Innovation Sweep

Technological Insights for the Murray–Darling Basin Authority

Presented by:
The University of Adelaide

About this report

This work was undertaken as a part of the Murray–Darling Basin Water and Environment Research Program (MD–WERP). 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 Holder and the Murray–Darling Basin Authority.

The Murray–Darling Basin Authority commissioned the University of Adelaide to prepare this report. The report encompasses the results of an innovation scan and stakeholder workshop to identify opportunities and emerging technologies that could enhance monitoring and research capability within the Murray–Darling Basin.

Contributing authors

Isabelle Onley (University of Adelaide)

Ehsan Abbasnejad (University of Adelaide)

Lingqiao Liu (University of Adelaide)

Qi Wu (University of Adelaide)

Jamie Sherrah (University of Adelaide)

Tyler Dornan (University of Adelaide)

Scott Bainbridge (Australian Institute of Marine Science)

Melanie Olsen (Australian Institute of Marine Science)

Steve Delean (University of Adelaide)

Ramesh Segaran (University of Adelaide)

Thomas Prowse (University of Adelaide)

Micha Jackson (University of Adelaide)

Paul Dalby (University of Adelaide)

David Hamilton (Griffith University)

Matthew Hipsey (University of Western Australia)

Ben Sparrow (TERN)

Qifeng Ye (SARDI)

Chris Bice (SARDI)

Justin Brookes (University of Adelaide)

Acknowledgements

We acknowledge and pay respect to the Traditional Owners of this land, the Nations of the Murray–Darling Basin and their Elders past and present who have a deep cultural, social, environmental, spiritual and economic connection to their lands and waters.

Contents

Tables

Figures

Executive Summary

The Murray–Darling Basin Authority (MDBA) Innovation Sweep is a report intended to provide a summary of emerging conservation technologies and their potential applications for the ongoing management of the Murray–Darling Basin (MDB). These technologies have been selected based on their advanced development, innovation, and increasing uptake in the fields of conservation biology and natural resource management. This report is targeted towards on-ground managers of natural areas within the MDB seeking to further develop their monitoring toolkit by incorporating cutting-edge technology and research.

The MDB is a large system of interconnecting rivers and lakes encompassing over one million square kilometres in south-eastern Australia, including the 2 main tributaries, the River Murray and the Darling (Baaka) River. The MDB has significant cultural, environmental and economic value, supporting tourism, agriculture, internationally significant wetlands and sites of spiritual heritage. The MDBA was established to manage and operate waterways across state jurisdictions, with the aim of achieving a healthy working Basin for the benefit of all Australians.

The MDB faces a number of challenges that can be exacerbated by competing requirements for water. These include drought, salinity, fish deaths, algal blooms, acidic soils, blackwater events and the increasing pressures of climate change. Such issues make the maintenance of good water quality and healthy ecosystems difficult, particularly when operating on large spatial scales. Therefore, this report aims to provide a synthesis of emerging technologies that may provide solutions to the challenges of environmental monitoring within the MDB at scale and guide proponents towards skilling a workforce in these applications via technology-based training.

To identify emerging technologies and gain insight into their current and potential applications, experts from various agencies and institutions were engaged to contribute working knowledge of innovations within their industry. The results of this Innovation Sweep are a curated selection of technologies that are of high relevance and potential for the MDBA. These technologies have been summarised in 3 main areas: sensors, data analysis, and models (Table 1). Opportunities for managing and sharing the large data sets generated by these technologies are also explored.

The technologies described in the report present exciting potential pathways towards monitoring techniques that are high sensitivity, high resolution, automated, and in many cases relatively low cost. These technologies can be applied at broad spatial and temporal scales within the MBD to increase the power of monitoring efforts and inform future management initiatives working towards improved water quality and environmental condition.

Table 1 *Table of key technologies, their applications for the MDBA and relevant considerations.*

Category	Technology	Brief Description	Examples of Usage	Scale of Application	Timeframe of Application	Maintenance Requirements	Skills Required	Cost Breakdown ($-$$$)
Sensors	eDNA	Isolates trace DNA from environmental samples to inform on organisms present in the environment	• Detecting presence of cryptic species in waterways • Monitoring fish spawning events • Detecting presence of harmful organisms (e.g. invasive species, algal blooms, parasites)	Local	1 day – several weeks. DNA analysis can be performed in the field or via commercial platforms	Regular sampling may be required	• Wet lab skills • Bioinformatics	$-$$
	Autonomous Platforms	Platforms and vessels able to navigate air, land or water to undertake pre-planned missions using provided mission plans and logic	• Autonomous surface vessels are able to navigate waterways and survey physical and biochemical parameters • Mobile and static platforms can be tasked with detecting and	Local and Regional – scale is only limited by number of units available	Fine-scale spatial data must first be collected to develop survey paths, and this may take several months. Survey timeframe is limited only by power source of systems.	Systems must be deployed and collected regularly and survey paths updated to incorporate obstacles or changes to areas of interest. Permits and	• Advanced coding and data processing skills	$$-$$$

