Demonstration of the climate adaptation toolkit workflow

A report prepared for the
Murray–Darling Water and Environment Research Program

Report T1.TK3

Georgia K. Dwyer[1], Galen Holt[1], David Robertson[2], Joel Bailey[3], Martin Job[3], Lara Palmer[3] and Rebecca E. Lester[1]

[1]Centre for Regional and Rural Futures, Deakin University, [2]CSIRO Land and Water, [3]Murray–Darling Basin Authority

ACKNOWLEDGEMENTS

This work was undertaken as a part of the Murray–Darling Basin Water and Environment Research Program (MD–WERP) Climate Adaption Theme. The MD–WERP is an Australian Government initiative to strengthen scientific knowledge of the Murray–Darling Basin. It is managed through a partnership between the Department of Climate Change, Energy, the Environment and Water, the Commonwealth Environmental Water Office and the Murray–Darling Basin Authority. The Climate Adaptation Theme brings together researchers from CSIRO, Deakin University and eWater.

The authors pay respect to the Traditional Owners and their Nations of the Murray−Darling Basin. We acknowledge their deep cultural, social, environmental, spiritual and economic connection to their lands and waters.

Location:

Geelong Waurn Ponds Campus

Locked Bag 20000

Geelong Victoria 3220 Australia

For more information about CeRRF, Deakin University, visit the website www.deakin.edu.au

EXECUTIVE SUMMARY

The impacts of climate change will affect the social, economic, environmental and cultural values of the Murray–Darling Basin. Some elements will be more vulnerable to climate change than others. There will be many options for avoiding, ameliorating or adapting to these impacts. There will be trade-offs, synergies and uncertainty associated with each option.

To enhance the ability of the Australian Government and stakeholders to assess likely future impacts of climate adaptation across social, environmental, cultural and economic values in the Basin, we have developed a demonstration toolkit of new and existing information, knowledge and models. This will enable transparent, repeatable assessments of impacts and adaptation to future climates.

The toolkit enables end-users to choose future climate scenarios, identify possible adaptation options and then model the flows associated with those scenarios. These can then be used to understand the resultant response of identified values, be they social, cultural, environmental or economic. This new capability will enhance the ability of the Australian Government to account for the impacts of future climate change on water supplies and Basin assets in water planning. It will support non-flow-related decisions.

This report outlines the demonstration of the key functionality of the toolkit. For the toolkit, 3 major work components were identified. These were:

1. develop the ability to analyse and run scenarios (ModelArch)
2. develop modules
3. develop the causal network.

The first activity, to develop ModelArch, is the backbone of the toolkit. Here we demonstrate the basis for ModelArch, the foundational architecture for the toolkit, which will now accept inputs, undertake a number of functions and provide outputs. ModelArch has been developed as a series of elements to undertake those functions: the Scenario Controller, the Indicator Assessor, the Objective Translator, the Aggregator and the Comparer. Sitting within ModelArch will be modules describing responses of each of social, cultural, economic and environmental values. To date, an environmental module has been included for demonstration purposes.

The second activity will continue to develop these modules to enable a repeatable method for defining the response of values and assets to hydrology and climate.

The final activity, developing the causal network, involves capturing the rationale for links between hydrology, climate and the response of values and assets. A demonstration causal network is presented that helps to identify synergies and trade-offs among values represented to date. It is also then able to assist in the development of methods to scale the simulated responses in space and time.

In addition, we also present a clear decision tree to enable transparent choices regarding aggregation of components within ModelArch. The decision tree presents the rationale for different methods to aggregate in space, through time and across different elements within the quadruple bottom line of social, cultural, economic and environmental values.

This demonstration illustrates a minimal but functioning version of ModelArch and provides the foundation for additional development. As such, it provides the basis of a toolkit to assess the impact of climate adaptation measures on social, cultural, environmental and economic values and assets in the Murray–Darling Basin. It will provide the foundation for populating that toolkit over the lifetime of the Murray–Darling Water and Environment Research Program.

1 CONTENTS

2 FIGURES

3 TABLES

4 INTRODUCTION

The impacts of climate change are affecting the social, economic, environmental and cultural values (the quadruple bottom line) of the Murray–Darling Basin. Some elements are more vulnerable to climate change than others. There are many options for avoiding, ameliorating or adapting to these impacts that could be implemented in isolation or in combination. There are trade-offs, synergies and uncertainty associated with different adaptation options and the resulting outcomes on Basin values and vulnerabilities. Many impacts and adaptation options are outside the Australian Government's control – the states, industry and communities all play important roles. Thus, the task of assessing the potential impacts of climate change on a healthy, working Basin is complex.

As a part of the Murray–Darling Water and Environment Research Program (MD–WERP), this project aims to enhance the ability of the Australian Government and other stakeholders to identify and assess the likely future impacts of climate change across social, environmental, cultural and economic values in the Basin, and to identify potential adaptation options and assess the outcomes associated with adopting them. This will involve developing a toolkit of new and existing information, knowledge and models to enable transparent, repeatable assessments of impacts and adaptation to future climates.

This toolkit will enable end-users to select future climate scenarios, model the flows associated with those scenarios and incorporate other relevant information to understand the response of identified values, be they social, cultural, environmental or economic. Users will then be able to access simulations of a range of possible adaptation options and assess their impacts on those values.

The toolkit will be applied to a series of adaptation investigations across the Basin. This will result in new capability. It will enhance the ability of the Australian Government to account for the impacts of future climate change on water supplies and Basin assets in water planning, and it will support decision-makers involved in non-flow-related decisions (*see Architecture design for the Climate Adaptation toolkit workflow* (Holt et al. 2022) for a full description of the toolkit capability).

The purpose of this report is to provide a demonstration of the toolkit, including the current functionality, the components used to create that functionality, and the plan to develop each component. Where required, we have documented the rationale for the choices that have been made.

5 MAJOR COMPONENTS

The toolkit architecture has 3 major components that have been the focus of development, each of which has additional internal divisions. These components, though they are highly inter-related, are conceptually distinct and can be developed in parallel. It is important to note that the 3 work components are not equally complex or time-consuming.

The 3 major work components are:

1. Develop the ability to analyse and run scenarios (ModelArch)
2. Develop modules
3. Develop causal networks

1. Develop the ability to analyse and run scenarios (ModelArch)

This work component is the backbone of the toolkit. It provides the functional architecture for the toolkit and ensures that scenarios are assessed in repeatable ways that use each of the components in the toolkit, and can then be scaled, compared and assessed. Creating the ability to analyse and run scenarios is predominantly a code-building task, but it will also include significant development of the science that is captured by that code.

ModelArch accepts input data, undertakes a number of functions capturing responses to those inputs, and then provides outputs. It is, therefore, best conceptualised as a model itself. As a result, modelling terminology such as ‘running’ scenarios will be used in this document and others relating to the toolkit. It is important to note that ModelArch ingests hydrological data from a hydrological model rather than incorporating the hydrological model itself. These hydrological data are then run through the workflow described in this document to model their impacts on social, economic, environmental and cultural outcomes. These modelled impacts are then used to assess the effects of various adaptation options. Comparisons among those scenarios are the final output.

This work component forms the primary focus of the toolkit demonstration described here.

2. Develop modules

This work component focuses on defining the responses of values and assets to hydrology and climate. These responses will be defined based on the best available science and using formats usable for assessment by the ModelArch component. Each element in the quadruple bottom line (environmental, social, cultural, and economic) will be represented, although to date the flow-to-outcomes science for some elements is much better developed than for others.

Initially, the focus of this work has been to add needed functionality to existing tools defining relationships between hydrology and environmental outcomes – for example environmental watering requirements (EWR) – while developing these relationships for other elements (social, cultural, and economic). Reflecting the state of the hydrological modelling, these tools initially reflect responses to hydrology alone, but they may be extended to include other impacts influenced by future climate (for example heatwaves, bushfire risk, water quality).

To date, we have focussed on adapting the existing environmental watering requirements tool for use in the demonstration, but we have not further developed other elements of the quadruple bottom line (see Holt et al. 2022 for a full description of the planned future development).

3. Develop causal networks

This work component captures the causal relationships between climate, adaptation options and outcomes for environmental, cultural, social and economic values and assets (that is, the elements of the quadruple bottom line) (adapted from Peeters et al. 2022). The focus is currently on those relationships described by the modelling, but others may be included as well, aiding in the identification of blind spots. The causal network identifies and demonstrates those relationships and captures characteristics such as the degree of certainty associated with each link. The causal network has the capacity to be spatially explicit, capturing differences in causal relationships among locations. This work component ensures that the links within the model are appropriate and will aid communication of the approach externally.

A demonstration of the causal networks and development to enable them to be used to support the toolkit is included in this report.

The 3 work components are illustrated in Figure 1, including the various links among the components. Our approach to tackling this overall body of work has been to develop the components in parallel, initially focusing on having a functional structure and then iterative improvement. This approach means that no single component needs to be complete for others to proceed. For example, we built our demonstration of ModelArch based on the existing EWR tool, without needing to develop other modules first. Similarly, whilst updating the EWR tool or adding new modules requires new runs to obtain updated output values, it does not need fundamental changes to ModelArch itself.

5.1 Rationale

Developing the ability to analyse and run scenarios (via ModelArch) represents the core functionality of the toolkit (shown in blue in Figure 1). This element enables users to model the links from climate and adaptation options to outcomes presented in a useable manner. This component takes many scenarios (for example of climate or hydrology), feeds them to a driver-indicator-response model (such as the existing EWR tool), reports outcomes and enables comparisons among scenarios.

The second major component is module development (shown in teal in Figure 1). This component will capture the relationships that link the outcomes arising from the hydrology to the elements of the quadruple bottom line. This component is essential for ModelArch, as it creates the basis of the driver-indicator-response model.

Finally, the third major component is the causal network development (shown in grey in Figure 1). This component creates a visual representation of the complex inter-relationships between river-related outcomes that are represented in the toolkit. This aids transparency, reducing the risk that the toolkit becomes a ‘black box’, and it improves communication about the toolkit and its outputs.

Each of these components is discussed in further detail below.

DEMONSTRATION OF THE CLIMATE ADAPTATION TOOLKIT WORKFLOW

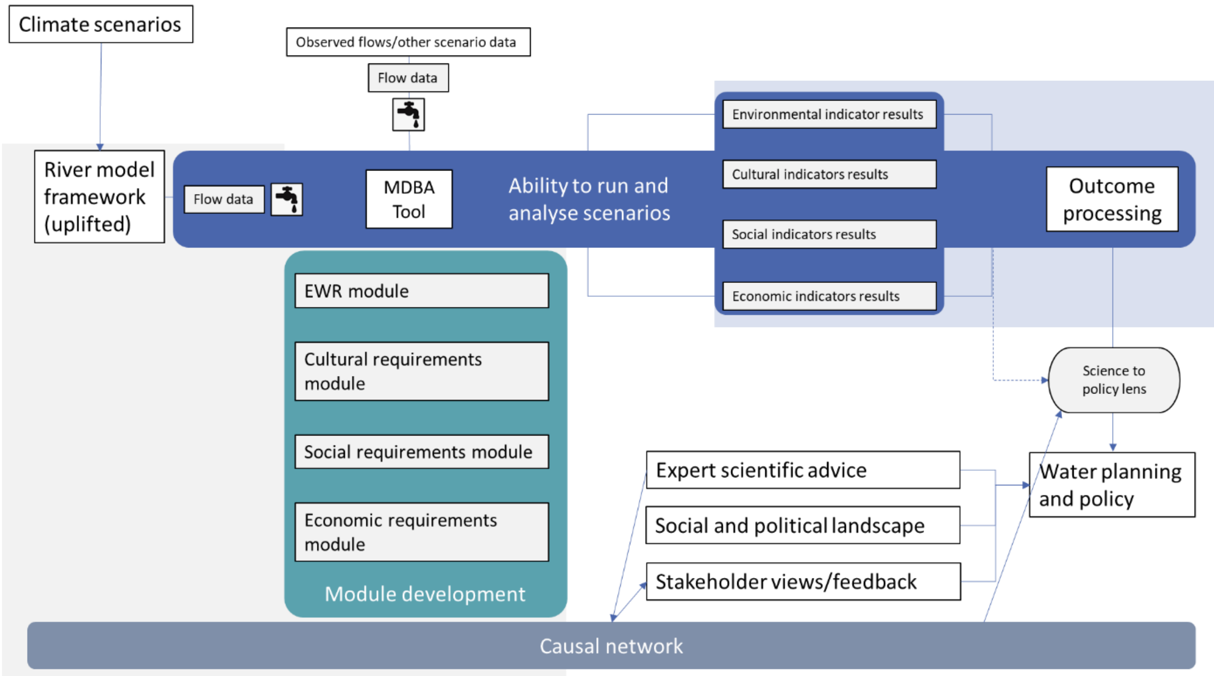

Figure 1. Conceptualisation of the near-term components of work in the MD–WERP toolkit (filled boxes). This sits downstream of the climate scenarios, adaptation option definitions and hydrological modelling. It feeds out results in a format useful for the science-to-policy lens.

6 RUN AND ANALYSE SCENARIOS (MODELARCH)

The primary aim of the toolkit is to assess responses of elements in the quadruple bottom line (environmental, cultural, social and economic) to changes in climate and possible climate adaptation options. To do this, we require capability to run and analyse various scenarios of climate and climate adaptation. These analyses must be scientifically robust, and the outcomes must be interpretable and usable for the MDBA and other stakeholders at the science/policy interface. The ModelArch component of the toolkit provides the core modelling functionality for the toolkit project to achieve these needs.

The ModelArch component consists of a series of modelling modules that step through from hydrology scenarios to local outcomes at the quadruple-bottom-line element levels and process those outcomes for presentation and use. Each of these modules will capture specific steps in the process and will be based on the best available science and needs. Each step in the process includes uncertainty, from the climate scenarios themselves through to the effects of local changes on larger-scale outcomes. The ModelArch components will be developed to propagate this uncertainty through, in order to assess the confidence in the final comparisons among scenarios (although this propagation does not form part of the initial demonstration).

The modular approach outlined here allows each step in the modelling process to be developed as a discrete element, simplifying and clarifying development. Each component incorporates the best available relationships defining that particular step. Updating a module does not, therefore, require redevelopment of the architecture or other modules. Rather, the toolkit would only need to be re-run to obtain updated outputs. This approach allows iterative improvement in terms of the knowledge captured, capability enhancement and output needs. For example, adding new responses, changing aggregation statistics, or producing new visualisations of the results will all be straightforward.

6.1 Overview

Taken holistically, the architecture comprises 4 modelling components (blue boxes in Figure 2) and the links between them, along with hydrological inputs (Figure 2).

