eco Evidence Database for storing and sharing evidence items (Webb *et al.*, 2012a), and the desktop Eco Evidence Analyser (Webb *et al.*, 2011).

The database uses a standardised list of terms for classifying causes and effects, each of which has a definition available to assist users. The definitions help users to classify causes and effects when study authors use inconsistent terms across different studies. They promote repeatability in the classification of cause and effect across many studies and among different reviewers. The list presently contains 229 items that cover the scope of applications to which Eco Evidence has already been applied (mostly aquatic ecology). It can be expanded further as new use cases arise. Entries in the list are categorised by a Term (an entity) and an Attribute (a property of the entity), structured as 'term (attribute)', for example 'fish (abundance)'. There are also term-only entries, for example 'fish'. Standard terms facilitate classification, search and retrieval of the evidence in the

database. Restricted field types (e.g. drop-down lists) are used to facilitate classification of study design, replication, and the nature of the association between hypothesised cause and effect. Free text fields are used to describe different components of the evidence item more fully. Evidence items entered into the database by one user are available for reuse by future users. The evidence used in this study (revised classification only) can be found by searching the question field for the tag 'P&Z'. Sharing evidence in this way reduces the amount of work required for a new systematic review.

The analysis tool uses a wizard-style interface (a sequence of dialogue boxes that lead the user through a series of well-defined steps) to guide users through the 8-step analysis framework, prompting the reviewer to input the question and context, conceptual model, literature search strategies they used, etc. It uses web services to connect to the Eco Evidence Database, and search for evidence relevant to the cause–effect hypotheses being tested, retrieving evidence entered separately by the reviewer and/or entered by other users. Following an assessment by the reviewer of the relevance or otherwise of the evidence items retrieved, it collates the evidence for and against the hypotheses and returns the conclusions. Lastly, it produces a standardised report that details the question asked, the process used to gather evidence and the conclusion reached. This maximises transparency and repeatability of the assessment.

Data sets

We extracted evidence from the 165 references provided in Appendix 1 of PZ2010, collating two data sets to classify cause and effect. The first data set used the original classification of PZ2010 for flow components (magnitude, frequency, duration, timing, rate of change, not specified) and organism groupings (aquatic, riparian). In the second data set, we used a revised classification of flow components based on the standard terms list from Eco Evidence (Table 2). We also classified ecological response at the greatest level of conceptual resolution possible using combinations of available taxonomic groups (e.g. fish, macroinvertebrates, etc.) and attributes of those groups (abundance, diversity, etc.). In the revised classifications, flow components were sometimes reassigned for one of two reasons: (i) Eco Evidence has eight flow component definitions compared to the six used by PZ2010 and (ii) we interpreted the flow component slightly differently to the original review. Any reassignments were based on a consensus view (Table S1). We also identified the additional information required

Table 2 Flow components from PZ2010 versus equivalent flow components from the Eco Evidence standard terms list. Several of the terms do not have an equivalent term in the opposing list. Four other non-flow related cause terms were also used in the Eco Evidence classification of studies (Table 4)

| PZ2010	Eco Evidence
Magnitude	Surface water (volume)
Frequency	Surface water (frequency)
Duration	Surface water (duration)
Timing	Surface water (seasonality)
Rate of change	–
–	Surface water (area)
–	Surface water (depth)
–	Surface water (velocity)
–	Flow regulation (dam)
Not specified	Flow regulation

(experimental design and replication) to weight the evidence and assess hypotheses. Two other features of the two data sets differed from PZ2010. First, Eco Evidence requires primary data; that is, papers that cite other research – 21 papers from PZ2010 – do not contribute evidence. Second, Eco Evidence allows one to record a lack of association between a putative cause and effect, whereas PZ2010 recorded all papers as either showing a positive or negative association (12 papers; Table S1).

Results

Question 1: Does an algorithm for weighting and combining evidence lead to stronger conclusions than counting papers?

We used the original classification (data set 1) to test very broad hypotheses relating aquatic or riparian organism responses to overall changes (increases and decreases pooled) in flow regime components (Fig. 2). We compared the Eco Evidence results to the counts of papers presented in Table 1 of PZ2010.

