Integrative modelling framework to evaluate multi-scale impacts of environmental watering

Commonwealth Environmental Water Holder's Science Program: Flow Monitoring, Evaluation and Research Program

April 2023

Australian Government
Commonwealth Environmental Water Holder

FLOW Monitoring Evaluation Research

Citation

Holt G, Macqueen A, Lester RE (2023) Integrative modelling framework to evaluate multi-scale impacts of environmental watering. Flow-MER Program. Commonwealth Environmental Water Holder, Department of Climate Change, Energy, the Environment and Water, Australia. 62pp.

Acknowledgements

The project team and the Commonwealth Environmental Water Holder (CEWH) respectfully acknowledge the traditional owners of the land on which this work is conducted, their Elders past and present, their Nations of the Murray–Darling Basin, and their cultural, social, environmental, spiritual and economic connection to their lands and waters.

The Flow-MER Basin-scale Project is led by CSIRO in partnership with the University of Canberra, and collaborating with Charles Sturt University, Deakin University, University of New England, SARDI, Arthur Rylah Institute, NSW Department of Primary Industry, Australian River Restoration Centre and Brooks Ecology & Technology. The Program delivers to the Commonwealth Environmental Water Office, Department of Climate Change, Environment, Energy and Water. The authors thank all partners and project staff for their support and interest in the work reported herein.

The authors would also like to thank the staff of the CEWH who joined the project's Modelling Advisory Group (MAG). This Group provided the CC2 modelling team with opportunity to demonstrate progress and receive feedback.

Copyright

© Commonwealth of Australia 2023
With the exception of the Commonwealth Coat of Arms, partner logos and where otherwise noted, all material in this publication is provided under a Creative Commons Attribution 4.0 International Licence https://creativecommons.org/licenses/by/4.0/. The Flow-MER Program requests attribution as '© Commonwealth of Australia (Flow-MER Program, https://flow-mer.org.au)'.

Document history

June 2022	Draft prepared
November 2022	Submitted for review
December 2022–January 2023	Review period
April 2023	Final version prepared and accepted, release pending publication of accompanying journal article

Cover photograph

Lake Cargelligo Monitoring Team members conducting water quality, vegetation and waterbird surveys with the Department of Planning and Environment – Environment and Heritage Group (DPE EHG), in the Lake Brewster inflow wetland; Lachlan River system Selected Area. Photo credit: Mal Carnegie, Lake Cowal Foundation and Department of Planning and Environment – Environment and Heritage Group

Important disclaimer

The information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, Flow-MER (including its partners and collaborators) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Flow-MER is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document, please contact CEWOmonitoring@environment.gov.au.

Summary

Management of resources is often a large-scale task addressed using many small-scale interventions. The range of scales at which organisms respond to those interventions, along with the many outcomes which management aims to achieve can make determining the success of management complex. Key challenges include addressing the interactions and dependencies in space, time and among species. Uncertainty often increases with scale, and accounting for it usefully is non-trivial. Further, considering drivers that may affect outcomes but are not under management control is important. Environmental flows is an example of management that encompasses many of these challenges, and where there is a recognised need for managers to integrate information about types of ecological responses. Thus, there is an opportunity for a new approach to supporting decision making surrounding providing water for the environment.

Here, we build and describe a modelling framework to address these challenges (eFlowEval). It has the capacity to capture best-available knowledge, to scale it in space and time, explore interactions among species, compare scenarios and account for uncertainty. Thus, it provides a basis for including multiple target groups in a common system. The framework is readily updateable as new information becomes available and can identify where data are insufficient to be scientifically robust.

We demonstrate the eFlowEval framework using 2 very different environmental responses: metabolism, which is a measure of the energy produced and then used in an ecosystem, and suitability for a bird species of interest (royal spoonbill *Platalea regia*). These demonstrations are intended to illustrate the capability of the eFlowEval framework and so the outputs shown here should not be used to assess ecological responses to management. Advice from water managers has helped shape the outputs that are shown here, attempting to encompass those most valuable to aid decision making.

These demonstrations illustrate the capacity of the eFlowEval framework to provide assessments across a range of scales, from local wetland to whole of Basin and from short time frames (weeks to months) through to multi-year assessments. They illustrate the ability to scale responses, vary driver-response model types, represent uncertainty and compare scenarios. The framework's ability to accommodate variable parameter values at different locations and present drivers alongside outputs to facilitate transparency is also illustrated.

The eFlowEval framework extends the capacity of previous similar models. It allows for interactions among species or processes to be incorporated, as well as in space and time. A large degree of flexibility is offered by the framework, in terms of driver-response model types, input data and aggregation methods. Thus, the eFlowEval framework provides a mechanism to enhance the transparency of environmental watering decision, capture institutional knowledge, enhance adaptive management and undertake evaluation of the impact of environmental watering at a range of spatial and temporal scales.

Key messages

- We successfully developed a framework, eFlowEval, for the evaluation of impacts of environmental watering at a range of scales from individual wetland to Basin scale.
- Input from water managers has shaped the capability and outputs developed from the framework to maximise the utility of the framework.
- We demonstrate the ability of eFlowEval to scale responses, vary the types of responses included, compare scenarios and represent uncertainty.
- eFlowEval has novel capacity to allow interactions among species and processes, in space and in time, and unprecedented flexibility in the types of models, inputs and aggregations that are possible.
- eFlowEval will enhance transparency in environmental watering decision making, capture institutional knowledge, enhance adaptive management and enable evaluation of environmental watering across spatial and temporal scales.

Key findings

- The eFlowEval framework provides a transparent basis for evaluating response of multiple target groups in a common system and offers the capability to capture and update best-available knowledge gained from in-depth research and monitoring.
- The framework has the capacity to provide assessments across a range of scales, from local wetland to whole of Basin, and from short time frames (e.g. weekly/monthly) through to multi-year assessments.
- We incorporated advice from water managers regarding the types and scales of outputs that would be of most use, and so have developed example output at a scale and resolution that will be suitable for natural resource management.
- We present 2 demonstrations of the eFlowEval framework to illustrate its functionality, including the ability to scale responses from local to Basin scales, vary driver-response model types, represent uncertainty, compare scenarios, vary parameters at different locations, present drivers with responses and illustrate possible outputs.
- The eFlowEval framework extends capability compared to other similar models. Specifically, functionality to allow interactions among species or processes, in space and in time are novel, as is the degree of flexibility in driver-response models, inputs and aggregation methods.
- The eFlowEval framework provides a mechanism to enhance the transparency of environmental watering decisions, capture institutional knowledge, enhance adaptive management and undertake evaluation of the impact of environmental watering at a range of spatial and temporal scales.

Overview of Flow-MER

The Commonwealth Environmental Water Holder (CEWH) invests in monitoring, evaluation and research activities delivered through an integrated program called the Monitoring, Evaluation and Research (Flow-MER) Program. This program builds on work undertaken through the Long-Term Intervention Monitoring (LTIM) and Environmental Water Knowledge and Research (EWKR) Projects (2014–2019) to monitor and evaluate the contribution of Commonwealth water for the environment to environmental outcomes in the Murray–Darling Basin. The Flow-MER Program:

- monitors and evaluates ecological responses to Commonwealth environmental water in 7 Selected Areas and at basin-scale using established metrics and methodologies
- undertakes best-practice science in 7 Selected Areas and at basin-scale to research ecological processes and thus improve capacity to understand and predict how ecosystems respond to water management
- demonstrates outcomes from Commonwealth environmental water and documents these via a regular reporting schedule and engagement and extension activities
- facilitates a regular, timely and effective transfer of relevant knowledge to meet the adaptive management information requirements of Commonwealth environmental water decision-makers.

Up-to-date information on and outcomes from the Flow-MER Program are available from the Flow-MER website[1].

Flow-MER research

The Flow-MER Program is the primary means by which the CEWH undertakes research to deliver improved methods and a richer evaluation of environmental outcomes from Commonwealth environmental water. Flow-MER Research aims to improve basin-scale understanding of the contribution of Commonwealth environmental water within and outside of Selected Areas, develop new approaches to evaluating outcomes, support adaptive management and develop a richer understanding of ecological processes and responses to Commonwealth environmental water.

The Research Plan has evolved from the LTIM and builds on the EWKR research priorities together with a large body of previous work, resulting in 13 research projects: Flow-ecology (BW2), Condition response (E2), Non-woody plants (V1), Woody plants (V2), Fish population models (F1), Fish movement (F2), Waterbirds (E1), Refugia (BW1), Scaling and condition (E3), Bioenergetics (BW3), Visualisation (CC1), Modelling (CC2) and Indigenous engagement (CC3).

This report is the final report from the Modelling research project (CC2) team.

[1] https://flow-mer.org.au/

Contents

Figures

Tables

1 Introduction

1.1 Challenges in natural resource management

Management of resources, and particularly water, is inherently a large-scale problem with many small-scale interventions (Gawne et al., 2018a). The success of management is often assessed at a range of scales, from local to very large. The range of scales at which freshwater systems operate (from metres to 1000s km, depending on the process and species) indicates that managers must understand their ability to influence those processes with relevant management levers at each scale (van den Belt and Blake, 2015; Soranno et al., 2010). Further, the outcomes against which that success is measured are often many, responding to different factors and for different reasons, and these outcomes are likely to interact. Thus, determining how best to manage for multiple objectives, potentially in multiple locations at multiple times, and then determining the success of that management can be highly complex.

Natural resource managers are often tasked with maintaining the integrity of ecological systems as a whole (e.g. 'restore' or 'improve long-term health' are common words in management planning objectives; see DPIE, 2020). Similarly, best-available science is often intended to be the basis for that management (Ryder et al., 2010). However, best available science is often developed for individual components of ecosystems (e.g. a species of interest) and for one or perhaps a couple of drivers of interest, and under a limited set of conditions in space and time. For example, most published studies of environmental flows assess a single biotic response in one individual river, often in a single reach (Olden et al., 2014). This presents a mismatch between the information available to aid decision making and the decisions that need to be taken. One approach to deal with this mismatch is by the use of indicator sites and species (e.g. umbrella species) but challenges remain. Indicator species and sites are rarely rigorously chosen (e.g. see Downes, 2010) for a discussion on the issues associated with the use of 'representative' sites and (Seddon and Leech, 2008) for the criteria to select an umbrella species). This lack of rigour potentially results in sub-optimal ecological outcomes as approaches intended to simplify decision making obscure variability or fail to adequately capture important drivers.

Addressing the issues of spatial and temporal scale is also non-trivial. Challenges in scaling arise increasingly as the connectivity and interdependence of processes increase due to globalisation and large-scale environmental change (van den Belt and Blake, 2015). Water management is particularly susceptible to issues of scale because of the connectedness of hydrologic processes (van den Belt and Blake, 2015). While the connectedness of those hydrological processes is increasingly understood and accounted for in management processes, the related biological connectedness (e.g. fish movements and hydrochory in plants) is less well understood (Gawne et al., 2018a). Systems are often data poor about how ecosystems function at large scales, let alone how they may respond to management actions. There are limited opportunities to assess the impact of changes in flow (including managed flows) at large spatial and temporal scales (Gawne et al., 2018a). Monitoring of the response to flows is often focused on the effects of individual (or a few) short-term events rather than long-term responses (Olden et al., 2014). Ecological theory is often also limited at large scales (Heffernan et al., 2014). Thus, our understanding of small scale and fast responses is far better developed, but scaling that understanding to catchment and basin scales is difficult (e.g. see de Vente and Poesen, 2005) for an example using erosion processes). Nonetheless, our large-scale knowledge is constantly updating and improving, so it is necessary to provide mechanisms for that knowledge to be incorporated into decision making (Gawne et al., 2018a).

Another important element that is often overlooked in decision making (or ignored to reduce complexity) is that multiple outcomes respond to a range of factors, some of which are under management control, and that these responses will likely interact (Thompson et al., 2018). Best available science and resultant tools which focus only on those levers under management control potentially overlook other factors that could de-rail management efforts and prevent desired responses from being realised (e.g. impact of feral animals on the regeneration of floodplain vegetation following environmental flows). Carefully selecting those drivers that are critical to the response of species or processes under management can assist to identify interacting factors that should be considered in concert with proposed actions (Lester et al., 2020). Interaction among drivers and the species and processes that respond to those drivers is also a casualty to simplification (Thompson et al., 2018). Again, this can be warranted but has the potential to obscure the mechanisms behind responses and lead to failure of intended management actions. As for any model or analysis, parsimony is important, and so the components explicitly included need to be considered carefully and relevant covariates included, among species and processes, in space and in time so that the model is fit-for-purpose (Lester, 2019).

Dependence in time has been explored under the term 'cumulative effects'. Cumulative effects science attempts to integrate disturbances to develop an understanding of past, current and future impacts on the components of a system (Venier et al., 2021). There is general agreement that cumulative effects science be conducted at larger, more holistic scales than most experimental studies allow (Venier et al., 2021). Processes occurring at large spatial scales also tend to occur over longer time frames than those at smaller scales (Poff et al., 2017). Different species also respond at different rates with short lived species, for example, often responding more quickly to environmental change than long lived species (e.g. Couet et al., 2022). Understanding how the past conditions of an ecosystem, lags in response time and long-term trends affect species and processes would greatly enhance our understanding of the likely impact of management actions (Thompson et al., 2018). This suggests a need to explore interactions in space and time that has been largely lacking to date.

The inherent variability of ecological systems combined with limited knowledge yields high uncertainty. That uncertainty is often unaccounted for in ecological models and decision making, but that does not diminish the ability of that uncertainty to affect outcomes in ecological systems (Lester, 2019). Some forms of uncertainty can be quantified and explicitly accounted for, while others remain qualitative (Ascough et al., 2008). Ignoring uncertainty has real consequences for management and public confidence (Ascough et al., 2008) but identifying it provides opportunity – for example, understanding the range of possible outcomes and their likelihoods helps plan management actions, communications that identify a range of possible responses are more trustworthy, understanding uncertainty enables explicit hypothesis testing of other drivers and enables areas requiring further study to be identified and prioritised. Uncertainty tends to increase with scale, with the greatest levels of uncertainty associated with catchment or basin-scale processes and responses (Gawne et al., 2018a) and so accounting for that uncertainty is increasingly important at those large scales.

Environmental flows are an example of management that encompasses the challenges identified. Environmental objectives, along with the associated social and economic objectives, for freshwater ecosystems are not currently being met, given increasing demand and the impacts of climate change globally (Vörösmarty and Sahagian, 2000), suggesting a need to continue to improve management practices and understanding. Basin-wide management is required due to the connectivity inherent in river basins and the dependence of downstream processes and communities on upstream management (Gawne et al., 2018a). Much institutional knowledge exists at small scales to establish relationships between environmental conditions and outcomes, and water managers are already exploring opportunities for incorporating large scale processes in their management practices (e.g. CEWO 2017) and monitoring programs (Gawne et al., 2018a). There is a recognised need for managers to integrate information across types of ecological responses, drawing disparate responses together in a holistic manner. Thus, opportunity is ripe for a new approach to supporting decision making surrounding providing water for the environment.

1.2 Benefits of a new approach

Fundamentally, a solution to the issues associated with modelling outcomes at a basin scale include i) a framework that can be scaled, such that the outcomes of environmental water can be explored across the Basin, ii) considering outcomes over one or more years and iii) comparing scenarios such as climate change or watering with no watering. It should be suitable for representing the flow response of species, processes or habitats and be able to relate those to the relevant management objectives, but also incorporate drivers that are not under direct managerial control (e.g. climate) so as to understand the likelihood of other factors influencing outcomes (Walker et al., 2002). A model should be based on relationships that are well understood such that there is sufficient knowledge available to describe those relationships and where relevant evidence is documented such that relationships are robust and results can be replicated (Grimm et al., 2014). The model also needs to be representative of the range of flow requirements and responses of biota and processes in water-dependent ecosystems, recognising that the expected outcomes of environmental watering can be nested in time and space (Gawne et al., 2018a). Finally, model outcomes need to be consistent, so that assessment of ecological outcomes accounts for any biases in methodological approach, scale of application or scenarios considered (Grimm et al., 2014). These types of considerations have shaped other large-scale programs, including in the Murray-Darling Basin (e.g. Sustainable Rivers Audit; Norris et al., 2001).

The benefits of such an approach, scientifically and to guide management, are many. When developed in consultation with managers, this approach offers a mechanism to capture the institutional knowledge and mental models that are the norm in much natural resource management. Such mental models often incorporate years of experience and local expertise, but understanding is not captured explicitly and can lack transparency and repeatability (Greca and Moreira, 2000). Where culturally appropriate, Indigenous knowledge systems can be included to provide greater influence on decision-making than is currently usual (Tengö et al., 2014). An approach such as eFlowEval offers a repository for such institutional and local knowledge and model outcomes can be validated against that knowledge through the use of hindcasting, for example.

Natural resource management often relies on 'best available science' for decision making (Ryder et al., 2010). It is necessary to make decisions immediately, or within short time frames, and it is not possible to wait for a scientific consensus as to how best to proceed. Adaptive management, or learning by doing, is a common method for incremental improvement and learning during this process (Williams, 2011). The use of models (or frameworks such as eFlowEval) to assist with adaptive management provide a mechanism for including specific hypothesis testing in the adaptive management cycle – thus providing a mechanism for improving scientific understanding as a result of management actions (Williams, 2011) and also enabling implementations of the framework to be incrementally updated as knowledge and data availability improve.

Such an approach also has the advantage of enabling disparate knowledge to be combined in a standard and comparable way. Managers are tasked with achieving many objectives in space and time (e.g. those in the long-term watering plans for the Murray–Darling Basin[2] and it can be difficult to identify synergies and trade-offs. Having a single repository for disparate knowledge (e.g. potentially including Indigenous and local knowledge; Tengö et al., 2014) and a transparent mechanism for scaling responses reduces the ad-hoc nature of decisions and enables natural resource managers to more easily grasp the big picture across groups, responses and locations. Similarly, the single repository enables explicit understanding of how small-scale interventions affect larger-scale outcomes and uncertainty to be quantified (or acknowledged) (Gawne et al., 2018a). Hindcasting can be used for evaluation purposes (i.e. what was the benefit derived

---

[2] environment.nsw.gov.au/topics/water/water-for-the-environment/planning-and-reporting/long-term-water-plans

from the use of environmental watering) while forecasting can aid planning and, in some cases, enable the use of environmental watering to be optimised.

1.3 Objectives

Here, we build and describe a modelling framework (eFlowEval) to flexibly scale and integrate responses to environmental flows. Our framework provides the capability to capture best-available knowledge gained from in-depth research and monitoring, to scale it in space and time, explore interactions among species, compare scenarios and to account for uncertainty, providing the basis for including multiple target groups in a common system. We have aimed to ensure that it is able to be readily updated as new knowledge becomes available and that outputs are at a scale and resolution useful for natural resource management. In addition, it also identifies where data are insufficient to be scientifically robust.

We demonstrate our eFlowEval framework using 2 very different types of environmental responses: metabolism, based on a quantitative regression model of gross primary productivity and ecosystem respiration, and suitability for a bird species of interest (royal spoonbill *Platalea regia*), using a threshold model. In developing these demonstrations, we do not suggest that the specific response models used are definitive – they were selected to enable the demonstration of the features of the eFlowEval framework. We identify where modifications have been made that would possibly make specific findings from those models less robust ecologically and we caution against using those specific outputs for decision making purposes. Instead, the demonstrations are intended to illustrate the approach, including outputs that target management needs, such as evaluation, scenario comparison and planning at multiple scales.

We used the Murray–Darling Basin (the Basin) as a test case for the development of the eFlowEval framework. The Basin is Australia's largest river system, covering an area of more than 1 million km^2; and supports 50% of Australian irrigated agriculture production, 2.2 million people and a diverse mosaic of habitats and species (Hart et al., 2021). Due to a range of stressors including water abstraction and habitat loss, the water-dependent ecosystems of the Basin have declined through time (Hart et al., 2021). In response, Australian Governments developed the Murray-Darling Basin Plan ('Basin Plan'), which seeks to optimise social, economic, cultural and environmental outcomes via integrated water management across jurisdictions (Gawne et al., 2018a). Under the Basin Plan, long-term adaptive management frameworks have been developed, including the monitoring of outcomes and evaluation of actions to assess their contribution to achievement of Basin Plan objectives (Hart et al., 2021). As a result, there is a legislative requirement for natural resource managers to be assessing the impact of their actions at the basin scale.

The Basin operates as a partially connected and integrated system that supports meta-populations and ecosystem function across broad areas in space. The Basin is hydrologically connected, but with large variability relative to similar rivers elsewhere (McMahon et al., 2007) which can be obscured via the reporting of averages (e.g. average annual flow; Stewardson et al., 2021). Nevertheless, hydrology is the primary management lever in the Basin, with the purchase of water and its use to support environmental outcomes ('environmental flows') a key plank in the Basin Plan management strategy (MDBA, 2011). Understanding of the flow regimes required to support environmental values at a basin scale is based on the cumulative impacts of anthropogenic change, usually established from monitoring at smaller scales (Frissell et al., 1986; Gawne et al., 2018b; Englund and Cooper, 2003). Thus, there is a need for tools and frameworks to assist with the scaling and aggregation of such small-scale monitoring data to the larger spatial and temporal scales at which decision-making occurs (Gawne et al., 2018a).

2 eFlowEval framework

To meet our objectives, we developed the eFlowEval framework to accept disparate data sources, process those data to be appropriate as inputs to ecological response models, enabling scaling, integration and characterisation of uncertainty and to present outcomes in a manner relevant for management use cases (Figure 2.1).