The architecture ingests flow timeseries at one or more locations. For example, in the case of the EWRs, these are in the form of hydrographs at gauges. Climate and adaptation scenarios are represented by different hydrographs for each scenario, developed and specified outside the toolkit architecture described here (for example in hydrological models). The following steps are carried out:

1. Each scenario at each gauge is fed into the Indicator Assessor module via the Scenario Controller. The Indicator Assessor module will check the hydrograph against a database matching hydrological indicators (for example flow requirements) to quadruple-bottom-line element-based objectives to determine which, if any objectives, are met.
2. The hydrological indicators are defined because they have causal relationships with quadruple-bottom-line element-based objectives (for example, bankfull flows of a given duration are needed for successful fish breeding). Taken with step 1, these represent simple models for a range of local outcomes associated with elements of the quadruple bottom line, dependent on hydrological conditions. Thus, the outcomes of the hydrological indicators are mapped back to these element-based objectives in the Objective Translator.
3. Local outcomes from each gauge can be scaled up to larger areas and time steps to meet user needs via the Aggregator. The spatial scale at which results are needed is likely to vary, depending on end-user needs, but it will typically be larger than a single gauge,

encompassing perhaps a valley or a statistical area. This scaling can be done using a range of methods tailored to the specific indicators and the use case.

4. Finally, aggregated outcomes are compared among the scenarios in the Comparer module. These comparisons allow visualisation and assessment of the differences among adaptation options and climate trajectories. The method of comparison can be tailored to the specific use case and required types of outputs.

Each of these components is described in more detail below.

All components will be version-controlled, including current software environments, to ensure that all steps of the toolkit workflow are repeatable. Strict reproducibility is significantly more difficult, but archiving of foundational data (to the extent possible) combined with versioning at major analyses will provide near-reproducible results. Unit tests with minimal standard data will be used to ensure repeatability and expected behaviour, particularly across versions with updated functionality. This version control functionality has not been developed in the initial demonstration.

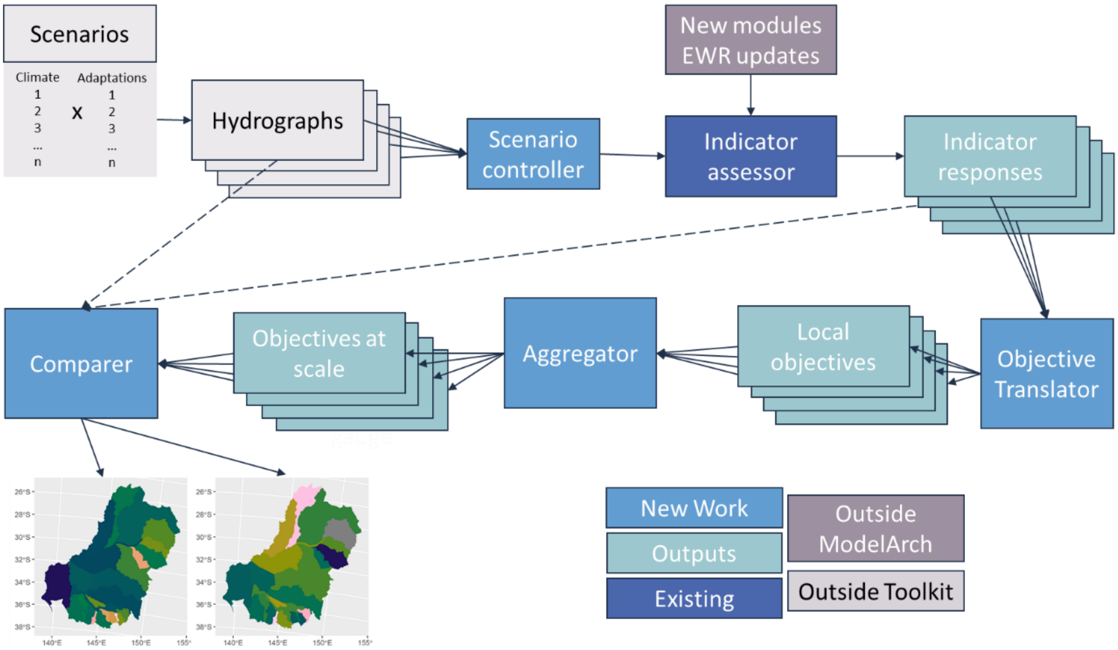

Figure 2. Flow of the ModelArch. This work provides the modelling functionality for the toolkit. The goal is to produce useful analyses and comparisons of how climate scenarios impact the 4 elements of the quadruple bottom line. To do so, the architecture ingests scenario hydrographs, tests whether they meet hydrological thresholds, translates those hydrological thresholds into local management objectives relevant to the elements (for example bird breeding, cultural values), then aggregates the results to a desired scale and compares the scenarios in a way that is useful for the MDBA science-policy interface.

6.2 The Scenario Controller

Assessments of the effect of adaptation options under climate change often require many scenarios. These scenarios each represent one adaptation option and one example future climate or build understanding of stochastic variability. To effectively manage multiple scenarios, a Scenario Controller is needed.

The Scenario Controller is the interface between the externally generated hydrographs defining the scenarios and the toolkit architecture described here. Specifically, its function is to take those hydrographs and systematically provide them to the indicator modules for modelling their impact on the quadruple bottom line (social, cultural, environmental, and economic) outcomes.

The Scenario Controller identifies the location of data representing relevant hydrology scenarios, loads them (or points to them) and feeds them to the indicator modules. To accomplish this, a user needs to define the scenarios and their locations and control which indicator modules to run. The core functionality is delivered via a lightweight script and simple functions that loop over those scenarios and call the indicator modules.

The Scenario Controller provides a record of each run and its setting to enhance repeatability, ensure correctness and increase comprehension. This includes the run-specific setup of the Scenario Controller itself (that is, the files called). In addition, the Scenario Controller also includes metadata explicitly, attaching information about the scenarios that will be carried through each step of the architecture. This ensures that outputs at every stage are tagged with information about the run that created them. Further, the Scenario Controller sets up standardised output locations, including records of the run settings, to aid recordkeeping and future usability of assessments.

The implementation of this component is via a user-editable notebook in Jupyter to define the scenarios and make any other necessary settings. The notebook is an interface that combines text, code and output, enables user input, and can be version controlled. Once edited, the notebook then calls a small set of standard functions to tag the run with metadata and control sending the hydrographs to the indicator modules. There is clear opportunity to enable these functions to send scenarios in parallel to improve processing of multiple scenarios simultaneously.

Further development may include additional input data other than hydrographs. For example, if the indicator modules are expanded to include factors other than hydrology that define responses, such as temperature or local socio-economic status. As testing and development progresses, this interface may also be used to define the analyses performed on the output (for example, to select spatial scale units and aggregation statistics), making the user-editable interface of this component an overall controller for a particular set of scenarios.

6.2.1 Demonstration

The core functionality of the Scenario Controller has been developed using existing hydrographs and the EWR tool for development, testing and demonstration. The build has consisted of identifying the components that need to be in the user-editable interface, developing functions to send sets of hydrographs defining scenarios (for example, Figure 3) to the indicator modules, including the ability to attach run information (for example, Table 1) and enabling storage and accessibility of assessments. The demonstration used here consists of 5 gauges across two MDBA planning units, as illustrated in Figure 3.

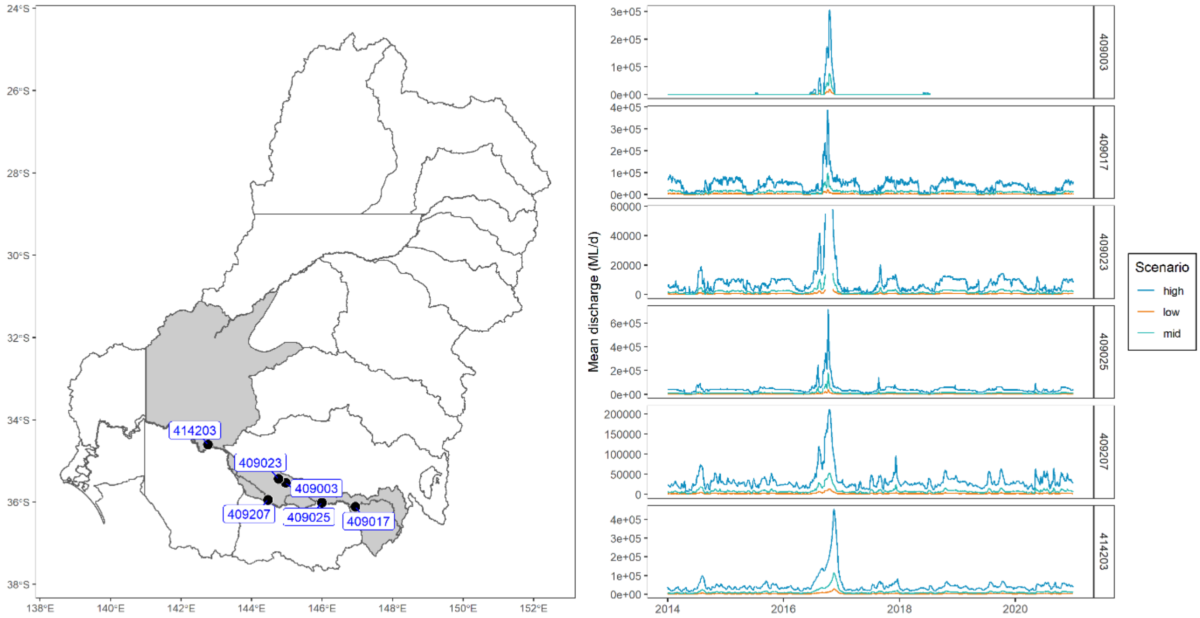

Figure 3. The Scenario Controller ingests flow timeseries at specific locations. In this example, these are in the form of hydrographs (right) at specific gauges (indicated on the map, with the associated water resource plan areas shown in grey). Hypothetical high (blue), mid (green) and low (orange) climate scenarios are represented by different hydrographs for each scenario, which were developed for demonstration purposes only, outside the toolkit architecture.

Gauge number	Scenario	Data present (years)	Data missing (days)
409003	high	2013-2020	901
409003	mid	2013-2020	901
409003	low	2013-2020	901
409017	high	2013-2020	0
409017	mid	2013-2020	0
409017	low	2013-2020	0
409023	high	2013-2020	0
409023	mid	2013-2020	0
409023	low	2013-2020	0
409025	high	2013-2020	0
409025	mid	2013-2020	0
409025	low	2013-2020	0
409207	high	2013-2020	0
409207	mid	2013-2020	0
409207	low	2013-2020	0
414203	high	2013-2020	0
414203	mid	2013-2020	0
414203	low	2013-2020	0

Table 1. The Scenario Controller attaches run information such as that shown here in this table, sets up the storage structure for all outputs, and sends the input data to the selected Indicator Assessor module.

6.3 The Indicator Assessor

The impact of adaptation options and climate on quadruple-bottom-line element-based objectives will be assessed based on whether specified water requirements (that is, hydrological indicators) are met for a given scenario. Thus, hydrographs for each scenario need to be assessed to determine whether those requirements are met. The Indicator Assessor will perform this task.

The existing EWR tool already fulfills the purposes of the Indicator Assessor tool, with the original interface replaced by the Scenario Controller. There will likely need to be some minor modifications as we add requirements beyond the environmental element of the quadruple bottom line, but we expect the structure to remain the same. Currently, the EWR tool provides quantitative indicators (for example duration of bankfull flows) and pass/fail indicators that link hydrological conditions to environmental objectives, so both these formats are illustrated in the toolkit demonstration. Newly developed modules will be able to incorporate both formats at the outset.

The indicator modules themselves are the databases defining water requirements for quadruple-bottom-line element-based objectives. The development of these water requirements is described in the Model Architecture document (Holt et al. 2022). The Indicator Assessor tool is the software that ingests these databases and the hydrographs, calculates the necessary hydrometrics and checks whether each requirement is met. That assessment for each of the hydrological requirements is then sent to the Objective Translator module to assess the implications for the quadruple-bottom-line element-based objectives.

6.3.1 Demonstration

For the ModelArch demonstration, the existing EWR tool is sufficient. Outputs from the EWR tool (for example see Figure 4) are stored in the output folder structure created by the Scenario Controller and are then passed to the Objective Translator. The EWR tool works well with the Scenario Controller and Objective Translator and no modifications were required to the functions of the EWR tool. However, the Scenario Controller now runs these functions rather than using the EWR tool interface, reflecting the different use cases. As new quadruple-bottom-line element requirements are added or updates made, they will be incorporated and tested.

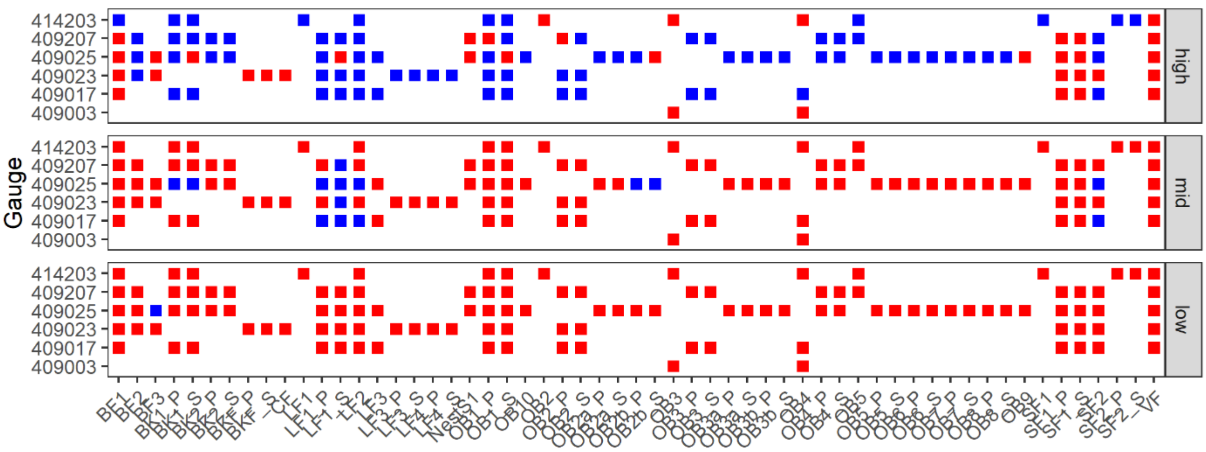

Figure 4. Outputs from the EWR tool are in the form of tables but can be visualised for the demonstration by functions provided in the Comparer. This tile plot illustrates the pass (blue)/fail (red) results for relevant EWRs (which are listed across the bottom – for more detail see Appendix A). Irrelevant EWRs are not shown (represented as missing tiles) at each gauge under each scenario.

6.4 The Objective Translator

Objectives are defined based on environmental or other values (the quadruple bottom line) and are modelled based on hydrological indicators. To determine the impact on those objectives, translation of those hydrological indicators is required. The Objective Translator links hydrological indicators to the expected requirements of environmental (or other) factors.

