International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Eco Evidence Database: a distributed modelling resource for systematic literature analysis in environmental science and management

J. Angus Webb[1], Siobhan C. de Little[2], Kimberly A. Miller[2], Michael J. Stewardson[2]
[1]*Department of Resource Management & Geography, The University of Melbourne, Australia*
[2]*Department of Infrastructure Engineering, The University of Melbourne, Australia*
siobhan.delittle@unimelb.edu.au

Abstract: Balancing the needs of the environment with human uses of natural ecosystems relies on understanding causal relationships between environmental stressors and ecological responses. Causal relationships are difficult to demonstrate in natural environments due to environmental heterogeneity, lack of replication, and confounding influences. Published scientific literature is a currently under-used source of evidence for such analyses. We describe the Eco Evidence Database – a web database application for storing and sharing the atomized findings of research studies from the literature. This 'evidence' can then be used in synthesis studies to test cause-effect hypotheses. The database provides a permanent, machine-independent repository for causal evidence. Moreover, it allows users to access evidence items entered by previous users, thereby reducing the burden of extracting evidence from the literature. The database provides quick and open access to the summarized findings of research to researchers and managers. It employs standard lists of causes and effects to avoid semantic uncertainty between researchers and to facilitate searching. The Eco Evidence database supplies evidence to a desktop analysis tool – the Eco Evidence Analyser, which synthesizes the evidence using causal criteria analysis. However, the database also contains fields not currently used by the Eco Evidence method, and hence can export evidence for other types of causal analyses. Such assessments can be used to improve scientific understanding and the transparency and rigour of decision making by environmental managers.

Keywords: *Systematic review; Database; Analysis software; Environmental flows*

1 INTRODUCTION

Ecological studies are almost always carried out in heterogeneous, stochastic environments, making it difficult to reach clear, robust conclusions about hypothesised causes and effects [Beyers 1998]. 'Natural' experiments often suffer from poor experimental design, such as lack of proper replication in space and/or time, large-scale effects, lack of randomization of 'treatments' and observations, and the presence of one or more confounding factors, thereby weakening researchers' abilities to make strong inferences [Johnson 2002]. While this presents a conundrum for ecological researchers, the implications are more immediately serious for management of natural environments. Environmental managers are required (often by legislation) to use 'best-available science' to help develop 'evidence-based policy' [Ryder et al. 2010]. The scientific literature is large, complex, and often inaccessible for managers, who seldom have subscriptions to journals, nor the time or expertise to wade through numerous studies. However, when management decisions involve large expenditure of taxpayer funds [e.g. DEWHA 2011], these decisions must be transparent and defensible in their use of scientific information [Ryder et al. 2010]. In contrast, given the lack of empirical evidence, decisions are often informed by the experience of the manager or by expert advice [Stewardson and Webb 2010], and hence lack transparency. Clearly, tools that can help environmental scientists and managers quickly summarise and synthesize scientific information to address questions of causality are needed to help inform the decision-making process.

Facing similar issues of weak inference, epidemiologists developed ‘causal criteria’ in the 1960s to assess causality in the study of disease. Causal criteria analysis is a standardised methodology developed for assessing cause-effect hypotheses in the face of weak experimental evidence, and is widely used in medical research [Tugwell and Haynes 2006]. Causal criteria analysis provides a method to assess the evidence for and against cause-effect hypotheses. Relationships supported by sufficient ‘evidence’ can inform transparent and robust management decisions.