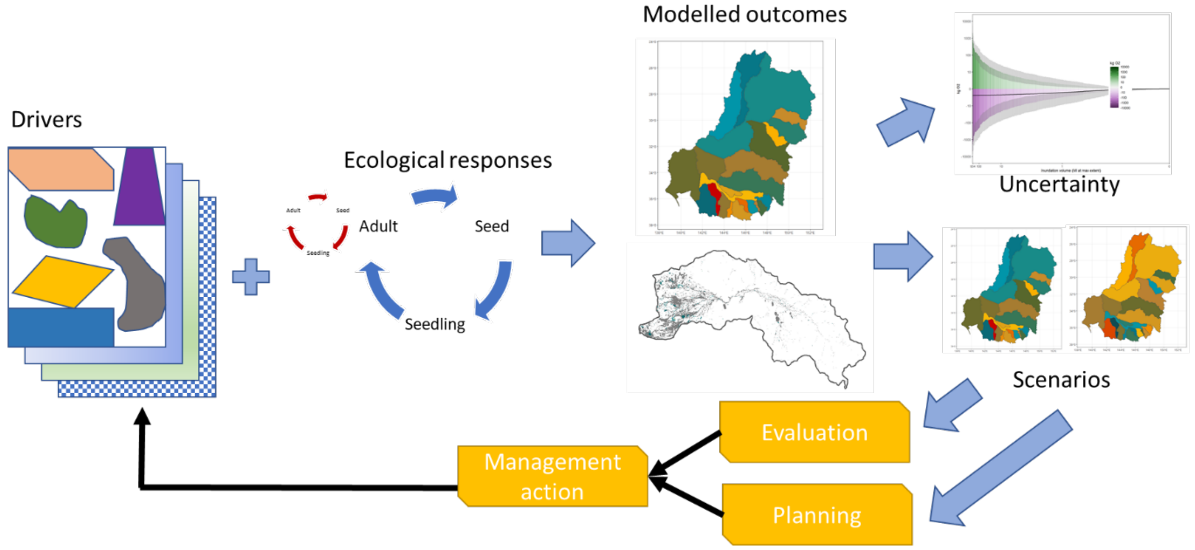

Figure 2.1 eFlowEval framework illustrating the input spatial layers of drivers to be used as inputs for ecological response models for species or processes of interest

Those models can be included in a variety of formats, including threshold and regression models as demonstrated below. Modelled outcomes are derived from scaled and integrated local scale outputs, with uncertainty captured and scenario model comparison possible. These outputs are designed to be used for evaluation of the impact of environmental watering and to assist with planning watering and other management actions.

2.1 Framework components

2.1.1 Response models

At its core, eFlowEval utilises driver-response models. To enable consideration of multiple ecological responses of interest, we recommend the use of simple ecological response models focused on factors that have the potential to limit population growth or suitability (e.g. strictures and promoters framework; (Lester et al., 2020) to capture critical responses without adding greatly to complexity and processing times. This approach is also well suited to the common situation of low information when modelling at large scale – for example, it is rare that detailed population dynamics models (or the input data for them) are available at such large scale. However, eFlowEval has been designed to take any driver-response model (e.g. regression or other) where drivers are available at an appropriate scale.

The development of driver-response models is intended to be external to eFlowEval. Instead, eFlowEval incorporates existing models and uses them in a consistent and repeatable manner. Thus, the driver-response models can be developed or guided by experts in those particular responses, while the eFlowEval provides the consistent framework for those models and their outputs. This structure enables flexibility in the development of those driver-response models as well as ensuring that the framework is updatable as new models are developed or existing models are improved.

2.1.2 Base polygons and driver data

The choice of driver-response model determines the type and nature of the required input data. For example, in one of the demonstrations below, gross primary productivity is a function of air temperature and inundation, so those are the datasets required as input for that application of eFlowEval. The intention of eFlowEval, to provide input at broad spatial and temporal scales to assist with natural resource management planning and evaluation, leads us to an emphasis on datasets available at those broad scales (e.g. satellite imagery). eFlowEval can either house static layers of input data or point to cloud repositories of data updated regularly and so use the most recent dataset available.

The eFlowEval framework incorporates driver input data into a common format based on meaningful polygons. Using meaningful polygons enables the later integration and scaling of responses to categories that are relevant for management decisions. In the demonstrations below, those meaningful polygons are derived from habitat mapping for the relevant region (Australian National Aquatic Ecosystem Classification Framework [ANAE]; Brooks et al., 2014) because environmental watering decisions often target specific habitat types. However, other classification systems can be used to identify those polygons. Using meaningful polygons as the base unit of modelling provides several advantages over uniform gridded data such as rasters, primarily because polygons vary in size and shape. Large areas of similar conditions can be captured in a single large polygon that would otherwise require many raster cells, while areas where conditions change rapidly in space can be represented by numerous small polygons. Thus, this approach yields a spatial scale for the modelling that flexibly responds to the scale at which relevant conditions change. Moreover, large areas of the Basin may not be relevant to modelling aquatic outcomes, and so the use of polygons allows those to be ignored, yielding vast reductions in necessary processing power.

Input data to eFlowEval other than the base polygons can include many formats, including polygons, rasters, points, and other types. Once incorporated, it is usual for raw datasets to need to be processed to constitute the input variables needed to act as drivers for the response models. Typically, this consists of incorporating them into the base polygons using some mathematical aggregation function (e.g. rasters often will use an area-weighted mean). Polygons may also be used to split the base polygons if interactions are important. eFlowEval has flexible processing capacity to develop relevant statistics including weighted means, rolling averages, time since particular events and custom statistics such as depth above a threshold, for example.

These statistics and variables are then used as inputs to the selected response model(s). Responses are produced for each at the level of the individual polygons. Furthermore, eFlowEval has the capacity to include interactions among response models. So, it is possible for a bird response model to rely on a fish response model (as a food resource, for example) as well as a tree response model (as nesting habitat, for example). This dependency is possible based on previous time steps or the current timestep.

2.1.3 Results scaling and presentation

Local outcomes are produced as close as possible to the scale of the process, provided the meaningful polygons are chosen carefully. At a minimum, this could be daily outputs at the scale of individual polygons. These results are then able to be scaled to management-relevant scales. Again, eFlowEval enables flexible approaches to scaling, so that appropriate metrics can be used to achieve that scaling including means, sums and others including custom statistics. This flexibility is important because different approaches will be appropriate in different cases. For example, in the demonstration below, maximum inundation extent over a 2-month time step is an input variable. Scaling using a mean would be inappropriate given that maximum inundation variable, and so a maximum in space and time is used instead. Scaling metrics are thus able to be selected on a case-by-case basis and to be tailored to different response models and different goals for the eFlowEval application.

In addition to the capability to scale outcomes, eFlowEval has the ability to handle multiple scenarios and compare outputs across scenarios, as well as other comparisons. Outputs have been developed to be useful for management evaluation of actions, planning and simulation of plausible future states. Ideally, there would be a link between the response models and feedback from management actions to continue to improve the ability of eFlowEval to adequately represent the best-available knowledge base and meet specific management needs.

A detailed description of the framework architecture can be found in Appendix A, providing additional details as to the development and application at each step, along with a link to the relevant code.

2.2 Rationale for the eFlowEval approach

The rationale for developing the eFlowEval framework in this way was based on creating a process-based but simple approach. Using process-based response models provides a robust basis for extrapolation to areas outside the spatial and temporal resolution of the data used initially to develop the response model (Yates et al., 2018). This extrapolation, to larger spatial scales such as a Basin scale, as well as through time (both the past and plausible futures) were key requirements for eFlowEval.

One of the demonstrations below utilises a strictures and promoters approach to developing a response model (Lester et al., 2020). This approach seeks to identify dependencies throughout the life cycle of species of interest. It relies on identifying possible inhibiting (strictures) and supporting (promoters) factors, emphasising the need to support and protect all life history stages to achieve environmental objectives such as population maintenance or growth (Lester et al., 2020). Initially, strictures models can be quite simple (e.g. threshold based) minimising computation times, but the driver-response units can be updated to be as complex as data permit. Maintaining simple models has advantages beyond simple data availability, though; complex models can be difficult to interpret, and interactions quickly become erroneous even if the individual model components are well supported (Lindenschmidt, 2006).

In developing the framework, we recognised the need to integrate across multiple scales to allow for multiple potential uses of the framework. Multiple dimensions were of importance – in space, through time and among species. Much has been written about the need to scale in space and time to adequately utilise small-scale monitoring data for large-scale decision making (e.g. Gawne et al., 2018a) but less focus has been given to the need to integrate across species and processes (although such interactions are modelled in ecosystem models, for example; Geary et al., 2020). Many ecological models instead incorporate multiple species as independent entities (e.g. existing tools for the Murray-Darling Basin such as the Murray Flow Assessment Tool, Young et al. 2003), and Ecological Elements (Overton et al., 2014) or model assemblages and communities as single units (e.g. using alternative stable states; Lester and Fairweather, 2011) which do not allow for interactions, synergies and trade-offs in the responses of multiple species to be explored. This is a common limitation in the ability to manage systems for sets of interacting species and assemblages. As a result, we sought to include the ability for multiple response models to interact in space and time.

To integrate in space and time, initially, we were largely focused on the challenges associated with developing capability at a basin scale over multiple years, but discussions with natural resource managers to workshop possible use cases and explore early outputs highlighted the benefits of smaller scales, for individual wetlands or wetland complexes, and at scales of weeks to months. As a result, we expanded our focus to span the scales of interest.

When undertaking extrapolation in space and time, we calculated ecological responses locally and then scaled those outcomes, rather than scaling inputs and processing overall. Non-linearities in response are highly likely at large scales (e.g. catchments and basins) where many ecosystem functions respond at multiple scales (DeFries et al., 2004). These non-linearities can create large biases (i.e. Jensen's inequality;

Ruel and Ayres, 1999). Calculating responses locally enables us to model at the scale at which the relevant ecological processes vary (Yates et al., 2018). This is usually local, and is likely determined by the environment, as well as the life history and characteristics of the species in question. From here, there are established methods to scale in space and time (e.g. Englund and Cooper, 2003). Similarly, we created the ability to increase the resolution of the models to respond to the scale of the process or response in question. Responses are modelled at the spatial and temporal scale most relevant to that response and based on the input data used to calculate it. This approach, rather than a fixed grid for spatial resolution, for example, minimises the computational power required for the model – a key consideration for large-scale models that include multiple species and processes.

While hydrology is the primary management lever in the Basin for achieving environmental outcomes, the species and processes of interest often response to multiple interacting drivers, including but not limited to, hydrology (Poff and Zimmerman, 2010). As a result, we explicitly account for multiple interacting drivers including those that are not able to be actively managed. This approach is effectively the same as including covariates in statistical models. In the demonstration below for metabolism, for example, both inundation and temperature are drivers. This enables us to identify how drivers interact, whether management actions are likely to succeed or fail given (and sometimes because of) those interactions and where simulated responses are unrealistic, giving us the opportunity to identify additional drivers that may be affecting outcomes. As for the species response models, we suggest keeping the number of drivers included small to limit complexity. Nonetheless, we have developed the capability to increase the complexity of interactions and the number of drivers as needed and as appropriate data and knowledge are available.

Thus, we have developed a modelling framework that takes a consistent approach to handling and processing driver data, specifying responses, assessing those responses and then calculating and scaling outcomes. The specifics for each response can vary to reflect knowledge regarding that species or process – while we seek consistency in the basic structure and methods, vegetation does not respond to the same drivers as fish, for example (and certainly not in the same way) and so flexibility is a key part of the approach.

3 eFlowEval capability demonstration

We have developed 2 demonstrations of eFlowEval to illustrate the capability of the framework as developed. We selected demonstrations specifically to illustrate elements of that capability, focusing on the application of eFlowEval to support management decision making, planning and evaluation. The demonstrations span a range of management outcomes and response types. They are intended to illustrate the framework's flexibility and its ability to shape disparate input data and response types to a common format. An additional outcome is the ability to identify where desired outputs are not possible and why. This will inform future monitoring and research when those outputs are critical to supporting decision making.

It is important to note that the response models that underpin these demonstrations have been modified to enable their use in eFlowEval, and so are not reliable estimates of ecological responses. For example, these demonstrations should not be used to estimate kg O_2 of gross primary productivity based on the figures provided. The reliance on bimonthly maximum inundation and simple regressions precludes the use of results in that way. Instead, these demonstrations illustrate the importance of ongoing work to develop appropriate driver-response models and that existing models may need additional development to be appropriate for use by a framework such as eFlowEval, which is focused on integration and extrapolation.

Additional detail regarding the development of the demonstrations, including the driver-response models used, are given in Appendix B.

3.1 Targeting local scales in space and time: wetland metabolism demonstration

Management decisions such as those regarding environmental watering decisions are often made at local spatial scales (e.g. individual wetland or groups of wetlands) for the weeks or months ahead. Event planning and decisions targeting specific ecological goals in important locations are both likely to occur at this local scale. eFlowEval models processes as close to the scale at which they occur as allowed by the data, and so has the capacity to illustrate both driver and responses at the smallest scales within the datasets (Figure 3.1).

The regression models used here for gross primary productivity (GPP) and ecosystem respiration (ER; refer to Appendix B for model details) define quantitative relationships between the drivers and the outcomes. Here, those drivers include those able to be influenced by managers (i.e. inundation extent) and those that are outside management control (i.e. temperature). The inclusion of both provides managers with the capacity to understand how factors outside their control (here temperature) may alter the impact of a specific action (e.g. changing inundation). Further, the inclusion of seasonality and catchment terms can help assess the effect of altering the timing or location of water delivery.

An R/Shiny application linked to the model outputs enables users to identify times of interest for the spatial scale and location of their choice (Appendix D). This provides the capacity for users to interact with the data, explore trends of interest or investigate specific events in the past that may have similarity to a likely future event, or a desired outcome.

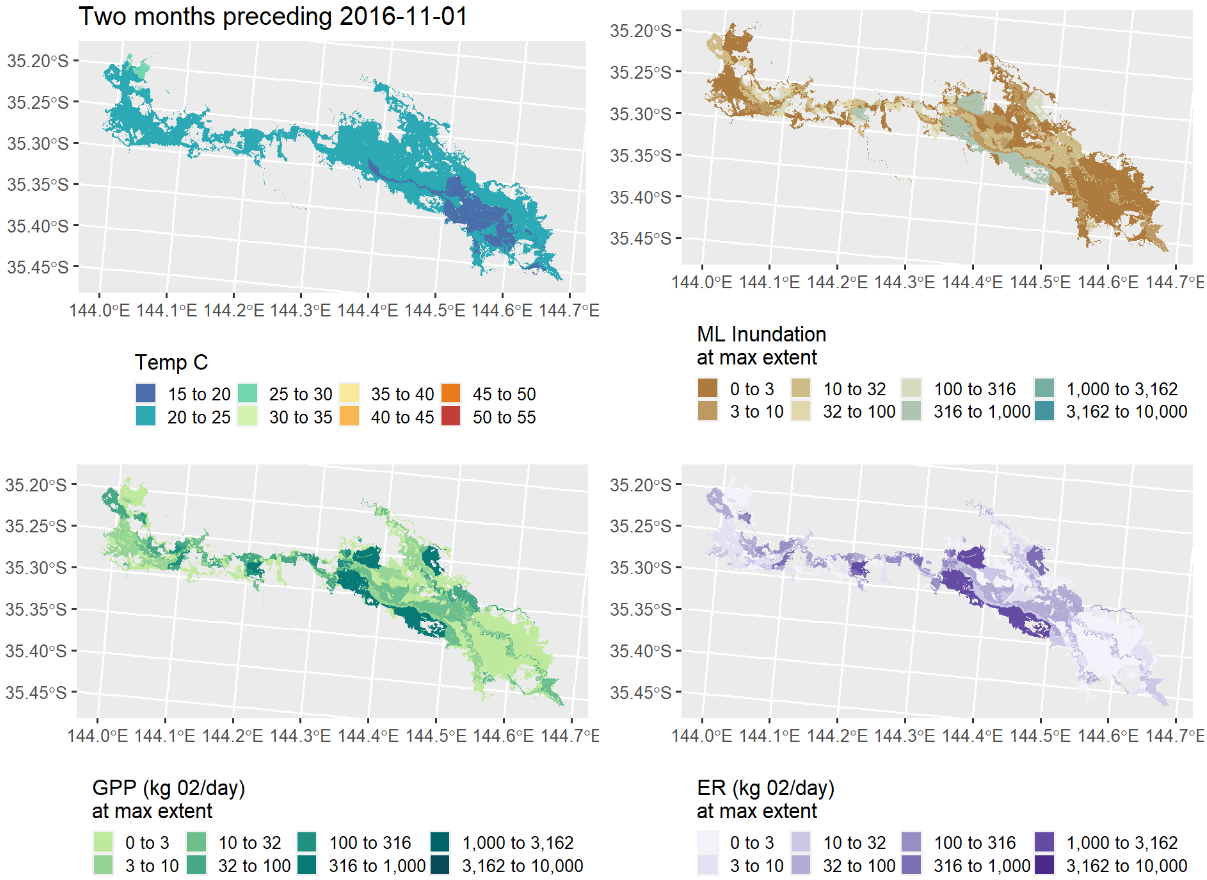

Figure 3.1 Example of drivers (top row) and predicted outcomes (bottom) for wetland metabolism in individual wetlands within the Werai Forest

The modelled period is the 2 months preceding 1 November 2016, and reflect the maximum inundation extent during that period, as that is the scale of the inundation dataset. A Shiny app has been developed to allow users to select the 2-month period of interest and update the outputs in a browser.

Annual summaries are also often useful for management reporting (Figure 3.2). For this demonstration, we have generated those summaries by summing the metabolic activity for each wetland over all the bimonthly periods of a water year to generate total yearly metabolic activity (calculated at the maximum inundation extent of each period). Other methods of combining through time (e.g. means) are also possible where appropriate.

Comparing across multiple years illustrates the effect of warmer wetter years (e.g. 2016) where both gross primary productivity and ecosystem respiration were relatively high compared with cooler drier years (e.g. 2015, 2017). The local scale also enables hotspots of productivity to be identified.

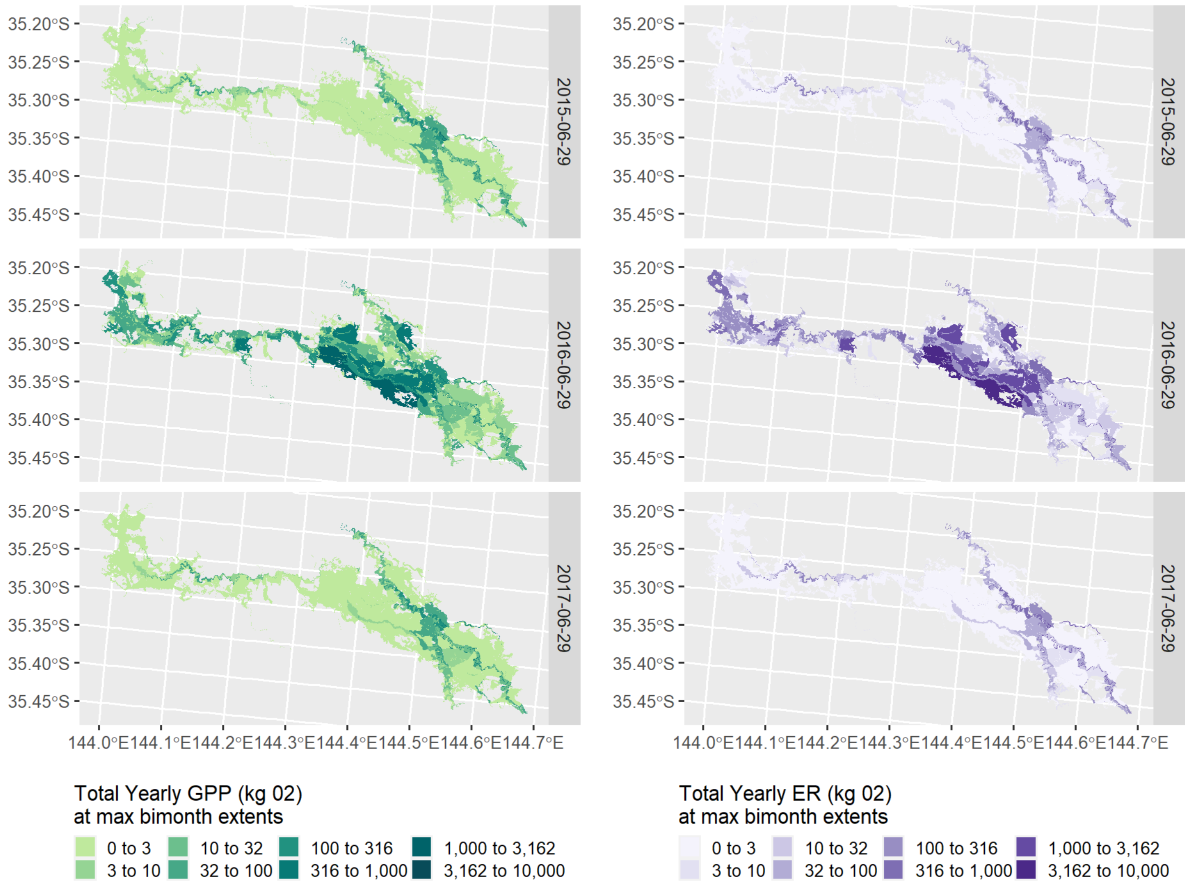

Figure 3.2-GPP and ER for each wetland in the Werai Forest summed over bimonthly inundation predictions for 3 water years to yield total metabolic activity for the year (at the maximum inundation extent in each bimonthly period)

More quantitative visualisations than maps can show the expected metabolic activity across all wetlands, as well as the uncertainty around those estimates. For example, in Figure 3.3, we see that the majority of wetlands contribute no metabolic activity, while others are responsible for a disproportionate share. Prediction intervals are large, but we can be confident that some wetlands are cycling 100–1000 kg O_2 at maximum inundation extents. This understanding may assist managers to pinpoint productivity hotspots in space and time to enable investigation of drivers in those hotspots, or to target actions to those locations.