Indicator modules define the hydrological indicators (for example bankfull or cease-to-flow requirements), and the Indicator Assessor checks whether the requirements for given quadruple-bottom-line element-based objectives are met. These requirements are defined by their causal relationship to the element-based objectives (see Causal Network Development section below). For example, long-term watering plans (LTWPs) set hydrological indicators based on the expected requirements of ecological groups. The Objective Translator makes these links, taking the pass/fail of the water requirements at each gauge into the element-based objectives (for example no loss of native fish species).

The Objective Translator is similar to the Indicator Assessor: it ingests a database (here, matching environmental water requirements to their respective environmental objectives) and the outputs of the Indicator Assessor tool (here, table of pass/fail results for each environmental water requirement). Hence, there are 2 parts to the Objective Translator. The first is the database which matches the objectives to the water requirements; the second is the tool that matches the 2 and returns the outcomes. As new modules are created, objective mapping can be incorporated from the initial build and so database creation can occur simultaneously for both water requirements and objectives.

The Objective Translator returns an output indicating whether each element-based objective is met, based on whether its water requirements are achieved. Some objectives (for example no loss of native fish species) have multiple associated water requirements (for example cease-to-flow requirements and bankfull) at a single gauge. Aggregating these into a single native fish assessment will occur in the Aggregator tool; the Objective Translator returns whether each water requirement for each objective has been met. This approach keeps the functions of the 2 tools separate and ensures that each is doing one task to increase the clarity, robustness, and modularity of ModelArch.

6.4.1 Demonstration

For the database construction, the first stage has been to extract the links between environmental objectives and the EWRs for each gauge from the LTWPs. This has been achieved using output from the MDBA’s EWR scraper tool, which was processed using a Deakin-developed R script for the extraction of those links. This script currently works effectively for catchments in New South Wales, as confirmed by initial manual quality checking for consistency in the database compared to the LTWPs. Quality checks are in progress, and thus far have been targeted to catchments used in the demonstration and those that will likely be used in the case studies. As quality checks progress, we continue to update the R script to resolve discrepancies in the database, rather than manually fixing errors in the database. This will enable the MDBA to use that script in internal functions and hold the point of truth for linking EWRs to environmental objectives.

The Objective Translator tool is a lightweight script that matches the water requirement outcomes from the Indicator Assessor to the environmental objectives in the database, and then returns pass/fail for each environmental objective at each gauge (Figure 5). It also links environmental objectives to target species, and the 5-year, 10-year and 20-year planning targets (see Aggregator section).

Figure 5. Pass/fail outcomes from the EWR tool are linked to the environmental objectives (listed by code on the left hand side; see Appendix A for meaning) by the Objective Translator for each gauge under each scenario according to the LTWPs. Here, pass/fail outcomes for each EWR are illustrated with the EWR box colour and the linking line colour (pass = blue, fail = red). Many environmental objectives depend on multiple EWRs, some of which may pass while others fail. This multiple dependence is why the environmental outcomes are not given a pass/fail at this stage, but instead each of the links (lines) is carried forward into the aggregator.

6.5 The Aggregator

One of the key tasks for the toolkit will be to aggregate responses in space and time and across environmental objectives (or water requirements, or multiple species). Best available science indicates that a flexible approach to aggregation is needed. Specific objectives are best combined in different ways, depending on the intent of those objectives. Thus, a standard approach to all objectives could produce misleading results. As a result, we have built in several options for aggregation, and we will expand this list as toolkit development progresses. We have also included the ability to tailor new aggregation options by users. This enables users to select how best to aggregate in 3 dimensions: space, time and across objectives.

To illustrate the need to develop a flexible approach to aggregation, an example is useful. Consider 2 different environmental objectives, 'no species loss' and 'successful bird breeding'. The most appropriate method to aggregate each is likely to be very different. To meet the 'no species loss' objective, all species must persist in all locations at all times (otherwise species loss will have occurred at one or more gauges). From an aggregation perspective, this means that locations, time steps and species cannot compensate for one another: the loss of a species at any point or in any location would cause the overall objective to fail. Given the manner in which the EWR tool functions, a mathematical minimum is likely to be the most appropriate method of aggregation to represent this set of circumstances.

In contrast, the 'successful bird breeding' objective should be aggregated using quite a different approach. Here, different species will have different requirements regarding a minimum interval between breeding events. These might range from one to 5 years, for example. For a given species, successfully breeding once within that interval would qualify as success, and so breeding is not necessarily required in all years or locations for the objective to be met.

Different degrees of spatial aggregation will likely be required for different species in the bird breeding example. For example, for some species, successful breeding at one location in the Basin will be sufficient whilst, for others, localised breeding will not be sufficient. As with the 'no species loss' objective, successful breeding in one species can not compensate for a lack of breeding in another species. Thus, no compensation is possible across species, but for each species success in some locations and some times can compensate for a lack of success at other locations or times. As a result, a mathematical maximum could be an appropriate aggregation method for the outputs of the EWR tool for the identified spatial extent and time frame for breeding, with each tailored to the relevant species' requirements. Following this spatial and temporal aggregation, a geometric mean may be a useful way to combine species into a single aggregated 'successful bird breeding' objective. To reduce complexity for users, we have begun to develop a decision tree to assist with the selection of aggregation methods in space and time and across objectives. We plan to incorporate a default set of mappings between objectives and appropriate statistics (see the Decision Tree section below).

The need for the Aggregator arises because of the way quadruple-bottom-line element-based objectives are defined. For example, the LTWPs define local environmental objectives at each gauge, because that is as close as possible to the scale of the causal relationships with hydrology. Parallel modules for other elements of the quadruple bottom line (social, cultural, and economic) are expected to be similar, and there is a requirement to report against basin-scale objectives (for example as per Murray–Darling Basin Authority 2019). Most uses of the toolkit are expected to require aggregation from the scale of the causal drivers of the objectives to some larger scale, whether the whole Basin, individual catchments, sustainable diversion limit (SDL) units or statistical areas. The particular scale of this aggregation will depend on the specific use of the toolkit as well as the objectives themselves (as described in the example above). For example, economic or social objectives may best be assessed at the statistical area level, while environmental objectives may be more appropriate to consider in catchments. Work related to the Basin Plan may aggregate to the entire Basin, while community engagement might require assessing responses at the scale of catchments or culturally important sites.

Aggregation of objectives is also valuable. For many objective types (for example maintaining populations of native fish), multiple water requirements are defined (for example bankfull and low flow). Aggregating these into a single local objective outcome for native fish populations would be ideal, but this requires understanding the causal relationships driving each of those water requirements in detail. As illustrated above, the most appropriate method of aggregation is linked to the objectives themselves and whether compensation is possible.

To illustrate, for maintaining populations of native fish, a bankfull requirement may provide spawning opportunities while a low-flow requirement may ensure refugia during drought. Aggregation would require assessing how these relationships fit together to affect fish populations – if all adults die because there are no refugia, then subsequent spawning opportunities are moot. There are several potential pathways. A strictures and promoters approach (Lester et al. 2020) is particularly well suited, as it captures dependencies between the requirements (typically related to life cycles, though these could also relate to economic, cultural, or social outcomes) but it does not require complex models. Another approach is a multi-criteria decision analysis style of assessment. Similarly, aggregations across elements may be valuable for decision-making. For the purposes of the toolkit demonstration, the necessary detail for this sort of aggregation does not exist in the LTWP. Thus, the Aggregator component developed here focuses on the separate local objectives, while ensuring that future aggregation of those objectives can be readily integrated.

Finally, to ensure robust aggregation and reduce bias, we aggregate the responses (local objectives), not the drivers (hydrology). By scaling the responses, we avoid the large potential biases created by nonlinear relationships between drivers and responses (Jensen's inequality: Ruel and Ayres 1999). This approach is particularly important when using pass/fail outcomes because this is an extreme sort of nonlinearity and so is particularly susceptible to large errors.

Thus, the purpose of the Aggregator is to scale from individual quadruple-bottom-line element-based objectives at local gauges to a flexible set of larger scales in space and time and across outcome types in a robust, general way that minimises bias in the outcome.

6.5.1 Demonstration

From a technical perspective, the core of the Aggregator is a function that takes quadruple-bottom-line element-based objectives, spatial polygons (if aggregating in space), timeframes and an aggregation statistic (for example minimum or geometric mean in the example above), and returns an aggregated outcome. Surrounding code prepares those polygons, times and statistics. We use spatial aggregation functions in existing spatial packages and incorporate additional error-checking and data handling to tailor inputs and outputs to specific needs. The objective inputs come from the Objective Translator, but the polygons into which the outcomes will be aggregated and the statistics are specified by the user.

The Aggregator has been built and tested using the EWR tool outputs. Currently, we have incorporated sets of default spatial units, including the Basin boundary, catchment boundaries and MDBA planning units; but the Aggregator could incorporate other spatial units such as ABARES statistical areas. Likewise, the mathematical functions to achieve aggregation (as described above) are included as default options, including the arithmetic mean, geometric mean, limiting factor and compensating factor summarisation (See***Error! Reference source not found.*** section below for details). However, bespoke mathematical functions can also be easily created and applied.

A subsequent step will be to develop code to parallelise the aggregation process. This code will manage the set of objectives to aggregate and their statistics, feeding them in parallel to the aggregation function and processing the results, including checking that the spatial information is correct following aggregation. The resultant spatially-referenced aggregated outcomes will then be output for use by the Comparer component.

In Figure 6 and Figure 7, 2 quantitative indicators (overbank flows 1 [OB1] and very low flows [VLF]) at individual gauges have been aggregated to the planning unit scale using 2 different aggregation methods (arithmetic mean and compensating factor summarisation). We have used a hypothetical high, mid and low water scenario for illustrative purposes. Aggregation of EWR pass/fail outcomes is illustrated in the ***Error! Reference source not found.*** section below, as is how outcomes are influenced by the choice of spatial aggregation method.

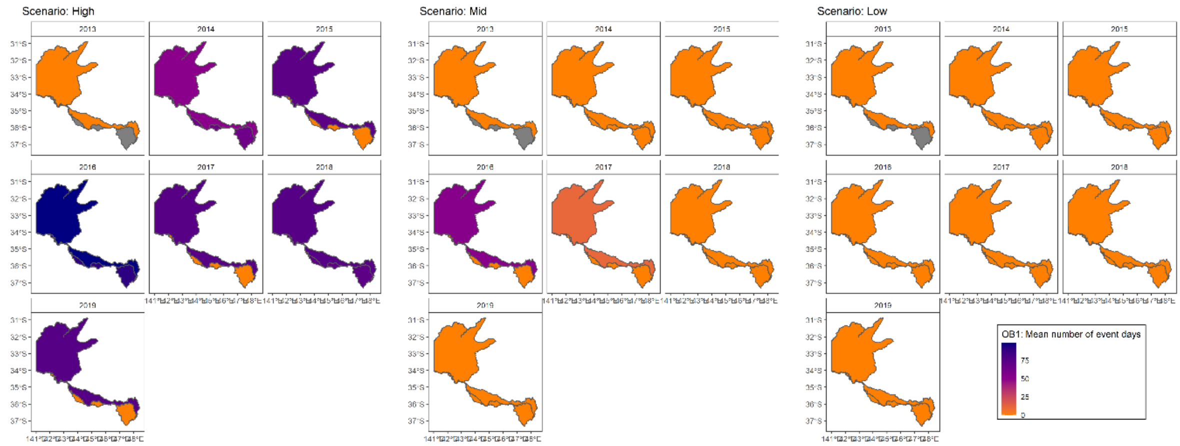

Figure 6. Spatial aggregation of a quantitative indicator (days of overbank flows) using the arithmetic mean. Time has been aggregated to the yearly scale, as is achieved internally by the EWR tool. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Refer to text for details. Grey indicates NA values, and occurs here due to a missing value for one gauge in 2013.

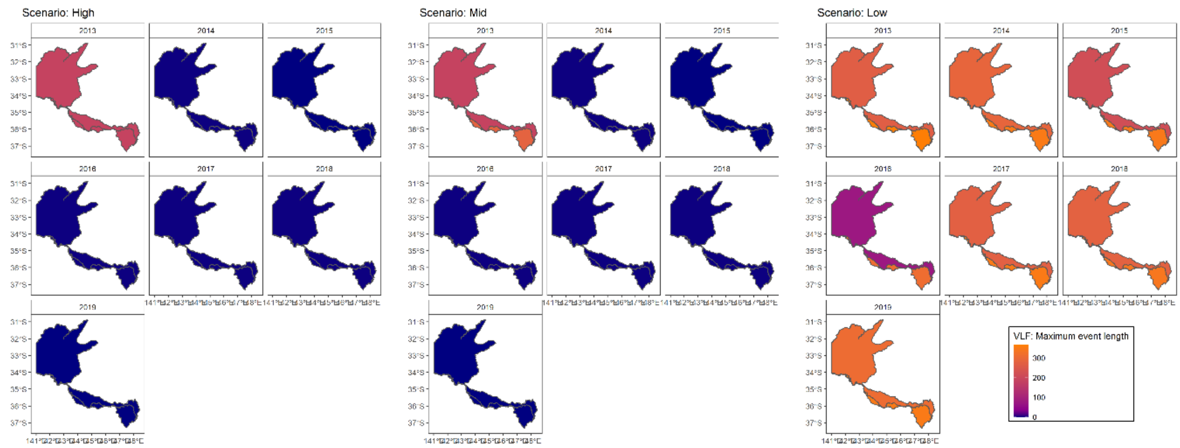

Figure 7. Spatial aggregation of a quantitative indicator (event length of very low flow days) using the compensating factor method. Time has been aggregated to the yearly scale, as is achieved internally by the EWR tool. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Refer to text for details.

6.6 The Comparer

Comparisons are essential to assess climate scenarios and adaptation options. The Comparer is designed to make comparisons between scenarios, allowing assessment and visualisation of their differences. Each comparison will characterise the simulated impact of different climate scenarios or of adaptation options under given climate scenarios. The best method for comparing will vary depending on the intended use of the comparison so, as for the Aggregator, a number of common default options have been developed, with more continuing to be developed as the toolkit progresses. Flexibility is provided to enable users to define alternatives.

Comparisons provide distinct advantages over reporting absolute values from modelling. Difficulty in accurately simulating a complex system means that comparing the relative outcomes between scenarios is an effective way to identify differences between them. For example, any bias in the baseline assumptions applies to both. In short, it is safer to say that scenario A is twice as beneficial as scenario B than to specify the absolute level of benefit for both.

The relative importance of differences must also be considered. Differences may exist that are not likely to be ecologically meaningful (or meaningful to other elements in the quadruple bottom line). For example, all scenarios will have different outcomes provided the ecological (or other) responses are sensitive to the differences between them, but the size of those differences determines their importance. Thus, the real-world impact of differences needs to be considered. Moreover, if the difference in predicted outcomes is small relative to the uncertainty around them, the true differences between scenarios will be hard to predict even if the difference between the average outcome is large.