			responding to particular events			licensing must be maintained for operation		
	Animal Tracking (passive integrated transponder (PIT) tagging)	Devices are triggered to record presence when the receiving device detects a PIT tag.	• Record organism movements and behaviours	Local	Limited only by power source to receiver and lifespan of tagged animal	Receiver devices require a power source – if batteries are used, these must be replaced	• Basic training in technology and data management • Animal welfare training	$-$$ (Low-cost options available)
	Animal Tracking (satellite telemetry)	Animal is marked with unique biologging tag equipped with a satellite geolocation sensor.	• Record organism movements and behaviours	Local or Regional	Limited by lifespan of biologging tag	Generates large amounts of data that needs to be "cleaned" to eliminate errors	• Basic to advanced training in technology and data management • Animal welfare training	$$
Data Analysis	Machine Learning-Based Analysis	AI technique that uses algorithms to learn the characteristics of input data to build models that can be used to predict content in new	• Automatically identify pest or threatened species in images • Classify biomass of aquatic vegetation in images • Use time-series	Local, Regional or Basin-wide	Data collection may take several months. Developing an algorithm can require 6-12 months, while training a model	Continual provisioning of annotated data to increase accuracy	• Advanced coding and machine learning skills • Advanced data processing and	$$-$$$

		data or images.	data to recognise patterns in particular events such as fish spawning		typically takes several days.		bioinformatics • Ecological knowledge for image annotation and curation	
	Visual Question Answering and Visual Language Navigation	AI techniques that can facilitate communication and commands between humans and machines. This allows machines to collect images or footage of interest and answer direct questions about the content.	• Collect images of birds and conduct bird counts • Estimate vegetation coverage • Identify features of interest in images • Species identification	Local, Regional or Basin-wide	Data collection may take several months. Developing an algorithm can require 6-12 months, while training a model typically takes several days.	VQA systems require continual data provisioning to improve accuracy. VLN costs and maintenance are typically associated with the procurement and running of drones.	• Advanced coding and machine learning skills • Advanced data processing and bioinformatics	$$$
	Data Efficient Learning for Vision and Audio Recognition	Machine learning techniques in which models are trained using varying data requirements to output information on the input image or sound	• Identify species based on images or audio recordings • Count objects or organisms in images or videos • Search multiple images for	Local, Regional or Basin-wide	Data collection may take several months. Developing an algorithm can require 6-12 months, while training a model typically takes	The majority of investment is required in collecting the data required to train the model, however continual data provisioning is	• Advanced coding and machine learning skills • Advanced data processing and bioinformatics	$$$

			organisms with the same identity		several days.	required to improve accuracy.		
	Semantic Change Detection in Images	Machine learning technique that compares a series of images of the same scene or object over time to identify changes of interest	• Count populations and monitor demographic changes over time • Delineate and measure wetland flooding extent, tree death, and vegetation growth rates	Local, Regional or Basin-wide	Data collection can take several months to years, depending on the timescale of interest. Developing an algorithm may require 6-12 months, while training a model typically takes several days.	Repeated collection of data/images is required over the designated timescale.	• Advanced coding and machine learning skills • Advanced data processing and bioinformatics	$$$
Models	Sequential Decision-Making and Reinforcement Learning	Artificial Intelligence that uses deep learning to repeatedly play out and optimise a scenario and its outcomes, developing a “best-practice” policy.	• Weather and rainfall prediction • Bushfire prediction • Autonomous driving and robotic navigation • Sports strategy and coaching	Local, Regional or Basin-wide	Data collection may take several months. Developing an algorithm can require 6-12 months, while training a model typically takes several days.	Continual provisioning of updated data to train the model	• Advanced coding and machine learning skills • Advanced data processing and bioinformatics	$$$

	Advanced Hydrological Models	Incorporates ecological requirements and responses into hydrological models	• Evaluate and interpret ecological monitoring using relationships between discharge and velocity statistics	Local, Regional	Data collection may take several months. Developing an algorithm can require 6-12 months, while training a model typically takes several days.	Continual provisioning of updated data to train the model	• Advanced coding and machine learning skills • Advanced data processing and bioinformatics	$$

Introduction

Water managers in Australia have responsibility for large river reaches, management of resources and ensuring ecological outcomes from allocation of environmental water. The Murray–Darling Basin (MDB), in south-eastern Australia, is a natural resource of immeasurable cultural, ecological and economic value. However, prolonged periods of drought combined with competing requirements for water and the additional pressures of climate change have resulted in a number of threats and challenges to the health of the MDB (Figure 1). Consequently, the Murray–Darling Basin Authority (MDBA) implemented an Environmental Watering Strategy designed to restore ecosystem health to the Basin. To determine this strategy's effectiveness and track progress towards restoration targets, ongoing monitoring of key indicator species and systems is required.