The potential utility of causal criteria analysis for addressing environmental issues was recognised more than 20 years ago [Fox 1991]. There have been several related attempts to develop causal criteria analysis methods for use in ecology [Fox 1991, Suter et al. 2002, Plowright et al. 2008], but they have not been heavily adopted. One hypothesized reason for this is that the methods were not supported by tools to facilitate implementation [Norris et al. 2012]. The Eco Evidence framework is another standardized method developed to facilitate causal criteria analysis in environmental science [Norris et al. 2012]. However, this method is supported by a freely-available software suite [Webb et al. 2011] that consists of two components: the Eco Evidence Database and Eco Evidence Analyser. The database is a web application for storing and sharing ‘evidence items’ (the information necessary for the analysis). The analyser is a desktop tool that uses evidence shared via the web application to assess causal hypotheses

Eco Evidence employs the scientific literature as an underused source of information for causal assessments. It uses the summary information from individual studies as the ‘data’ in a systematic review that employs the principles of causal criteria analysis. The Eco Evidence method has been fully described elsewhere [Nichols et al. 2011, Norris et al. 2012], as has its execution using the Eco Evidence software [Webb et al. 2011].

By summarizing and synthesizing the literature, tools like Eco Evidence can help to combat the ‘data deluge’ [Attwood et al. 2009] that currently confronts scientists and managers alike. However, extracting evidence from the literature is labour-intensive [Webb et al. 2012]. The open-access Eco Evidence database can reduce this burden of evidence extraction by allowing users to reuse evidence items previously entered by other users. This paper aims to better describe the Eco Evidence Database (EED) as an example of a distributed modelling resource for environmental science and management, and looks beyond its current applications to a potential future where the database is driven and managed by a self-assembling user community.

2 THE ECO EVIDENCE DATABASE

The Eco Evidence Database provides a permanent archive for extracted causal evidence, and is accessible by simply linking to the internet. The EED and Eco Evidence Analyser (EEA) are accessible through the eWater Cooperative Research Centre’s ‘Toolkit’ website (www.toolkit.net.au/tools/eco-evidence). This site offers a range of software packages for water research and management. Users must register with the Toolkit site and log in to undertake a self-approval process, after which they can access the Eco Evidence web application and the desktop analysis tool. The EED is accessible via any web browser and requires no installation or system customization, making it easily accessible for new users.

The database contains 'atomized' information extracted from scientific studies; that is, the methods and results of the study are classified according to standardized criteria, and then stored in standard fields within the database record. This information is subsequently available for analysis.

2.1 Database fields

The key data items managed in the EED are 'literature items' (e.g. journal papers) and evidence items. Literature items are similar to the entries in bibliographic databases (e.g. Web of Science). For each citation, the database contains standard bibliographic information (i.e., author, title, source, abstract, keywords). It also contains basic metadata that must be defined and entered by the user (i.e., region in which the study took place, climatic classification, ecosystem type, spatial and temporal scale of the study, broad class of study type). These metadata classifications can be used to filter the evidence returned from a database search to, for example, a specific regional context. Attached to each literature item, is one or more evidence items. The evidence item, a subset of the data contained in a study assessing a cause-effect linkage, consists of a set of standard attributes (Table 1).

Table 1. Fields and their organisation in an Eco Evidence evidence item.

Grouping	Drop-down, list, or restricted type fields			Free text fields
Cause	Trajectory*	Term & attribute*		Description
Effect	Trajectory*	Term & attribute*		Description
Design & replication	Type*	# impact units*	# control units*	Description
Association	Type*			Description
	Dose-Response+	Form of D/R		Description
Analysis	Appropriate?*			Description
	Cause in biota?+	Suitable for meta-analysis?		
Strength of Association	Type	Effect size	Variability	Description
Time Order	Type			Description
Coherence	Type			Description
Predictive Performance				Description
Information for other users				Question asked Design pp Results pp Discussion pp