As for aggregation, comparison can occur in many ways. The particular method should be selected based on the intended use and the underlying data being compared. Example options include using an arithmetic difference (that is, subtraction), multiplicative differences, or change relative to a baseline, among others. To illustrate, multiplicative differences can be particularly useful when comparing disparate objectives (for example, birds increased 3-fold while fish declined 5-fold), while an arithmetic difference between scenarios is better when large changes in small areas may obscure larger-scale patterns. In general, the relative comparisons assessed by multiplicative differences are likely to be preferred for most situations because they are less reliant on the magnitudes of the values, and so are less sensitive to uncertainty and bias in the assumptions common to all scenarios. Similarly, investigation of the effect of adaptation options at a catchment scale is likely to best occur via assessing multiplicative change from a baseline (because it eliminates the effect of catchment size). As for aggregation, we will develop a decision tree with worked examples to assist end-users.

Options to present and visualise comparisons may include maps, graphs, tables and narrative descriptions. Just as the values within scenarios must be chosen depending on the underlying data and the intended use, the form of presentation will vary as well. For example, maps are particularly useful for visualising geographic patterns, but they can be difficult to interpret if changes are small or have a temporal component. Tables and graphs typically provide more precise ability to assess values, but they cannot clearly show geographical relationships. Timeseries plots, particularly those relative to a historical baseline, are particularly useful for visualising climate trajectories, but they are most useful for quantitative data without too many locations. Different uses require different sets of outputs (for example community engagement versus internal planning).

Each of the preceding steps in the ModelArch is designed to generate quadruple-bottom-line element-based objective outcomes for a single scenario in a scientifically robust way in the most useful format. The Comparer then makes the comparisons between scenarios, allowing assessment and visualisation of their differences. The Comparer module includes 3 major components to enable flexibility:

- a set of mathematical functions that calculate the comparison values (for example relative differences)
- a set of functions that yield different sorts of visualisation outputs
- a user interface notebook (incorporated into the Scenario Controller or standalone) that allows choice as to which outputs to generate for a particular use.

Separate functions have been developed and will continue to be developed for common analyses and use cases, with careful version control to ensure repeatability of analyses. Visualisations and output types have been built that are most relevant to the demonstration. They will be extended to include others identified in consultation with end-users. Comparisons will incorporate uncertainty, allowing assessment not only of the expected differences, but also of how confident we should be in those assessments and how likely the outcomes are.

6.6.1 Demonstration

We have identified a set of initial data presentations to develop, including maps, timeseries, bar charts or similar and tables. The set of initial presentations have been chosen to capture a range of potential uses, including accentuating different scales, geographic information or temporal trends. The interface to determine which outputs to produce will likely be a user-editable Jupyter notebook, combined with the Scenario Controller (but this interface notebook is not yet implemented). Outputs are saved as universally readable files (for example JPEG, PDF) for integration into reports or presentations. Future development may include shifting the notebook format into a dashboard or auto-generation of reporting.

Much like aggregation, comparison may occur across multiple dimensions: time, space and quadruple-bottom-line element. For the temporal dimension, scenario comparisons including yearly time steps of quantitative EWRs indicators are seen in Figure 6, Figure 7 and Figure 10. The capability to observe EWR pass/fail results through rolling time steps (time since occurrence, or value over preceding 5 years) has not yet been developed.

Scenario comparisons at different levels of spatial resolution can currently be achieved with the toolkit. This demonstration, however, is limited by the implemented data set, which consists of only 5 gauges across 2 planning units. Subsequently, for purposes of this demonstration, we present example outputs for scenario comparison at different levels of element resolution. These include gauge-specific (no spatial aggregation) indicator pass/fail results relating to a target outcome (Figure 8). We also present comparisons of scenarios spatially aggregated to the planning unit scale, including indicators by quantitative (Figure 9A) or pass/fail results (Figure 9B), environmental objectives (Figure 11, Figure 12), and target-scale outcomes (Figure 13). We welcome feedback to allow iterative improvement using best practices from data visualisation and in targeting MDBA needs and uses.

For each scenario, we illustrate the indicator pass/fail results from the EWR tool relating to the native fish target (Figure 8) for each gauge individually (no spatial aggregation). This figure is a good illustration of the direct outputs from the EWR tool. However, with a more complex series of scenarios, more spatial units, and greater number of gauges, this method of comparison is unlikely to be useful for scenario comparison. Indeed, it is this large number of locations, times, targets and scenarios that makes aggregation necessary.

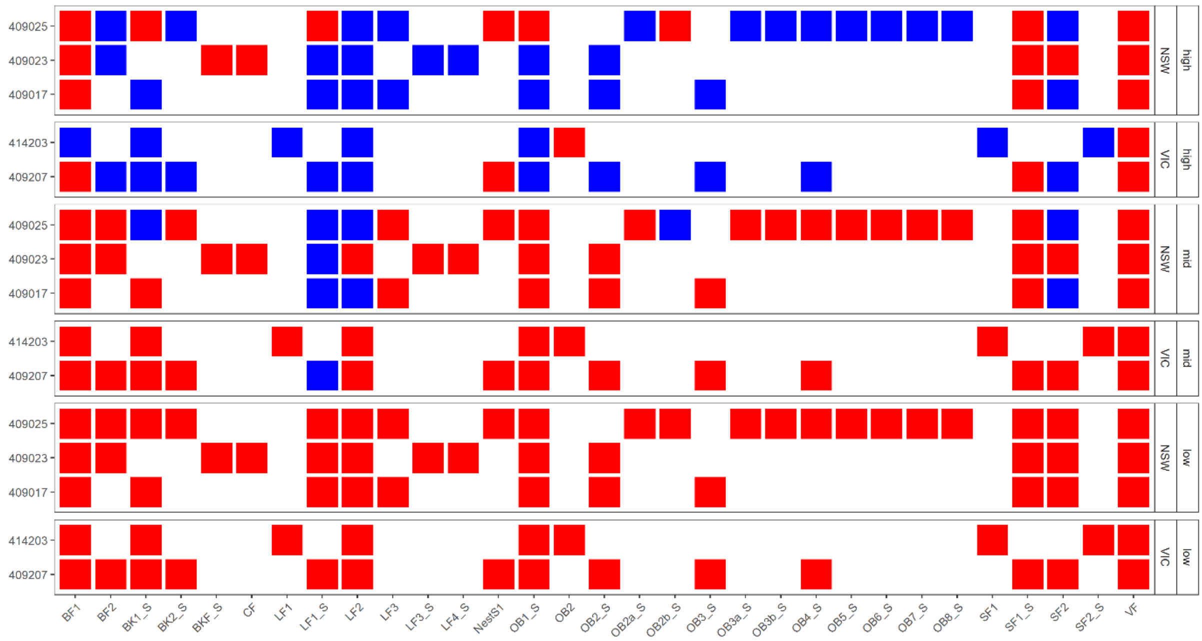

Figure 8. Comparison of all EWR pass (blue) and fail (red) results relating to native fish target for each gauge (which are listed across the bottom) shown without any spatial aggregation. Gauges for this demonstration are located in both New South Wales and Victoria, reflecting locations within different planning units and enabling demonstration of spatial aggregation. These are listed on the left hand side. Irrelevant EWRs are not shown (represented as missing tiles) at each gauge under each scenario.

To demonstrate spatial aggregation of quantitative outcomes from multiple gauges into larger spatial units, we find the mean over all gauges in each planning unit. This spatial aggregation for each planning unit and scenario provides visual comparisons of the differences between scenarios. Here, we demonstrate spatial aggregation for a quantitative indicator (mean length of time overbank flows are achieved at each gauge each year, Figure 9A) and a pass/fail indicator asking whether a target frequency of overbank flows is met (Figure 9B). In both cases, the outputs are aggregated to the planning unit scale using the arithmetic mean, with the quantitative indicator yielding the average number of overbank flow days, while the average of the pass/fail indicator yields the proportion of gauges meeting the target. This demonstration is developed at very fine detail (a single EWR) to show how the spatial aggregation works and the sorts of outputs that can be created with the minimal requirements (values at gauges). Any other value at a gauge can be aggregated similarly. While this fine level of detail will remain accessible in the toolkit, this level of detail for one EWR is unlikely to be a primary output. Instead, decision-making will likely require additional EWRs and the flow-on effects on environmental objectives.

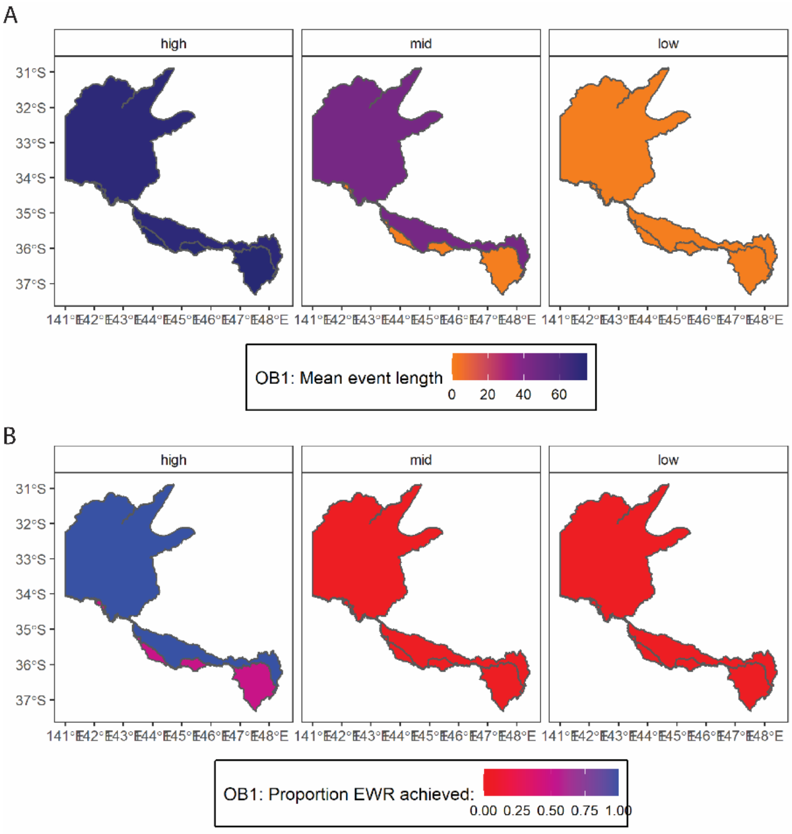

Figure 9. A) Comparison of a quantitative indicator (OB1: Overbank flows 1). Values are the mean number of overbank flow days per year at each gauge, averaged to the planning unit using the arithmetic mean over the gauges in the unit. B) Comparison of an indicator pass/fail result (OB1: Overbank flows 1). Values are the proportion of gauges that pass the target number of overbank flow days per year in each planning unit. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Refer to text for details.

Outcomes at any element and spatial scale do not occur at a point in time, but vary through time – for example, whether an EWR passes at a gauge will vary between years, as will native fish responses in a planning unit. Timeseries are ideal to visualise these changes. They can be used at any level of spatial and element aggregation (Figure 10), and they will also be important in understanding the effects of extreme events. As a demonstration using the same spatial aggregations as in Figure 9A (arithmetic mean of the quantitative OB1 [overbank flows] EWR over space), timeseries show how the overbank flow days fluctuate over time, including differences between planning units and scenarios. Distinct differences between years are noticeable, as well as different patterns in the different scenarios, likely representing the threshold nature of overbank flows.

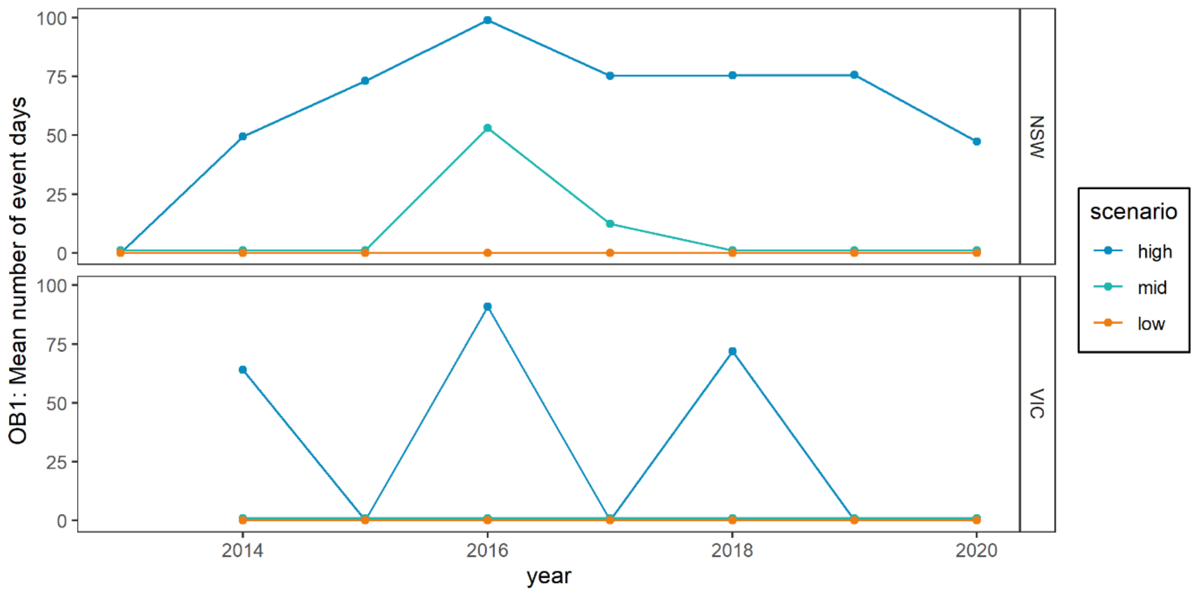

Figure 10. Comparison of a quantitative indicator (OB1: Overbank flows 1) through time. Values for each year are the mean number of overbank flow days at each gauge, averaged to the planning unit using the arithmetic mean over the gauges in the unit. Lines connect points for ease of visualisation.

To aggregate multiple hydrological indicators relating to a single environmental objective – for example, native fish 1 (NF1), no loss of native fish species – we first calculate the proportion of passing EWRs contributing to NF1 at each gauge. We then spatially aggregate by finding the mean proportion of passing EWRs over all gauges in the planning unit (Figure 11). Presenting the outcome for each scenario as a map allows rapid visual comparison, though other visualisations are possible (for example, Figure 12). This shows the proportion of EWRs that are achieved averaged over the gauges in the planning units, which gives an indication of how likely it is that the environmental objective is achieved under each scenario.