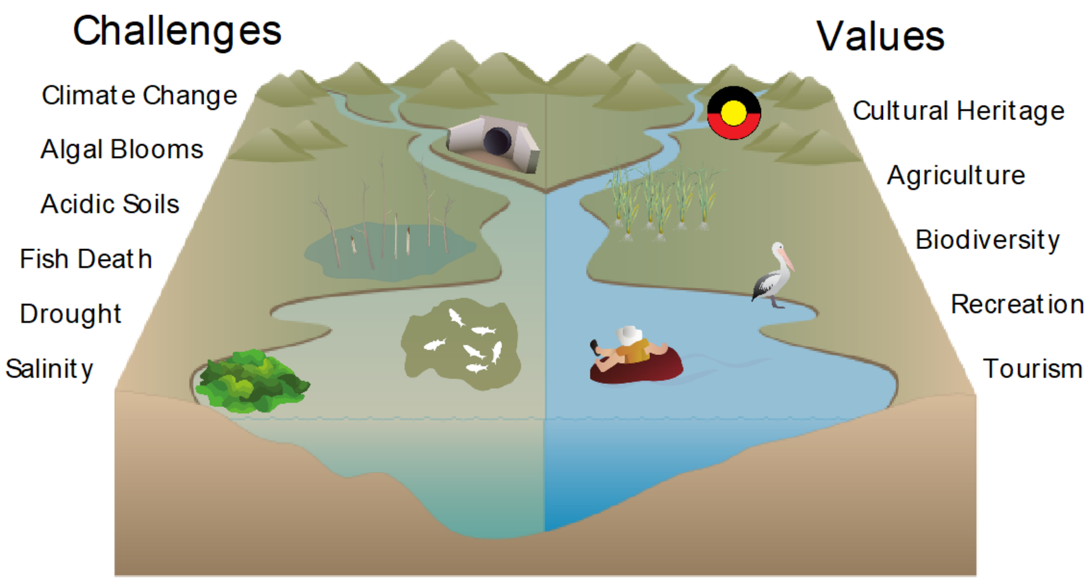


Figure 1 *Key challenges faced by the MDBA in the management of the MDB, and a summary of the MDB's key values.*

Assessing ecological outcomes at scale is challenging. However, technologies that can be applied in an ecological context for monitoring and assessment have experienced rapid development in recent years. The advent of tools such as remote sensing, artificial intelligence (AI) and modelling have transformed the way managers approach ecological issues. Monitoring can now be conducted at landscape scales with less effort, at a lower cost and with higher resolution than traditional methods. This has the potential to significantly improve monitoring approaches and outcomes for large and complex environmental systems such as the MDB.

Emerging technologies and opportunities for adoption

Technology to collect data is becoming cheaper and more ubiquitous, and emerging "techno-ecology" hardware is spawning the next generation of ecological data. These physical technologies (Table 1) include:

- Biologgers
- Drones and drone-mounted technologies for recording data
- Autonomous vehicles and data buoys
- Radio-frequency identification (RFID) tagging
- Low power networks
- Peer-to-peer data transfer networks (e.g., used to create smart environments to integrate recording arrays of recorders, such as camera traps)

Data from these technologies are recorded at high spatial and temporal resolutions, creating large volumes of data that cannot be processed manually. Multidisciplinary collaborations are therefore critical to developing ecoinformatics approaches to curate, store and analyse these data. Technology to process this information is becoming rapidly more powerful, more accurate in predictive capabilities and more able to collaborate with humans without the need for programming intermediaries. Some examples include (also see Table 1):

- Image classification and labelling (e.g., to species or individuals)
- Audio classification (e.g., to species)
- Geospatial image interpretation
- Natural language interactions with data and AI
- Machine learning of time series data to predict future outcomes
- Machine learning to analyse very large numbers of images such as satellite images
- Reinforcement learning to learn winning strategies for environmental improvement from simulators
- Environmental DNA analysis

Beyond identifying and classifying measurements into useful data, new approaches for quantifying ecological relationships that integrate statistical inference and machine learning are developing in parallel. We have assembled a multidisciplinary team with expertise at the forefront of these disciplines for the Innovation Sweep.

Data-sharing platform

Data-sharing is key to extracting the greatest value from the investment into collecting the data in the first place. Data-sharing, however, also creates challenges around confidentiality and ownership. Single, standardised data lakes are one solution, but are expensive and time-consuming to negotiate and maintain. Contemporary solutions such as distributed databases and federated learning may offer faster, lower-cost solutions to data sharing. Options for curating and sharing the large datasets generated by next-generation ecological technology will be explored in this report.

Skilling a workforce to adopt and implement technologies

Technology-based training is critical for the uptake and development of this “new ecology”. This Innovation Sweep will map the skills necessary to adopt the identified emergent technologies above and identify bottlenecks in skills development and educational and professional pathways to resolve them.

Objectives

The purpose of the Innovation Sweep is to form a watch-list of emerging technologies that have the potential to be used to monitor ecological outcomes of interest to the MDBA at the landscape scale. It is not designed to be prescriptive, but instead provides a scope of areas of interest. Each emerging technology will be briefly explained and its potential application outlined, so that the report can also be used as the initial step towards incorporating new technologies into the MDBA's business. The report will also provide insights into the advantages and opportunities of implementing an advanced data-sharing system that will enable the large datasets collated by these technologies to be readily accessed and shared across the MDBA network. In addition, the report will guide proponents towards skilling their workforce with the tools and capabilities required to implement these technologies on the ground. Ultimately, the end goal of this report is to provide a framework to lower the cost of monitoring in the MDB, improve data accuracy, inform watering strategies to improve outcomes and make data and information more available within the MDBA network and beyond.