* Necessary field for analysis using Eco Evidence method and EEA tool

+ Not necessary for analysis using EEA, but will be used in analysis if provided

For many of the attributes, drop-down lists are used because the various options have specific interpretations and weightings applied within the EEA. These attributes were determined over two years through user input, data usage and discussions with users and collaborators. Only a small subset (cause, description of cause, effect, description of effect, study design) of the fields in an evidence item are compulsory, but a number of others are necessary if the evidence item is to be used in a subsequent analysis using the Eco Evidence method and/or software. Fields not necessary for the Eco Evidence analysis exist for two purposes. First, the free-text fields describe that part of the evidence item more fully (e.g. 'describe

the nature of the association'). These descriptions are designed to help users to determine whether the evidence item is pertinent to the question being asked. Such information is especially important if the evidence item is to be considered for re-use by a subsequent user. The 'Information for other users' section of the evidence item fulfils a similar purpose. Second, the group titles Strength of Association, Time Order, Coherence, and Predictive Performance refer to causal criteria not currently employed by the Eco Evidence method [Susser 1991]. Such fields may be useful for other kinds of causal criteria analyses, and possible future revisions of the Eco Evidence method [Webb et al. 2012].
The standardised list of causes and effects currently consists of 229 items that cover the scope of applications to which Eco Evidence has already been applied (mostly aquatic ecology), and beyond this to hypothesized applications. It is also open to further revision and expansion. For each cause/effect, a definition is available to help users classify cause and effect for a particular evidence item. The terms are structured so that they consist of a Term (an entity) and an Attribute (a property of the entity) – e.g. "Fish (Abundance)". This standard list means that the evidence in the database is organized as best as possible. Otherwise, semantic differences between individual users would lead to poor classification of evidence, restricting search power.

2.2 Data entry

Registered users can add citations and evidence items to the database. As a basic means of quality control over the changes made, all citations and evidence items are tagged with the user name. 'Editor' users are able to edit and refine only their own contributions, while 'power' users are able to edit all citations and evidence items, regardless of whether they contributed them or not [Webb et al. 2011]. To add a new literature item to the database, the user first tests whether the title is already present in the database (to prevent duplication of literature items). If the title does not match an existing one, the user is taken to a screen where the bibliographic data and study metadata are entered. Once the literature item exists, the user can begin to add evidence items to it. Users can also add new evidence items to existing literature items, regardless of who originally entered the literature item. To date, the database has been populated with over 1000 evidence items (from over 400 citations).

2.3 Search and export

The database has powerful search capabilities that allow users to search for evidence items using any or all of the following criteria (Figure 1):

- Bibliographic information from the original literature item
- Multiple causes and/or effects from the standard terms list
- The 'question' recorded by the original extractor of the evidence
- Study-level meta data to restrict scope of search
- Search only with the 'my evidence' list for the user to restrict scope of search to evidence entered / modified by that user

The causes and effects are identified by typing key words into the "Choose cause(s)" (or "Choose effect(s)") fields. The standard list is immediately cut down to those terms that match the key words typed, either in the term itself, or in its definition (Figure 1). Results from a search like this can be exported to a .csv (comma separated values) text file for further use. When exporting evidence from the EED, a 'shopping cart' is available for users to build up a list of relevant evidence items and citations to export for different questions, search types, etc. Evidence can also be retrieved from the EED by running cause-effect searches in the Eco Evidence Analyser tool. This process uses Microsoft 'Windows Connectivity Framework' web services [Webb et al. 2011].