In this demonstration, we have not weighted the EWRs, and so each is assumed to be as important as any other for native fish outcomes. As noted above, additional (but not currently available) ecological information about how the different EWRs relate to fish life history may suggest different weightings or a calculation other than the mean. In this example, all planning units have similar responses under each scenario, but assessing additional planning units may suggest that some areas of the Basin are more vulnerable to loss of fish species. Moreover, even in this example, we can see that the scenarios are quite different from each other, with a much higher proportion of EWRs achieved in the high scenario than low.

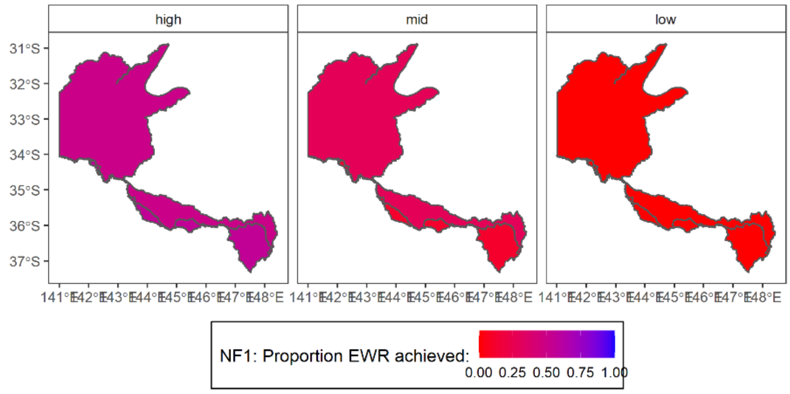

Figure 11. Comparison of all EWR indicator pass/fail results relating to an objective (NF1: native fish 1) and aggregated to the planning unit. Values are the proportion of indicators that pass at each gauge, aggregated spatially to the planning unit scale with the arithmetic mean. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Refer to text for details.

Examining similar data as a bar chart yields more quantitative detail, providing clearer visualisation of the extent of differences between scenarios (Figure 12). Further, this approach provides visualisation of all environmental objectives relating to the native fish target (NF1 to NF10 – see Figure 5). The aggregation in each bar parallels that seen in Figure 11; the multiple EWRs contributing to each objective are used to calculate a proportion of met EWRs at each gauge, and this is then averaged over the planning unit with the arithmetic mean. The bars for NF1 are therefore the same values plotted in Figure 11. By presenting these results for all objectives for the native fish target, we gain a more holistic view of fish outcomes, and the presentation also accentuates the comparison between scenarios.

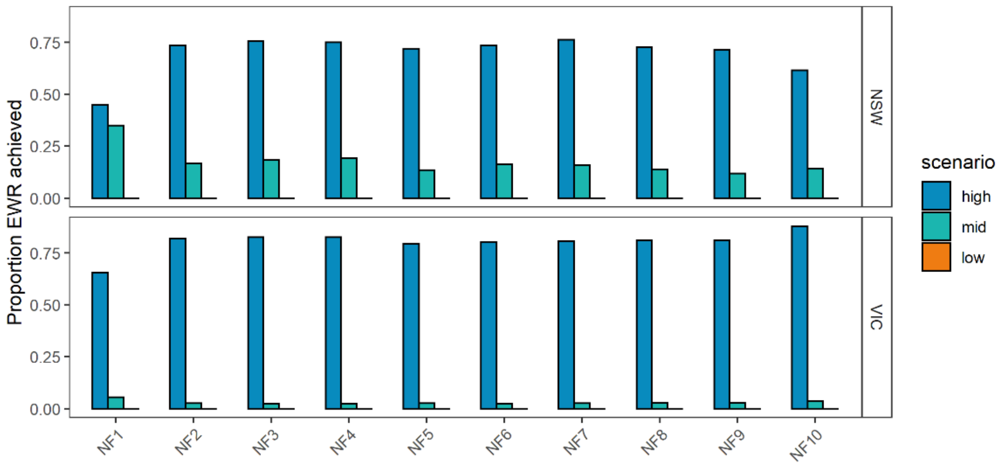

Figure 12. Comparison of all objective outcomes (NF1 to NF10) relating to a target (native fish). Values are the proportion of indicators that pass each environmental objective using the arithmetic mean to aggregate gauges spatially to the planning unit scale. All outcomes for the 'low' scenario are zero, so no orange bars are visible.

While bar charts in this instance clearly show that successful native fish outcomes are unlikely under the 'low' scenario, with a more complex set of scenarios or more spatial units this method of comparison may be less immediately interpretable, so aggregation to the target scale may be warranted. While likely sufficient in this case, the assessment of overall native fish outcomes from

Figure 12 relies on an internal mental aggregation across the objectives (bars). The toolkit provides the ability to more rigorously aggregate across objectives to the target, codifying and clarifying the rapid 'gut' assessment. Moreover, explicitly defining the aggregation over objectives can capture important quantitative relationships or dependencies in a more rigorous way than the qualitative view, particularly in more complex situations.

One possibility, which potentially closely mimics the 'gut' impression gained from Figure 12, is to take the average (arithmetic mean) of the values for each separate objective, that is to average the bars for each scenario. Presenting this outcome spatially (Figure 13) provides a useful way to both compare scenarios and detect particular areas in the Basin that are more vulnerable to negative outcomes for native fish species. Other aggregations may be specified if appropriate, as described in the *Decision tree for aggregation* section below. If, for example, NF1 is a more important outcome than the others, it could be weighted more heavily; or a minimum could be used instead of a mean in a case where fish life history cannot compensate for low (or zero) objectives and so the lowest objective controls the outcome. These decisions would require additional knowledge (outside the EWRs) about target dependence on objectives.

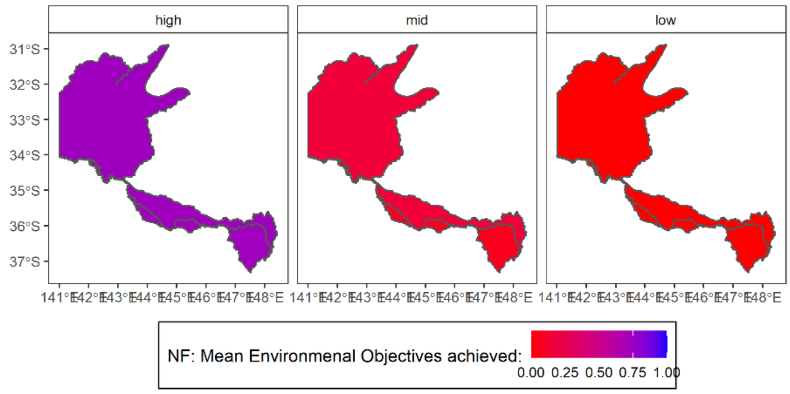

Figure 13. Comparison of one target scale outcome (native fish). Values are the mean over environmental objectives for the target (native fish, NF), where the environmental objectives are themselves quantified as the proportion of passing constituent EWRs. The values here are the average across objectives of the values shown in Figure 12 and using the arithmetic mean to aggregate gauges spatially to the planning unit scale. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Refer to text for details.

Scenarios will often be defined quantitatively, for instance as degrees of warming, flow changes or cover of farm dams. Here, our simple demonstration scenarios are deviations in flow. In this case, much insight can be gained by examining how outcomes change across the quantitative range of scenarios (Figure 14), particularly when the responses are nonlinear. Moreover, these sorts of comparisons can also provide quantitative comparisons of other categories, such as planning units (Figure 14) or target groups (Figure 15). The shift in the proportion of achieved objectives with these scenarios is clearly nonlinear, and visualisation is aided by adding lines, which can simply connect the output values (Figure 14) or be nonlinear fits to those output data to mathematically characterise the relationship (Figure 15), noting that in general we would require more than 3 points for these fits to be well-supported.

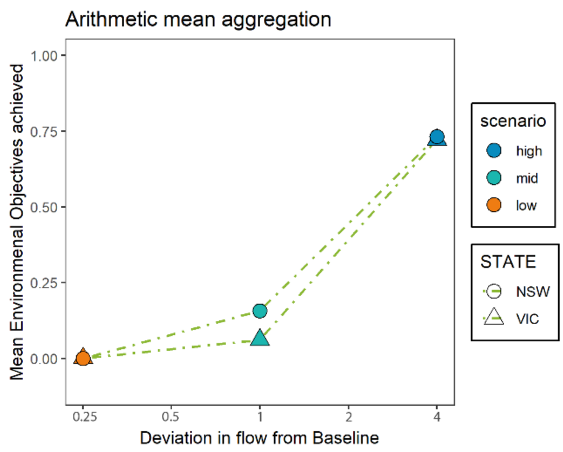

Figure 14. Comparison of one target scale outcome (native fish) with scenarios presented on a quantitative axis. Values are the mean proportion of indicators that pass each environmental objective, that is the average of the values shown in Figure 12, using the arithmetic mean to aggregate gauges spatially to the planning unit scale. Lines are used to visualise the change in outcome for each planning unit between scenarios defined as deviations in flow from the baseline (0.25x to 4x).

Showing the change in outcome on an arithmetic (or proportion) scale, as in Figure 14 and Figure 15A, is likely to be best for at-a-glance interpretation. This scale clearly shows the nonlinearity of response to the scenarios and the reduction of all outcomes to zero in the ‘low’ scenario. However, questions about sensitivity to change, such as which groups are likely to be most responsive, or which locations are most likely to be affected, are better addressed on a relative scale. In that case, the baseline differences are removed (such as might arise from planning unit area or consistent differences between fish and birds in the proportion of met requirements), and the focus is on the relative impact of the scenarios for each group (Figure 15B).

These relative changes are multiplicative, so a value of 10 means the outcome has increased 10-fold, while a value of 0.1 is a 10-fold reduction relative to baseline. Typically, relative changes would be plotted on a log-scale to make increases and reductions equivalent magnitudes. Here, the reductions to zero in the ‘low’ scenario yield infinite reductions, so we plot only the increases as multiplicative change. The high scenario increases outcomes relative to the baseline, but the different target groups respond very differently. Waterbirds (WB) and other species (OS) respond dramatically, while the other groups (EF, NF, and NV) increase but are less sensitive to the changes in flow (Figure 15B).

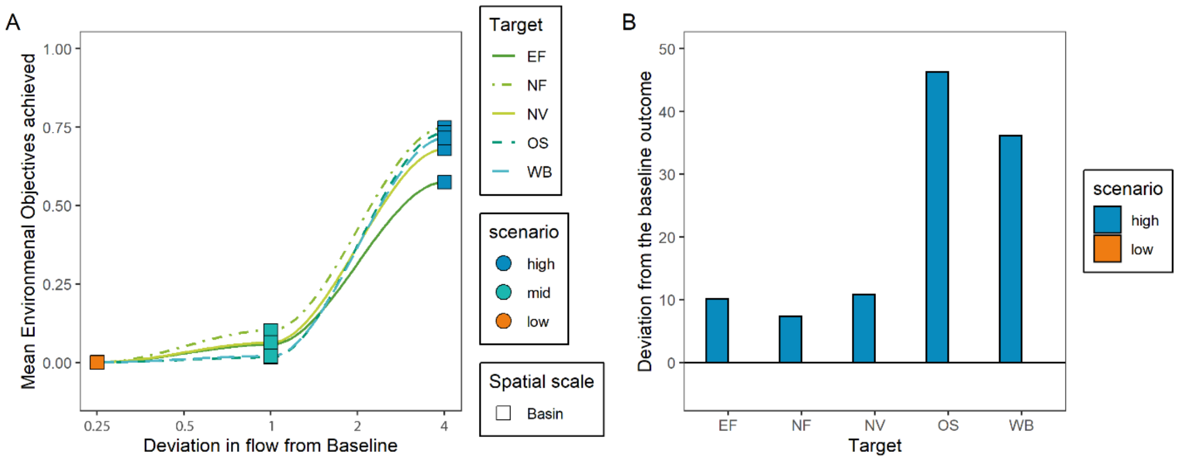

Figure 15. A) Comparison of all target-scale outcomes (priority ecosystem functions (EF); native fish (NF); native vegetation (NV); other species (OS); and waterbirds (WB); with scenarios presented on a quantitative axis defined as deviations in flow from the baseline (0.25x to 4x). Values for the points are the mean proportion of indicators that pass each environmental objective using the arithmetic mean to aggregate gauges spatially to the planning unit scale. A loess fit (local polynomial regression) is used to visualise which targets are most influenced by deviations in rainfall from the baseline at the Basin scale (one fit for all planning units). B) Comparison of all target-scale outcomes (priority ecosystem functions (EF); native fish (NF); native vegetation (NV); other species (OS); and waterbirds (WB). Values are the relative change from the baseline 'mid' scenario for the high or low scenario outcomes. This is a multiplicative scale, so a value of 10 means 10 times greater. Bars for the low scenario do not appear because they are zero or nearly zero, and so occur at the origin. Outcome values are the mean proportion of indicators that pass each environmental objective using the arithmetic mean to aggregate gauges to the planning unit scale and the arithmetic mean to aggregate planning units to the basin scale.

6.7 Decision tree for aggregation

A challenge for water policy and management is to integrate and aggregate information across scales (Saintilan and Overton 2010). A decision tree aims to help with choosing aggregation methods across multiple dimensions, that is space, time and quadruple-bottom-line element resolution (for example indicators, objectives). The complexity of the problem (highlighted in Figure 16) is in the number of steps required to aggregate in these 3 dimensions, with a decision required at each step. As each aggregation step can influence outcomes and the appropriate aggregation methods at subsequent steps, the decision tree aims to help avoid unintended interactions based on ecology of the species of interest (Lester 2019).

To simplify the choice of aggregation method at each step, we have created a decision tree based on the influence of components on outcomes and whether they can compensate for one another (Figure 17). Where full compensation allows a good outcome (pass or high values) in one component to offset poor outcomes (fail or low values) in all other components, only the best outcome determines the aggregated value. Partial compensation allows good outcomes in one or multiple components to offset poor outcomes in other elements, but not completely – all outcomes contribute to the aggregated value. No compensation means that a poor outcome in one component cannot be offset by good outcomes in other components – only the worst outcome determines the aggregated value. In context-specific instances, this may require custom mathematical functions, but in most cases we expect that one of the 4 default aggregation methods that we have incorporated into the Aggregator will be most appropriate.

The default set of aggregation methods includes:

1. Arithmetic mean

- Proportion passed for pass/fail results
- Average outcomes for quantitative results

Using this approach, times, places or objectives are equally influential or can be weighted. Under this method, times, places or objectives can partially compensate for one another.

2. Geometric mean
    - Overall pass or fail for pass/fail results without compensation
        - passes if all pass, otherwise fails
    - Average outcomes for quantitative results for multiplicative processes

    Using this method, times, places or objectives are equally influential or can be weighted. Under this method, times, places or objectives cannot compensate for one another

3. Limiting factor
    - Single pass/fail for pass/fail results without compensation
        - if any element fails, the aggregation result is fail
    - Minimum outcome for quantitative results

    Using this method, each single time, place or objective is essential. Under this method, times, places or objectives cannot compensate for one another and the method effectively measures the ‘weakest link’.