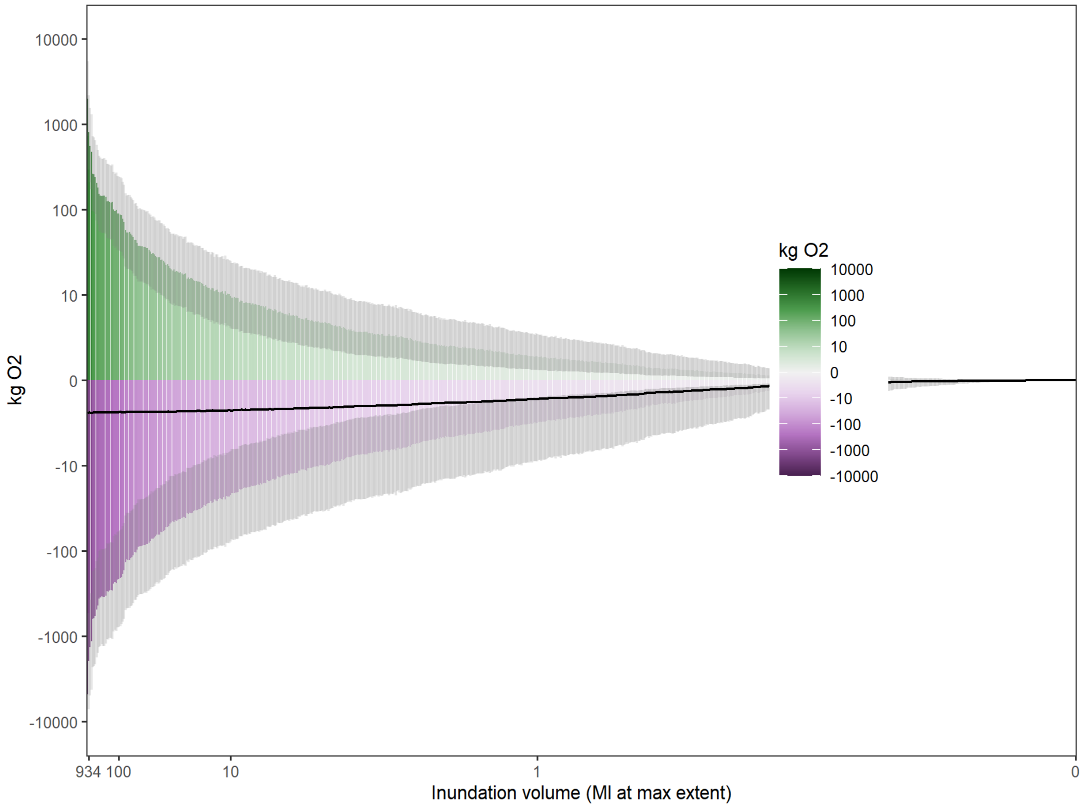

Figure 3.3 GPP (green, positive values) and ecosystem respiration (purple, negative values) in kg O_2 produced (or respired, if negative), represented as bars for each of 1,163 ANAE wetlands in Werai forest in the 2 months ending 2016-11-01

Grey lines are 95% prediction intervals for each wetland, indicating uncertainty in the predictions of metabolic activity due to temperature and inundation. Black roughly horizontal line is net metabolism and indicates that those wetlands with appreciable metabolism were heterotrophic during this period. Because bar heights represent metabolic contributions of individual wetlands, the areas under the curves represent the total metabolic output of all wetlands in the Werai forest.

3.2 Targeting catchment and basin scales in space and time: wetland metabolism demonstration

While local scale outputs are highly relevant for informing individual watering decisions or targeting particular wetland complexes, long-term and broad-scale planning and evaluation are also needed (e.g. to describe the effect of environmental watering under the Murray-Darling Basin Plan; Gawne et al., 2018a). Often, such planning and evaluation will mean that it is desirable to be able to report ecological response for periods of a year or more or across broad areas like catchments or basins. eFlowEval offers the functionality to aggregate local-scale results to those larger scales (Figure 3.4).

A range of visualisation options are available within eFlowEval. Maps are easy to grasp, present broad patterns well and put data in a geographic context that many managers and the public should find intuitive (e.g. top panels, Figure 3.4). Maps easily enable managers to assess *where* environmental water is (or is not) having an effect. One clear advantage of this approach is to identify catchments that are disproportionately important for the ecological response of interest or areas of the basin that tend to respond in concert. Maps are also extremely powerful communication tool for a broader audience.

Maps, however, have limitations. They are necessarily limited in the precision at which data can be reported, particularly when there is a very large range of values. In the examples here, discrete colour bins prevent identifying precise values and lump catchments into groups while providing bounds on their values. Continuous colour ramps are another possibility, that would yield different colours in every catchment but, while the values represented are more precise, they are more difficult to assess other than a general impression of relatively high or low. Other sorts of graphs (e.g. lines, bars, or scatterplots) are better at more rigorous assessment and comparison of the values.

We illustrate one option here with a bar chart of inundation and metabolism for each catchment (e.g. middle panel, Figure 3.4), showing values of each with more precision and allowing visualization of differences between catchments that have the same colour on the map. This visualization provides managers with the ability to better assess small but potentially important differences.

Another option are line charts, and particularly timeseries, which can show the trajectory of values (e.g. bottom panel, Figure 3.4). We illustrate this here for the maximum inundation and metabolism summed over the whole of the Basin, where we can see that some values (ecosystem respiration, in particular) respond differently across years. This sort of approach is also particularly important for identifying temporal trends, as might be the case in assessing the net effect of climate change across the Basin.

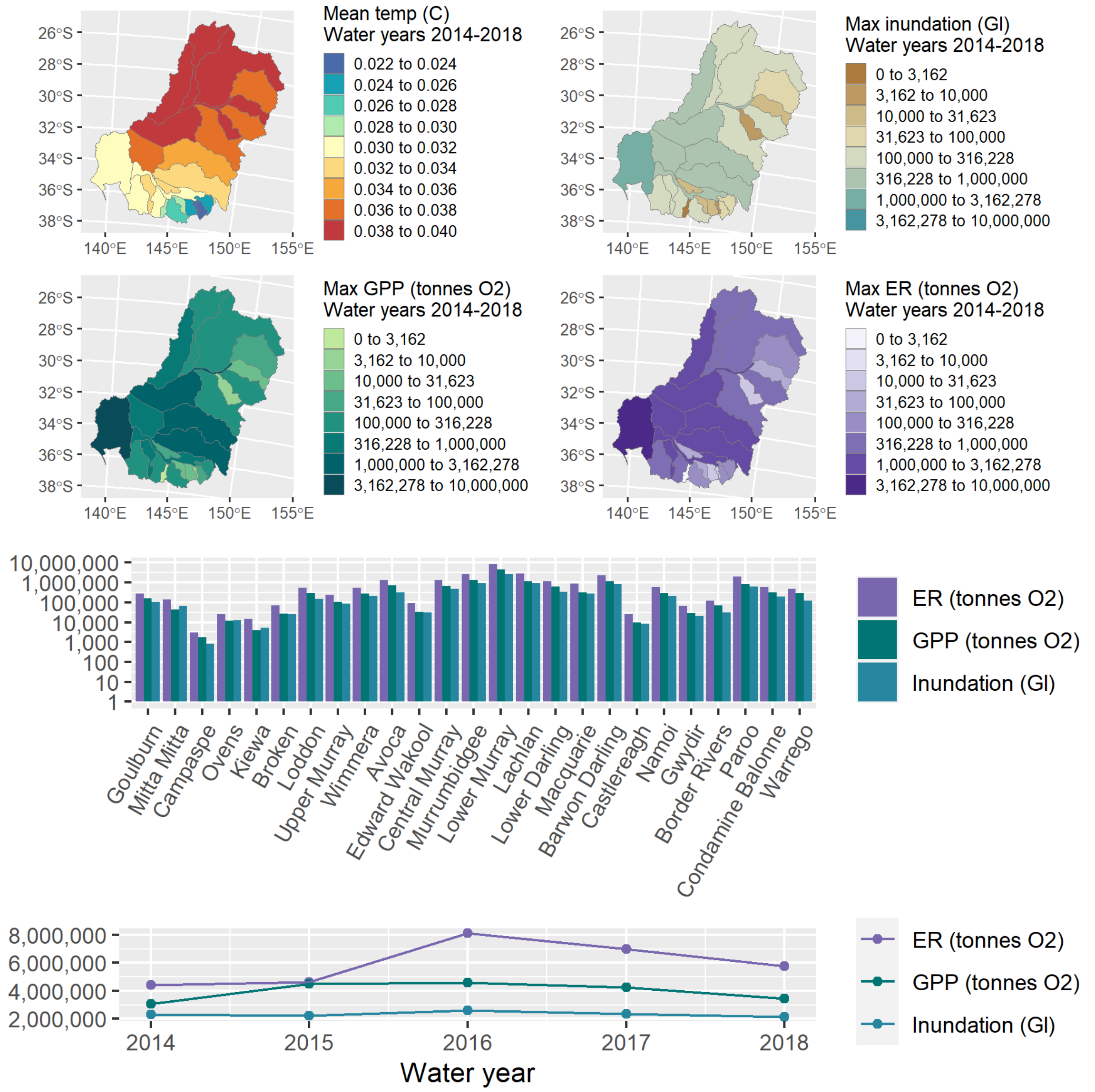

Figure 3.4 Aggregated inputs (mean temperature and maximum inundation volume) and outputs (GPP and ER) across space and time for the period of 2014 to 2019 water years

Spatial aggregation is done by averaging (temperature) or taking the sum of the maximum (Inundation, GPP, ER) of all wetlands in each catchment in the Basin at each timestep. Temporal aggregation is done by calculating the same statistic for the full 5-year period. Maps at top represent the inputs and outputs within each catchment, but the large range of values across catchments hides nuance. Bar charts (middle) can better capture small differences between catchments. Spatially aggregating to the basin scale instead of catchments shows the timeseries of total inundation and metabolism over the five years aggregated in upper panels.

Timeseries can illustrate trends for the whole Basin (e.g. bottom panel, Figure 3.4), but managers are also likely to need to understand how catchments might differ through time, perhaps to see whether different areas of the Basin are being more or less affected by changing climate. Once again, maps can provide at-a-glance striking results that accentuate the geographic differences as well as temporal change (e.g. top panel, Figure 3.5). To investigate these outcomes with more nuance, line timeseries could be produced for a selected set of catchments (e.g. bottom panel, Figure 3.5) or, indeed, all catchments, allowing managers to assess whether the catchments respond similarly through time. Matching such detailed outputs with maps (as shown here) provides the benefit of each sort of visualisation of how conditions change through the Basin in time.

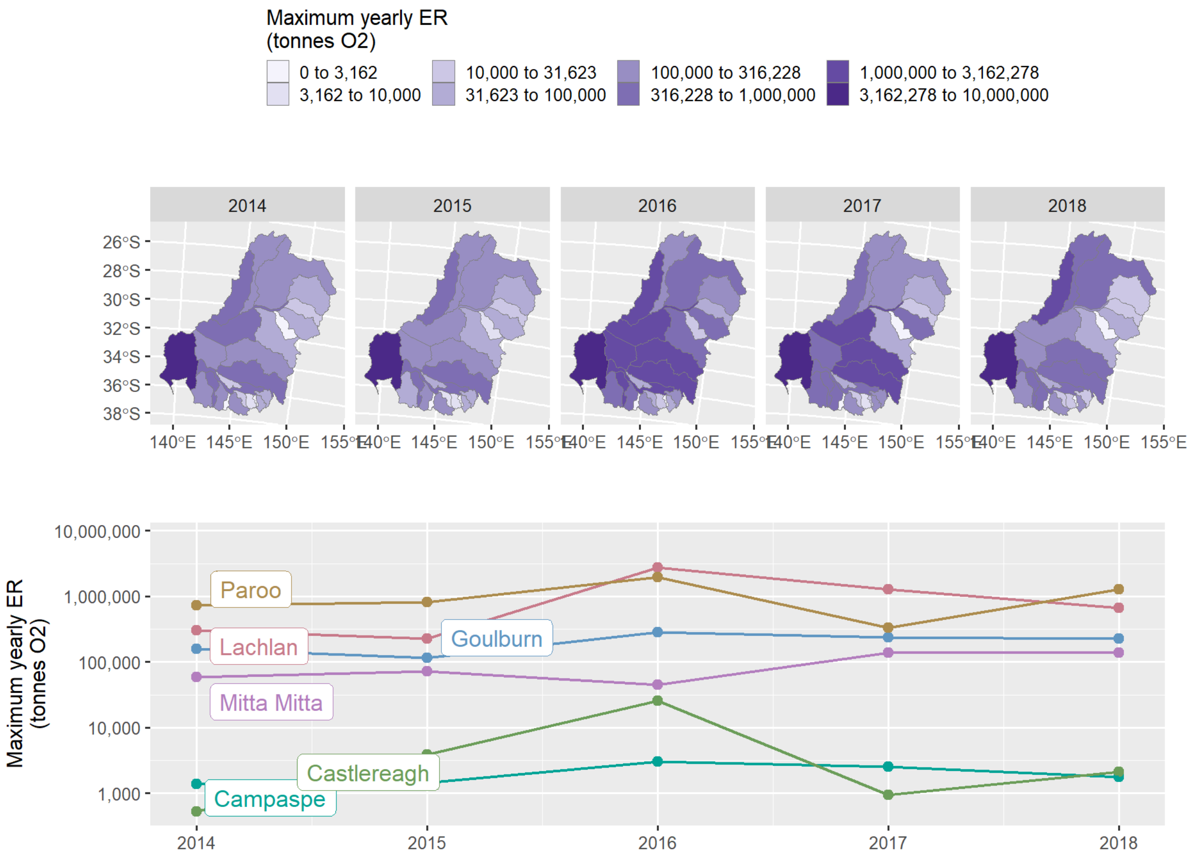

Figure 3.5 Timeseries of ecosystem respiration in each catchment within the Basin represented as maps (top) and a more traditional line graph (bottom)

Similar plots could be produced for all inputs and outputs with eFlowEval if desired.

3.3 Scenario modelling: wetland metabolism demonstration

A powerful tool for management planning is the ability to simulate ecological response to plausible scenarios. These can simulate the effect of proposed management actions (e.g. adding environmental water) or unmanaged environmental changes (e.g. changes in temperature due to climate change). Scenarios can also be used to visualise trajectories of change, either retrospectively (e.g. warming in the Northern Basin) or prospectively (e.g. onward climate shifts in temperature, inundation or changes in environmental water allocation or management). The eFlowEval functionality described above can be used to produce such scenarios showing either broad-scale patterns with maps (Figure 3.6) or specific comparisons using line and bar charts (Figure 3.7).

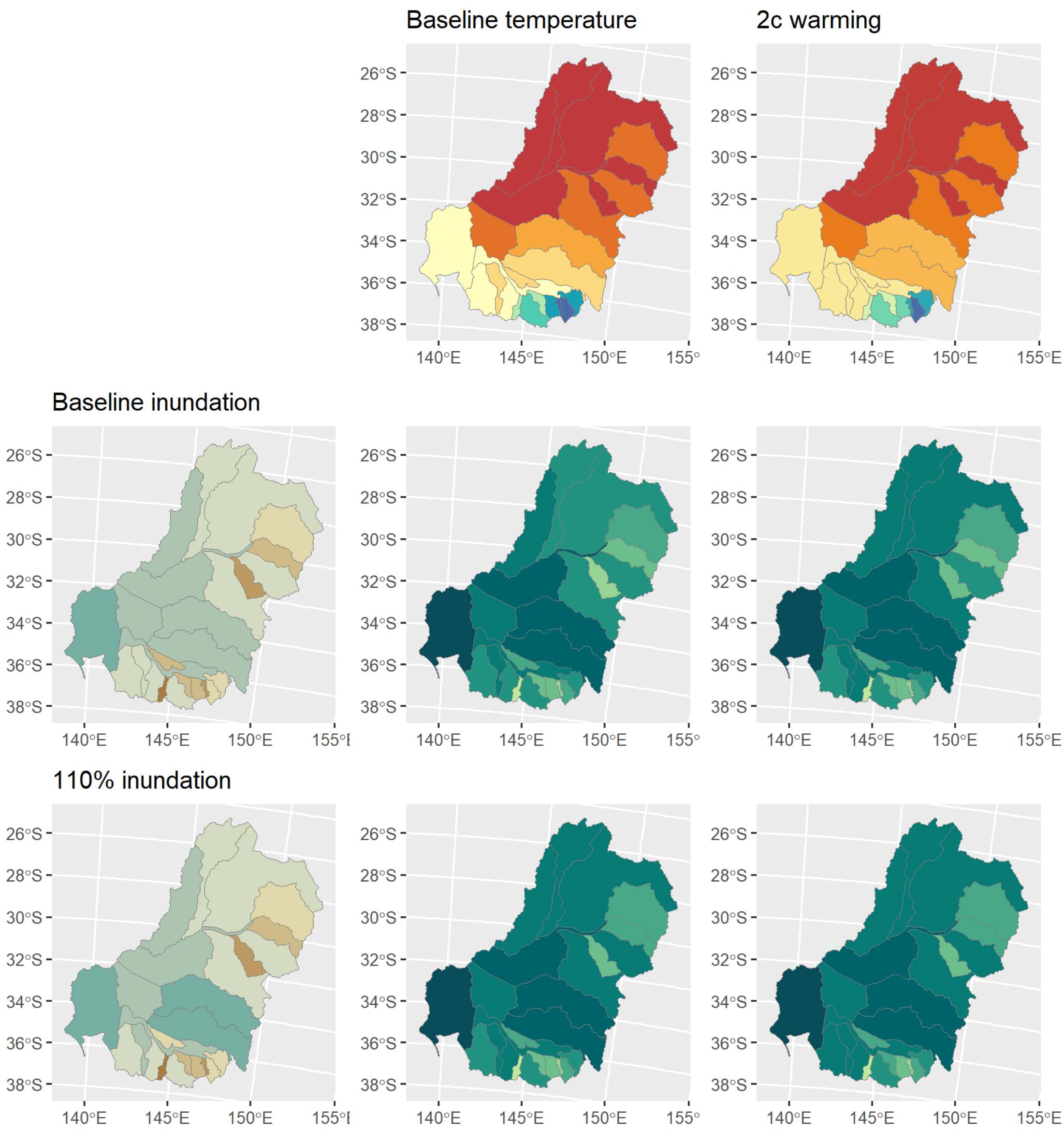

Figure 3.6 Demonstration of baseline temperature and inundation, along with scenarios of uniformly increasing temperature by 2 degrees C and inundation volumes by 10%

This is a demonstration of how results could look, and how they can be useful for management, not a representation of realistic changes in either inundation or temperature. Results are for the aggregated water years 2014-2019.

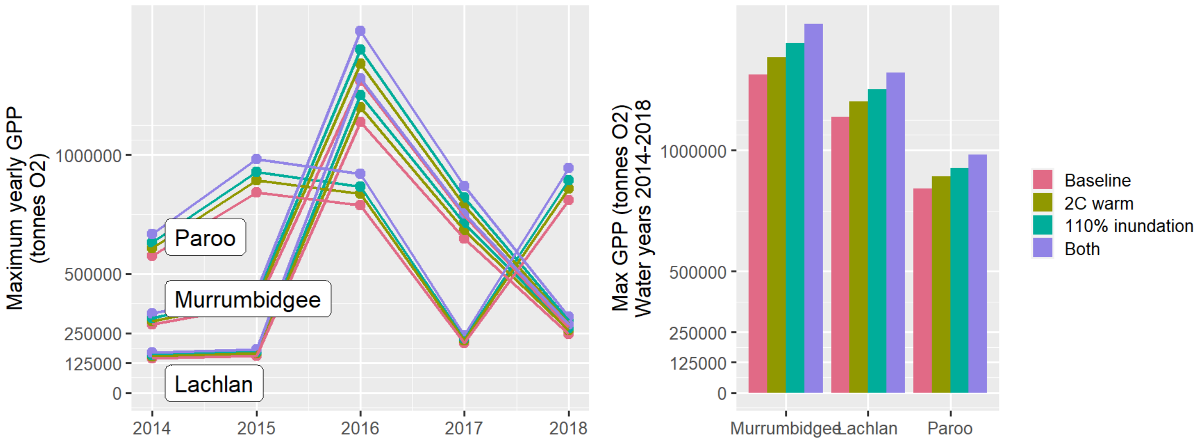

Figure 3.7 Scenario comparisons of GPP for a selected subset of catchments
Timeseries (left) and bar plots (right) provide more quantitative discrimination of differences between scenarios and catchments than may be visible with chloropleth maps.

While results can be presented at large spatial scales (e.g. Basin scale as shown above), parallel visualisations can also be done at relevant local scales (Figure 3.8). The ability to simulate ecological response under a range of scenarios would be extremely useful for local water managers to target particular outcomes and compare potential alternative actions. In addition, it can assist with understanding likely future trajectories for regions of high conservation status, for example. Finally, local scale scenario modelling is likely to add to our ability to understand and investigate relationships in the model, as this is the scale where management knowledge is best developed, and anomalous results are most likely to be identified, leading to opportunities to refine the underlying driver-response models.