Methodology

The Innovation Sweep Report identifies and informs on emerging technologies in conservation and natural resource management that have the potential to be applied in the MDB at scale. To identify these innovations, a five-step process was undertaken; discovery, synthesis, consultation, evaluation and conclusion (Figure 2).

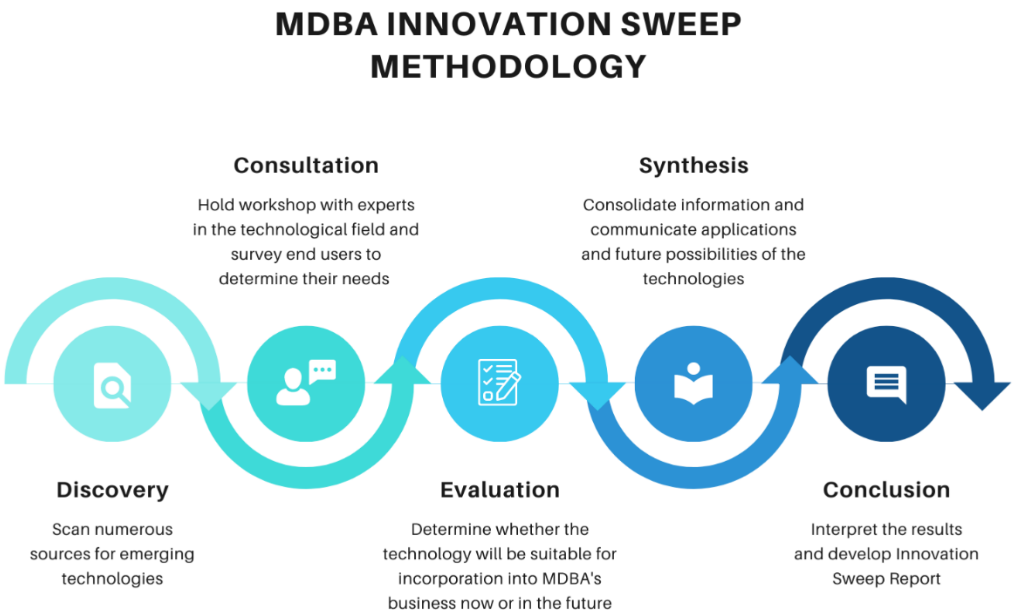


Figure 2 *The five-step approach undertaken to complete the MDBA Innovation Sweep.*

A scan of emerging technologies was undertaken to identify and shortlist those innovations of particular interest or application to the MDBA's business. A synthesis of those shortlisted technologies was then developed, and experts from a range of institutions were engaged to provide insight and specialist knowledge. Ongoing communication with the target audience of this report (managers within the MDBA network) identified the requirements of these end-users for emerging technologies. The shortlisted technologies were subsequently evaluated for suitability and grouped into 3 categories according to their relevant application; sensors, data analysis and models. Finally, technical reports containing a brief description of the technology, its current uses and future directions, and the potential applications for the MDBA's business, were compiled. To ensure that the Innovation Sweep Report provided relevant and practical information for the MDBA, key pathways for each technology were reviewed, including opportunities for adoption in the MDB, available platforms for data sharing, analysis and collaboration, and requirements for the skilling of a workforce to adopt and implement new technologies.

Technologies

Here follows a summary of the key technologies identified by the Innovation Sweep, grouped into 3 categories: sensors, data analysis and models.

These categories are designed to address each stage of the technological pipeline (Figure 3).

Each summary includes the following;

- A brief description of each technology
- SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis
- Current applications and potential uses in the MDBA's business
- Costs and limitations
- Technological developments
- Identification of leading experts in the field