We found support for the hypotheses that changes in flow peak duration and timing negatively affect riparian organisms. The hypotheses for which we found most evidence (changes in magnitude negatively affect aquatic and riparian organisms; changes in frequency negatively affect aquatic organisms) returned findings of Inconsistent Evidence. There was insufficient evidence to reach conclusions for the majority (6 of 11) of the hypotheses (Table 3). The summed evidence weights for the individual hypotheses broadly reflected the numbers of papers originally classified by PZ2010 as showing negative and positive ecological impacts (Table 3), but the conclusions differed. Our findings of Inconsistent Evidence for aqua-
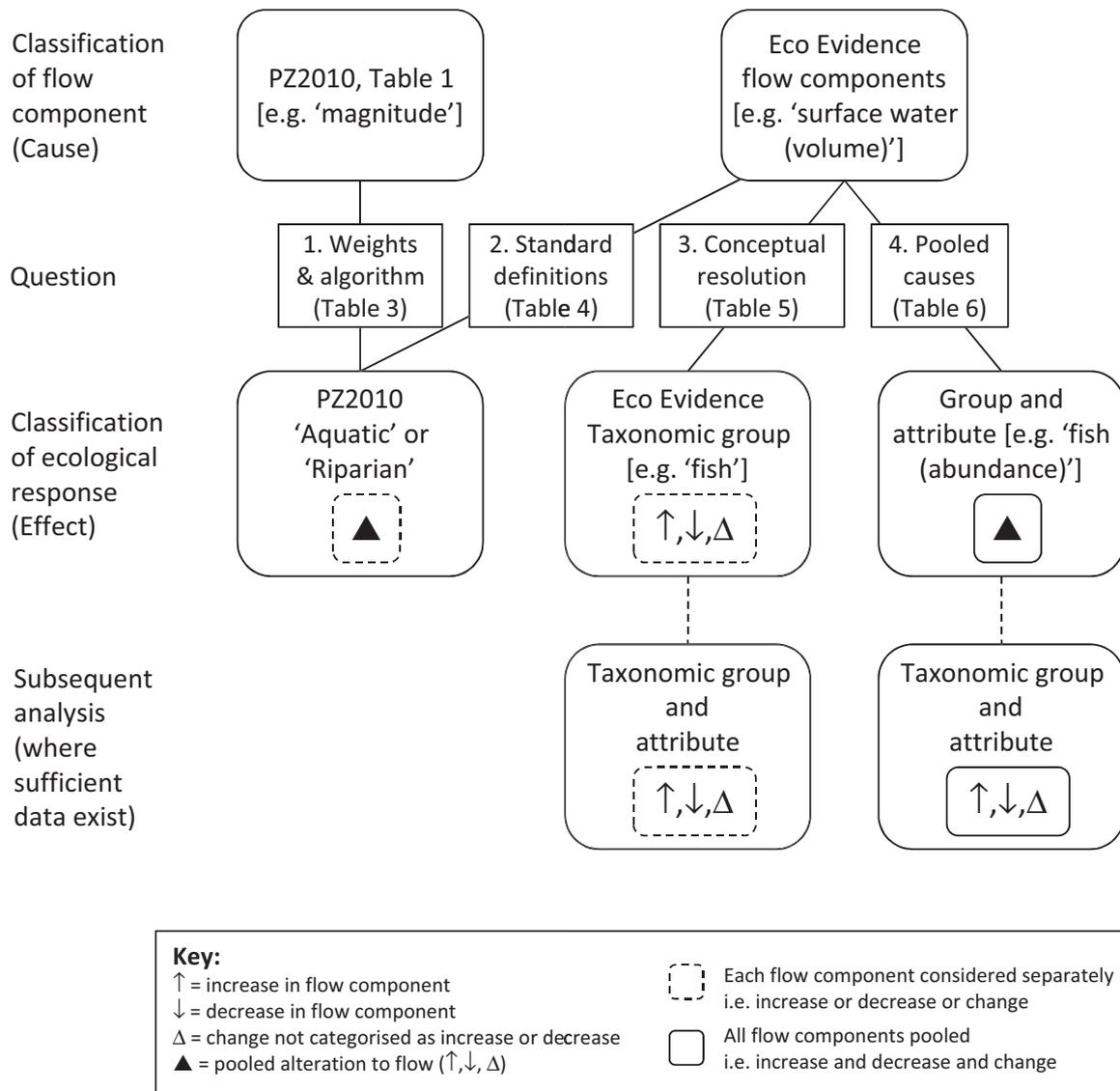
Fig. 2 Schematic representation of the four separate Eco Evidence analyses (see Methods). Solid lines linking flow component boxes to ecological response boxes show the combination of cause and effect classification used to address the question intercepted by that line.

tic species contrast PZ2010's classification of over 90% of papers as showing a negative impact. Of the two supported hypotheses reported above, the former reflects the original classification of PZ2010. The latter was not reported by PZ2010, but because the summed evidence weight only just reaches the threshold value of 20 points, the result should be interpreted with caution. Our findings of Insufficient Evidence included the hypothesis that changes in flow timing impact aquatic organisms, despite PZ2010 classifying all 12 relevant studies as in favour of that hypothesis. In summary, while PZ2010 were only able to use counts of papers to gain an indication of the level of support for the hypotheses in Table 3, Eco Evidence allowed us determine

which hypotheses were supported, which were falsified, and which had insufficient evidence to reach a conclusion.