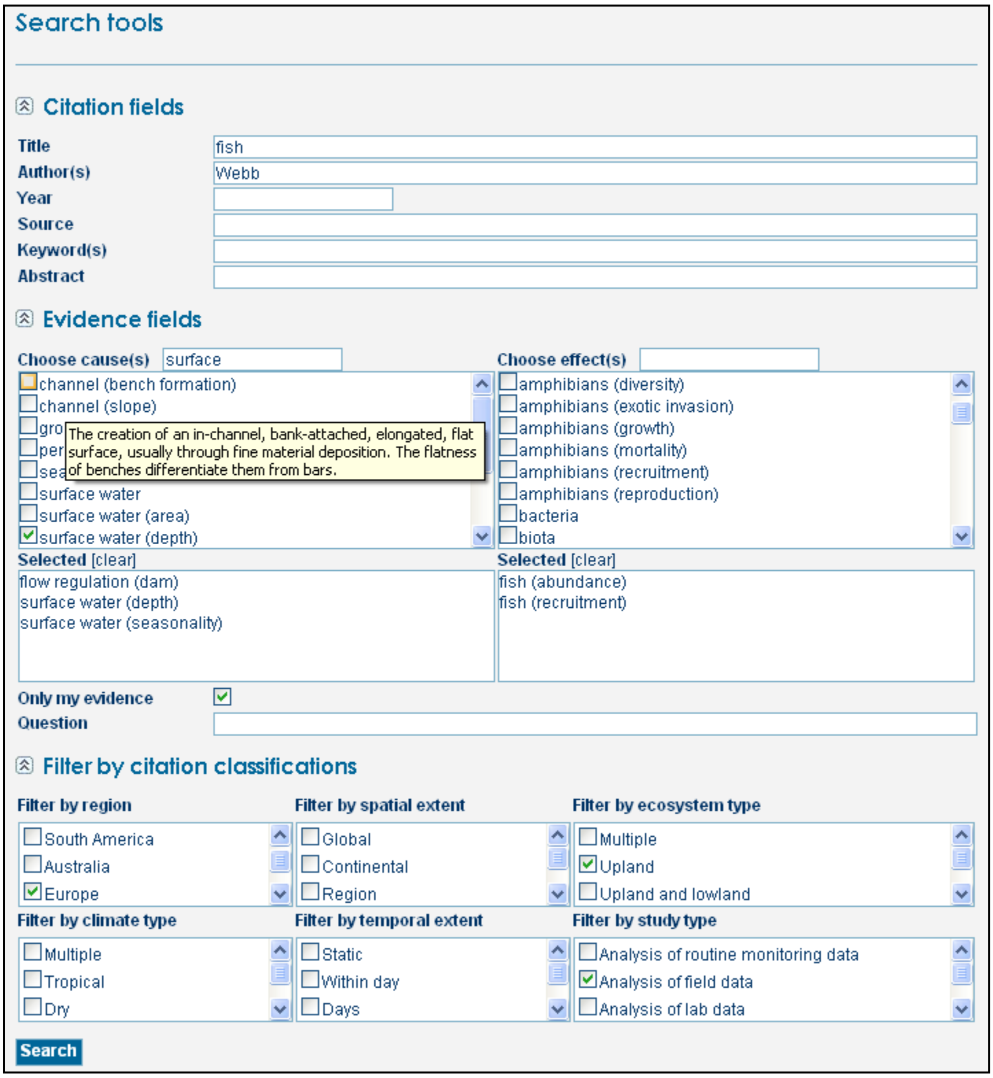

Figure 1. Screen shot of the EEA search interface.

3 ILLUSTRATIVE CASE STUDY

It is not the purpose of this paper to provide an in-depth case study to illustrate evidence synthesis. Nevertheless, it is important to show the types of results obtainable with the evidence that has already been entered in to the EED. Here, we present an updated result set for an Eco Evidence analysis presented in greater (but still brief) detail in Webb et al. [2011]. For fully-developed Eco Evidence analyses, the reader is directed to Greet et al. [2011] and Grove et al. [2012].

Environmental flow assessments commonly predict that restoring a more natural level of base flow in winter/spring and/or re-instigating higher flow events in rivers will reduce encroachment of terrestrial vegetation into channels. This question was depicted as a conceptual model that defined a set of five detailed cause-effect hypotheses, which were then tested using the Eco Evidence software suite. For the 5 causal linkages we examined, we used 6 terms to describe inundation, or flow, and 4 terms to describe vegetation abundance (Table 2).

Table 2. Terms used to search Eco Evidence database for relevant evidence items relating to cause-effect relationships between flow and terrestrial vegetation.