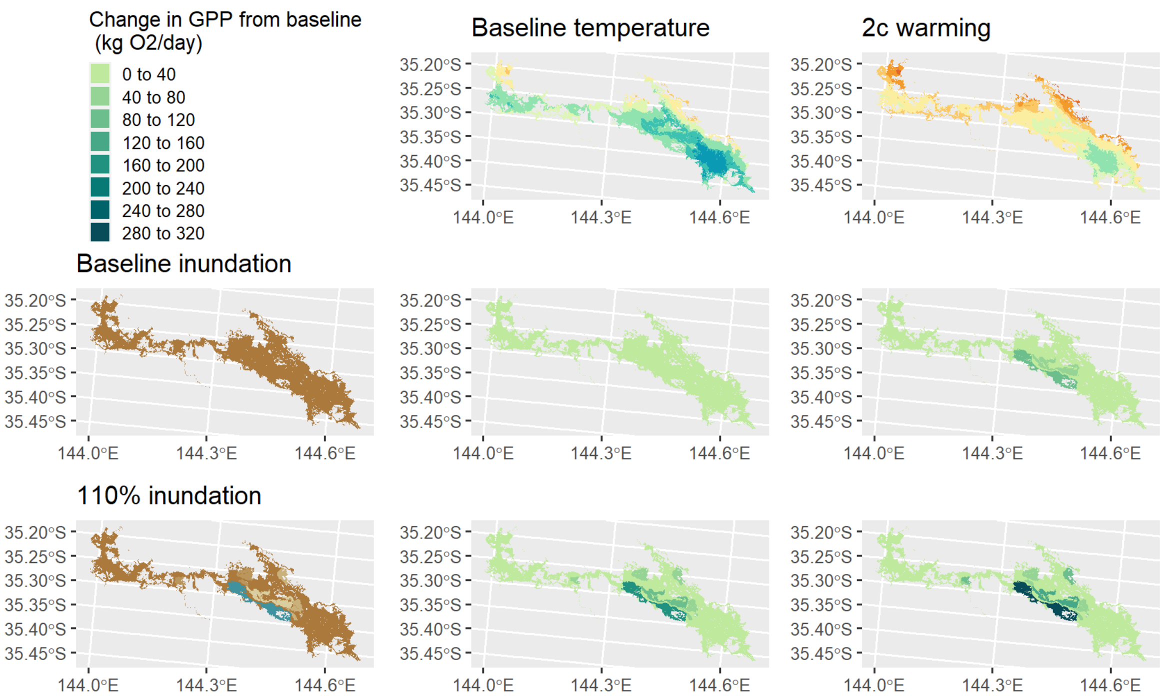

Figure 3.8 Scenario comparisons for each wetland in the Werai forest
Colours show difference in production from the baseline scenario (top left). As above, scenarios are defined by a uniform temperature increase of 2°C or inundation increase of 10%. Values are aggregated over the water year ending 2016-06.

3.4 Assessing critical strictures in life cycles: royal spoonbill demonstration

Capturing dependencies in ecological response is fundamental to ensuring that management decisions have the intended impact. Dependencies can occur in the life histories of species of interest (e.g. nesting can be supported to allow fledging but, if juveniles are unable to successfully forage, then they will not recruit to the adult population), in space (e.g. adults must have access to suitable foraging locations within a certain distance from breeding locations to enable them to feed chicks in the nest) or through time (e.g. suitable foraging habitat for juveniles must follow suitable breeding habitat in sequence to be useful). Failure to consider and capture such dependencies is one possible explanation for increasing evidence that short-term success does not always lead to long-term gain (e.g. high nest survival in royal spoonbills has not arrested adult population decline; McGinness et al., 2019; Kingsford et al., 2012).

Dependencies among life history stages and in space and time can be captured using the eFlowEval framework. Here we demonstrate those for the breeding and foraging stages for royal spoonbill, using a strictures and promoters approach to identify critical thresholds (Appendix C; Lester et al., 2020). Lester et al. (2020) set out a large number of strictures for waterbirds, using royal spoonbill as an example. Here we focus on a subset of those which are best supported by current knowledge and are thought to be most critical, and particularly those that are water driven. How well this subset captures critical life-history processes is as yet untested but is sufficient for the purposes of this demonstration.

Breeding habitat is identified as occurring within ANAE types that are known to support breeding based on an analysis of breeding event records (McGinness et al., 2020). Conditions suitable for breeding occur when those polygons are inundated to a depth of between 0.5 and 1.5 m for a minimum of 6 months (assessed as a minimum of 3 bimonthly inundation periods using the same inundation data layer as for the metabolism demonstration). Inundation needed to occur during the breeding season, which was implemented differently for the Northern and Southern Basin due to observed variation in breeding, where the breeding season in the Northern Basin was September to April inclusive and October to March in the Southern Basin. For a wetland complex to be identified as being likely to support a breeding event within the demonstration, a minimum area of that wetland needs to meet all strictures as a hurdle to assuming initiation of a breeding event. Here, for demonstration purposes, that has been set at 70% of the maximum area in the available record (1988–2020) that meets the inundation requirements for breeding. This is intended to represent the hurdle of sufficient suitable habitat within a wetland needing to be available before a breeding event commences.

Foraging habitat was also limited to ANAE types that are known to support foraging. These were defined as suitable for foraging when they were less than 0.4 m deep with no constraints on timing of availability and less than 10 km from a breeding site (i.e. for use during breeding events). It is recognised that feeding associated with breeding can also occur in the wetland complex supporting the breeding event, but inclusion of nearby sites enabled demonstration of the spatial look-around function within eFlowEval. This demonstration occurs at the scale of wetland complexes (i.e. groups of ANAE polygons that are known to support breeding events).

We demonstrate the ability of the eFlowEval framework to capture dependencies among life-history stages and in space using 3 wetlands in the Southern Basin: Koondrook and Perricoota Forests, Millewa Forest and Werai Forest. These wetlands have previously been suitable for breeding by royal spoonbill. Based on the strictures described above, we calculated the area of suitable breeding habitat for the period of 2013 to 2017 (Figure 3.9).

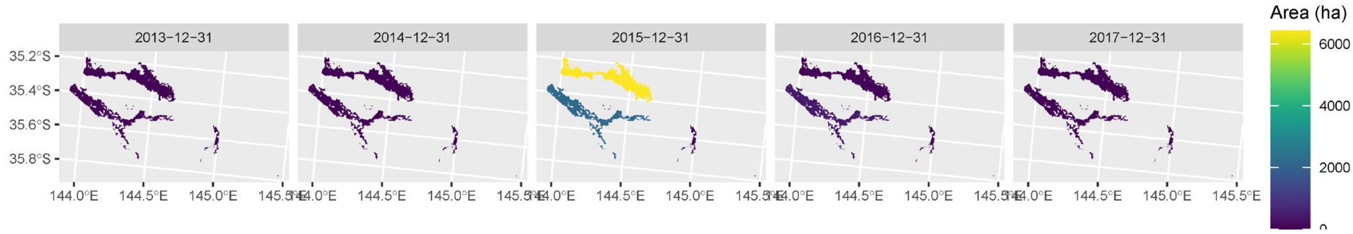

Figure 3.9 Area of breeding habitat for royal spoonbill within 3 selected wetlands in the Southern Basin
For each of 3 wetlands identified as suitable breeding habitat in the Southern Basin, the area of suitable breeding habitat for the period of 2013–2017. The area is a sum across individual polygons that have met the required strictures.

To assess the initiation of a breeding event, the area of breeding habitat available in each year for each wetland was compared against a minimum threshold, set at 70% of the maximum extent of suitable habitat in the historical record. This converts the continuous measurement of area shown for selected years in Figure 3.9 to a binary assessment of a likely breeding event initiated (Figure 3.10).

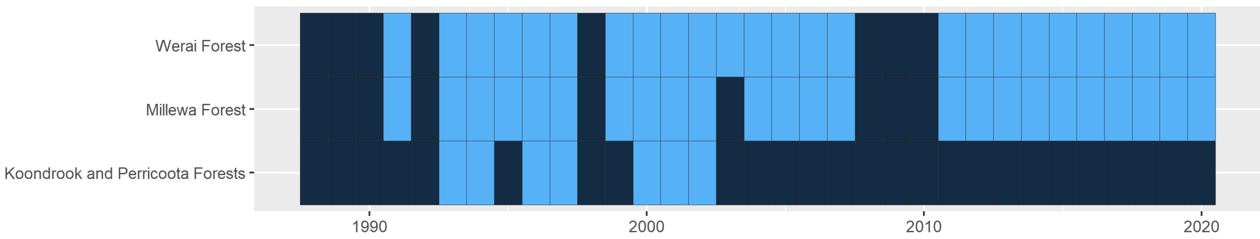

Figure 3.10 Years when the total area of inundated habitat within the 3 selected wetlands passing the required strictures, exceeded the required minimum threshold for breeding
For each of 3 wetlands identified as suitable breeding habitat for royal spoonbill in the Southern Basin, the years in which 70% of the maximum area observed historically for breeding exceeds the required strictures described in text are shown in light blue. The maximum area observed historically is assessed based on the available record of inundation (1988–2020) and captures the maximum area which meets the inundation requirement for breeding in that timeframe.

To incorporate feeding requirements, for each of those 3 wetlands, a kernel with a 10-km radius was then constructed. Within that kernel, the area of suitable foraging habitat was then identified through time. Figure 3.11 illustrates the foraging habitat within the breeding wetland (a), the kernel around that breeding wetland in which suitable foraging habitat may be found (b) and then the total foraging habitat for the wetland, combining within-wetland and the additional habitat in that foraging kernel (c). The amount of foraging habitat within that kernel is illustrated in Figure 3.12. As for initiation of breeding events, a threshold was required to describe sufficient available foraging habitat. Here that was set at 5% of the historical maximum extent, for demonstration purposes. Little information is available regarding a likely true value for that threshold in this system (McGinness, pers. comm.).

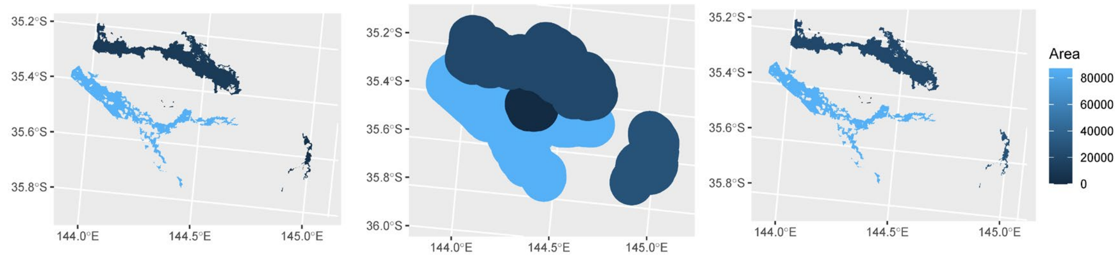

Figure 3.11 Total foraging habitat available for breeding wetlands within the wetland itself (a), within 10 km of each wetland (b) and then combined (c)

For each of 3 wetlands identified as suitable breeding habitat in the Southern Basin, the area of suitable foraging habitat within a 10-km radius has been identified. In this example, the additional area of foraging habitat within 10 km is minimal and insufficient to change the colour displayed in (c).

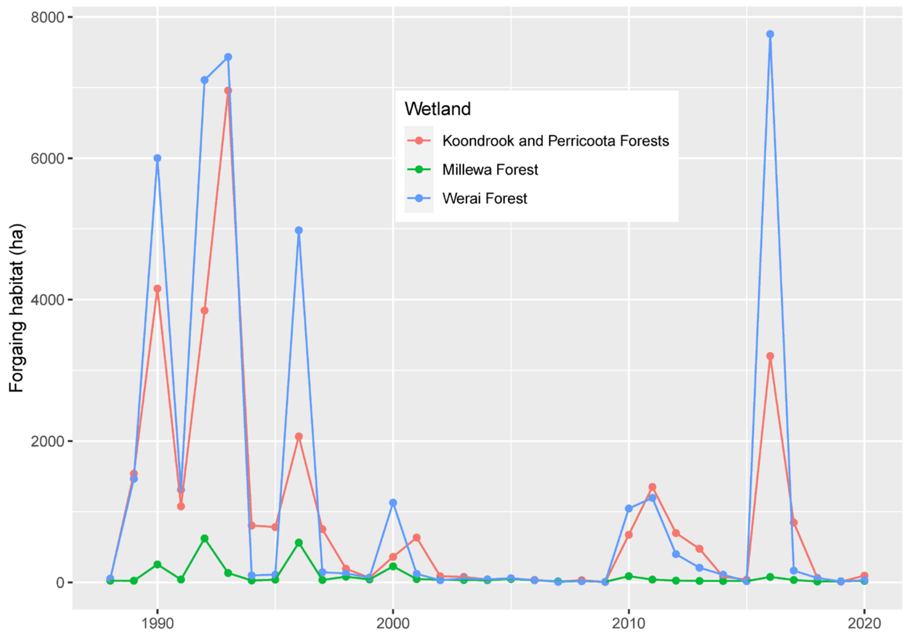

Figure 3.12 Area of foraging habitat available within 10 km of each wetland

For each of 3 wetlands identified as suitable breeding habitat in the Southern Basin, the area of suitable foraging habitat within a 10-km radius has been identified (ha).

Finally, requirements for foraging are combined with the earlier breeding initiation requirements to illustrate wetlands that were likely to be able to support successful royal spoonbill breeding across the available record (1988 to 2020) (Figure 3.13). We stress that these outputs are for the purposes of illustrating the capability of the eFlowEval framework, rather than intended to be an accurate assessment of breeding success.

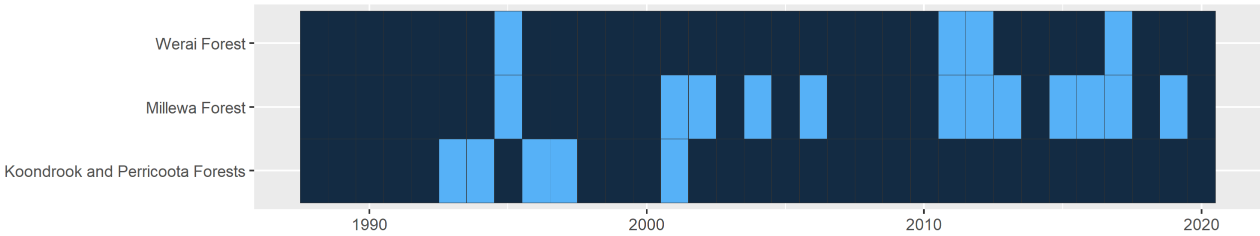

Figure 3.13 Suitability of 3 wetlands to support royal spoonbill breeding based on habitat requirements for breeding and foraging

For each of 3 wetlands identified as suitable breeding habitat in the Southern Basin, whether there is sufficient area (i.e. above the specified thresholds) for both breeding and foraging, indicating areas of likely breeding success through time.

4 Representing uncertainty

Dealing with uncertainty in frameworks such as eFlowEval can be fraught. Assessment of uncertainty is an important but often overlooked aspect of model development. Many models do not explicitly present the associated uncertainty and this can create the erroneous impression that they are precise, which can lead to suboptimal decisions relying on that precision (Chatfield, 2001). All models are uncertain and that uncertainty can take multiple forms (Ascough et al., 2008). There are 3 types of uncertainties relevant to ecosystem management: unknowable outcomes arising from complex and dynamic systems; poor or deficient understanding of physical or ecological principles upon which the model is based; and poor data quality, sampling bias and analytical errors (Christensen et al., 1996; Hilborn and Stearns, 1987). Some types of uncertainty are able to be quantified or characterised, such as sampling error and modelling error (e.g. associated with regression used here) or stochastic uncertainty (e.g. uncertainty associated with what the weather will be for the next year; Ascough et al., 2008). However, there are also sources of error that are unquantifiable and unmeasured, such as missing data or assumptions in model development.

Mathematical constructs can also present challenges. For example, at such large scales, the number of individual data points in any application is so large as to invoke the law of large numbers. Under that scenario, the effect is for uncertainty to disappear if dealt with naively by assuming the uncertainties are independent. This situation is not reflective of our level of confidence in the estimates, partly because of the lack of independence among the data points. Alternatively, simply propagating maximum possible uncertainty through time and space (by assuming uncertainty is perfectly correlated) can lead to estimates that are so broad as to also be effectively meaningless, especially for management purposes and decision support. Thus, careful consideration is required to deal with and represent uncertainty.

There are multiple ways to attempt to capture uncertainty in a meaningful and useful manner. Sampling and modelling error can be calculated and incorporated. This approach is relevant for the estimation of terms in process-based models or for the regressions used here for the wetland metabolism demonstration, for example, which can then be illustrated as prediction intervals (e.g. Figure 4.1, Figure 4.2) (Chatfield, 2001). For stricture-based models, error can be estimated at the level of each stricture, or in rates of change.

Confidence intervals are a common way to express uncertainty but are also dependent on sample sizes and so shrink with very large sample sizes such as those used here. This dependence on sample size is because confidence intervals assess confidence in an *estimated statistic* (e.g. regression fit or mean), not the range of potential *data values*. In contrast, prediction intervals provide a range between which the real values are likely to occur with some probability (Chatfield, 2001). So, by illustrating prediction intervals, we can show the range of potential outcomes to better assess what happened (or might happen) across a landscape. This approach is valuable both for evaluation, where we might want to ask what the likely metabolic activity was as a result of past conditions, or for future forecasting. In both cases, we use prediction intervals because we are interested in the plausible range of metabolism values, not the quality of the relationship between temperature, volume, and metabolism.

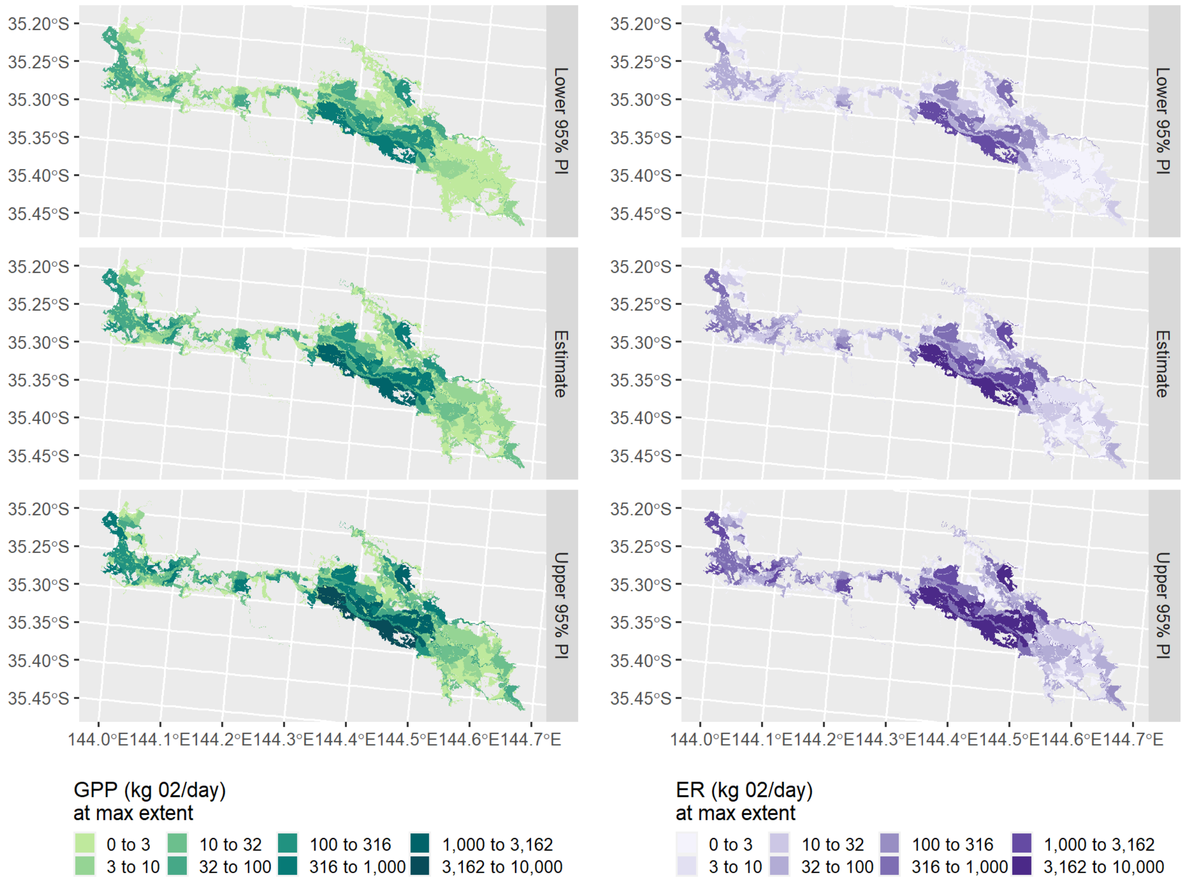

Figure 4.1 Estimated metabolism given temperature and inundation volume for wetlands in the Werai forest (middle panels), with lower and upper 95% prediction intervals in top and bottom panels, respectively
Because these are prediction intervals, not confidence intervals, they mean that 95% of the time we would expect the true metabolic activity to be between them for these temperatures, with the most likely outcome shown in the middle panels. The modelled period is the 2 months preceding 1 November 2016, and reflect the maximum inundation extent during that period, as that is the scale of the inundation dataset.

Once calculated, these sampling and modelling errors can be propagated in space and time. Each individual wetland has a prediction and as a prediction interval that captures 95 % of the potential outcomes based on the temperature and inundation. When these predictions are spatially aggregated to the catchment (or other larger scale), the overall prediction interval for the catchment requires additional information – the magnitude and distance of how errors are autocorrelated. Do model outputs for many nearby wetlands all ‘miss’ in the same direction (i.e. are they all too high or too low)? How nearby and by how much? In the absence of this information, we can only assess 2 extreme and unrealistic endmembers – either all wetlands are independent and so, where a given wetland falls in the prediction interval, it has no bearing on its neighbours, or all wetlands in the catchment are perfectly correlated and so, all wetlands in the entire catchment are always at the same percentile in the prediction interval. The former case yields effectively no uncertainty (due to the law of large numbers) and the latter unrealistically wide ranges for uncertainty (e.g. Figure 4.2). A Bayesian approach incorporating autocorrelation and uncertainty would then produce more realistic uncertainty estimates because, if one wetland was close to the lower 95 % prediction limit, say, then others around it are likely to be low as well. Similarly, summarising uncertainty through time relies on understanding how errors are correlated through time. At the moment, these correlations are unknown for the demonstration case of wetland metabolism but could be measured in the future.