Question 2: Do standardised definitions of flow components lead to stronger conclusions?

We used the revised classification (data set 2) to specify change in the flow components, but retained the broad organism groupings (aquatic, riparian) and compared the findings to those from the Eco Evidence analysis for Question 1 (Fig. 2).

There were two noticeable effects of reclassifying the flow components (Table 4). First, we found support for

Table 3 Results for Question 1

Flow component	Organism group	PZ2010 count of papers		Eco Evidence analysis	Summed evidence weights		
		Negative impact	Positive impact	# evidence items	In favour	Against	Conclusion
Magnitude	Aquatic	66	5	71	246	88	Inconsistent
	Riparian	24	7	29	63	46	Inconsistent
Frequency	Aquatic	12	1	12	48	29	Inconsistent
	Riparian	4	2	5	12	15	Insufficient
Duration	Aquatic	7	1	7	9	14	Insufficient
	Riparian	17	1	18	48	16	Support
Timing	Aquatic	12	0	12	18	19	Insufficient
	Riparian	4	0	4	20	0	Support
Rate of change	Aquatic	2	2	3	6	0	Insufficient
	Riparian	2	0	2	8	0	Insufficient
Not specified	Aquatic	2	2	4	11	13	Insufficient

Inconsistent, Inconsistent Evidence; Insufficient, Insufficient Evidence; Support, Support for Hypothesis; Alternate, Support for Alternate Hypothesis (Norris *et al.*, 2012); # evidence items, number of evidence items assessed; In favour, evidence weight in favour of research hypothesis; Against, evidence weight not in favour of hypothesis.
These abbreviations are also used in Tables 4–6. See methods for the rules for calculating evidence weights and combining evidence to reach conclusions.

Table 4 Results for Question 2. Where applicable, the equivalent flow component from PZ2010 is shown

Flow Component		Organism group	# evidence items	Summed weights		
Eco Evidence	PZ2010			In favour	Against	Conclusion
Surface water (volume)	Magnitude	Aquatic	52	177	68	Inconsistent
		Riparian	16	29	19	Support
Surface water (frequency)	Frequency	Aquatic	22	71	39	Inconsistent
		Riparian	9	12	31	Alternate
Surface water (duration)	Duration	Aquatic	5	17	13	Insufficient
		Riparian	8	22	5	Support
Surface water (seasonality)	Timing	Aquatic	16	37	12	Support
		Riparian	13	45	8	Support
Flow regulation	Not specified	Aquatic	2	0	8	Insufficient
		Riparian	2	7	0	Insufficient
Flow regulation (dam)		Aquatic	8	25	21	Inconsistent
		Riparian	3	9	4	Insufficient
Surface water (area)		Riparian	1	8	0	Insufficient
Surface water (depth)		Riparian	1	0	4	Insufficient
Surface water (velocity)		Aquatic	3	8	7	Insufficient
Ground water (depth)		Riparian	6	21	5	Support
Physical habitat		Aquatic	2	18	0	Insufficient
Water quality (temperature)		Aquatic	1	0	2	Insufficient
Water quality (turbidity)		Riparian	1	0	3	Insufficient

twice as many (four) of the original hypotheses (changes in surface water volume and frequency affect riparian organisms, and changes in seasonality affect both aquatic and riparian organisms). In contrast, we found insufficient evidence for only three of the nine hypotheses that had a direct analogue in Question 1. We again found inconsistent evidence for the hypotheses with the greatest number of evidence items (changes in surface water volume and frequency affect aquatic organisms). We also rejected the hypothesis that changes in inundation frequency negatively impact upon riparian organisms. Second, we classified 26 of the studies with causes that had no direct equivalent in the original data set. For one of these classifications, we found sufficient evidence

to support the hypothesis that change in groundwater depth negatively affects riparian organisms. In summary, compared to the original classification of flow components used in PZ2010, we found that the standardised definitions in Eco Evidence led to a higher proportion of informative conclusions.