4. Compensating factor
    - Pass/fail results with full compensation
        - if one element passes the aggregation result is pass
    - Maximum outcome for quantitative results

    Using this method, no single time, place or objective is essential, but any one can determine the outcome. Here, times, places or objectives compensate for one another.

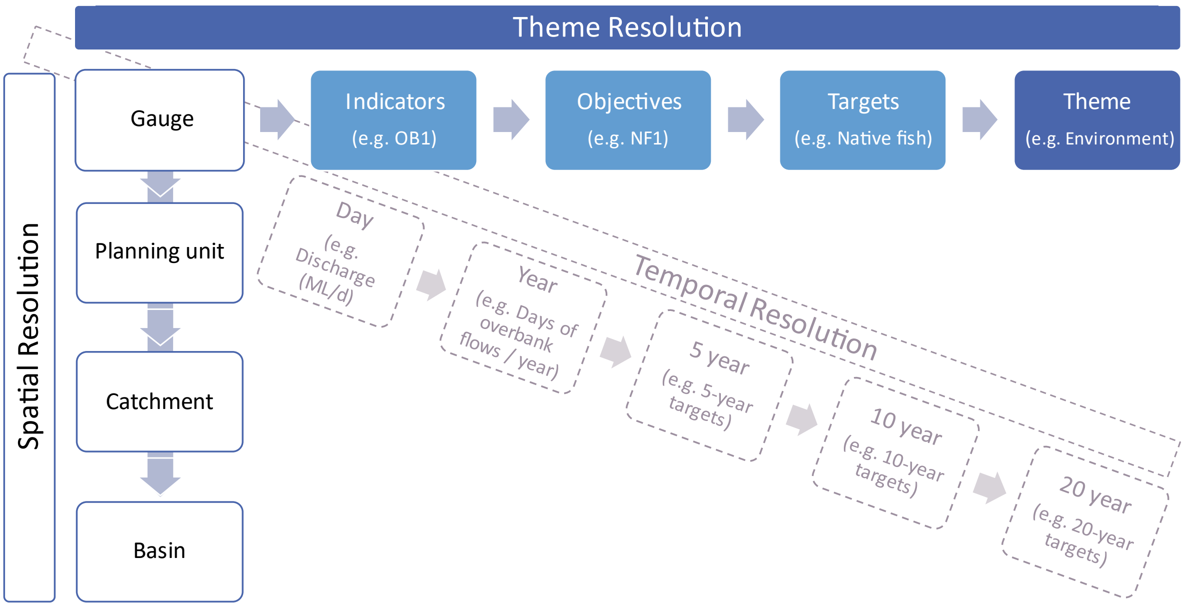

Figure 16. Selection of aggregation methods requires choices to be made across multiple dimensions: time, space and quadruple-bottom-line element (for example indicators, objectives) resolution. In this demonstration, the EWR tool largely handles the first steps of temporal aggregation, with outputs of quantitative indicators (for example days of over bank flows) returned yearly, and output of EWR pass/fail results returned for the time-period of the data. Future development may allow the temporal resolution to be selected which maps on to the 5-year, 10-year and 20-year targets outlined in the long-term watering plans, for example. To present results at an appropriate element resolution, the gauge-specific results need to be aggregated stepwise through each scale of element resolution. For example, gauge-specific EWR results are aggregated to assess results at the environmental objectives scale; these are then aggregated to assess results for individual target groups, and so forth to reach the quadruple-bottom-line element scale. Each of these aggregation steps requires a choice of method. Likewise, a choice of method is required to obtain an appropriate spatial resolution. This can be achieved stepwise if different methods are required for each step. However, if the method is consistent, this can occur in one step. The Aggregator can apply weightings where there are different numbers of gauges per planning unit, for example. Further, intermediate scales can be skipped entirely if desired, aggregating over all gauges in a catchment or the Basin, for example. The decision tree aims to assist with selection of aggregation methods across multiple dimensions.

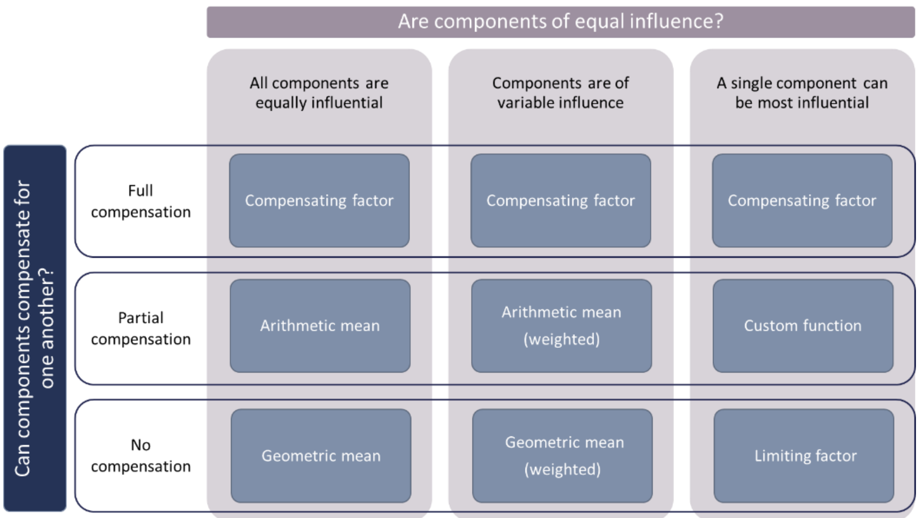

Figure 17. At each step shown in Figure 16, evaluation is required to determine the aggregation method, based on whether all components have equal importance and whether they can fully, partially or by no means compensate for one another. This matrix may help to identify which method is appropriate given the answers to those questions.

6.7.1 Demonstration

For the demonstration we have focused on aggregation to broaden the spatial and quadruple-bottom-line element resolution of outputs, as the initial steps of temporal aggregation are largely handled by the EWR tool. Here, we provide 3 examples, where the EWR tool outputs are aggregated to the indicator (Figure 18), the objective (Figure 19) and the target element (Figure 20) resolutions. While we have already demonstrated these increasing levels of aggregation in demonstrating the Aggregator and Comparer modules, in that case we used arithmetic means for simplicity. Here, our focus is on comparing the different sorts of aggregation and choosing between them. For examples we apply spatial aggregation of individual gauges to reach the planning unit scale using different spatial aggregation methods. This illustrates both the reasoning behind the choices and how the outcomes are influenced by the choice of spatial aggregation method. At all aggregation steps, the choice of method is important.

The simplest spatial aggregation is of a single EWR indicator. In this case, the only aggregation that occurs is spatial, which can be demonstrated for each of the four default methods: arithmetic mean, geometric mean, limiting factor and compensating factor (Figure 18). Though we would rarely want to spatially aggregate a single EWR indicator, this provides a clear illustration of how the aggregation functions differ and the information they capture. Here, the indicator passes at all locations in the New South Wales planning unit, but at only one of the two locations in the Victorian unit. This single failure in Victoria causes Victoria to fail (that is, to be zero) with limiting factor and geometric mean aggregation. For compensating factor aggregation, the one passing gauge compensates for the failure, yielding a pass for Victoria. Finally, the arithmetic mean quantifies the proportion of gauges that have passing values, yielding 1 for New South Wales and 0.5 for Victoria. As described above, each of these choices may be appropriate, depending on the outcome of interest.

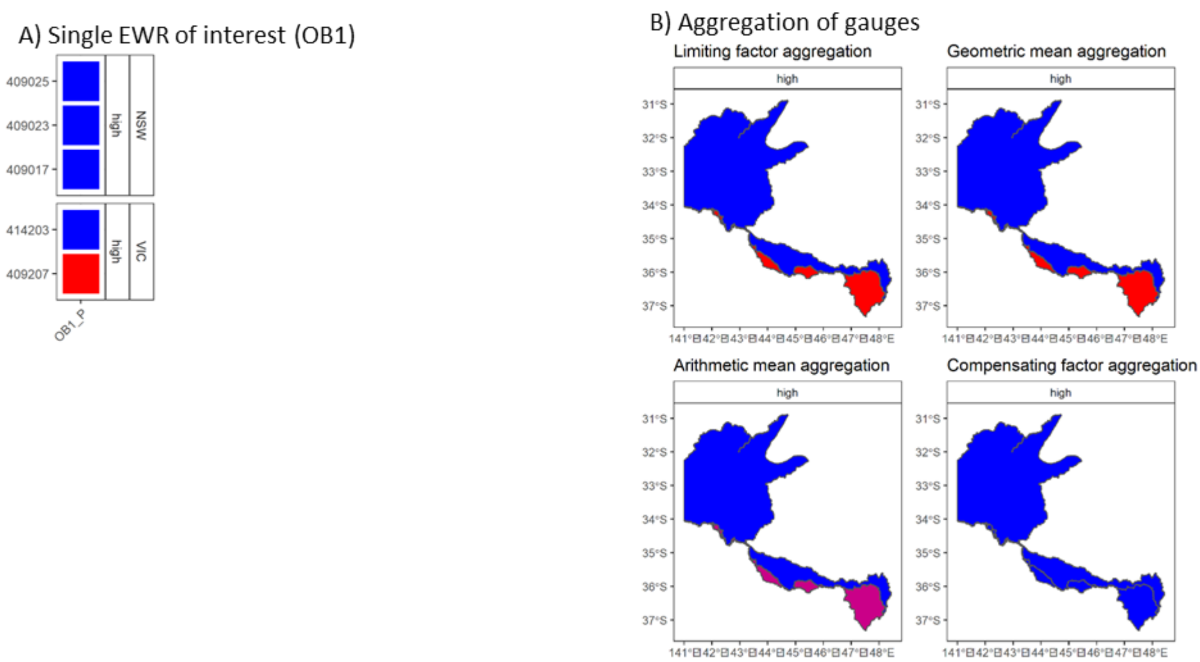

Figure 18. (A) One indicator of interest (OB1_P; overbank flows 1 with preferred timing) (B) aggregated to the planning unit spatial scale using the 4 default aggregation methods: 1) Limiting factor (one fail equates to all failing)- blue: pass, red: fail; 2) Geometric mean: blue: pass, red: fail; 3) Arithmetic mean, proportion passing, blue: 1, red: 0, purple: 0.5; and 4) Compensating factor (one pass equates to all passing) blue: pass, red: fail. Data used are from the high scenario only. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Gauges are listed on the left hand side of A. Refer to text for details.

Aggregating to the second level, objectives, requires aggregating the multiple EWR indicators that contribute to each objective. We can do that at each gauge, again with the 4 demonstration aggregation functions. The arithmetic mean yields the proportion of EWRs that are met for that objective at each gauge, while the geometric mean is similar but does not allow for compensation, so if any EWR fails the objective fails at that gauge. Similarly, limiting factor aggregation yields a failure if any EWR fails, perhaps representing the situations where all EWRs are essential for survival. In contrast, compensating factor aggregation means the objective passes at the gauge (or the gauge receives the maximum EWR value), representing a situation where successful EWRs can make up for failures in others.

These gauge-scale objective outcomes can then be scaled spatially according to any of the aggregation functions (they do not have to match the functions used for the EWR to objective aggregation).

We demonstrate the case where the functions do match (Figure 19). For limiting factor and geometric mean aggregations, no gauge passes for all EWRs, so the objective fails at every gauge and across both planning units. For compensating factor, each gauge has at least one EWR that passes, so the objective passes at each gauge and in each planning unit. The arithmetic mean first calculates the proportion of EWRs that pass at each gauge, each of which is somewhere between 0 and 1. Then, the arithmetic mean of these proportions is taken across each planning unit, yielding intermediate outcomes that are similar, but not the same to those calculated for each gauge.

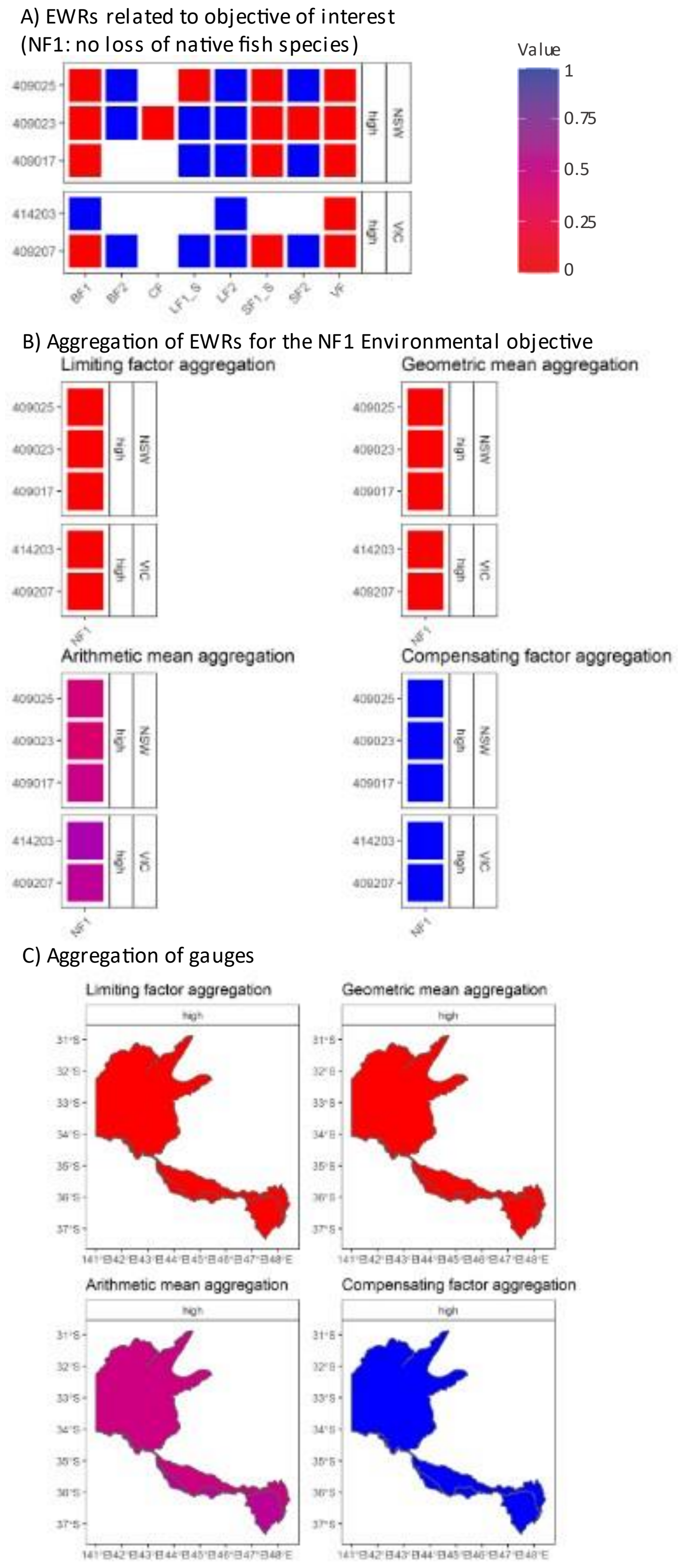

Figure 19. (A) Indicators (VF, BF1, BF2, SF1_S, SF2, LF1_S, LF2, for each of the demonstration gauges (listed by number), and CF) (B) aggregated to one objective of interest (NF1) and (C) then to the planning unit spatial scale using the 4 default aggregation methods: 1) Limiting factor (one fail equates to all failing); 2) Geometric mean; 3) Arithmetic mean, and 4) Compensating factor (one pass equates to all passing). The same aggregation method is used for both steps; compensating factor aggregation environmental objective outcomes for each gauge in (B) are aggregated to the planning unit scale in (C) using compensation factor aggregation. Data used are from the high scenario only. Irrelevant EWRs are not shown (represented as missing tiles) at each gauge under each scenario. Legend applies to all plots, as each displays values from 0

(red) to 1 (blue), though their meaning differs as indicated by the type of aggregation. Other than Arithmetic mean aggregation, all values are either 0 or 1, while the arithmetic mean gives the proportion of EWRs achieved, and so appears as shades of purple.