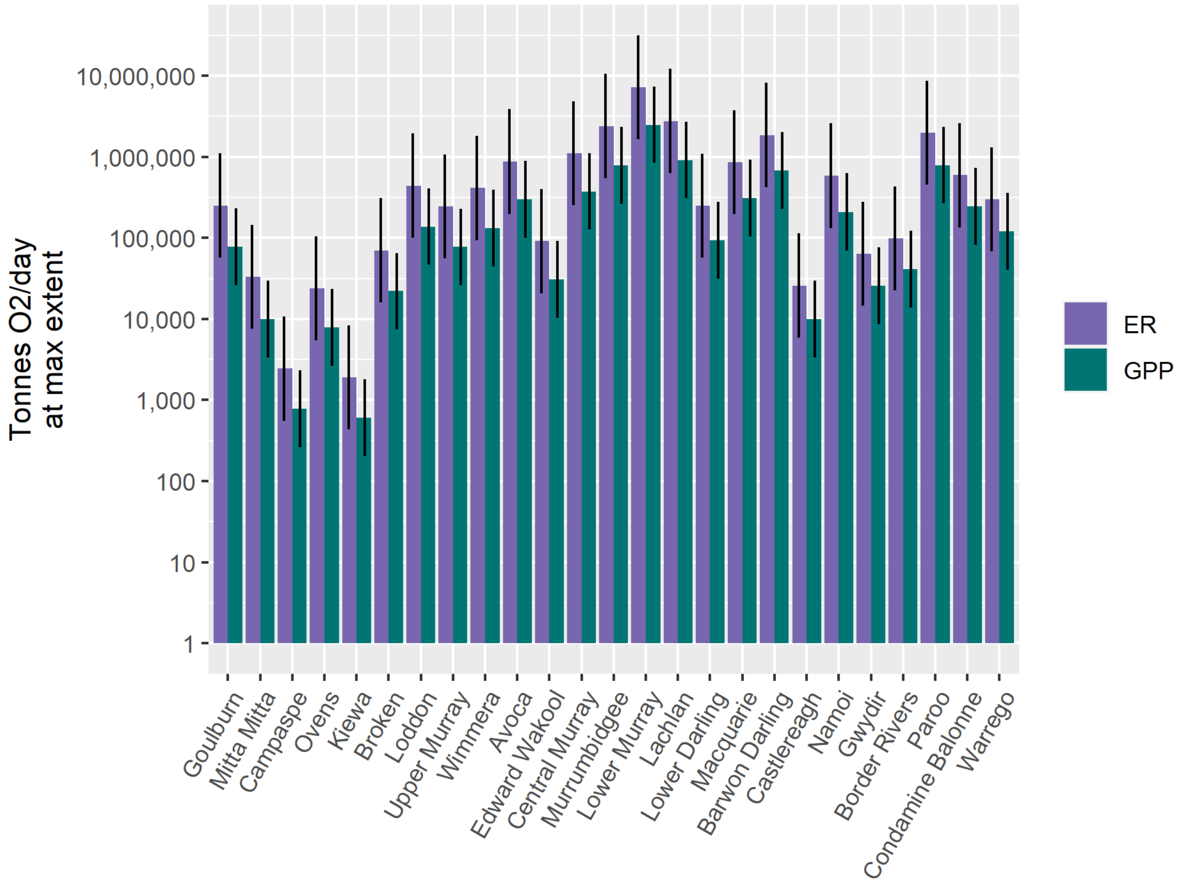

Figure 4.2 Predictions (bars) and 95% prediction intervals (black lines) for each catchment with perfectly correlated errors

With independent errors, the law of large numbers means that the prediction intervals are undetectable - the mean prediction is the nearly identical whether all wetlands are at their separate mean predictions, or if they independently vary within their separate prediction intervals. The intervals may appear smaller and more symmetrical than expected, but this is because of the use of a log y-axis.

Whether or not uncertainty can be quantified or propagated, a common issue with management-relevant data is that there are often some locations or situations where a relationship is better established than others. Extending beyond those locations requires (more) extrapolation. Thus, there is a trade-off – including the whole basin requires accepting greater model uncertainty, but it can be done. In contrast, if managers are not willing to make that trade, better predictions are available, but only for a subset of catchments. This trade-off and associated uncertainty changes can be visualised explicitly by limiting outputs to areas where data are more available (e.g. catchments for which wetland metabolism regressions have been specifically parameterised) or where colour intensity could be used as a qualitative indication of the uncertainty associated with the estimates (e.g. paler colours could indicate more uncertainty; Figure 4.3). Note, however that this is only one of many potential sources of uncertainty in these predictions, with others described in this box also relevant.

The eFlowEval framework is able to represent uncertainty in each of these ways. At present, we provide demonstrations for a very limited set (e.g. wetland metabolism with and without a catchment term) because the data and information to enable more sophisticated treatments of uncertainty do not exist (e.g. the scale and level of autocorrelation among catchments). Despite this, we are still able to provide useful estimates of the degree of certainty that exists for the various outputs, to ensure that is transparent and able to be accounted for in decision making processes.

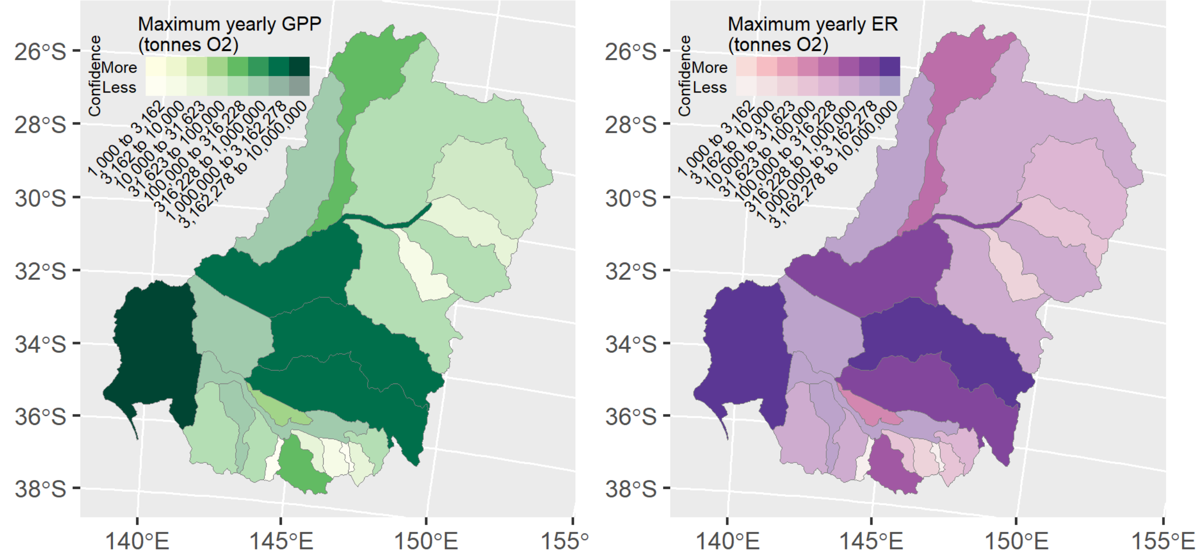

Figure 4.3 Estimated metabolism given temperature and inundation volume for all catchments in the Murray–Darling Basin, shaded to indicate confidence in the estimates

More vibrant colours are used for only those catchments for which data were available to parameterise regression models, while faded colours indicate catchments with extrapolated predictions.

The level of certainty regarding estimates is higher for individual catchments used for regression parameterisation. The modelled period is the 2 months preceding 1 November 2016, and reflect the maximum inundation extent during that period, as that is the scale of the inundation dataset. The regression models used differ between the vibrant and faded catchments. For catchments with vibrant colours, each catchment-specific regression is used. In catchments with faded colours, a global regression without a catchment term is used. Refer to Appendix B for additional detail.

5 Discussion

We successfully developed a modelling framework to enable the flexible scaling and integration of responses to environmental flows. The eFlowEval framework provides a transparent basis for evaluating response of multiple target groups in a common system and offers the capability to capture and update best-available knowledge gained from in-depth research and monitoring. The framework has the capacity to provide assessments across a range of scales, from local wetland to whole of basin, and from short time frames (e.g. weekly/monthly) through to multi-year assessments. We have incorporated advice from water managers regarding the types and scales of outputs that would be of most use, and so have developed example output at a scale and resolution that will be suitable for natural resource management.

Across the 2 demonstrations presented, we illustrate the functionality of the eFlowEval framework. The metabolism demonstration illustrates the range of spatial and temporal scales at which eFlowEval outputs can be generated. This involved developing the capability to scale responses from the local scale at which processes occur (e.g. at the scale of meaningful polygons or smaller) to larger scales (Gawne et al., 2018a). Scaling responses in this way necessarily increases uncertainty as scales become larger because of the range of ecological processes responding to multiple scales (DeFries et al., 2004), but the eFlowEval framework can represent that uncertainty either quantitatively or qualitatively. The use of prediction intervals ensures that any quantitative estimates of uncertainty will be realistic and focused on the range of likely values, rather than the uncertainty associated with the underlying model estimates (Chatfield, 2001).

Feedback from water managers suggested that the range of scales was useful across multiple management goals – across evaluation and planning, scales from local to basin were relevant. For example, one feature that was suggested by water managers was the value in matching information about drivers as well as the environmental responses themselves. Thus, we developed a presentation option including those drivers. We illustrated other presentation options as well. Many others are possible and could be developed to suit specific use cases and are best developed in collaboration with the intended end users.

The metabolism demonstration illustrated the ability of eFlowEval to assess quantitative responses and to compare responses across scenarios of plausible futures. These could include different future climates or different options for watering actions, or other natural resource management (e.g. as per Lester and Fairweather, 2011). Further capability, to develop synthetic outcomes from the environmental response, are also possible. For example, gross primary productivity and ecosystem respiration are able to be converted into a measure of heterotrophy/autotrophy for each polygon (not shown), and other processing of responses to variables of interest are possible. These could include measures designed to be readily interpreted by the general public (e.g. gross primary productivity could be converted to equivalent numbers of cows fed or equivalent hamburgers to assist with communication).

The royal spoonbill demonstration illustrates different capability in the eFlowEval framework. The driver-response model was a contrast to that presented in the metabolism demonstration. There, a quantitative regression model linked drivers to primary productivity and respiration. Here, for royal spoonbill, a life-history based threshold approach was used, based on the strictures and promoters framework (Lester et al., 2020). This type of driver-response model is extremely parsimonious – incorporating the critical thresholds that are most likely to affect environmental outcomes at each life history stage. The risks associated with using such an approach are associated with ensuring that the most appropriate thresholds are identified and included to maximise the likelihood of outcomes occurring as modelled. For example, previous thresholds identified by experts have not matched those selected in a subsequent data-derived identification of drivers (Lester, 2019), so caution is warranted. However, the focus of eFlowEval is in the use of existing driver-response models, rather than in their development.

The royal spoonbill demonstration also illustrated some of the novel capability of the eFlowEval framework compared with other similar tools. In particular, the ability to include areal look-arounds (e.g. identifying whether there is suitable foraging habitat surrounding breeding sites) is new capability, as is the ability to include multi-scale driver polygons. The demonstration also included the ability to include different parameters at different locations (e.g. different windows for breeding to occur in the North and South of the Basin), as well as incorporates a minimum area threshold for the initiation of a breeding event, or for available foraging habitat.

The eFlowEval framework offers a number of other improvements on similar previous efforts. The framework has been developed to include significant levels of flexibility in the types of driver-response relationships, inputs, aggregation and comparison methods possible. The focus on capturing the most important drivers across life history maximises the likelihood that limiting factors will be identified without unduly increasing model complexity or computational requirements (Lester et al., 2020). The scaling method is robust, given that outputs (rather than drivers) are scaled and aggregation methods have been selected with care (noting that the scaling inherently incorporates uncertainty; Englund and Cooper, 2003). The ability to include inter-dependence across species, time steps and spatial locations means that the framework does not assume independence across those components – an assumption that we know is false but one that is common in ecological models to support natural resource management (Lester et al., 2020). Thus, important and realistic interactions can be explicitly included and effects accounted for. The eFlowEval framework also provides unprecedented flexibility to investigate patterns from local scales at which processes occur to the whole of basin.

One of the potential advantages of such a framework is that it provides the ability to capture institutional knowledge. Many decisions in natural resource management are made by local managers with extensive experience in a particular location (e.g. a wetland or wetland complex) and that expertise may not be captured (Hilborn, 1992). This creates risk of a loss of institutional knowledge should that experienced local manager move on and also tends to result in a lack of transparency if the rationale for decisions is not captured, regardless of the success of the action (Hilborn, 1992; Greca and Moreira, 2000). A framework like eFlowEval, if implemented carefully, can be a repository for that institutional knowledge and create transparency and repeatability in decision making. The implementation of the framework could occur via a two-way exchange of knowledge, where initial driver-response models are developed using best-available science that is then validated against the expertise of local managers. Enabling side-by-side comparison of drivers and responses enables users to better visualise how outcomes arise, enabling direct interface with expert opinion and making the model itself transparent for its users. This would also enable local idiosyncrasies in response to be incorporated to minimise uncertainty arising from factors not explicitly modelled. Developed in this way, models are more likely to be accepted and trusted by practitioners but also more likely to adequately represent responses (Lester et al., 2020). This means that models such as an implementation of the eFlowEval framework could provide new insights into what might happen under a given set of conditions and, importantly, why, enabling explicit hypotheses about mechanisms to be tested (Sutherland, 2006).

The results shown here are not intended to provide projections of metabolism or bird outcomes *per se*. The driver-response models used (particularly for metabolism) have been adapted to enable them to be used as demonstrations of eFlowEval (Appendix B, Appendix C). One of the outcomes of approaching the demonstration in this way is that it highlights the utility of the framework to demonstrate elements of our ecological understanding that are missing and assist in the direction of future data collection (Lester, 2019; Polasky et al., 2011). For example, quantitative estimates of uncertainty are not possible to provide without understanding the autocorrelation of errors in space and time. Similarly, the inundation input data (as a maximum extent on a bimonthly timestep) limit our ability to estimate total metabolism. This is because there is not a method to estimate total metabolism in space and time from a maximum extent. So, we report the maximum metabolism associated with that maximum inundation over the aggregated unit (i.e. time or space). The outputs are therefore constrained and are, perhaps, less useful to managers than total metabolism would be, but still provide a value that can be compared in space and time. Thus, the outcome represents best available science at this time and is preferable than no providing an outcome until data deficiencies are addressed (e.g. by finer resolution inundation data or explicit relationships developed between total metabolism and maximum bimonthly inundation extent; Ryder et al., 2010). Furthermore, identifying which information would be needed to improve our ability to project responses would be a key output from the framework as it is implemented.

Future work could extend the application of the framework to assessing differences relative to a counterfactual of no environmental watering, for example. Counterfactuals are scenarios that are the same as the scenario of interest in the absence of the policy choice (so here would exclude environmental water; Thomas and Koontz, 2011). Counterfactuals enable the testing of putative causal mechanisms, and also evaluation of management actions (Thomas and Koontz, 2011).

Another development could allow the creation of an ability to identify sequences in the past with similar flow events, or to construct future environmental flow delivery; this would enhance the decision-making focus of the tool, moving it towards a tool capable of actively assisting decision-making at a range of scales from local to basin-wide. Then, for example, managers would be able to determine whether the most appropriate watering actions in a wet year were the same as those in a dry year, or if strategies that were the most favourable under the current climate would continue to be so under a drier future climate. Strategies that involved watering for a particular biotic group (e.g. fish) could be assessed for identifying synergies across groups. Other questions that could be addressed using this approach include:

- What difference does environmental water make compared to a scenario without those flows (i.e. the counterfactual)?
- What flow management strategies provide the best outcomes?
- What are the potential impacts of climate change and extreme events?

The framework could then form part of an adaptive management cycle by: formalising conceptualisations of system processes; creating and testing hypotheses from known environmental outcomes; aiding decision-making for future environmental water allocation; and identifying where knowledge and data collection can be improved (Lester et al., 2020; Sutherland, 2006).

6 Conclusion and recommendations

6.1 Contribution to Flow-MER objectives

The development of a common framework for linking environmental outcomes to hydrology provides a mechanism for evaluating the outcomes of management across the Flow-MER Themes and Selected Areas. The modelling framework enables extrapolation from Selected Areas to unmonitored locations and enables transitioning from considering outcomes at local scales to the basin scale. The framework facilitates comparisons across taxonomic groups to articulate specific relationships and responses, enabling the identification of potential synergies for environmental water use. Key questions that could be addressed by implementation of this framework could include:

- What are the likely outcomes of Commonwealth environmental water in monitored and non-monitored areas considering a range of biota and their needs?
- How can environmental outcomes be evaluated at a basin scale, considering both local and basin-scale outcomes?
- How can the benefits of environmental watering be communicated to a range of audiences?

6.2 Recommendations

We recommend that the eFlowEval framework be implemented for a broader number of species or processes. Additional development, in collaboration with water managers, would continue to enhance the utility of the framework. Additional functionality, enabling users to identify sequences in the hydrologic record that share characteristics with forecast conditions, to enable different methods of environmental flow delivery to be compared and to add bespoke environmental flow delivery. Integration of the eFlowEval framework implementation with other visualisation products from Flow-MER would also be desirable.

Appendix A Model platform and data

The eFlowEval framework provides a general platform for assessing ecological and other outcomes at the basin scale by providing a consistent modelling approach and platform to incorporate data, assess response and scale, and present results. To achieve these outcomes, the framework has a defined series of steps that puts the necessary data into standardised formats and follows a consistent workflow. The workflow specifies a response model, creates standardised driver data, feeds that data to the model to assess predicted responses, and then typically aggregates those responses in space and time and presents results.

An essential feature of the framework is the ability to compare scenarios (e.g., compare to counterfactual or predict future outcomes under different conditions). These scenarios typically involve using the modelling capacity in the framework in parallel with different sets of inputs representing the scenarios. The different scenarios would run through the framework separately, yielding outcome results in the standard framework format, which are then compared to each other.

A.1 Base polygons and driver data

Driver-response models identify the necessary input data, based on the best-available system knowledge. Driver data should be available at the basin scale, or the largest scale at which the model is intended to be run. Driver data can include many formats, e.g. spatio-temporal rasters of temperature, point records of rainfall or flow at gauging stations, or inundation extent polygons. For consistency between workflows data type is standardised by merging formats into a standard set of base polygons, to which additional driver data is attached.

Using polygons as the base of the framework at the local scale provides several advantages. First, the shape and size of polygons are flexible, and so they can be both detailed and data-efficient. When conditions change over short distances, small polygons can capture those changes, while large polygons can be used when conditions are broadly similar. Flexible shapes allow them to reflect landscape characteristics, rather than a rigid grid structure. Further, large areas can be outside polygons entirely, avoiding the need to process areas of the basin unaffected by watering or irrelevant to the groups of interest.

Here, the examples use ANAE wetland polygons (Brooks, 2021) as these represent wetland habitat types, however any polygon may be used. ANAE polygons delineate water-influenced areas and capture other characteristics of those areas, and these wetland types are used throughout MER and beyond as covariates for many ecological responses. Wetland polygons and their attributes were obtained from the ANAE (v3) (Brooks, 2021), and the ANAE geodatabase (Brooks, 2017) was used to obtain catchment and basin polygons and other relevant information. Minor processing was performed on the ANAE data. First, LTIM (Long-term intervention monitoring) catchments were exported after removing the Northern Basin polygon to ensure the basins do not overlap. Second, the ANAE wetland polygons were intersected with the Koppen climate classifications to investigate interactions between ANAE type and climate. Finally, the ANAE polygons were intersected with catchment boundaries to allow processing and analysis in distinct catchments. The resulting polygons were geohashed to create unique identifiers and formed the basis for all further analysis.

The eFlowEval platform includes several methods to add additional environmental drivers as attributes to the polygons. These additions return polygon datasets with some relevant value for the environmental driver (e.g. temperature) in each polygon, potentially at many timesteps. Briefly, raster data is intersected with the polygons to calculate an area-weighted statistic (e.g., mean, maximum) that accounts for the area of each raster pixel that intersects the polygon. Incorporating data from other polygons proceeds similarly with area-weighted statistics on the intersection, though the platform also has the capability to split polygons to generate a new set of base polygons incorporating information from both sets if interactions are thought to be important. Point data (e.g. gauged rainfall) can simply be matched to polygons, and statistics calculated across the points if multiple fall in a polygon. Whether incorporated from other polygons, points, or rasters, the platform provides the ability to define custom functions for these statistics, for example the volume of inundation greater than 10cm depth or rolling windowed operations such as time since last temperature above 30 C. In many cases, the data being integrated into the base polygons will be timeseries (e.g., daily surface temperatures). In that case, output polygons with the integrated data attributes are generated at that same timestep.

Incorporating driver data into the base polygons yields a standard data format (polygons with driver data), but necessarily requires different processing for different data. Each response model will identify a specific set of drivers, possibly using different base data or data transformations. Moreover, different input data will have different formats and may require different statistics to aggregate into the base polygons. Thus, each set of input data has a unique processing script, though all follow a consistent workflow. In other words, eFlowEval takes a consistent approach to the processing of the drivers, while allowing flexibility for the specific drivers to reflect theme knowledge.

In use, the input data and base polygons are read in and given matching coordinate systems. Functions for how to aggregate into the polygons are defined, which include standard statistics (e.g., mean, maximum) or custom based on particulars of the response model (e.g., depths over a threshold). These functions can also include temporal rolling, capturing quantities including time since inundation or maximum temperatures over a specific timeframe. The base polygons and input datasets are then divided into multiple chunks (typically 100 per catchment) with the statistics calculated in parallel. The output is a dataset of base polygons with the appropriate calculated values of the input data in each, often at many timesteps. These are then fed into the response model.