Question 3: Does greater conceptual resolution in the review lead to stronger conclusions or dilute the evidence pool?

We used the revised classification of both causes and effects to test whether changes in individual flow components cause ecological impacts for individual taxonomic groups (fish, macroinvertebrates, vegetation; Question 3a). We also separated increase, decrease and non-directional change in each flow component. Where sufficient evidence existed, we tested for effects on the attributes of the taxonomic groups (e.g. 'macroinvertebrate (abundance)'; Question 3b; Fig. 2). This analysis is most comparable to the attempts by PZ2010 to derive quantitative relationships between taxonomic groups and changes in flow volume.

Posing hypotheses at this finer level of conceptual resolution resulted in far less evidence being available to test individual hypotheses compared to Questions 1 and 2. We found evidence for 36 cause–effect hypotheses when considering effects at the scale of taxonomic group (i.e. not considering attributes; Table S2). However, the evidence was sufficient to reach a conclusion for only nine of these. At the finer scale of taxonomic group attribute, there was sufficient evidence for a further four hypotheses (Table 5). For fish, we found negative effects

of a decrease in volume and rejected the hypothesis of a negative effect of increased volume. Decreased volume also negatively affected fish assemblage composition. There were negative effects of any change in flow peak frequency on fish. Macroinvertebrates were negatively impacted by increased flow volume, but there were inconsistent effects of a decrease. Macroinvertebrate abundance was negatively impacted by any change in flow volume. We rejected the hypothesis that increased flow peak frequency has negative impacts upon macroinvertebrates, with this conclusion extending to assemblage diversity. Lastly, we found vegetation to be impacted negatively by both decreased flow seasonality and groundwater depth (i.e. ground water being further below the surface). In summary, while increasing the level of conceptual resolution did dilute the evidence for any single hypothesis, there was still sufficient evidence to reach conclusions for a large number of specific flow–ecology relationships. This contrasts the attempts by PZ2010 to derive relationships for taxonomic groups, which only produced a clear qualitative result for fish sensitivity to changes in flow magnitude.

Question 4: Does pooling individual causes result in stronger conclusions?

Using the revised classification, we pooled all flow components into changes in 'surface water' (the parent term for river flow in Eco Evidence) and tested for effects on ecological responses (individual combinations of taxonomic group and attribute, Question 4a). Where evidence was sufficient, we tested whether individual trajectories (increase, decrease, non-directional change)

Table 5 Results for Question 3. Analysed attributes of taxonomic groups are shown in brackets. Cause–effect hypotheses for which there was insufficient evidence to reach a conclusion are reported in Table S2

Flow component				Summed weights		
Name	Trajectory	Taxonomic group	# evidence items	In favour	Against	Conclusion
Surface water (volume)	Increase	Fish	11	19	26	Alternate
		Macroinvertebrates	8	32	12	Support
		(abundance)	5	21	4	Support
	Decrease	Fish	15	87	0	Support
		(assemblage)	9	56	0	Support
		Macroinvertebrates	15	50	29	Inconsistent
		(abundance)	7	30	7	Support
Surface water (frequency)	Increase	Fish	7	25	4	Support
		Macroinvertebrates	9	19	28	Alternate
		(diversity)	4	7	24	Alternate
	Decrease	Fish	5	22	0	Support
Surface water (seasonality)	Decrease	Vegetation	9	33	0	Support
Ground water (depth)	Decrease	Vegetation	6	21	5	Support

in surface water were plausible causes of a decrease in the ecological response (Question 4b; Fig. 2). This analysis was designed to test whether a focus on individual flow components could be obscuring results for broadly sensitive ecological responses.

We found evidence regarding the ecological impact of any change in surface water (i.e. not considering trajectory) for 37 detailed ecological responses (Table S3), but only nine of these had sufficient evidence to reach a conclusion. At the finer scale of a specific trajectory of change in surface water, there was sufficient evidence to reach a conclusion for a further 11 hypotheses (Table 6). We rejected the hypotheses that pooled changes or reductions in surface water impact negatively upon in-stream algae. We found consistent negative effects of pooled changes in surface water on fish abundance, assemblage composition and diversity. Fish assemblage structure was also affected by both increased and decreased surface water, and abundance and diversity were impacted by decreased surface water. For macroinvertebrates, we found more variable results, with negative impacts of either increased or decreased surface water on abundance and also of decreased surface water on diversity. However, we rejected the hypotheses that any changes in surface water impacted assemblage structure. We also found inconsistent effects of pooled changes in surface water on abundance and diversity, but only one of these hypotheses remained inconsistent when trajectory of surface water change was taken into account (effects of an increase on diversity). For vegetation, we found negative effects of pooled changes in surface water on diversity and mortality, with mortality also responding to decreased flow. Overall, there were four responses (fish diversity, macroinvertebrate assemblage structure, vegetation diversity and mortality) for which informative conclusions were reached in this analysis, but for which there had been insufficient evidence to reach any conclusions for individual flow components (Question 3). Thus, pooling flow components into surface water did identify sensitive ecological responses that were obscured by a focus on flow components, a focus that was also employed in PZ210.