The next step in aggregating along the quadruple-bottom-line element dimension is from objectives to target (for example native fish, NF, has several defined objectives, often relating to different parts of the life cycle for ecological targets). Now, we must choose an aggregation method from EWR to objectives and from objectives to targets. We do this element aggregating at the gauge before spatially or temporally aggregating in general, though it is possible to interleave the steps if necessary and appropriate. Figure 20 illustrates this process. First, we take the set of EWR indicators contributing to all of the environmental objectives related to native fish and aggregate those indicators for each objective (central panels). The leftmost column (NF1) in these panels is the same set of aggregations illustrated above for the EWR to single objective aggregation (Figure 19). We then aggregate all objectives to the target level (NF) at each gauge (across the rows of the panels in the centre), and then spatially-aggregate those gauge-scale target outcomes to obtain overall native fish outcomes in each planning unit.

There is no requirement to use the same aggregation method at each step, and indeed it may be appropriate to use different methods depending on the values being aggregated, particularly along the different aggregation axes of Figure 16. Perhaps an objective or target-level aggregation would need limiting factor aggregation at a gauge, because all EWRs must be met for a population to persist at that location, while population persistence at the planning unit scale would use compensating factor aggregation of those gauge-scale values because persistence in the planning unit only requires persisting in at least one gauge location.

In Figure 20, we calculate the EWR to objective aggregations using all 4 default methods. NF10 is the only objective in which all constituent EWRs pass for one gauge (409017) and fail for another (409023). Thus, NF10 at gauge 409017 is the only passing objective when aggregation is limiting or geometric mean. Conversely, NF10 at gauge 409023 is the only failing objective when aggregation is compensating (using the arithmetic mean or compensating factor aggregation). All other gauges have at least one passing EWR for each objective. With arithmetic mean aggregation, the result for each objective at each gauge is the proportion of passing EWRs.

To demonstrate different aggregation methods at different stages, we then take the objective aggregations using just the arithmetic mean of the EWRs at each gauge (the lower left of panel B in Figure 20), and aggregate to the target level using all 4 default methods. As all EWRs failed for NF10 for gauge 409023, this is carried through to failure at the target level for that gauge. We spatially aggregate the gauges to planning units using each of the 4 methods that match the method used for the target level aggregation. Since there is a strict failure in one gauge in New South Wales, New South Wales yields a failure for the spatial aggregation with limiting factor and geometric mean aggregation. However, these methods yield slightly different outcomes in Victoria because the minimum value is low but not zero. There, the limiting factor simply takes the minimum, while the geometric mean finds the multiplicative average, and so accounts for values other than the minimum. Compensating aggregation takes the maximum of the target value across gauges in each planning unit, so it is high in both Victoria and New South Wales. The arithmetic mean is the mean of the target value across the gauges in each planning unit, which are in turn the arithmetic mean across objectives of the proportion of EWRs met.

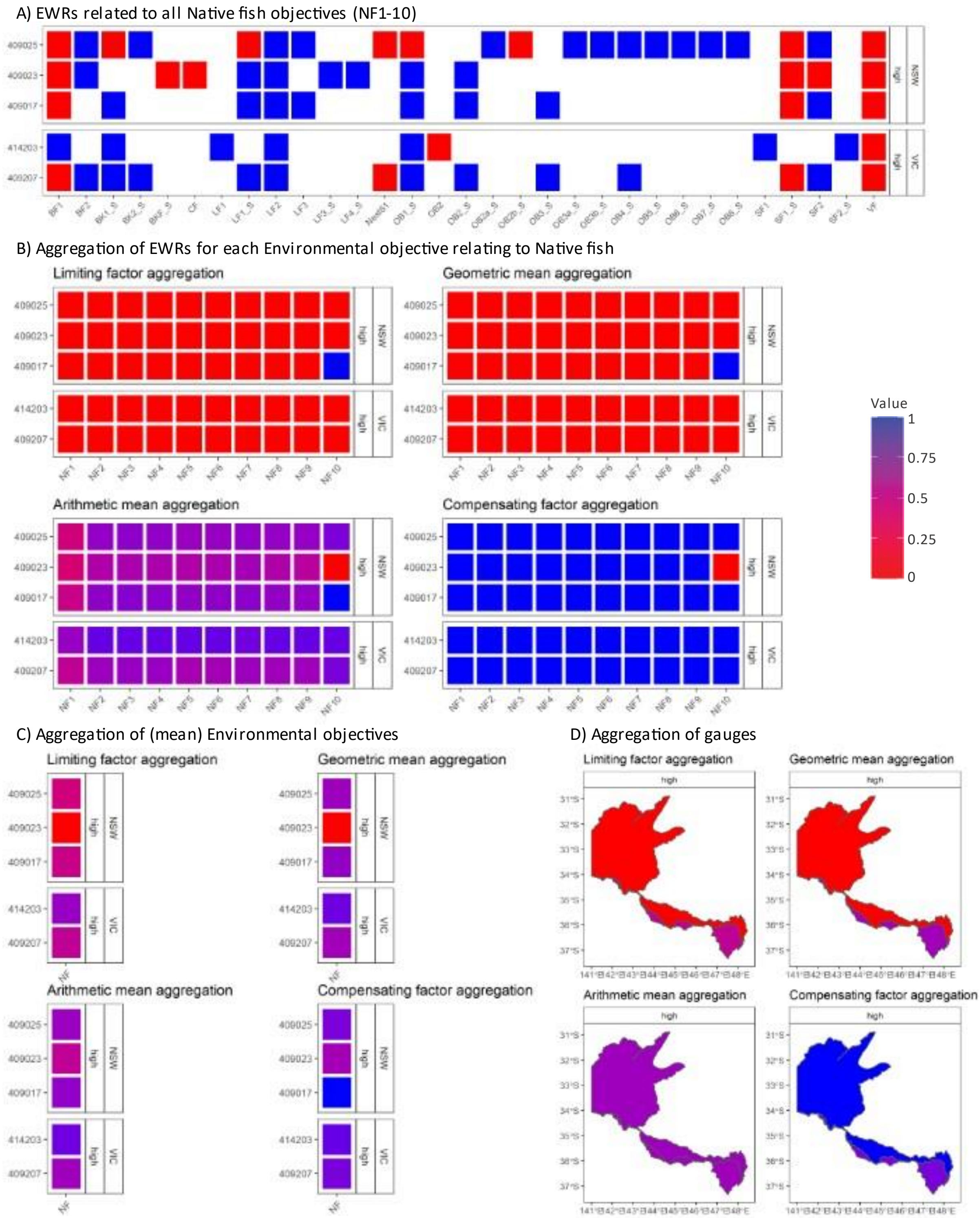

Figure 20. (A) Indicators (VF, BF1, BF2, …., etc.) aggregated to (B) objectives (NF1-10) relating to the target of interest (native fish) using the 4 default aggregation methods: 1) Limiting factor (one fail equates to all failing); 2) Geometric mean; 3) Arithmetic mean; and 4) Compensating factor (one pass equates to all passing). (C) The arithmetic mean for each environmental objective is then aggregated to target scale resolution. (D) Target scale outcomes (from each aggregation method) for each gauge are aggregated to the planning unit spatial scale using the same method: compensating factor aggregation targets for each gauge in (C) are aggregated to the planning unit scale in (D) using compensation factor aggregation. Data used are from the high scenario only. Irrelevant EWRs are not shown (represented as missing tiles) at each gauge under each scenario. The shape shown is the planning units used for the demonstration, as highlighted in grey in Figure 3. Gauges are listed on the left hand side of A, B and C. Refer to text for details. Legend applies to all plots, as each displays

values from 0 (red) to 1 (blue), though their meaning differs as indicated by the type of aggregation. Because the aggregations in C and D are based on arithmetic mean scaling to the objectives, which are not strictly 0 or 1, the limiting, compensating, and geometric mean aggregations now yield values other than 0-1.

7 CAUSAL NETWORK DEVELOPMENT

The aim of this component is to identify and characterise the causal links from climate through to outcomes. Capturing a network of those causal links turns the toolkit from a 'black box' to a transparent assessment of specific processes that link climate to outcomes for valued assets in the Basin. This network provides a visual and conceptual representation of the relationships modelled in the ModelArch component. The network is, therefore, the rationale that underpins the toolkit.

For example, during a dry period overbank flows are likely to be less common. This may make floods of a given size less frequent, which may result in less extensive floodplain inundation. Floodplain inundation may be required to cue recruitment in some vegetation species (for example river red gum), so recruitment may decline for those species. Rationales such as this will be captured within the causal network. Moreover, these links can help define mathematical aggregation functions – the amount of floodplain inundation is likely best represented by a mean, while requirement for reproduction may suggest limiting factors. Thus, for the toolkit, capturing causal links involves characterising how water availability or elements of the hydrograph lead to the state of defined indicators, and then how the states of those indicators lead to specific outcomes. Where available, the rationale will reflect that within the long-term watering plan used to develop the indicators included in the EWR tool.

Finally, the causal network will also need to characterise how local-scale outcomes can lead to larger-scale outcomes or can be synthesised with other groups of outcomes to provide holistic assessments.

The causal network concept aims to create a network representation of cause-and-effect relationships to illustrate likely relationships between activities (here, adaptation options and climate change) and environmental impacts (or responses) (Peeters et al. 2022). An example is illustrated in Figure 21.

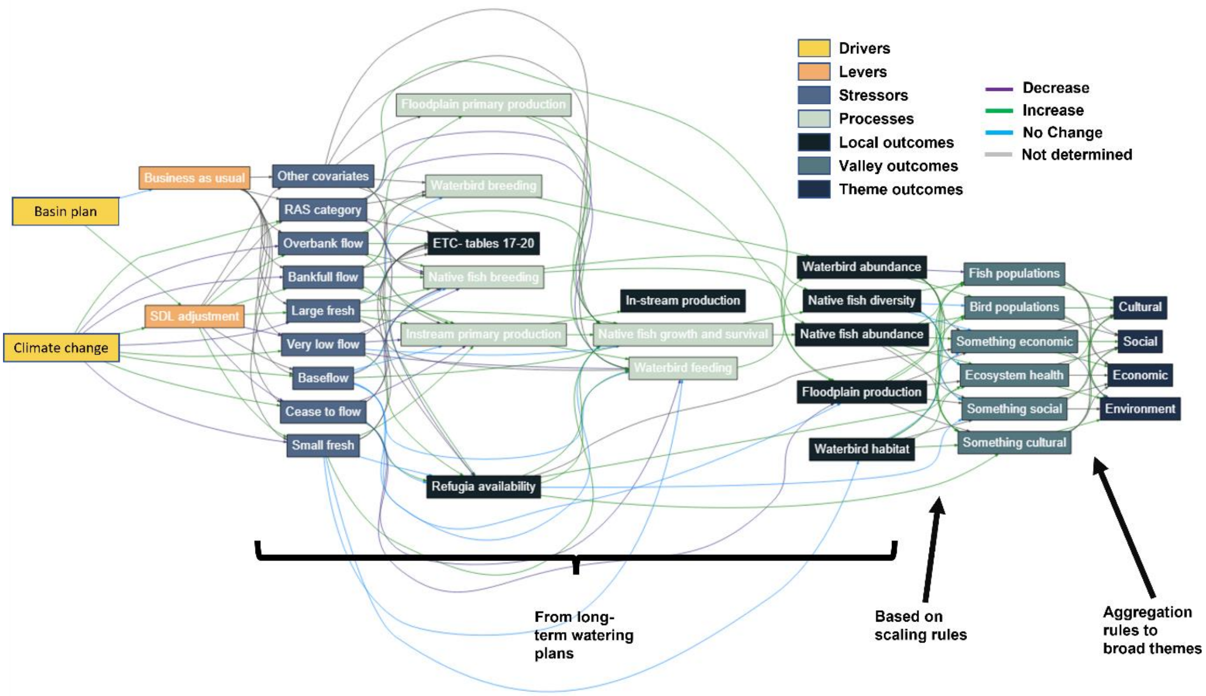

Figure 21. Example causal network capturing links from drivers (yellow) and policy responses (orange) to hydrology (indigo) to ecological processes (sage) to local objectives for the outcomes of those processes (black) to valley-scale aggregations (blue-green) and finally element-scale outcomes. Link colours indicate directionality but are currently only illustrative. This example is built from a small subset of the NSW long-term watering plans.

The network is based on a graphical representation of pathways represented by nodes connected by links (that is, in a conceptual model). The nodes represent the drivers, levers, stressors, processes and outcomes at multiple levels in the system. The links represent the causal relationships among the nodes. Each node is clearly defined in terms of the current knowledge base, relevant knowledge gaps and key assumptions, to provide transparency.

7.1.1 Demonstration

To demonstrate the causal network approach, we have used the long-term watering plans (LTWPs) developed by each state. These are already based on causal links between watering events and environmental outcomes. Early discussions have commenced with some Traditional Owners to explore opportunities to include cultural outcomes. Should this be appropriate, those cultural relationships could be included (provided Traditional Owners agree and see value in that approach).

The relationships contained within the LTWPs have been obtained using output from the MDBA's EWR scraper tool and the Deakin-developed R script for the extraction of those links. These have then been quality checked (see the *Objective Translator* section above). The relationships map environmental water requirements to environmental objectives, environmental objectives to target species, and target species to the 5-year, 10-year and 20-year planning targets. These are specific to catchments, planning units and/or gauges.