A.2 Driver-response model

Much like the data input process, the driver-response model assessment involves a consistent workflow, but each driver-response model must be developed individually, given their different driver data and responses to it. The specification of the response models proceeds in consultation with Theme experts to translate best-available ecological knowledge for responses to environmental drivers into functions that can be assessed with available driver data. These response models are typically strictures and promoters, but can be statistical (e.g. the metabolism example) or more complex process-based models. The development of these driver-response models is discussed in detail in Appendix B (metabolism demonstration) and Appendix C (waterbird demonstration).

In use, the eFlowEval framework reads in the relevant driver data in the standardised format and uses the specified response models to calculate a response for each polygon at each timestep. The responses are then stored in the same polygon structure. In other words, this is a spatio-temporally explicit model with spatial units defined by polygons. By representing the responses in the same standard data structure, the responses themselves can be interdependent and multi-staged. For example, in the case of strictures and promoters, one life stage might respond to temperature, while the next life stage responds to inundation and the presence of the first life stage at some earlier time point (e.g. seed set requires previous successful germination and a growth period).

To achieve dependencies between life stages or other species often involves further calculation of derived values in the response model. Indeed, dependencies often involve treating previous responses as drivers for subsequent responses. By continuing to use the base polygons, the processing and analysis workflow remains consistent despite differences in the calculations done to reflect each response model or dependencies between response models. These calculations and derived datasets follow the same consistent approach as the original driver data, with the outcomes being some function (possibly user-defined) of the set of data in base polygons. These might include both 'snapshot' measures (stricture passing at each point in time) or rolling calculations on previous life stages to capture quantities like days since germination. As with the data processing, these response models perform many calculations at the scale of the base polygons at many timesteps, and so are broken into many chunks and run in parallel, yielding sets of output values in each base polygon and timestep.

A.3 Results processing

The outputs from the driver-response model are the simplest form of result from the eFlowEval framework, representing predicted outcomes at each polygon and timestep. The eFlowEval framework provides several next steps with these results. First, the standard results format allows for consistent and general extensions and development of new result presentations. Second, eFlowEval provides some standardised types of results presentations, including mapping, various quantitative plotting options, and tables. Finally, aggregation of the results to management-relevant spatial and temporal scales, combined with scenario comparisons, provide critical results interpretation.

A.4 Aggregation

In some cases, results are useful at the local scale (i.e., individual base polygons) when assessing the outcomes for a small area of the basin, as illustrated in the text for metabolism in the Werai forest and bird breeding. In this case, the response model outcome data can be accessed directly and plotted as maps or other graphical or tabular forms. In most cases, however, the eFlowEval framework will be used to assess outcomes at larger spatial and temporal scales, and so outcomes must be scaled up (aggregated) appropriately.

The eFlowEval framework provides a consistent method to scale up from base polygons to large spatial units such as whole water catchments and basins (e.g., the Murray–Darling Basin). To achieve this, larger polygons of interest are identified, and all outputs from the base polygons are aggregated into the larger polygons. Similar to the ability to choose and define functions for inputting driver data into base polygons, the aggregation function can be defined, which should differ based on the values inside the base polygons. For example, if the base polygons contain a total inundation volume or metabolic output, then the simple sum of those polygons within the larger polygon would yield the total wetland inundation or metabolic output in the catchment. However, if we desire a mean at the catchment scale (e.g., mean temperature in wetlands across the catchment), the mean should be calculated weighted by the areas of the base polygons. The aggregation outputs and a set of variable values in the new, larger polygons, typically at a set of timepoints, very similar to the base polygon outputs.

Aggregation may also occur across a temporal scale, for example, to combine daily outputs into water years or other scales relevant to management decisions. The consistent process for this aggregation is to provide a set of timepoints defining the breaks between time sections, and a function to calculate on all the polygons within that time period. The temporal aggregation can be conducted on the base polygons or after a spatial aggregation to larger polygons. In either case, the result is contained within the same set of polygons, now aggregated to the new coarser time units.

In both spatial and temporal aggregation, the results are aggregated, never the inputs (driver data). This allows us to model ecological processes as close to the scale at which they occur as possible and minimises the impact of Jensen's inequality. For example, if we model bird breeding as dependent on inundation in every wetland across a catchment and then calculate the total breeding in the catchment, the answer will be much different (and more correct) than if we calculated breeding output at the mean inundation level over all wetlands in the catchment.

A.5 Results presentation

The eFlowEval framework provides standard functionality to generate maps and graphics in a consistent way. These outputs form a set of default outputs, though the details of these presentations should be modified for different responses, for example changing scales or colour ramps.

All themes use similar standard plot types (e.g. maps, bar charts, timeseries plots), but the particular plots differ between them, depending on what is being plotted. For example, a map of metabolic output needs to handle very different values and accentuate different comparisons than a map of bird breeding. However, a set of plots can be largely standardised within themes. The approach we have taken with eFlowEval is to provide a consistent set of plot types generally, and then define consistently-defined plots within theme to achieve comparisons between plots, maintain colour choices, etc. These plots are built using plotting functions for consistency and called in Rmarkdown notebooks for final adjustments and presentation (with examples throughout the text). These outputs apply to local results and to larger spatial scales post-aggregation.

Maps are a natural output format, as nearly all the data and results are spatial, consisting of variable values in polygons. The eFlowEval framework provides mapping capability as static maps, zoomable interactive html maps, gifs, and Shiny apps. In general, these maps present both driver and response data, to illustrate the reason why responses occurred. They are also useful for showing differences between scenarios or uncertainty, as seen in the text. They will typically need to be adjusted for a given set of outputs, particularly colour ramps that best represent the range of values and the types of variables. These settings are developed in plotting functions, which then auto-apply them to each desired set of maps.

Most data and outputs are spatio-temporal, and so outputs are developed to present results through time. These may be static maps at different time points, gifs, or interactive Shiny apps with selectable time periods. Beyond mapping, plotting capability is developed to provide timeseries of the variables, either on a per-polygon basis or further aggregated to the basin-scale. These sorts of presentation can be ideal for visualising quantitative changes though time or between locations, particularly if they are too small to be effectively captured by a chloropleth map. For any particular set of outcomes, decisions need to be made about the scaling applied, including simply presenting the data or relativising to a baseline (as may be necessary with very large differences in values between catchments, for example). These plots are also developed within plotting functions, and enable maintaining consistent plot 'look', including colour ramps, with maps and other figures.

Additionally, consistent methods for scenario comparisons have been developed. At the simplest, these include side-by-side maps or other plots of the data for the different scenarios. More analytically, the eFlowEval framework provides a standard set of capabilities to directly compare outputs between scenarios, including differences or relative changes in values between them in point, line, and bar charts. The structure of these comparisons are made consistent by the framework, while providing the flexibility to define the most relevant comparisons. These can then be plotted and presented in various ways paralleling other results, including maps, bar graphs, and timeseries, as seen in the text.

Appendix B Metabolism demonstration

Metabolism data was applied to the model, using knowledge gained from a regression of metabolic rates on temperature and seasonality, developed by the Food Webs and Water Quality theme (Ryder et al., 2021, Darren Giling, pers. comm). The relationship found in the Food Webs and Water Quality theme is based on in-stream metabolism sensors, using water temperature measurements from the same sensors and flow volume calculated from gauges (Ryder et al., 2021).

Here, we demonstrate eFlowEval for wetlands across the Murray–Darling Basin (the Basin), using a similar regression. This approach uses different input data, namely remotely sensed temperature and inundation, in order to reach the basin-scale and analyse wetland conditions instead of in-stream. We identify sources of increased uncertainty in these new relationships. This approach demonstrates the framework and achieves basin-scale outcomes, while acknowledging that the values for metabolic rates are highly uncertain without additional research into models for wetland metabolism.

B.1 Data inputs

To establish the relationships used here, instead of in-stream temperatures at monitoring locations, we obtained remotely-sensed surface temperatures across the Basin, acquired from NASA MODIS (Wan et al., 2015) for the period 1 Jan 2014 to 31 Dec 2020, which are raster grids at 1km resolution and daily timesteps.

To allow predictions across all Basin wetlands based on temperature, the surface temperature raster was used to calculate an average temperature in each ANAE polygon in the Basin on each day using an area-weighted mean for the raster intersections. We calculate the area-weighted mean because pixels that intersect polygon edges will not have their full area contained within the polygon, and so their contribution to the polygon-average temperature should be down-weighted. This process, like nearly all processing steps operating on the full set of ANAE polygons in the Basin, is intended to run on an HPC system, where it is broken up into 100 pieces within each catchment to run in parallel. The subsequent outputs are then re-combined into files containing all ANAEs in each catchment.

Metabolic rates are defined per-litre, and we are interested in predicting aquatic metabolism. Thus, we include inundation data to provide the volume of water in each wetland. The best-available data at the basin scale gives maximum inundation extent during 2-month periods in 30 m rasters (Teng, 2021). This dataset was used to find the inundation volume and the maximum inundation extent for each ANAE polygon for each 2-monthly period. The analyses for metabolism required the simple volume of inundation, and so the inundation data were incorporated into the base ANAE polygons by calculating the area-weighted sum of inundation volume for the pixels intersecting each polygon. As with temperature, the area-weighting is included to account for the varying amounts by which pixels intersect the edges of the polygons. Processing is intended to run on the HPC for many parallel chunks with concatenated output.

To predict metabolism from temperature given inundation, we developed a regression of metabolism on temperature. This regression used the temperature data from MODIS, which were spatially matched with the locations and dates of the metabolic sensor data collected during LTIM and MER from the CEWH’s Monitoring Data Management System. These operations yielded a spatial point dataframe with columns for date, location, metabolic values, temperature, and catchment for each datapoint in the MER metabolism sensor file. Additionally, the in-stream temperature data from the sensors themselves were retained, in order to assess differences between regressions using in-stream versus remotely sensed temperatures. Data were excluded from the sensor at *1km upstream Wynburn Escape* because the sensor location coordinates were incorrect and so could not be correctly matched to either temperature data or catchment. Following this data matching, we developed regressions similar to those from the Food Webs and Water Quality theme.

B.2 Response model

Regressions were developed for both log(GPP) (gross primary production) and log(ER) (ecosystem respiration), capturing the relationships identified by (Ryder et al., 2021). Model selection (likelihood ratio tests where models with nested terms were considered, otherwise AIC (Akaike Information Criterion)) was used to choose between several potential models for each dependent variable. In all cases temperature (in degrees C) was included. Additional terms considered were the *daysAwayFromWaterYear* (a continuous measure of seasonality), *ValleyName* (catchment), and *bimonthlyPeriod* (capturing seasonality in discrete chunks matching the inundation data). Water Year was included as a random intercept to capture differences between years. The seasonality term *daysAwayFromWaterYear* is defined as the number of days (in either direction) a data point is from the water year breakpoint (July 1), and so has a maximum of 183 in a normal year, equating to Jan 1. This construction was chosen to maximise information and power after extensive testing against other options, including day of year (to capture differences between spring and autumn, for example) and various discretisations, including bimonthly to match the inundation data. Seasonality is clearly correlated with temperature ($\rho = 0.81$), but also captures other seasonal effects such as insolation, life cycles, agricultural practices, etc. Most combinations of the independent variable terms, along with interactions between temperature and *ValleyName* and temperature and the seasonality measures were considered. Note, however, that our primary goal here was demonstration, and so we sought to produce a reasonable model of metabolism given the data available, not necessarily to find the best possible metabolism model. For both log(GPP) and log(ER), the best fit models included *temperature*, *daysAwayFromWaterYear, ValleyName*, and the interactions *temperature* daysAwayFromWaterYear* and *temperature*ValleyName*, and included the *WaterYear* random factor. Estimates for each term in these models, along with 95% confidence intervals and p-values are given in Table B.1 and Table B.2.

Table B.1 Predictors of log (GPP) for best-fit model

Includes terms for temperature (tempC), catchment (ValleyName), seasonality (daysAwayFromWaterYear)and the interactions tempC*daysAwayFromWaterYear and tempC*ValleyName. WaterYear included as random effect to capture unspecified differences between years. Marginal R2 is the R2 for the fixed effects alone, while conditional is R2 for the full model.

Predictors	Estimates	CI	p
(Intercept)	-2.66	-4.07 – -1.26	<0.001
tempC	0.08	0.04 – 0.11	<0.001
daysAwayFromWaterYear	0.01	0.01 – 0.01	<0.001
ValleyName [EdwardWakool]	1.71	0.31 – 3.12	0.017
ValleyName [Goulburn]	1.5	0.10 – 2.91	0.036
ValleyName [Lachlan]	2.21	0.80 – 3.61	0.002
ValleyName [LowerMurray]	0.94	-0.50 – 2.37	0.2
ValleyName [Murrumbidgee]	1.54	0.12 – 2.96	0.034
ValleyName [Warrego]	1.8	0.17 – 3.42	0.03
tempC * daysAwayFromWaterYear	0	-0.00 – -0.00	<0.001
tempC * ValleyName [EdwardWakool]	-0.05	-0.08 – -0.01	0.016

Predictors	Estimates	CI	p
tempC * ValleyName [Goulburn]	-0.05	-0.08 – -0.01	0.013
tempC * ValleyName [Lachlan]	-0.05	-0.09 – -0.01	0.009
tempC * ValleyName [LowerMurray]	-0.03	-0.07 – 0.01	0.115
tempC * ValleyName [Murrumbidgee]	-0.05	-0.09 – -0.02	0.006
tempC * ValleyName [Warrego]	-0.04	-0.08 – -0.00	0.046
Random Effects			
s2	0.38		
t00 wateryear	0.01		
ICC	0.03		
N wateryear	7		
Observations	8447		
Marginal R2 / Conditional R2	**0.296 / 0.320**		

Table B.2 Predictors of log(ER) for best-fit model

Includes terms for temperature (tempC), catchment (ValleyName), seasonality (daysAwayFromWaterYear)and the interactions tempC*daysAwayFromWaterYear and tempC*ValleyName. WaterYear included as random effect to capture unspecified differences between years. Marginal R2 is the R2 for the fixed effects alone, while conditional is R2 for the full model.

Predictors	Estimates	CI	p
(Intercept)	-1.73	-3.57 – 0.11	0.065
tempC	0.06	0.01 – 0.11	0.011
daysAwayFromWaterYear	0	0.00 – 0.00	<0.001
ValleyName [EdwardWakool]	1.68	-0.16 – 3.51	0.073
ValleyName [Goulburn]	2.15	0.31 – 3.99	0.022
ValleyName [Lachlan]	2.74	0.90 – 4.57	0.003
ValleyName [LowerMurray]	0.02	-1.85 – 1.89	0.983
ValleyName [Murrumbidgee]	0.77	-1.09 – 2.63	0.419
ValleyName [Warrego]	1.79	-0.27 – 3.85	0.088
tempC * ValleyName [EdwardWakool]	-0.04	-0.09 – 0.01	0.09
tempC * ValleyName [Goulburn]	-0.06	-0.11 – -0.01	0.011
tempC * ValleyName [Lachlan]	-0.06	-0.11 – -0.02	0.009
tempC * ValleyName [LowerMurray]	-0.03	-0.08 – 0.02	0.286
tempC * ValleyName [Murrumbidgee]	-0.05	-0.10 – 0.00	0.056
tempC * ValleyName [Warrego]	-0.05	-0.10 – 0.00	0.067
Random Effects			
s2	0.62		
t00 wateryear	0.04		
ICC	0.05		
N wateryear	7		
Observations	8447		
Marginal R2 / Conditional R2	**0.260 / 0.300**		

These models use surface temperature from MODIS and not in-stream temperature captured by the same sensors as metabolism, in order to gain spatial generality. We tested the impact of this shift, as it has the potential to increase uncertainty. Model performance is very similar to models using the in-stream temperatures (For GPP, surface temp: conditional R^2 = 0.32, RMSE = 0.61; in-stream temp: conditional R^2 = 0.35, RMSE = 0.60. For ER, surface temp: conditional R^2 = 0.30, RMSE = 0.78; in-stream temp: conditional R^2 = 0.32, RMSE = 0.77). Accepting this slightly poorer performance trades off with the value to basin-scale modelling of having temperature data for all locations in the Basin, rather than only at monitoring points. It still must be acknowledged, however, that while the fits are close at these measured locations, there will be other sources of bias from extrapolating to unmonitored areas.

The chosen models are difficult to visualise, due to the number of fixed effects and interaction terms. The top panels of Figure B.1 and Figure B.2 show the relationship between temperature and metabolism in the raw data points for each catchment, coloured by the *daysAwayFromWaterYear* variable to indicate seasonality. For the model fits (bottom panels in Figure B.1 and Figure B.2), the relationship between temperature and metabolism in each catchment are shown, with separate panels for the fit near the beginning and end of the water year (*daysAwayFromWaterYear* = 0), and another for the middle of the water year (*daysAwayFromWaterYear* = 184), noting that the predictions shift continuously from one to the other (and back again) throughout the year.

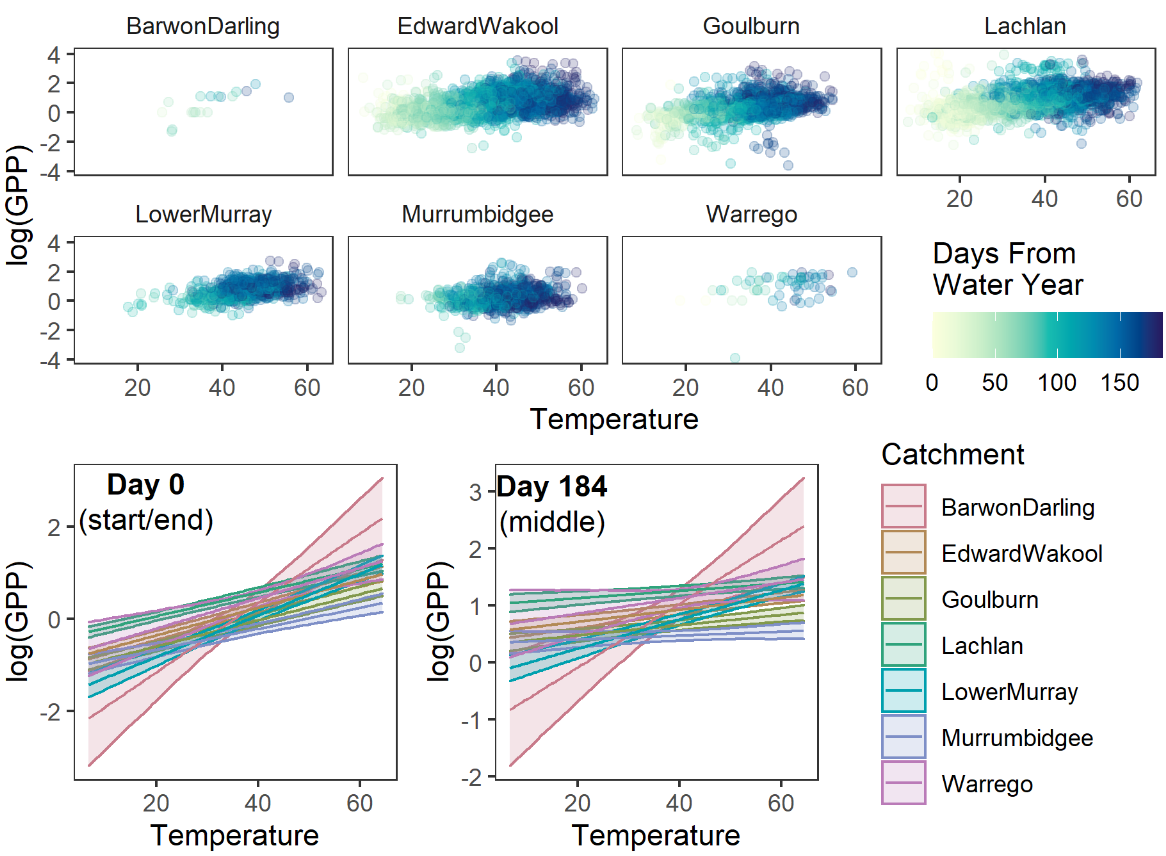

Figure B.1 Data points for log(GPP) and temperature from in-stream sensors (top panels) for each catchment, coloured by the days away from July 1. Bottom panels: best-fit models and 95% CI for the relationship between log(GPP) and temperature in each catchment near July 1

(Day 0- the end of one water year and start of the next) and near Jan 1 (Day 184, the middle of a water year). Days away from water year is a continuous variable, and so only the endpoints are shown.