Discussion

Our analyses reveal substantial evidence for ecological responses to human flow alteration in the published literature, supporting the original conclusions of PZ1010. However, the conclusions reached using the structured approach in Eco Evidence were more decisive and extensive than those reached in the earlier review. The additional insight gained can be attributed to the use of standardised terms and consistency in classification of evidence, which enables testing (and falsification) of

Table 6 Results for Question 4. Analysed attributes of taxonomic groups are shown in brackets. Cause–effect hypotheses for which there was insufficient evidence to reach a conclusion are reported in Table S3)

Effect	Flow Regime Trajectory	# evidence items	Summed evidence weights		Conclusion
			In favour	Against	
Algae	All	8	19	31	Alternate
	Decrease	6	19	23	Alternate
Fish (abundance)	All	13	48	2	Support
	Decrease	8	35	0	Support
Fish (assemblage)	All	17	102	7	Support
	Increase	6	37	7	Support
	Decrease	11	65	0	Support
Fish (diversity)	All	10	45	15	Support
Macroinvertebrates (abundance)	All	23	79	29	Inconsistent
	Increase	8	33	4	Support
	Decrease	12	46	6	Support
Macroinvertebrates (assemblage)	All	12	2	55	Alternate
	Increase	4	0	25	Alternate
	Decrease	7	2	26	Alternate
Macroinvertebrates (diversity)	All	15	55	38	Inconsistent
	Increase	9	29	33	Inconsistent
	Decrease	4	26	2	Support
Vegetation (diversity)	All	5	26	3	Support
Vegetation (mortality)	All	6	23	0	Support
	Decrease	5	20	0	Support

hypotheses at multiple levels of conceptual resolution. Eco Evidence also identifies when evidence is insufficient to reach a conclusion. We believe that the Eco Evidence approach to systematic literature review can improve synthesis of the burgeoning ecological literature, improving our general understanding in ecology and our consequent ability to manage complex issues in natural and modified environments.

Ecological impacts of flow alteration

The analysis revealed consistent evidence that fish are sensitive to changes in flow regime. When flow regime changes were pooled (Question 4), fish abundance, assemblage structure and diversity were all negatively affected by both increases and decreases in flow regime components. Pooled fish responses were also negatively affected by reductions in discharge and by both increases and decreases in frequency of high-flow events. Increases in discharge volume did not have a negative effect (Question 3), but there was considerable evidence both in favour of and against this hypothesis.

There was consistent evidence that flow alteration has a negative effect on vegetation, but sufficient evidence was not available to address as many hypotheses as for fish. Pooled flow regime changes negatively affected diversity and mortality, results that were matched by impacts of decreased flow seasonality and groundwater depth on pooled vegetation responses. Vegetation responses also made up the great majority of studies (50 of 62; Table S1) in the pooled riparian group (Questions 1 and 2). In those analyses, this group was negatively affected by changes in discharge volume and flow event duration.

Responses for macroinvertebrates were more complex, but clear results emerged for some analyses. Macroinvertebrate abundance was affected by both increases and decreases in discharge volume and pooled flow components. In contrast, macroinvertebrate assemblage structure and diversity were not consistently negatively affected by changes in either pooled or individual flow components. Assemblage structure, in particular, showed clear 'non-negative' (i.e. no change or a positive change) effects in response to changes in pooled flow components. Overall, this meant that pooled macroinvertebrate responses did not show consistent effects of changes in flow regime components.

Because we used the same literature as PZ2010, our results cannot be considered as an entirely up-to-date assessment of the state of knowledge. New research on ecological responses to changes in flow regime is appearing in the literature at a rapid rate (Stewardson &

Webb, 2010), and an updated review of such literature could find more evidence for the detailed hypotheses posed by Questions 3 and 4. However, other applications of Eco Evidence have shown that findings of Insufficient Evidence are relatively common despite exhaustive searches of the literature (e.g. Webb *et al.*, 2012b), indicating knowledge gaps requiring new research. Thus, Eco Evidence can provide a useful synthesis of existing literature, but also identify knowledge gaps when designing research programs (e.g. Greet *et al.*, 2011).