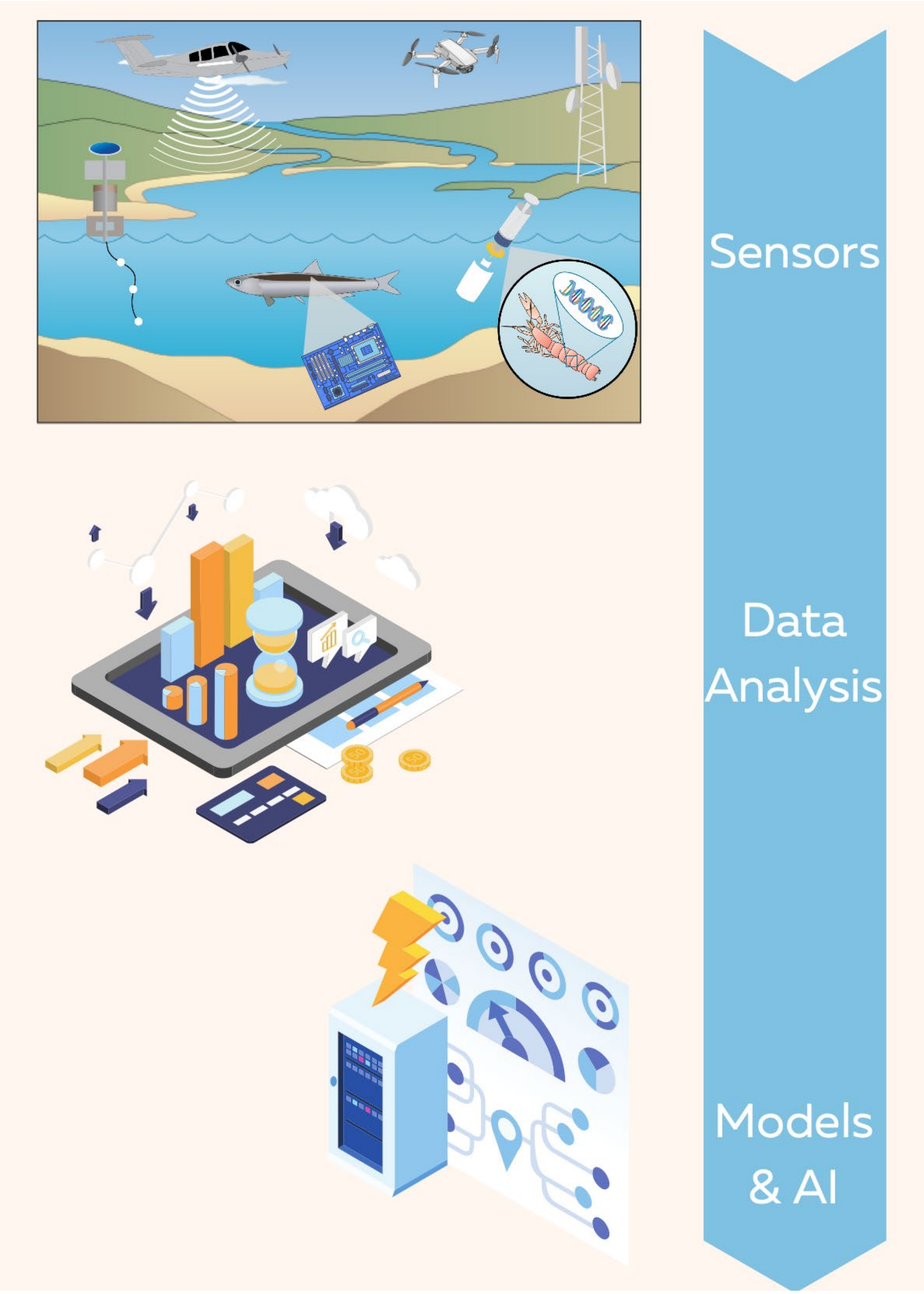


Figure 3 *The technological pipeline of data collection and analysis.*

Category 1: Sensors

An Introduction to Autonomous Sensing

In recent years there have been a number of new technology applications and breakthroughs in sensor arrays, such as camera traps and audio-recorders, that are useful for monitoring various environmental factors of interest to the MDBA. This may include species such as waterbirds, fish, amphibians, zooplankton and phytoplankton, as well as spatial changes in habitat through water inundation, riverine and floodplain vegetation and mudflats extents.

Sensor technologies have been commercially available for many years. However, the cutting edge that is relevant to applications at the types of geographic and temporal scales that can facilitate conservation outcomes are the improvements in existing technologies to produce products that are durable, lightweight, and cheap, as well as being power efficient, easy to deploy, element-proof, and allowing remote communication. It is here that tremendous gains can be made. Wireless sensor networks can expand traditional camera traps to accommodate other sensors (e.g., acoustic) with real-time data communication via satellites that are driven by low power networks. One particularly promising area may be in the deployment of SmallSat or nanosatellite constellations that have the potential to yield higher resolution data at 3-5 metres. These high-resolution sensors are more relevant to environmental monitoring than those currently available and can be deployed at a lower cost.

There is a myriad of potential applications of technologies with some potentially relevant to the MDBA listed in Table 2. The most obvious benefits are the potential to deploy sensors at large spatial scales and to record consistent spatial data at any temporal scale that was deemed appropriate. The costs will be significantly lower than that of human effort and hours for equivalent data and will be greatest where there is a need to develop and deploy wireless and/or low power networks to power and record data in real-time at scale. The Innovation Sweep will evaluate detailed costings and explore the status of technology development in these areas for specific biodiversity monitoring applications. Most of the base technologies are well developed and tested, so the risks to application will be in operationalising deployments and data retrieval at scale.

Table 2 *Examples of the innovative technologies that can be applied to monitoring aquatic ecosystems*

Feature	Application	Technology
Water Accounting	Measurement of inundated area in floodplain systems	Remote sensing and image analysis of open water
	Broadscale measurement of evapotranspiration and plant water use	Radar and optical remote sensing of evapotranspiration and soil moisture content
	Connectivity between rivers and floodplains	Remote sensing and image analysis of open water showing degree of connectivity between rivers and floodplains
	Volume of water contained in snow pack	Lidar measurement of snow pack depth (difference between height with and without snow)
Mudflat area	Measurement of potential foraging habitat for shorebird species	Machine learning analysis of remote sensing imagery
Riverine and floodplain vegetation	Plant performance in response to water regime	NDVI and hyperspectral analysis, onsite cameras and image analysis
Fish monitoring	Fish movement	Electronic tags and network of sensors
	Fish movement	Citizen science and innovative reward for return of tags by rec fishers
	Fish movement	Camera deployed at fish way recording fish passage and species using image analysis
	Fish movement	eDNA