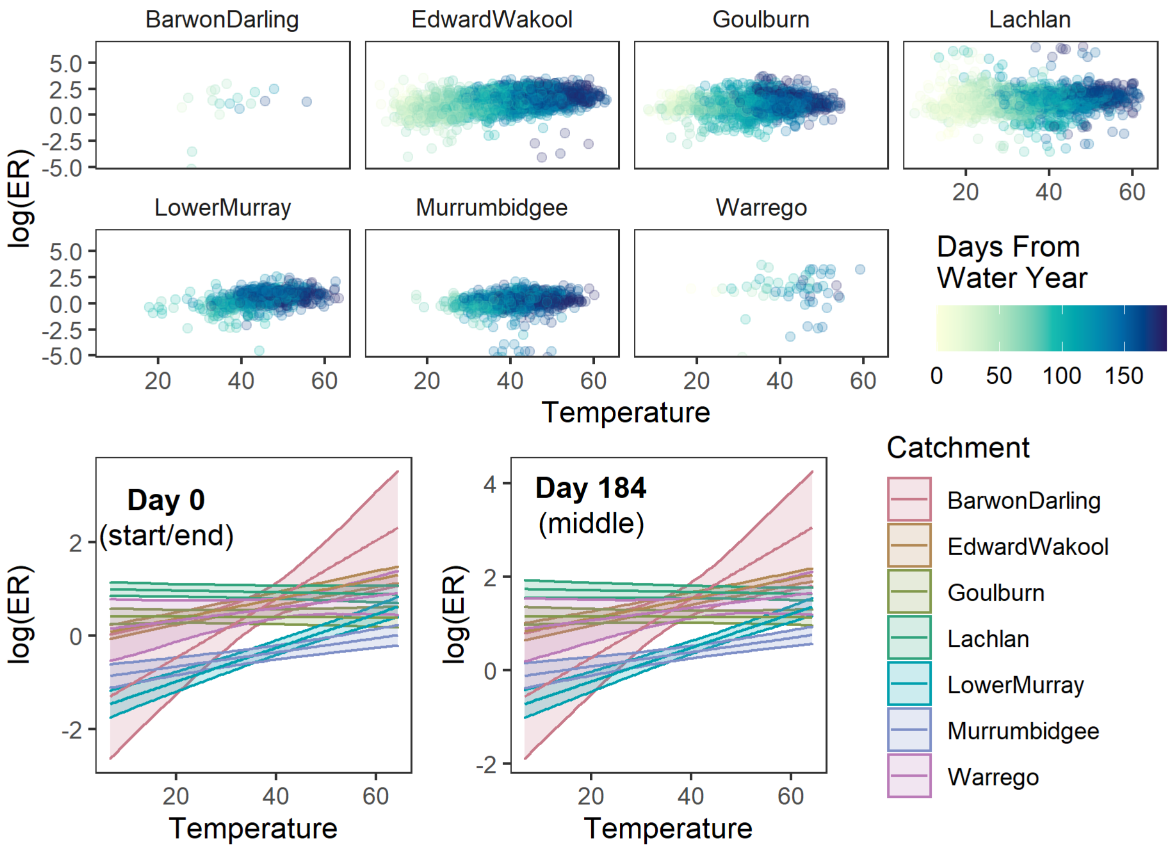

Figure B.2 Data points for log(ER) and temperature from in-stream sensors (top panels) for each catchment, coloured by the days away from July 1. Bottom panels: best-fit models and 95% CI for the relationship between log(ER) and temperature in each catchment near July 1
(Day 0- the end of one water year and start of the next) and near Jan 1 (Day 184, the middle of a water year). Days away from water year is a continuous variable, and so only the endpoints are shown

In addition to the best fit models, other options were also saved and moved forward to capture different sorts of uncertainty due to the limitations of available data. First, the regressions developed here only contained data from the subset of catchments with metabolic sensors (Murrumbidgee, Goulburn, Lachlan, Edward Wakool, Lower Murray, Warrego, Barwon Darling). Predictions in catchments without sensors are possible if we accept the additional uncertainty that comes from extrapolating beyond the spatial locations provided by the metabolic data. To do this, we developed a version of the model that dropped the *Valley* and *temp*Valley* terms. This model is used for the uncertainty characterisation in 4 Representing uncertainty, particularly Figure 4.3. Next, the inundation data is available only at the 2-monthly scale, and so the daily metabolism data (and daily predictions) cannot be exactly matched. One option (which we used throughout the text, because it maintains as much information as possible at each stage of the framework) is to find those daily predictions, and then temporally average them into bimonthly units. However, it is also possible to simply obtain the predictions based on regressions including a bimonthly temperature instead of the daily. We retained these models, but do not present them in this report.

For each regression model, the model objects were saved, and were then used for predictions of metabolic output across wetlands in the basin (or the subset of catchments for which the model was parameterised).

B.3 Predicting metabolism

The regression developed here defines a relationship between metabolic rates (in mg O_2 L^{-1} Day^{-1}) and surface temperature. We use this regression to predict the log(GPP) and log(ER) per litre in each ANAE wetland from their daily average temperatures (area-weighted from raster, as described above). In addition to the predicted value, 95% prediction intervals (PI) are also included, as these provide the plausible range of values for metabolic rates given that temperature. Predicted metabolism is found using regressions with and without a *Valley* term and for log(GPP) and log(ER). Predictions from the model with the *Valley* term are NA for catchments other than those with sensor data, while the model without the *Valley* term produces numeric predictions for all ANAEs in the basin. The predicted values and prediction intervals are saved together for all ANAE wetlands, but separately for each model.

As described in the text, we used prediction intervals instead of confidence intervals (CI) because we are interested in describing the range of metabolic rates, not confidence in the best regression fit. Note, however, that there are additional sources of uncertainty, most notably the extrapolation of relationships defined in flowing water to wetlands. As a consequence, these ranges represent the PI only if those extrapolations hold, and the true PI, given additional qualitative and unmeasured sources of uncertainty, is likely to be larger.

The per-litre rates predicted from temperature represent the metabolic potential for each wetland, but water is needed for that aquatic metabolism to occur. Thus, we multiply these rates by the inundation volume to obtain total predicted metabolism for each wetland. Note that 'potential metabolic output' is itself a potentially powerful measure, as it indicates the metabolic output that could be achieved if a wetland were inundated, and so may be quite useful for planning. Because the inundation data is the maximum inundation extent for 2-monthly periods, the daily per-litre predictions are first scaled to the 2-monthly timesteps to match. This process involves loading the predicted daily per litre metabolic rates in each ANAE wetland for each of the four models and assigning each date to a matching bimonthly time-period from the inundation dataset. Then, the mean of those daily per-litre values are taken over the 2 months to best represent metabolic rates during that period. Other statistics are possible, but without additional knowledge of the inundation sequence, the mean captures as much information as possible about the temperature sequence during the 2 months.

The regression predicts log(GPP) and log(ER), and so before taking the bimonthly mean, the antilog of the daily predictions of log(GPP) and log(ER) are calculated. This transformation is necessary to calculate mean GPP and ER, rather than their logs, as that would imply a multiplicative (geometric) average. While GPP and ER may be related to multiplicative processes (e.g., phytoplankton dynamics), the goal here is to characterise production and respiration in a wetland, and so we should average the values on the arithmetic scale. This process yields predictions for each ANAE that are still in per-litre units, they have just been scaled to the 2-monthly timestep.

Finally, these predicted mean bimonthly metabolic rates are multiplied by the maximum inundation volume in each polygon to obtain the predicted metabolic rates at maximum inundation extent. Because these predictions are at the maximum inundation extent, they almost certainly overestimate total production in that period. Thus, we must take care in interpreting these values or in further analyses and transformations to acknowledge that we are comparing maxima, not averages.

B.4 Spatial aggregation

As described in the text, presentation and interpretation of results will often best match management needs if scaled from the individual ANAE wetlands to larger spatial contexts (e.g., catchments or the whole basin). In the metabolism demonstration, 3 levels of spatial scaling are demonstrated: none (i.e., individual

ANAE wetlands), the catchment, and the basin. When aggregated to larger scales, we used the LTIM valleys polygons and Murray–Darling Basin boundary polygon included in ANAE v2 (Brooks et al., 2014). Briefly, each ANAE wetland is assigned to a larger-scale polygon based on location, and then relevant statistics are calculated. Aggregation to larger areas can consider many different potential statistics (e.g., minima, maxima, means). The choice of these aggregation functions should match both the desired use of the aggregated data as well as the nature of the data, with both sums and means demonstrated in the text. Note that although input data may also be aggregated for visualisation, aggregated inputs are not used for the modelling, as that could generate large errors due to nonlinear responses.

The sum over each catchment of the predicted metabolism in each ANAE wetland is used to aggregate GPP and ER. This gives the total metabolic activity for each catchment at the maximum inundation extent in the 2-month period. Note that this is not the predicted total metabolism over that period, because we only know maximum inundation extent. If we want to consider the potential metabolic activity (from only temperature, without knowing whether or not there was water present), we could calculate an area-weighted sum or mean of the raw daily predictions before multiplying them by inundation.

The input data (inundation and surface temperature) was also aggregated for clearer comparison with the outcomes. Surface temperature is aggregated using the area-weighted mean surface temperature of the ANAE wetlands in each catchment. Inundation volume is aggregated using the sum of inundation volume across all ANAE wetlands in each catchment. This sum gives the total volume of inundated wetland in each catchment at the maximum inundation extent during a 2-month period.

B.5 Plotting and figures

Investigating ANAE wetlands directly can be beneficial at times, particularly for assessing local outcomes. Although all wetlands are modelled, examining the outcomes is only practical for small subsets of the wetlands in the Basin. To do so, a relevant set of wetland polygons can be selected and viewed as-is. The Werai forest demonstration (described in this report) is an example of this approach. As all wetlands within the chosen region are viewed individually, no spatial aggregation occurs, only spatial clipping of the wetland datasets. For the Werai example, a boundary for the Werai Ramsar site is used to clip the full set of ANAE wetlands (both input and output data, i.e., temperature, inundation, GPP, and ER) for plotting. The mapped inputs and outputs are also developed into a Shiny app with user-selectable time periods to visualise the relationship between temperature, inundation, and metabolism at different time points (Appendix D). Local scale plots are consistently labelled, coloured, and scales adjusted for consistency with a set of plotting functions. The plots themselves are built in a notebook for final production, taking advantage of the notebook format's ability to preserve the output.

Basin-scale plotting proceeds in a very similar way, using the aggregated data instead of the individual wetlands. The same set of plotting functions are used, which now establish basin-relevant standard scalings and temporal aggregations. The approach of using plotting functions allows establishing standardised colour scales, holding plot dimensions and colour maps consistent across plots, and allows a consistent look to all the plots. Plots themselves are again generated in a notebook for ease of visualisation and production. A shiny app is also developed for the basin scale outcomes (Appendix D).

Both local and basin-scale plotting examine outcomes derived from the historical record of temperatures and inundation, as well as outcomes from demonstration scenarios. Scenario plots can be presented on their own, or calculations can be performed to visualise them relative to baseline.

B.6 Scenarios

One of the primary benefits of the eFlowEval framework is the assessment and comparison of different scenarios, whether those are around management actions, climate, or some other change. This capacity is demonstrated in the metabolism example. However, as no pre-established scenarios were available, 3 simple scenarios were developed to allow the demonstration of the process of running and comparing scenarios in the eFlowEval framework. The scenarios considered are a temperature increase of 2 °C, an increase of 10% in inundation, and a combination of the two. These scenarios are purely for demonstration purposes and should not be taken to be representative of expected change in the basin or potential management actions. For one, they are highly simplistic, neither temperature change nor inundation shifts will occur uniformly over the Basin. Further, temperature and inundation will interact, which is not captured here.

Scenario specification may occur at different points in the process, and so we demonstrate 2 potential methods. First, we created a temperature scenario with a uniform 2°C increase. To demonstrate the situation where a scenario is specified in the initial input data, we added 2°C to the temperatures immediately on reading in the original raster data. Subsequent processing was unmodified from the temperature data (averaging into ANAE polygons, use for predicting metabolism, aggregation, plotting). This approach parallels what we would expect if we received something like climate scenarios.

Second, we developed a scenario with a uniform 10% increase in inundation volume. To demonstrate specifying the scenario later in the process, we added 10% to the inundation in every ANAE wetland. This was done to approximate the situation where a manager might like to change the e-water delivery to a certain area. As with temperature, further processing exactly paralleled the baseline analyses, from multiplying by the predicted metabolic potential through the plotting.

The key aspect for comparing scenarios occurs at the very end of the framework, after all scenarios (including baseline or counterfactual) have produced outputs. All scenario processes are carried through in parallel, and then during the analysis and plotting phase they are compared to each other to assess how the different scenarios change potential outcomes. These outcomes can be compared using the same plotting functions as before, with plots simply positioned next to each other. The eFlowEval framework also includes additional plotting functionality to build similar plots based on differences or relative differences between scenarios, which is made possible by the consistent output formats of the framework. Additional scenario-specific plots are available to compare scenarios quantitatively in barcharts or timeseries (see text).

Appendix C Waterbird demonstration

To demonstrate the how the eFlowEval framework can be implemented for mobile organisms we selected the Royal Spoonbill (*Platalea regia*) as the target species. The Royal spoonbill was chosen because it is dependent on water for much of its lifecycle (Figure C.1), they nest in colonies, and have distinctive behavioural strategies that lend themselves to threshold requirements (Lester et al., 2020).

We worked with expert Heather McGinness to establish 1) the most important life-history stages, 2) the most important strictures applying to each life-history phase, and 3) a defensible value for that stricture/threshold. These were then compared to the available datasets that could be used to assess whether or not the strictures were met over time. We focused on water dependent strictures, as water is the primary management-relevant driver. Triggering a breeding event and maintaining suitable conditions for laying and survival of the young to fledging where considered to be the most critical stages. The availability of foraging resources within the natal wetland and the surrounding areas is important to support the raising of young.

C.1 Data and stricture processing

The royal spoonbill strictures depend on inundation characteristics (e.g. timing and spatial distribution) and how these overlay with vegetation communities (e.g. nesting habitat). The dataset underpinning the demonstration is the bimonthly maximum inundation dataset created by Teng (2021) at CSIRO, and also used for the inundation demonstration. This dataset provides 30m rasters with depth at the maximum inundation extent over a bimonthly timestep. These were processed into ANAE polygons using custom aggregation functions giving the area of each polygon meeting the depth requirements for breeding or foraging. The result was 2 sets of polygons, one for each activity, containing values for relevant inundation areas for 197 bimonthly timesteps (1988–2020). Strictures based on meteorological phenomena, such as large, rapid changes in temperature and heavy rainfall, could be implemented from Basin-wide datasets, but this is not part of the demonstration. There is evidence to suggest these play a significant role in spoonbill mortality (per comm. Heather McGinness).

The general approach for strictures was to evaluate whether conditions were met for the inundation values in each polygon giving a true (‘1’) or false (‘0’) values for each. Taking the product across all strictures and time steps within an activity period (e.g. breeding season) yields a total pass/fail measure for that period. The resulting value, zero or one, was then multiplied by the inundation area of each polygon such that all polygons failing to meet strictures are set to zero and those that pass retain an inundation area.

Birds are mobile, and so some strictures include values relevant to a wetland complex, rather than single wetlands. For example, breeding events require favourable conditions across a larger area, and foraging can occur some distance from the nest site. In these cases, wetland complexes were defined from RAMSAR sites and spatial buffers to capture nearby wetlands. Inundation areas of polygons with passing strictures contained within each wetland complex were summed and compared to wetland-scale thresholds. Combining these wetland-scale strictures gave us an estimate of whether a breeding event was likely each year by wetland complex (true/false).

C.2 Stricture definitions

Strictures driving breeding success included vegetation type and water depth, timing and distribution. Breeding requires water depths of between 0.5 and 1.5m. This stricture was implemented in the processing of bimonthly inundation dataset into the breeding ANAE polygons by defining a custom aggregation function that calculated area inundated given water depth was within the acceptable range. Some ANAE vegetation types have attributes, such as nesting sites, that support breeding. McGinness et al. (2020) ranked the ANAE types according to the frequency of nesting events recorded (Table C.1). We implemented the vegetation-dependence stricture by using this list to filter the ANAE to those that are suitable for nesting when flooded. Inundation must last for a least 6 consecutive months within the breeding season, which is between October and March in the Southern Basin and August and April in the Northern Basin. To realise these seasonality and duration strictures we calculated a rolling sum of areas within a 6-month window (3 bimonthly timesteps) and selected those that ended in March. This end date limits the applicability of the demonstration to the Southern Basin, though it is possible to create a more complex function that assesses whether a wetland is in the Northern or Southern Basin and adjusts the end date accordingly. Initiation of a breeding event requires inundation across large parts of a wetland complex, not just a single ANAE polygon, so we calculated a total area inundated in the complex. For demonstration purposes we used threshold of area equal to 70% of the maximum historical inundation meeting breeding strictures to implement the total inundation area stricture. Some analysis has been done on amount of instream discharge that initiates breeding events but given that our inundation datasets are spatial rather that point based, they did not translate (Bino et al., 2014; Brandis et al., 2018). For successful breeding, all of these strictures must be met (e.g. each gives a value of 0 or 1, and they are multiplied to determine breeding success, which is then used to calculate area of successful breeding).

Foraging strictures included vegetation type, water depth and spatial distribution relative to the wetland complex, and were implemented as described for breeding, with some modification to the details due to different needs in the different life stages. Specifically, foraging requires depths below 0.4 m to facilitate the wading technique of the spoonbill. Vegetation types for foraging also differ from those needed for breeding and were also drawn from McGinness et al. (2020). Foraging can occur beyond the bounds of the habitat complex so when summing total available foraging area we implemented a 10 km look-around, which reflects the maximum distance spoonbill travel (Figure 3.11). The additional area was not significant for the Warrego forest wetlands, but may be for other wetlands, or species with larger ranges. For this demonstration, the available area had to be greater than 5% of historical maximum forage area to pass the stricture.

For a successful breeding event, all of these strictures must be met (e.g. birds must successfully breed and then forage enough to keep those chicks alive to fledge). Each individual stricture gives a value of 0 or 1, and they are multiplied to determine success, which is then used to calculate area of successful breeding. The outcome is an overall estimate of whether a breeding event was likely for each wetland complex in each year, and the area over which it occurred (Figure 3.9).

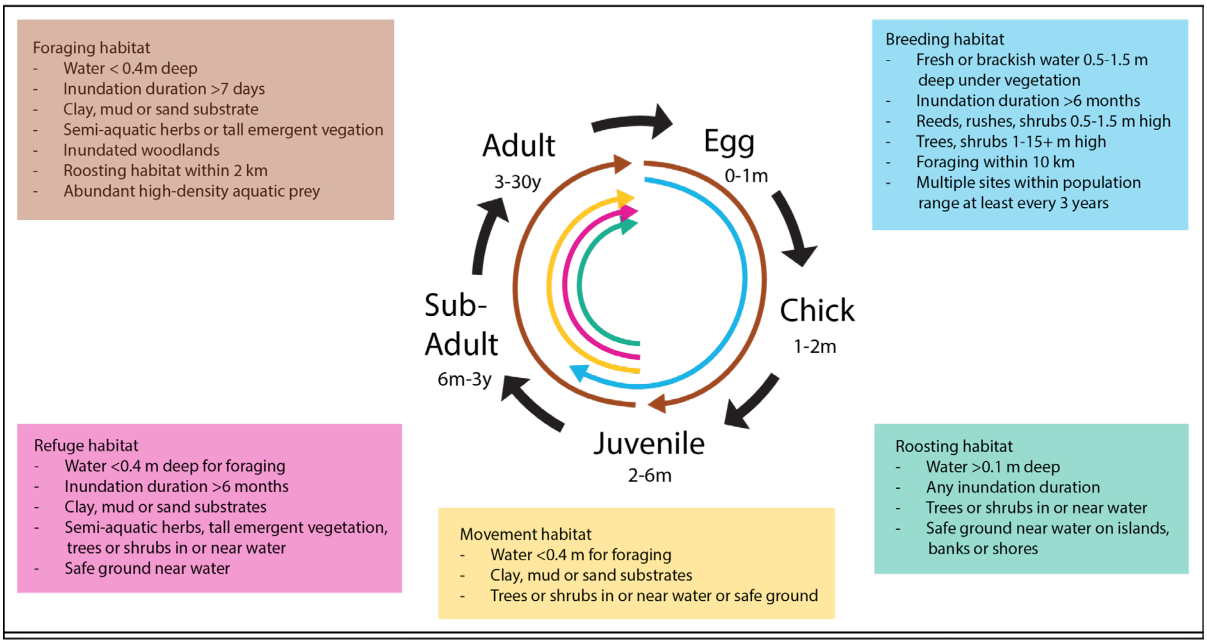

Figure C.1 Royal spoonbill habitat requirements (adapted from McGinness et al. 2020)

Table C.1 Top breeding and foraging habitat types for Royal spoonbill in the Murray–Darling Basin

The uppermost entry has the most recorded breeding or foraging observations. Any ANAE in these types was counted as success in the Vegetation stricture, while those in others were classed as failures.

Breeding	Foraging
Lp1.1: Permanent lake	Lp1.1: Permanent lake
F1.2: River red gum forest riparian zone or floodplain	Pp2.2.2: Permanent sedge/grass/forb marsh
Pp4.2: Permanent wetland	Pt3.1.2: Clay pan
Pp2.1.2: Permanent tall emergent marsh	Pp4.2: Permanent wetland
Pp2.2.2: Permanent sedge/grass/forb marsh	Pt2.2.2: Temporary sedge/grass/forb marsh
Pt1.1.2: Temporary river red gum swamp	F1.2: River red gum forest riparian zone or floodplain
Pt1.8.2: Temporary shrub swamp	Pt2.1.2: Temporary tall emergent marsh
Rt1.4: Temporary lowland stream	F1.8: Black box woodland riparian zone or floodplain
F2.2: Lignum shrubland riparian zone or floodplain	F1.4: River red gum woodland riparian zone or floodplain
F1.8: Black box woodland riparian zone or floodplain	Etd1.3.3: Tide dominated estuary
F2.4: Shrubland riparian zone or floodplain	Rt1.4: Temporary lowland stream
Pt3.1.2: Clay pan	Pt1.2.2: Temporary black box swamp
Lt1.1: Temporary lake	Pt2.3.2: Freshwater meadow
F1.4: River red gum woodland riparian zone or floodplain	F1.12: Woodland riparian zone or floodplain
Pt2.2.2: Temporary sedge/grass/forb marsh	Pt1.6.2: Temporary woodland swamp
Pt2.3.2: Freshwater meadow	Rp1.4: Permanent lowland stream
Rp1.4: Permanent lowland stream	Pt1.1.2: Temporary river red gum swamp
	Pt4.1: Floodplain or riparian wetland
	Lt1.1: Temporary lake
	Pp2.1.2: Permanent tall emergent marsh

Breeding	Foraging
	F1.10: Coolibah woodland and forest riparian zone or floodplain
	Etd1.2.1: Tide dominated saltmarsh
	F2.2: Lignum shrubland riparian zone or floodplain
	F2.4: Shrubland riparian zone or floodplain
	Lsp1.1: Permanent saline lake
	Lst1.1: Temporary saline lake

Appendix D Shiny app

Shiny apps are an ideal way to allow some user interaction with data analyses in R or python. Here, we have developed demonstration Shiny apps for our Werai metabolism example and for Basin-scale outcomes (Figure D.1 and Figure D.2, respectively). The user can select the bimonthly period of interest and the app fetches the driver data (volume of inundation and temperature) in each polygon and presents it as a map above the predicted values of GPP (gross primary productivity) and ER (ecosystem respiration). This approach allows the user to examine time periods of interest and compare the predicted outcomes for different temperature and inundation conditions that occurred during the modelled period (2014-2020). The values from the eFlowEval model sit atop fully zoomable basemaps using Leaflet, providing useful spatial context.