Improvements over PZ2010: benefits of the Eco Evidence structured approach to systematic reviews

In this study, we deliberately employed the same set of papers as PZ2010 because we wanted to assess whether the Eco Evidence approach could extract extra value from this data set compared to the original review. We contend that the structured approach indeed did lead to stronger conclusions, and a deeper, more nuanced, synthesis of general ecological responses to changes in flow regimes than PZ2010. We recognise five key reasons for this improved performance.

PZ2010 specifically noted the difficulty of synthesising sets of studies that employ different designs and models of reporting; this problem is common to many reviews in ecology. The weighting scheme in Eco Evidence allows one to combine results from studies of different designs by weighting the results according to the strength of the study. Studies that better control for potentially confounding influences (i.e. those that are less likely to reach spurious conclusions) get a higher weight. The issue of different models of reporting is also largely overcome because Eco Evidence does not require the paper to include specific information about the result (e.g. effect size); the statement of the result is enough for inclusion in an analysis. The step of refining hypotheses in the face of inconsistent evidence (see below) can be used to address inconsistencies among groups of studies examining different ultimate causes of environmental stress and/or different taxonomic groups. However, if studies from different regions on different organisms all lead to the same conclusion in an Eco Evidence analysis, then this is exactly the sort of general result that is of greatest value to research and management.

The Eco Evidence algorithm for weighting evidence and comparing summed evidence weights produced more informative results than the simple counting of papers undertaken by PZ2010. In particular, we were able to falsify hypotheses, a basic requirement of the

conjecture–refutation model of scientific progress under which nearly all science operates (Popper, 1983), but which is rarely used in literature reviews (Question 1). The hypotheses for which most evidence was available returned conclusions of Inconsistent Evidence, indicating that a substantial proportion of the evidence did not support the hypothesis. Normally, a finding of Inconsistent Evidence prompts the reviewer to refine the hypothesis (Norris *et al.*, 2012), which may identify the particular circumstances under which, for example, a change in flow magnitude negatively affects aquatic organisms. Conversely, a conclusion of Insufficient Evidence indicates that further literature review and/or field-based research is necessary before we can retain or falsify the hypothesis. We found insufficient evidence for the majority of hypotheses from PZ2010, but the original review did not recognise this weakness, only explicitly commenting on the low number of studies examining 'rate of change' in flow regime.

The standard terms list in Eco Evidence assists reviewers to classify consistently studies that do not use consistent terms to classify flow components or ecological responses. Reclassifying the studies using the standard terms list in Eco Evidence allowed us to reach conclusions (i.e. results other than Insufficient Evidence) for more hypotheses than with the original classification (7/9 versus 5/11; Questions 1 & 2). The larger number of 'informative' conclusions (i.e. Support for Hypothesis, Support for Alternate Hypothesis) suggests that more consistent results were being grouped together under the reclassified flow components. The original review did not have the benefit of standardised definitions of flow components, and papers were classified by more than one individual over a lengthy period (N.L. Poff and J.K.H. Zimmerman, unpubl. data). Therefore, it is unsurprising that some differences in classifications arose. The set step-by-step discipline of Eco Evidence promotes consistency of interpretation among studies and where there are multiple reviewers or reviews are undertaken over a lengthy period. This discipline thus promotes reproducibility of findings.

Partitioning responses (effects) and drivers (causes) allowed us to reach conclusions for hypotheses that are more likely to be useful for management. Environmental flow recommendations often make specific directional predictions concerning the effect of changes in flow components on individual taxonomic groups and their attributes. Therefore, hypotheses at this level of conceptual resolution are more likely to be able to inform management of natural environments. Classifying studies in greater detail, both in terms of cause (by separating out

trajectories of flow components) and effect (by classifying at the level of taxonomic group or finer; Question 3) inevitably reduced the evidence available to test any single hypothesis. However, more specific hypotheses should result in a greater proportion of informative results when sufficient evidence exists to reach a conclusion, and this is exactly what was found (Table 5). Of the 13 conclusive findings, only one was of Inconsistent Evidence. Testing hypotheses at this level of conceptual resolution thus resolved some of the apparent inconsistencies identified by PZ2010. For example, while macroinvertebrates showed an inconsistent response to reduced flow volume, macroinvertebrate *abundance* showed a consistent reduction.