	Fish health	Cameras deployed on fishways identifying and recording size/body condition of fish using image analysis
	Fish habitat	3D side scanner to evaluate snags, structure and fish presence
Waterbird monitoring	Waterbird abundance	Drone imagery with visual sensors or paired thermal and visual sensors and image analysis
	Waterbird nesting sites and abundance	Drone imagery coupled with automated image analysis
	Waterbird body condition	Image analysis and machine learning to determine body condition from photo images
Amphibian	Amphibian diversity	Monitor from call with audio recording and audio processing with machine learning to identify species.
Zooplankton	Zooplankton abundance and diversity	eDNA of samples and comparison with a genetic library of known species.
	Zooplankton abundance and diversity	Image analysis of microscope samples possibly in situ with FlowCam
Phytoplankton	Phytoplankton abundance	Image analysis of microscope samples
	Phytoplankton abundance	In situ Flow cytometry such as a Cytobuoy
	Phytoplankton abundance	In site microscopy and image analysis such as a FlowCam
	Cyanobacteria	Phycocyanin sensors

Data analysis	Automation of analysis and reporting	Machine learning Automated model run Dashboard driven displays
	Alert to breach of water quality standard	Automated alert when a sensor detects exceedance of a water quality indicator such as low DO, high DOC, cyanobacteria

With the adoption of these technologies, there is a need for the capacity to process the 'big data' that are generated from these remote sensors (i.e., 'data loggers') to transform the recorded signals into data. Automated data processing pipelines are integral to preserving the efficiencies gained by using remote sensor technology solutions for data collection. Developing these methods for the specific types of sensor data can transform the technology from simply recording signals, to providing data that can be analysed and interpreted. Further, these processing pipelines should be considered in the context of the development and implementation of federated data repositories for collating and sharing the data for analysis, reporting and visualisation.

Environmental DNA (eDNA)

What is eDNA?

Environmental DNA (hereafter eDNA) refers to genetic material that is extracted from environmental samples such as soil, sediment, water or air. eDNA may originate from faeces, hair, skin, or other biological material shed by organisms in the environment. Preservation times of eDNA varies significantly depending on the conditions the genetic material is exposed to; for example, eDNA may persist for several weeks in temperate water and hundreds of thousands of years in permafrost. A single environmental sample may contain genetic information on entire ecological communities. To extract this information, a polymerase chain reaction (PCR) is typically used to amplify any eDNA that may be present, in either a single- or multi-species approach using species-specific primers (also known as barcoding) or universal primers (also known as metabarcoding). eDNA is therefore an effective monitoring method for biodiversity at a landscape scale, particularly in freshwater ecosystems.

eDNA research has a number of benefits for ecosystem monitoring and management (Figure 4). Firstly, it is non-invasive, and can be performed without the destructive sampling of organisms or habitats. Secondly, it allows for sampling and assessment of multiple organisms with minimal cost and effort. Thirdly, eDNA sampling can facilitate the early detection of pathogens and invasive species in an environment, allowing managers to plan and carry out the appropriate intervention. Finally, it is an effective method of detecting and monitoring cryptic species that are otherwise elusive and difficult to survey, and so can be more effective than camera trapping and other manual survey methods.