Causes	**Effects**
surface water (frequency)	vegetation (abundance)
surface water (depth)	vegetation (mortality)
surface water (volume)	vegetation (reproduction)
surface water (duration)	vegetation (germination)
surface water (area)	
surface water (velocity)	

For the 5 hypotheses, the EED contained a total of 102 relevant studies and 122 evidence items, with some studies providing evidence for more than one causal hypothesis. The evidence items are weighted according to study quality, and the sum of evidence weights in favour of, and against the hypothesis are compared to a threshold value to reach a conclusion. A detailed explanation of this process, and how the default weights and threshold value were derived, is provided in Norris et al. (2012). Here, there was enough evidence to reach clear conclusions for 4 hypotheses (Table 3).

Table 3. Tally of evidence and results for the Eco Evidence anlaysis of cause-effect hypotheses relating flow regime to terrestrial vegetation encroachment. #LI = number of literature items, #EI = number of evidence items, ↑ = increase, → = causes, ↓ = decrease.

	Cause-effect hypothesis	**#LI**	**#EI**	**Conclusion**
1.	↑ ‘Scouring’ flow (velocity, volume) → ↑ Mortality	16	18	Support
2.	↑ Inundation (depth, duration, seasonality, volume) → ↑ Mortality	37	42	Support
3.	↑ Inundation (duration, volume) → ↓ Reproduction	12	13	Insufficient
4.	↑ Inundation (duration, frequency, volume) → ↓ Germination	15	22	Alternate
5.	↑ Inundation (area, depth, duration, frequency, volume) → ↓ Abundance (abundance, diversity)	38	44	Support

We found sufficient evidence to support the hypotheses 1, 2 and 5, and found sufficient evidence to falsify hypothesis 4 (i.e. the evidence suggested an opposite effect to that hypothesized). Only hypothesis 3 had insufficient evidence to reach a conclusion. However for the over-arching hypothesis: that increasing inundation decreases the abundance and/or diversity of terrestrial species in river channels, we found a high degree of support, and little evidence against (Table 3).

4 DISCUSSION

The Eco Evidence Database exists primarily to reduce the burden of evidence extraction on potential users of causal analysis tools, including the Eco Evidence Analyser. In our discussions with colleagues, the effort involved in extracting evidence from literature is frequently cited as a major impediment to the uptake of the Eco Evidence method.

Currently, population of the EED is driven by case studies investigating particular questions of cause and effect. However, as noted above, this is labour intensive. This also means that the database has good coverage of some specialist areas, but does not yet have the type of broad coverage of ecological domains necessary for it to be a useful tool to a wider audience [Webb et al. 2012].

We optimistically look forward to a time when the Eco Evidence database is a peer -produced and user-moderated resource that puts the atomized findings of research projects at the fingertips of researchers and managers. The ideal people to enter evidence into the database are the study authors themselves; they have a much deeper knowledge of how their study was carried out than can be conveyed within the confines of a journal paper. Such contributions could be motivated by the 'selfish' realization that studies in the database are more likely to be used in systematic reviews, and hence cited.

Quality control of such a resource is an area of significant concern. It would be possible to modify the database to include user ratings of evidence items, such as those employed by social media. Together with an ability to discuss the quality of evidence items, this could result in active community-driven review and revision of the database contents. All contributions and revisions to the database are directly attributed to the contributor on the page. This is more likely to facilitate constructive discussions and revisions than would an ability to contribute anonymously.

The Eco Evidence database is also not the only database of this type in existence, but it is the first to be publicly accessible. The USEPA maintain a database of ecological evidence to support their Causal Analysis / Diagnosis Decision Information System [CADDIS; Norton et al. 2008]. We are currently working with the EPA to build interoperability between these two databases. We have developed a standard set of fields employed by both databases (each database also has some unique fields). The two databases are also using the same standardised definitions of cause and effect. Our collective aim is that analysis tools (such as the EEA) will be able to access evidence from both databases for analysis. However, the Eco Evidence analysis of evidence that had previously been extracted from the published literature and entered into the EED showed that the database already contains sufficient evidence to address some questions of ecological cause and effect.