Pooling the flow components into changes in overall 'surface water' and concentrating on detailed ecological effects (Question 4) largely confirmed the results from Question 3 but also identified several further sensitive ecological responses. Examining the data in this fashion is possible with any evidence assessment, but the simple two-layer hierarchical structure of the standard terms list in Eco Evidence – that is Term (Attribute) – makes such an analysis a simple extension when many detailed cause–effect hypotheses have insufficient evidence to reach a conclusion. Therefore, while a finding of Inconsistent Evidence prompts a reviewer to pose hypotheses at a finer scale of conceptual resolution, multiple findings of Insufficient Evidence prompt the reviewer to pose fewer hypotheses at coarser scales (e.g. Grove *et al.*, 2012).

Our analyses do not achieve the original goal of PZ2010, to generate quantitative flow–ecology relationships from the published literature. However, Eco Evidence was not designed to do this. Quantitative meta-analysis (Gurevitch & Hedges, 2001) would be suited to this task, but is a more difficult analysis task. Meta-analysis also requires the reporting of summary statistics, which means that only a subset of relevant studies can typically be included in an analysis (Bekkering *et al.*, 2008). The Eco Evidence analysis provided us with far stronger qualitative conclusions than PZ2010, and it is tempting to speculate that the strongest results from the Eco Evidence analysis might suggest fruitful areas in which to search for quantitative relationships. However, such an analysis would require a new and expanded search of the literature to expand the collection of studies with the necessary quantitative results.

Our analyses show that the Eco Evidence method for systematic literature review can assess ecological issues of global importance across multiple environment types and taxonomic groups. Here, we found strong evidence

for ecological impacts of flow regimes across a large set of studies collected from across the world and many different types of river systems. Conclusions were stronger than those reached by the original study of Poff & Zimmerman (2010), which is proving extremely influential (257 citations Google Scholar, 131 citations Web of Science, August 1, 2013). In the move towards evidence-based environmental management, improved synthesis of the rapidly expanding scientific literature is of great importance (Attwood *et al.*, 2009). Eco Evidence provides a method for reaching strong conclusions for cause–effect questions relevant to complex environmental management problems.

Acknowledgments

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.

Attwood T.K., Kell D.B., McDermott P., Marsh J., Pettifer S.R. & Thorne D. (2009) Calling International Rescue: knowledge lost in literature and data landslide! *Biochemical Journal*, **424**, 317–333.

Bekkering G.E., Harris R.J., Thomas S., Mayer A.M.B., Beynon R., Ness A.R. *et al.* (2008) How much of the data published in observational studies of the association between diet and prostate or bladder cancer is usable for meta-analysis? *American Journal of Epidemiology*, **167**, 1017–1026.

Bunn S.E. & Arthington A.H. (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management*, **30**, 492–507.

CEBC (2010) *Guidelines for Systematic Review in Environmental Management*. Centre for Evidence-Based Conservation & Collaboration for Environmental Evidence, Bangor, Wales.

Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler D.J., Lévêque C. *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, **81**, 163–182.

Greet J., Webb J.A. & Cousens R.D. (2011) The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis. *Freshwater Biology*, **56**, 1231–1247.

Grove J.R., Webb J.A., Marren P.M., Stewardson M.J. & Wealands S.R. (2012) High and dry: an investigation using the causal criteria methodology to investigate the effects of regulation, and subsequent environmental flows, on floodplain geomorphology. *Wetlands*, **32**, 215–224.

Gurevitch J. & Hedges L.V. (2001) Meta-analysis: combining the results of independent experiments. In: *Design and Analysis of Ecological Experiments* (Eds S.M. Scheiner & J. Gurevitch), pp. 346–369. Oxford University Press, New York.

Harrison E.T. (2010) *Fine Sediment in Rivers: Scale of Ecological Outcomes*. PhD Dissertation, University of Canberra, Canberra, Australia.

Hill A.B. (1965) The environment and disease – association or causation? *Proceedings of the Royal Society of Medicine-London*, **58**, 295–300.

Johnson D.H. (2002) The importance of replication in wildlife research. *The Journal of Wildlife Management*, **66**, 919–932.

Khan K.S., Kunz R., Kleijnen J. & Antes G. (2003) Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, **96**, 118–121.

Lloyd N., Quinn G., Thoms M., Arthington A., Gawne B., Humphries P. *et al.* (2003) *Does Flow Modification Cause Geomorphological and Ecological Response in Rivers? A Literature Review from an Australian Perspective*. CRC for Freshwater Ecology, Canberra, Australia.