For example, Figure 22 illustrates how target species and ecosystem functions vary among planning units in the Murray–Lower Darling Catchment. For each gauge, a multitude of links exist; Figure 23 illustrates all the links for one gauge (409025) in the Murray–Lower Darling catchment, showing the high level of complexity that is inherent in these links. The complete network can be condensed down to visualise only the links pertaining to a particular environmental objective or set of objectives. For example, Figure 24 condenses the network in Figure 23 to only the links important for increasing waterbird breeding. A condensed network can highlight the species of interest and the overall targets for each environmental objective, which may help in making aggregation decisions as well as providing visual and conceptual representation of the relationships modelled in the ModelArch component.

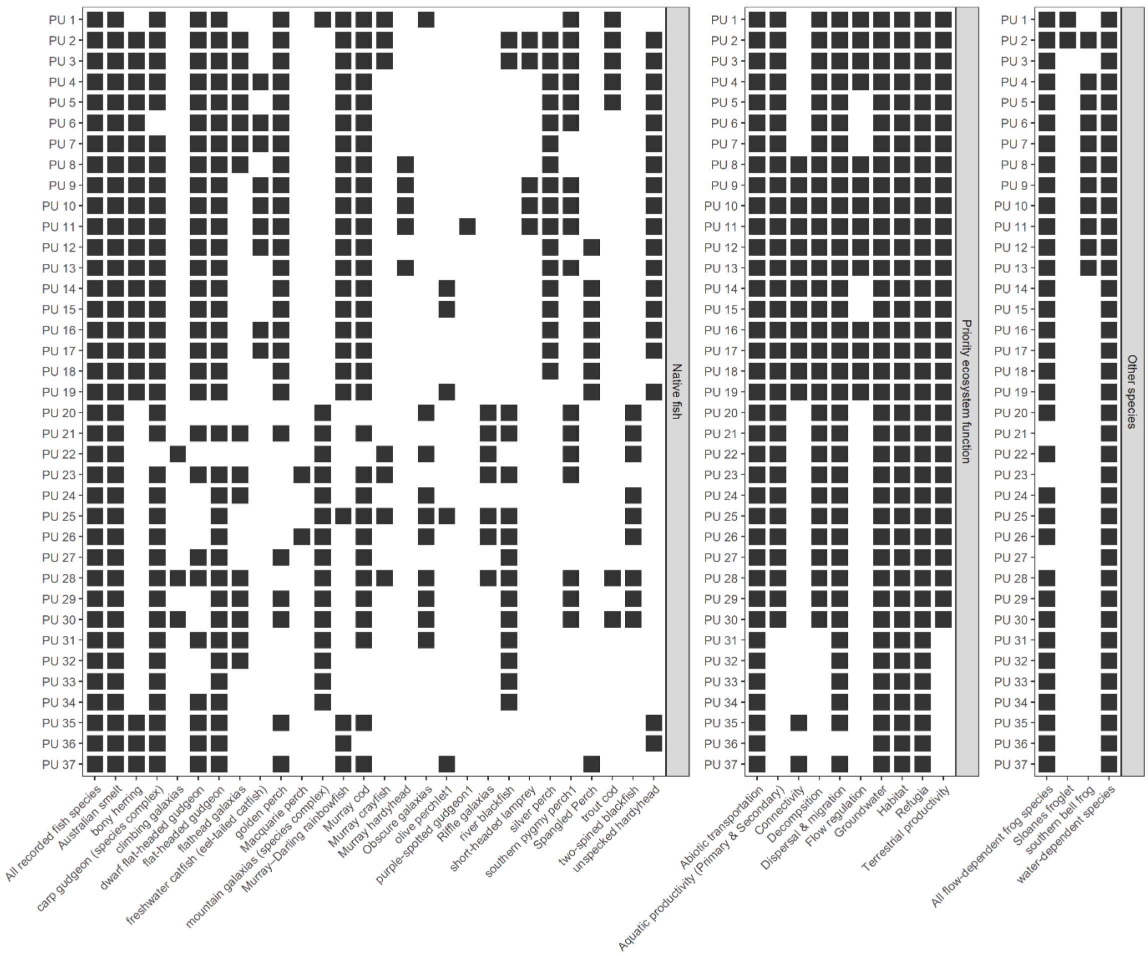

Figure 22. Target species and ecosystem functions vary among planning units in the Murray–Lower Darling catchment.

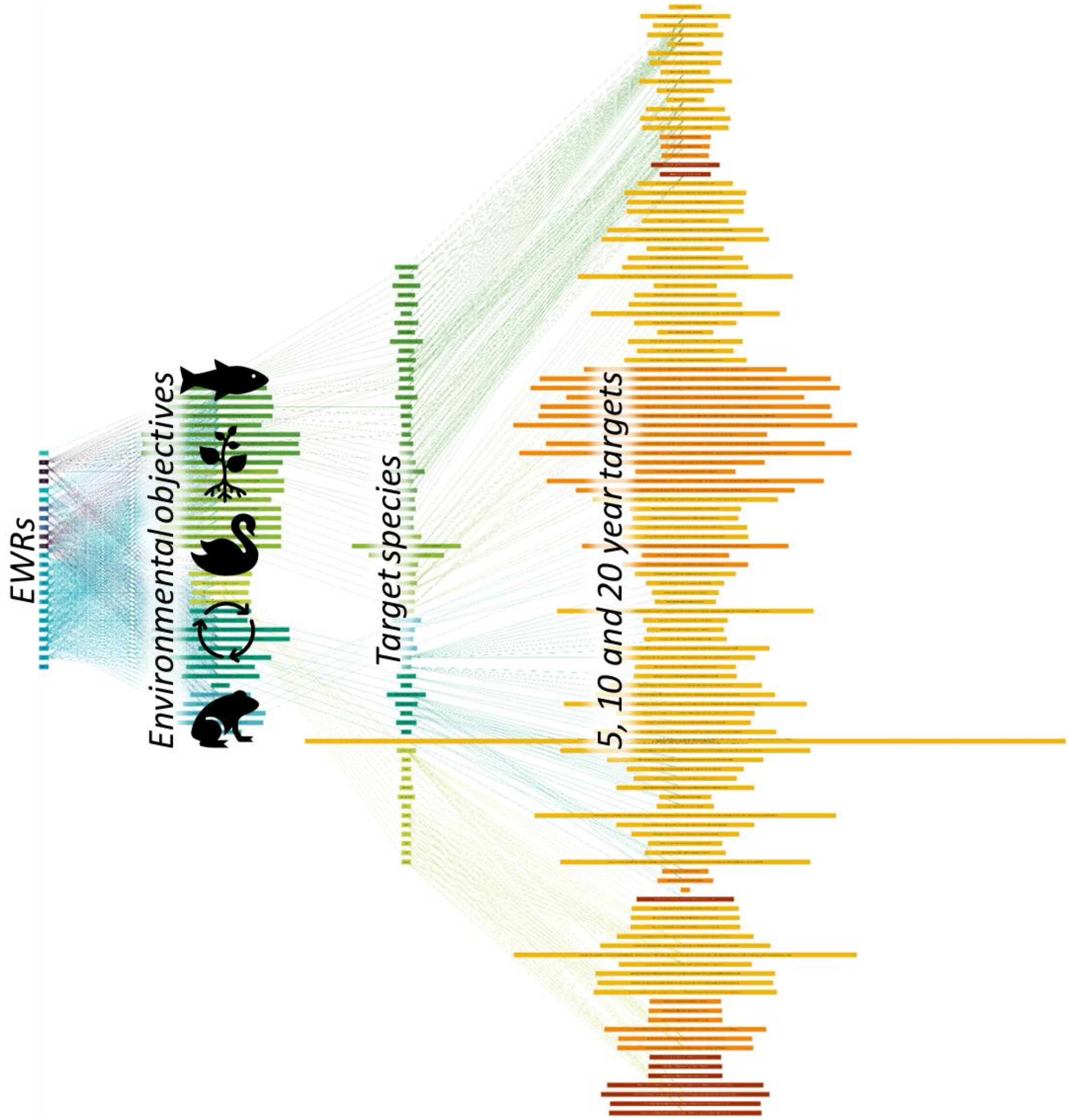

Figure 23. Complete causal network illustrating all the links in the LTWPs (environmental water requirements → environmental objectives → target species → 5-year, 10-year and 20-year planning targets) for one gauge in the Murray–Lower Darling catchment.

Figure 24. Condensed causal network illustrating links in the LTWPs important for increasing breeding in colonial and non-colonial waterbirds. Teal colour on EWR codes (left column; see Appendix A for meaning) indicate those that influence the waterbird breeding objectives illustrated in this example (WB3 & 4, green in column 2). Target species (third column) relevant to those objectives are shown in green. Colour in rightmost column indicates 5-year targets (yellow), 10-year (orange) and 20-year (red).

8 FUTURE DEVELOPMENT OF THE TOOLKIT

This report marks the completion of the toolkit demonstration, which is a minimum viable product. Additional work to continue to develop elements of the toolkit (as outlined above) will continue into Year 2 of MD–WERP. If appropriate and time permits, this work may include the following:

ModelArch development

- Plan for next steps: identify priorities for improved functionality, needed changes to the workflow or methods, module inclusion, or output types
- Update criteria for metadata and prioritisation of additional tools
- Integrate measures of uncertainty into the ModelArch
- Develop the Scenario Controller user-editable interface:
    - Incorporate selection of spatial unit settings, including the ability to select from existing options (the Basin boundary, catchment boundaries and planning units) and the ability to incorporate other spatial units (for example ABARES statistical areas).
    - Incorporate selection of aggregation settings, including the ability to select from existing options (for example arithmetic mean, compensating factor) and the ability to apply customised functions, with default settings for objectives.
    - Incorporate selection of comparison settings, that is the ability to select from existing options (for example maps, bar charts) with default settings for objectives.
    - Incorporate parallelisation capability.
- Continue to develop documentation of workflow and toolkit: how to use the architecture and reasoning for decisions made, with worked examples for common objectives for default options.
- If module development is sufficient, include additional elements from the quadruple bottom line.
- Develop a decision tree for comparisons: much like the decision tree for aggregation, incorporate choices for comparison made across multiple dimensions: time, space and quadruple-bottom-line element.

- Incorporate ability to observe EWR pass/fail results through time – currently quantitative indicators can be viewed though time, but pass/fail results are given for the whole block of time defined by the data.

Causal network development

- Complete quality check of links from environmental watering requirements to environmental objectives
- Develop user interface to automate collapse of network to desired sets of links

Module development

- Capture EWRs and related objectives from states other than New South Wales
- Develop indicators for social and economic values
- Explore options for appropriate inclusion of cultural values with Traditional Owners
- Develop inference among elements of the quadruple bottom line for indicators which link multiple elements: for example, engage with Traditional Owners to better understand how they value environmental indicators or consider the economic significance of environmental indicators, ideally in collaboration with MD–WERP Theme 4.

9 GLOSSARY

Bankfull flow	River flows at maximum channel capacity with little overflow to adjacent floodplains. Engages riparian zone, anabranches and flood runners and wetlands located within the meander train. Inundates in-channel habitats including all benches, snags and backwaters.
Baseflow	Reliable background flow levels within a river channel that are generally maintained by seepage from groundwater storage, but also by surface inflows. Typically inundates pools and riffle areas.
Driver-indicator-response model (or indicator module)	Module that checks hydrographs against a database of quadruple-bottom-line element-based hydrological indicators (flow requirements); for example the EWR tool for environmental outcomes.
Environmental water requirement (EWR)	Water required to support the completion of all components of a lifecycle of an organism or group of organisms (taxonomic or spatial), consistent with the objective/target, measured at the most appropriate gauge. Includes all water in the system including natural inflows, held environmental water and planned environmental water.
Hydrograph	Timeseries of flows, typically measured at a gaging station
Hydrological indicators (or flow requirements)	Flows that are required because they have causal relationships with quadruple-bottom-line element-based objectives; for example the EWRs for environmental outcomes.
Hydrology	The occurrence, distribution and movement of water.
Long-term watering plan(LTWP)	Plans required of Basin states by the Basin Plan. Long-term watering plans give effect to the Basin-wide Environmental Watering Strategy relevant for each river system and will guide the management of water over the longer term. These plans will identify the environmental assets that are dependent on water for their persistence, and match that need to the water available to be managed for or delivered to them. The plans will set objectives, targets and watering requirements for key plants, waterbirds, fish and ecosystem functions. The New South Wales Department of Planning and Environment is responsible for the development of nine plans for river catchments across New South Wales, with objectives for 5-year, 10-year and 20-year timeframes.
Objective	The defined goal for a state, condition or characteristic of an element-based asset or function, for example environmental objectives for the environmental element of the quadruple bottom line.
Overbank flow	Flows that spill over the riverbank or extend to floodplain surface.
Planning unit (PU)	A geographical division of a water resource plan area based on water requirements (in catchment areas in which water is actively managed), or a sub-catchment boundary (in all other areas).
Quadruple bottom line	Social, cultural, economic and environmental values
Water resource plan	Plans established by Basin state governments to meet requirements of the Basin Plan and address local requirements of water resource management.

Definitions here are attributed to the New South Wales Department of Planning and Environment Long Term Watering Plan.

10 REFERENCES

Holt G, Dwyer GK, Robertson D, Bailey J, Job M, Palmer L, and Lester RE (2022) Architecture design for the Climate Adaptation toolkit workflow. *MD-WERP Deliverable: T1.TK1*.

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

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*, 71:238–254.

Murray–Darling Basin Authority (2019) *Basin-wide environmental watering strategy*, Canberra.

Peeters LJ, Holland KL, Huddlestone-Holmes C and Boulton AJ (2022) 'A spatial causal network approach for multi-stressor risk analysis and mapping for environmental impact assessments', *Science of The Total Environment*, 802:149845.

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

Saintilan N and Overton I (2010) 'Conclusions: Lessons from ecosystem response modelling and future directions for the Murray–Darling Basin', *Ecosystem response modelling in the Murray–Darling Basin*, 411-419.

11 APPENDIX A – EWR CODES

The Environmental water requirements (EWR) codes used in this demonstration capture different aspects of the hydrograph, and are linked to environmental objectives in long term watering plans. Each code has a prefix that describes the type of flow it captures: see Table 2. There may be several numbered subsets of flow within each type, describing, for example, different subcategory of Small Freshes that reflect different outcomes for the environmental responses. The definitions for each category depend primarily on flow level, timing, duration, frequency, and inter-event periods. Requirements differ between gauges, and the details of each are available in the long-term watering plans.

Table 2. Meaning of EWR code prefixes. EWR codes then also contain numbers, which define different sorts of flows within that flow category. The details of the flows required to meet each category are defined for each gauge in the long-term watering plans.

Code prefix	***Meaning***
CF	Cease-to-flow
VF	Very-low flow
BF	Baseflow
SF	Small fresh
LF	Large fresh
BK	Bankfull
OB	Overbank
NestS	Nesting support