While we are not advocating using the EED as a replacement for all literature reviews, this tool provides a quick and effective resource for managers or others seeking to use existing scientific evidence to verify conceptual models and inform decisions and risk-benefit analyses. Legislative and social imperatives are prompting a move from an experience-based to an evidence-based model of environmental management. This will lead to more transparent and repeatable decisions, and potentially better decisions overall.

ACKNOWLEDGEMENTS

This research is funded through Australian Research Council Linkage Project LP100200170. We acknowledge the contributions of LeRoy Poff, Ian Rutherfurd and Andrew Sharpe to that project

REFERENCES

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

Beyers, D.W., Causal inference in environmental impact studies, *Journal of the North American Benthological Society*, 17(3), 367-373, 1998.

DEWHA, *Restoring the balance in the Murray-Darling Basin*. Department of the Environment, Water, Heritage and the Arts, Canberra. Available from http://tinyurl.com/mab3yg, 2011.

Fox, G.A., Practical causal inference for ecopidemiologists, *Journal of Toxicology and Environmental Health*, 33(4), 359-373, 1991.

Greet, J., J.A. Webb, and R.D. Cousens, The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis, *Freshwater Biology*, 56(7), 1231-1247, 2011.

Grove, J.R., J.A. Webb, P.M. Marren, M.J. Stewardson, and S.R. Wealands, 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(2), 215-224, 2012.

Johnson, D.H., The importance of replication in wildlife research, *Journal of Wildlife Management*, 66(4), 919-932, 2002.

Nichols, S., A. Webb, R. Norris, and M. Stewardson, *Eco Evidence analysis methods manual: a systematic approach to evaluate causality in environmental science*. eWater Cooperative Research Centre, Canberra. Available from http://tinyurl.com/Eco-Evidence-manual, 2011.

Norris, R.H., J.A. Webb, S.J. Nichols, M.J. Stewardson, and E.T. Harrison, Analyzing cause and effect in environmental assessments: using weighted evidence from the literature, *Freshwater Science*, 31(1), 5-21, 2012.

Norton, S.B., S.M. Cormier, G.W. Suter, II, K. Schofield, L. Yuan, P. Shaw-Allen, and C.R. Ziegler, CADDIS: the causal analysis/diagnosis decision information system, In Marcomini, A., G.W. Suter, II and A. Critto (Eds.), *Decision Support Systems for Risk-Based Management of Contaminated Sites*, pp. 351-374. Springer, 2008.

Plowright, R.K., S.H. Sokolow, M.E. Gorman, P. Daszak, and J.E. Foley, Causal inference in disease ecology: investigating ecological drivers of disease emergence, *Frontiers in Ecology and the Environment*, 6(8), 420-429, 2008.

Ryder, D.S., M. Tomlinson, B. Gawne, and G.E. Likens, Defining and using 'best available science': a policy conundrum for the management of aquatic ecosystems, *Marine and Freshwater Research*, 61, 821-828, 2010.

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

Susser, M., What is a cause and how do we know one - a grammar for pragmatic epidemiology, *American Journal of Epidemiology*, 133(7), 635-648, 1991.

Suter, G.W., S.B. Norton, and S.M. Cormier, A methodology for inferring the causes of observed impairments in aquatic ecosystems, *Environmental Toxicology and Chemistry*, 21(6), 1101-1111, 2002.

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

Webb, J.A., S.J. Nichols, R.H. Norris, M.J. Stewardson, S.R. Wealands, and P. Lea, Ecological responses to flow alteration: assessing causal relationships with Eco Evidence, *Wetlands*, 32(2), 203-213, 2012.

Webb, J.A., S.R. Wealands, P. Lea, S.J. Nichols, S.C. de Little, M.J. Stewardson, and R.H. Norris, Eco Evidence: using the scientific literature to inform evidence-based decision making in environmental management, paper presented at MODSIM 2011 - International Congress on Modelling and Simulation, Perth, Australia, 2011.