Nichols S., Webb A., Norris R. & Stewardson M. (2011) *Eco Evidence Analysis Methods Manual: A Systematic Approach to Evaluate Causality in Environmental Science*. eWater Cooperative Research Centre, Canberra.

Norris R.H., Webb J.A., Nichols S.J., Stewardson M.J. & Harrison E.T. (2012) Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. *Freshwater Science*, **31**, 5–21.

Poff N.L., Allan J.D., Bain M.B., Karr J.R., Prestegaard K.L., Richter B.D. *et al.* (1997) The natural flow regime. *BioScience*, **47**, 769–784.

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.

Popper K.R. (1983) *Realism and the Aim of Science: Postscript to the Logic of Scientific Discovery*. Hutchinson, London.

Pullin A.S. & Knight T.M. (2009) Doing more good than harm – building an evidence-base for conservation and environmental management. *Biological Conservation*, **142**, 931–934.

Pullin A.S., Knight T.M. & Watkinson A.R. (2009) Linking reductionist science and holistic policy using systematic reviews: unpacking environmental policy questions to construct an evidence-based framework. *Journal of Applied Ecology*, **46**, 970–975.

Pullin A.S. & Stewart G.B. (2006) Guidelines for systematic review in conservation and environmental management. *Conservation Biology*, **20**, 1647–1656.

Ryder D.S., Tomlinson M., Gawne B. & Likens G.E. (2010) Defining and using 'best available science': a policy

conundrum for the management of aquatic ecosystems. *Marine and Freshwater Research*, **61**, 821–828.

Skinner D.S., Webb J.A., Nichols S.J. & Stewardson M.J. (2012) Eco Evidence for systematic causal inference and knowledge transfer between science and policy. In: *Climate Change 2012: Water and Climate: Policy Implementation Changes*; Proceedings of the 2nd Practical Responses to Climate Change Conference. pp. 647–654. Engineers Australia.

Souchon Y., Sabaton C., Deibel R., Reiser D., Kershner J., Gard M. *et al.* (2008) Detecting biological responses to flow management: Missed opportunities; Future directions. *River Research and Applications*, **24**, 506–518.

Stewardson M.J. & Webb J.A. (2010) Modelling ecological responses to flow alteration: making the most of existing data and knowledge. In: *Ecosystem Response Modelling in the Murray-Darling Basin* (Eds N. Saintilan & I. Overton), pp. 37–49. CSIRO Publishing, Melbourne, Australia.

Susser M. (1991) What is a cause and how do we know one – a grammar for pragmatic epidemiology. *American Journal of Epidemiology*, **133**, 635–648.

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

Tugwell B. & Haynes R.B. (2006) Assessing claims of causation. In: *Clinical Epidemiology: How to do Clinical Practice Research* (Eds R.B. Haynes, D.L. Sackett, G.H. Guyatt & P.Tugwell), pp. 356–387. Lippincott Williams & Wilkins, Philadelphia.

Webb J.A., de Little S.C., Miller K.A. & Stewardson M.J. (2012a) Eco Evidence Database: a distributed modelling resource for systematic literature analysis in environmental science and management. In: *2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting* (Eds R.Seppelt, A.A. Voinov, S. Lange & D. Bankamp). International Environmental Modelling and Software Society (iEMSs).

Webb J.A., Stewardson M.J. & Koster W.M. (2010) Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biology*, **55**, 108–126.

Webb J.A., Wallis E.M. & Stewardson M.J. (2012b) A systematic review of published evidence linking wetland plants to water regime components. *Aquatic Botany*, **103**, 1–14.

Webb J.A., Wealands S.R., Lea P., Nichols S.J., de Little S.C., Stewardson M.J. *et al.* (2011) Eco Evidence: using the scientific literature to inform evidence-based decision making in environmental management. In: *Proceedings of MODSIM 2011 – International Congress on Modelling and Simulation* (Eds F. Chan, D. Marinova & R.S. Anderssen), pp. 2472–2478. Modelling and Simulation Society of Australia and New Zealand.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Comparison of the 'original data' gathered by Poff & Zimmerman (2010) and our 'reclassified data'.

Table S2 Complete results for Question 3 and 3a, using the standard terms in Eco Evidence for both the cause (flow components) and effects (taxonomic groups).

Table S3 Complete results for Question 4 and 4a, identifying the most sensitive effect (taxonomic group and attribute) to any change in flow regime (i.e. all Surface Water flow components pooled).

(Manuscript accepted 1 August 2013)