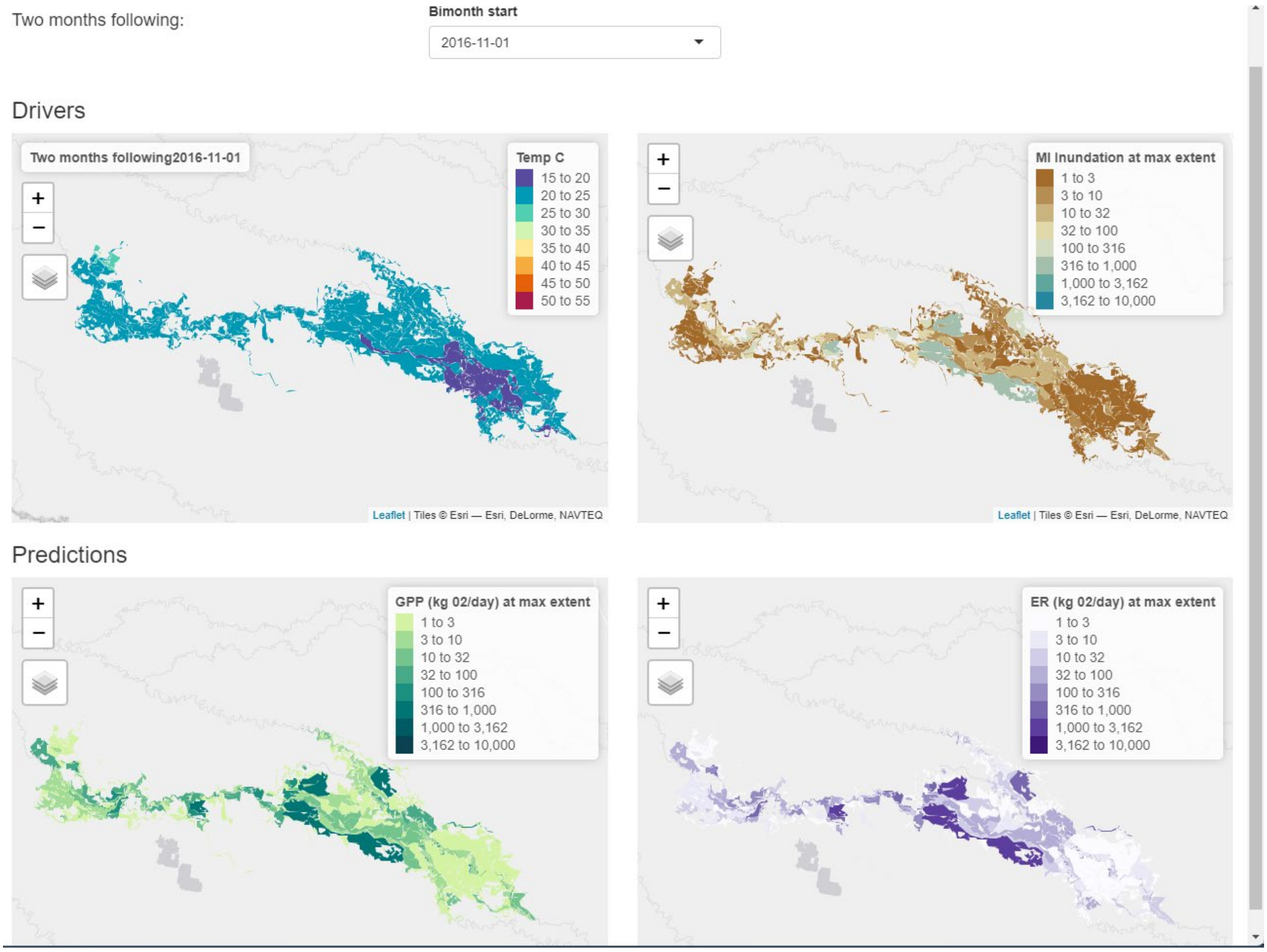

Figure D.1 Example screenshot of Shiny app presenting drivers and predicted metabolic responses for each wetland in the Werai forest

The pictured data matches Figure 3.1 but users can select any bimonthly period for which there is data from the dropdown menu, and all maps re-populate. Driver and prediction data are plotted on top of zoomable basemaps, of which there are several choices including some with topography or streetmaps.

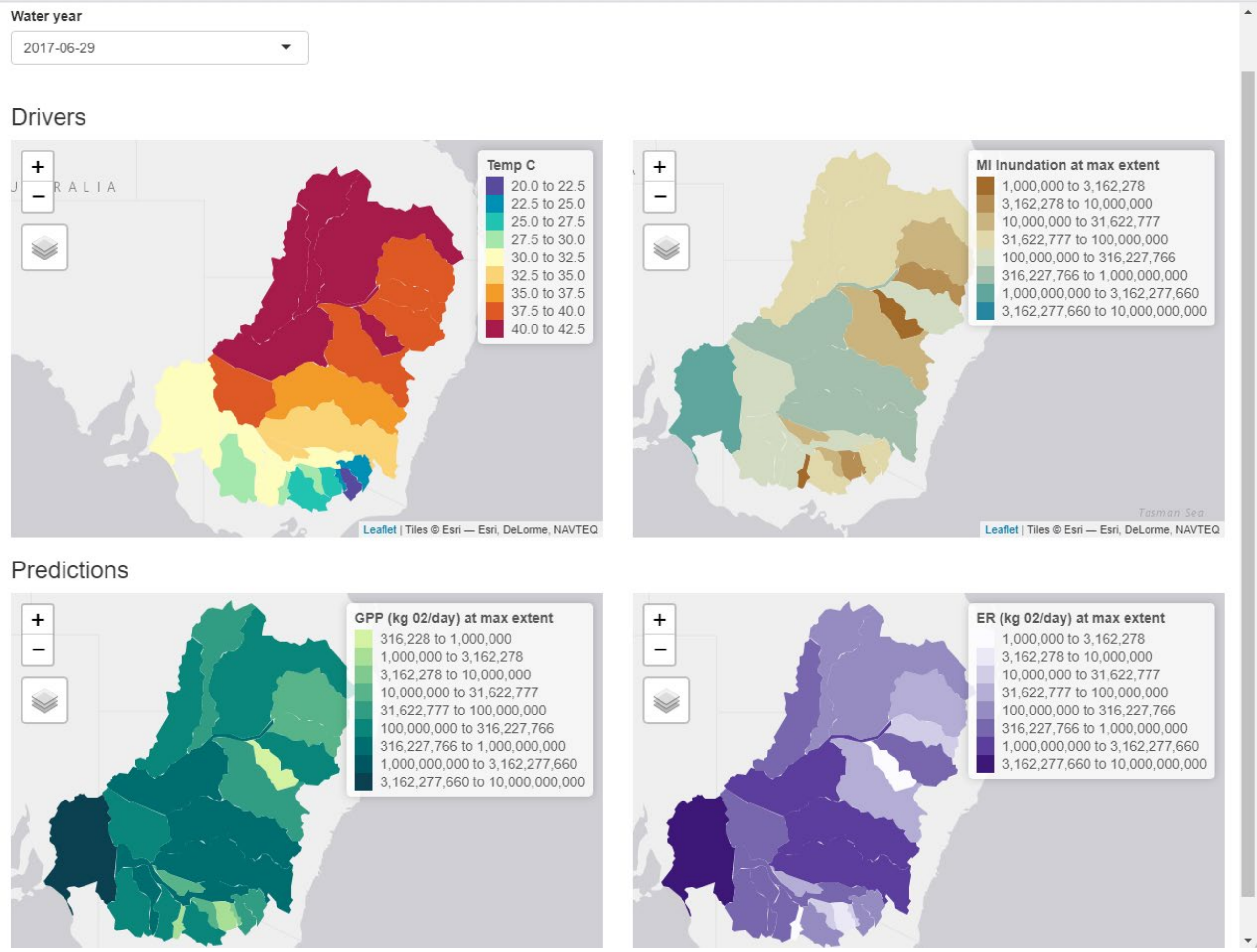

Figure D.2 Example screenshot of basin-scale Shiny app, with drivers and predicted outcomes scaled to catchment and aggregated across 2-monthly periods to the water year

As with the local example, the user can choose a water year to examine from a dropdown menu and all maps will repopulate. Maps are zoomable and draggable and have a choice of basemaps.

The developments here are 2 of many potential uses for Shiny's interface to allow user interaction. Many additional opportunities exist, such as allowing the user to select scenarios to compare or degrees of uncertainty. Implementation would proceed in consultation with water managers about the most useful direction.

References

Ascough J, Maier H, Ravalico J and Strudley M (2008) Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. Ecological Modelling 219(3-4), 383-399.

Bino G, Steinfeld C and Kingsford RT (2014) Maximizing colonial waterbirds' breeding events using identified ecological thresholds and environmental flow management. Ecological Applications 24(1), 142-157.

Brandis K, Bino G, Spencer J, Ramp D and Kingsford R (2018) Decline in colonial waterbird breeding highlights loss of Ramsar wetland function. Biological Conservation 225, 22-30.

Brooks S, Cottingham P, Butcher R and Hale J (2014) Murray-Darling Basin aquatic ecosystem classification: Stage 2 report. Peter Cottingham & Associates, Melbourne, Vic., Australia.

Brooks S (2017) Classification of aquatic ecosystems in the Murray-Darling Basin: 2017 update. Canberra, ACT: Murray-Darling Basin Authority.

Brooks S (2021) ANAE Classification of the Murray-Darling Basin Technical Report, Revision 3.0, March 2021. Technical report for the Commonwealth Environmental Water Office. Department of Agriculture, Water and the Environment, Australia.

CEWO (2017) Environmental Watering in the Mid-Murray Catchment in Previous Years. Commonwealth Environmental Water Office, Canberra, ACT, http://www.environment.gov.au/water/cewo/catchment/.

Chatfield C (2001) Prediction intervals for time-series forecasting. Principles of forecasting. Springer, 475-494.

Christensen NL, Bartuska AM, Brown JH, Carpenter S, D'Antonio C, Francis R, Franklin JF, MacMahon JA, Noss RF, Parsons DJ, Peterson CH, Turner MG and Woodmansee RG (1996) The Report of the Ecological Society of America Committee on the Scientific Basis for Ecosystem Management. Ecological Applications 6(3), 665-691.

Couet J, Marjakangas E-L, Santangeli A, Kålås JA, Lindström Å and Lehikoinen A (2022) Short-lived species move uphill faster under climate change. Oecologia 198(4), 877-888.

de Vente J and Poesen J (2005) Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models. Earth-Science Reviews 71(1), 95-125.

DeFries RS, Foley JA and Asner GP (2004) Land-use choices: Balancing human needs and ecosystem function. Frontiers in Ecology and the Environment 2(5), 249-257.

Downes BJ (2010) Back to the future: little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems. Freshwater Biology 55, 60-79.

DPIE (2020) Long Term Water Plan Part A: Murray–Lower Darling catchment.

Englund G and Cooper SD (2003) Scale effects and extrapolation in ecological experiments.

Frissell C, Liss W, Warren C and Hurley M (1986) A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10(2), 199-214.

Gawne B, Capon SJ, Hale J, Brooks SS, Campbell C, Stewardson MJ, Grace MR, Stoffels RJ, Guarino F and Everingham P (2018a) Different Conceptualizations of River Basins to Inform Management of Environmental Flows. Frontiers in Environmental Science.

Gawne B, Capon SJ, Hale J, Brooks SS, Campbell C, Stewardson MJ, Grace MR, Stoffels RJ, Guarino F and Everingham P (2018b) Different Conceptualizations of River Basins to Inform Management of Environmental Flows. frontiers in Environmental Sciences.

Geary WL, Bode M, Doherty TS, Fulton EA, Nimmo DG, Tulloch AIT, Tulloch VJD and Ritchie EG (2020) A guide to ecosystem models and their environmental applications. Nature Ecology & Evolution 4(11), 1459-1471.

Greca IM and Moreira MA (2000) Mental models, conceptual models, and modelling. International Journal of Science Education 22(1), 1-11.

Grimm V, Augusiak J, Focks A, Frank BM, Gabsi F, Johnston AS, Liu C, Martin BT, Meli M and Radchuk V (2014) Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecological Modelling 280, 129-139.

Hart BT, Bond NR, Byron N, Pollino CA and Stewardson MJ (2021) Introduction to the Murray–Darling basin system, Australia. Murray-Darling Basin, Australia. Elsevier, 1-17.

Heffernan JB, Soranno PA, Angilletta Jr MJ, Buckley LB, Gruner DS, Keitt TH, Kellner JR, Kominoski JS, Rocha AV and Xiao J (2014) Macrosystems ecology: understanding ecological patterns and processes at continental scales. Frontiers in Ecology and the Environment 12(1), 5-14.

Hilborn R and Stearns SC (1987) Living with uncertainty in resources management. North Americal Journal of Fisheries Management 7, 1-5.

Hilborn R (1992) Can fisheries agencies learn from experience? Fisheries 17(4), 6-14.

Kingsford RT, Porter J and Halse SA (2012) National waterbird assessment. National Water Commission, Canberra, https://webarchive.nla.gov.au/awa/20141215135501/http://archive.nwc.gov.au/__data/assets/pdf_file/0012/21720/Waterlines-74-National-waterbird-assessment.pdf.

Lester RE and Fairweather PG (2011) Ecosystem states: Creating a data-derived, ecosystem-scale ecological response model that is explicit in space and time. Ecological Modelling 222(15), 2690-2703.

Lester RE (2019) Wise use: using ecological models to understand and manage aquatic ecosystems. Marine and Freshwater Research 71(1), 46-55.

Lester RE, McGinness HM, Price AE, Macqueen A, Poff NL and Gawne B (2020) Identifying multiple factors limiting long-term success in environmental watering. Marine and Freshwater Research 71(2), 238-254.

Lindenschmidt K-E (2006) The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. Ecological Modelling 190(1), 72-86.

McGinness H, Langston A and Brooks S (2020) Royal Spoonbill (*Platalea regia*) requirements, distribution and habitat mapping.

McGinness HM, Brandis K, Robinson F, Piper M, O'Brien L, Langston A, Hodgson J, Wenger L, Martin J, Bellio M, Callaghan D, Webster E, Francis R, McCann J, Lyons M, Doerr V, Kingsford R and Mac Nally R (2019) Murray–Darling Basin Environmental Water Knowledge and Research Project – Waterbird Theme research report. Commonwealth Environmental Water Office, Department of the Environment and Energy, Canberra, ACT, Australia.

McMahon TA, Peel MC, Vogel RM and Pegram GGS (2007) Global streamflows – Part 3: Country and climate zone characteristics. Journal of Hydrology 347(3), 272-291.

MDBA (2011) Delivering a Healthy Working Basin – About the Draft Basin Plan. Canberra, ACT.

Norris RH, Liston P, Davies N, Coysh J, Dyer F, Linke S, Prosser I and Young B (2001) Snapshot of the Murray-Darling Basin river condition. Murray–Darling Basin Commission, Canberra.

Olden JD, Konrad CP, Melis TS, Kennard MJ, Freeman MC, Mims MC, Bray EN, Gido KB, Hemphill NP and Lytle DA (2014) Are large-scale flow experiments informing the science and management of freshwater ecosystems? Frontiers in Ecology and the Environment 12(3), 176-185.

Overton I, Pollino C, Roberts J, Reid J, Bond N, McGinness H, Gawne B, Stratford D, Merrin L and Barma D (2014) Development of the Murray-Darling Basin Plan SDL adjustment ecological elements method. Report prepared by CSIRO for the Murray–Darling Basin Authority, Canberra, Australia.

Poff NL and Zimmerman JKH (2010) Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology 55(1), 194-205.

Poff NL, Tharme RE and Arthington AH (2017) Evolution of environmental flows assessment science, principles, and methodologies. Water for the Environment. Elsevier, 203-236.

Polasky S, Carpenter SR, Folke C and Keeler B (2011) Decision-making under great uncertainty: environmental management in an era of global change. Trends in Ecology & Evolution 26(8), 398-404.

Ruel JJ and Ayres MP (1999) Jensen’s inequality predicts effects of environmental variation. Trends in Ecology & Evolution 14(9), 361-366.

Ryder D, McInerney P, Giling D and Hitchcock J (2021) Basin-scale evaluation of 2019–20 Commonwealth environmental water: Food Webs and Water Quality. Australia.

Ryder DS, Tomlinson M, Gawne B and Likens GE (2010) Defining and using ‘best available science’: a policy conundrum for the management of aquatic ecosystems. Marine and Freshwater Research 61(7), 821-828.

Seddon PJ and Leech T (2008) Conservation short cut, or long and winding road? A critique of umbrella species criteria. Oryx 42(2), 240-245.

Soranno PA, Cheruvelil KS, Webster KE, Bremigan MT, Wagner T and Stow CA (2010) Using landscape limnology to classify freshwater ecosystems for multi-ecosystem management and conservation. BioScience 60(6), 440-454.

Stewardson MJ, Walker G and Coleman M (2021) Chapter 3 - Hydrology of the Murray–Darling Basin. In: Hart BT, Bond NR, Byron N, Pollino CA and Stewardson MJ (eds), Murray-Darling Basin, Australia. Elsevier, 47-73.

Sutherland WJ (2006) Predicting the ecological consequences of environmental change: a review of the methods. Journal of Applied Ecology 43(4), 599-616.

Teng J (2021) MDB Flood Inundation Mapping. 1. GBA data repository

Tengö M, Brondizio ES, Elmqvist T, Malmer P and Spierenburg M (2014) Connecting diverse knowledge systems for enhanced ecosystem governance: the multiple evidence base approach. Ambio 43(5), 579-591.

Thomas CW and Koontz TM (2011) Research Designs for Evaluating the Impact of Community-Based Management on Natural Resource Conservation. Journal of Natural Resources Policy Research 3(2), 97-111.

Thompson RM, King AJ, Kingsford RM, Mac Nally R and Poff NL (2018) Legacies, lags and long-term trends: Effective flow restoration in a changed and changing world. Freshwater Biology 63(8), 986-995.

van den Belt M and Blake D (2015) Mediated Modeling in Water Resource Dialogues Connecting Multiple Scales. JAWRA Journal of the American Water Resources Association 51(6), 1581-1599.

Venier LA, Walton R and Brandt JP (2021) Scientific considerations and challenges for addressing cumulative effects in forest landscapes in Canada. Environmental Reviews 29(1), 1-22.

Vörösmarty CJ and Sahagian D (2000) Anthropogenic disturbance of the terrestrial water cycle. BioScience 50(9), 753-765.

Walker B, Carpenter S, Anderies J, Abel N, Cumming G, Janssen M, Lebel L, Norberg J, Peterson GD and Pritchard R (2002) Resilience management in social-ecological systems: a working hypothesis for a participatory approach. Conservation ecology 6(1).

Wan Z, Hook S and Hulley G (2015) MOD11A2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 1km SIN grid V006. Nasa Eosdis Land Processes Daac 10(10.5067).

Williams BK (2011) Adaptive management of natural resources—framework and issues. Journal of environmental management 92(5), 1346-1353.

Yates KL, Bouchet PJ, Caley MJ, Mengersen K, Randin CF, Parnell S, Fielding AH, Bamford AJ, Ban S, Barbosa AM, Dormann CF, Elith J, Embling CB, Ervin GN, Fisher R, Gould S, Graf RF, Gregr EJ, Halpin PN, Heikkinen RK, Heinänen S, Jones AR, Krishnakumar PK, Lauria V, Lozano-Montes H, Mannocci L, Mellin C, Mesgaran MB, Moreno-Amat E, Mormede S, Novaczek E, Oppel S, Ortuño Crespo G, Peterson AT, Rapacciuolo G, Roberts JJ, Ross RE, Scales KL, Schoeman D, Snelgrove P, Sundblad G, Thuiller W, Torres LG, Verbruggen H, Wang L, Wenger S, Whittingham MJ, Zharikov Y, Zurell D and Sequeira AMM (2018) Outstanding Challenges in the Transferability of Ecological Models. Trends in Ecology & Evolution 33(10), 790-802.

Young W, Scott A, Cuddy S and Rennie B (2003) Murray flow assessment tool: a technical description. Report to the CRC for Freshwater Ecology from CSIRO Land and Water, Canberra.

https://flow-mer.org.au

Australian Government
Commonwealth Environmental Water Holder

Partners

CSIRO

UNIVERSITY OF CANBERRA

Collaborators

Charles Sturt University

DEAKIN UNIVERSITY

BROOKS.eco ECOLOGY & TECHNOLOGY

Arthur Rylah Institute for environmental research

Government of South Australia
Department of Primary Industries and Regions

SARDI
SOUTH AUSTRALIAN RESEARCH AND DEVELOPMENT INSTITUTE

une University of New England

NSW GOVERNMENT Department of Primary Industries

ARRC Australian River Restoration Centre