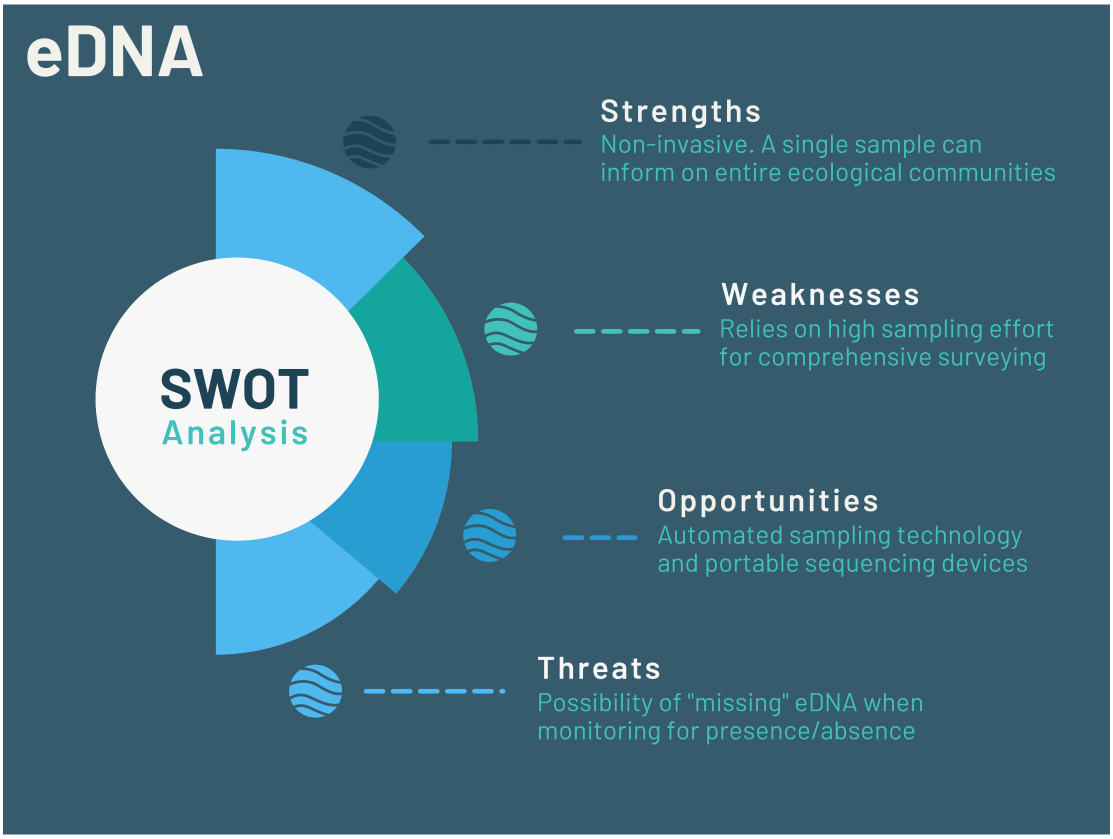


Figure 4 *Summary of the strengths, weaknesses, opportunities and threats of eDNA research.*

Current Applications

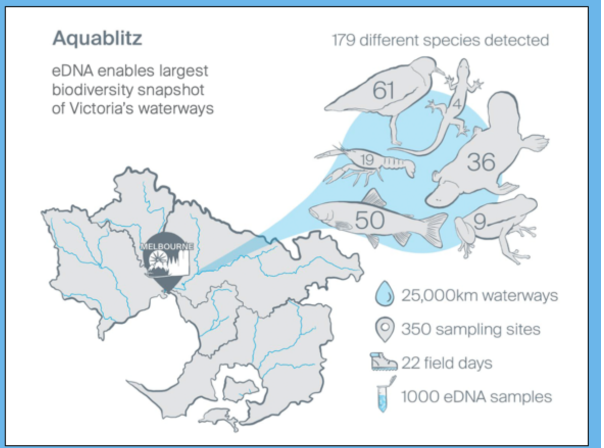
eDNA is being used across a range of industries, from fisheries management to conservation (Box 1). For example, eDNA has been successfully used to monitor the spawning activity of the endangered Macquarie perch (*Macquaria australasica*) (Bylemans et al. 2017), detect invasive species (Furlan et al. 2019), and monitor post-release survival of the endangered northern corroboree frog

(*Pseudophryne pengilleyi*) (Rojahn et al. 2018). eDNA is also being combined with citizen science to generate Australia-wide data on the presence of platypus (*Ornithorhynchus anatinus*) in waterways (Brunt et al. 2018; Lugg et al. 2018). Other recent applications of eDNA include assessing parasite loads in water for early disease intervention, mapping soil and marine microbiomes, and conducting biodiversity assessments of groundwater systems for Environmental Impact Assessments.

Box 1: Application of eDNA in aquatic monitoring - Aquablitz

The Aquablitz project, led by Dr Rhys Coleman (Melbourne Water) and Dr Andrew Weeks (EnviroDNA), was designed to capture snapshots of the biodiversity contained in the ~25,000 km of interconnected waterways that make up the city of Melbourne's catchments. eDNA was chosen as the preferred technique for this research, as applying traditional survey methods such as trapping was not feasible on such a large scale. eDNA enabled the researchers to bypass the financial, accessibility and time constraints that would normally hamper a study of this size.

Samples were collected from 350 sites across the study area and analysed in a laboratory using an eDNA metabarcoding approach, allowing researchers to identify species from multiple taxonomic groups from a single sample. Ultimately, 179 species were identified, both native and invasive, including 50 fish, 36 mammals, 61 birds and 19 decapods. Some species were elusive and difficult to find by sight, and others were non-aquatic species detected from genetic material shed during their interactions with waterways. This finding highlights the high sensitivity of eDNA. In comparison to previous surveys of the same catchments using traditional methods, this single-occasion survey was able to generate more biodiversity data at lower cost, survey effort and trapping intensity, suggesting a promising future for ongoing monitoring efforts using eDNA.



Images used with permission of EnviroDNA. Thanks to Dr Rhys Coleman (Melbourne Water) and Dr Andrew Weeks and Jacquie Murphy (EnviroDNA)

Potential Applications for the Murray-Darling Basin

There are a number of potential applications of eDNA that relate directly to management issues faced by the MDBA. eDNA can be used to evaluate species and community responses to environmental stressors, including drought and pollution. Further, eDNA has proven to be an effective method for the early detection of algal blooms, outperforming other techniques such as light microscopy and remote sensing. eDNA analysis can be used to identify the presence of harmful algae before blooms occur, as well as the presence and abundance of algicidal microorganisms that are capable of degrading toxins. Pathogens and parasites associated with fish and amphibian deaths can also be detected using eDNA, allowing managers to implement response plans in a timely manner prior to mass die-off events and preserve fish communities of ecological and economic value. In addition, increased concentrations of eDNA in the Murray-Darling Basin could be used to monitor fish spawning events. When detecting cryptic species, eDNA often demonstrates higher accuracy than conventional monitoring methods. eDNA, therefore, represents a promising technique to generate presence/absence data for rare and endangered species in the Murray-Darling Basin, such as the pouched lamprey (*Geotria australis*). eDNA can also be used to assess the microbiome of acid sulfate soils, an issue faced in the Murray-Darling Basin during times of drought, by quantifying the presence of acidophilic microorganisms. Finally, although not currently common practice, eDNA may prove useful as an early detection tool for blackwater events by measuring the concentration of plant genetic material in waterways.

Costs and Limitations

The cost of eDNA sampling varies with the sampling and sequencing effort, as well as the target species. Smart et al. (2016) determined a high-cost eDNA sampling scenario for a single species monitoring approach to be $86.06 per sample, with an additional $1,569.08 for the site setup. By comparison, a low-cost scenario for the same approach was determined to be $62.29 per sample and $569.08 for site setup. Therefore, it is possible to develop cost-efficient approaches to eDNA where necessary. Regardless, in many cases eDNA is cheaper to implement than traditional visual sampling methods.

eDNA is a low-risk approach to environmental monitoring because it does not require destructive analysis. The most notable limitations associated with eDNA are the limited persistence of eDNA in the environment and the uncertainty associated with sampling. It is difficult to confirm the "absence" of a species, as the genetic material may simply have been missed during sampling. However, these limitations can be reduced by increasing sampling effort and combining eDNA with other monitoring techniques.

Box 2: *Application of eDNA in aquatic monitoring – identifying spawning sites in aquatic systems*

Using species-specific qPCR assays, Bracken et al. (2019) were able to infer the spatial and temporal extent of spawning events for the sea lamprey (*Petromyzon marinus*) in two Irish river catchments.

During spawning season, the release of seminal fluid and ova, in addition to shed cells from nest building and decaying tissue from dead or dying individuals, leads to a considerable increase in *P. marinus* DNA in the water. By sampling at fixed points over three years, the authors were able to successfully outline the general spawning locations as well as the ecological distribution of the species, the extent of their upstream migration and the potential barriers to their migration within the catchments. This 'snapshot' sampling approach also revealed peaks in eDNA concentration in areas that were not considered important habitats for *P. marinus*, highlighting areas of interest for future spawning surveys.

The concentration of *P. marinus* DNA in the water column was also found to be significantly correlated with both the number of fish and nests that were observed, providing additional information on the population of *P. marinus* in the system. These results demonstrate that the power of eDNA is far more versatile than simple presence and absence analysis and can provide information crucial to the management of endangered species.

Image from Bracken et al. 2019, licensed under CC-BY 4.0 http://creativecommons.org/licenses/by/4.0/

Technological Development

There have been several technological advances in recent years that have made eDNA sequencing faster, cheaper and more convenient. Recent improvements in next-generation sequencing (NGS) technologies, including the establishment of commercial platforms (e.g., Illumina, Diversity Arrays Technology, Roche), have made the sequencing of large genetic datasets much more accessible and affordable for ecologists (Shokralla et al. 2012). Notably, portable sequencing devices are now available that can be taken into the field and perform sequencing at the source site (Tosa et al. 2021). Oxford Nanopore Technologies' MinION (starting at ~$1,400 AUD) can be plugged into a laptop or PC, while the company's alternative product, MinIONMk1C (starting at ~$6,700 AUD) is an all-in-one device. Researchers can load and sequence samples in real-time, as soon as they have been collected, drastically reducing processing time and negating the need to store and transport samples. Reducing sequencing effort further, researchers have recently developed a method of passive DNA filtration that can occur directly in the water column without the need to filter samples after collection. Automated monitoring stations designed to collect eDNA samples and quantify species diversity are also being designed, but are not currently commercially available.

Key Experts

In Australia, a number of organisations are working on refining eDNA techniques and applying them to research programs. For example, EcoDNA (https://www.ecodna.org.au) is a Canberra-based initiative that has used eDNA in a number of aquaculture and conservation research programs. Similarly, EnviroDNA (https://www.envirodna.com), a consulting company based in Melbourne, have combined eDNA research with citizen science to conduct biodiversity assessments of waterways and monitor cryptic species. The Australian Microbiome Project (https://www.australianmicrobiome.com) is currently using eDNA to map marine and soil microbiomes. Researchers are also employing eDNA to assess the biodiversity of groundwater systems, which was previously extremely difficult to quantify, to inform Environmental Impact Assessments.

Autonomous Platforms

What are Autonomous Platforms?

Autonomous platforms are platforms that are able to navigate and undertake pre-planned missions using provided mission plans and logic (such as obstacle avoidance and pre-loaded survey paths) combined with sensors and instruments to provide automated survey systems. These platforms include aerial drones (Unmanned Aerial Vehicles (UAV)), autonomous boats (Autonomous Surface Vessels (ASV)) and underwater vehicles (Autonomous Underwater Vehicles (AUV)). Of these, drones are now commercially available and fully operational while AUVs are still mostly in the development phase for riverine systems. ASVs are somewhat in-between these stages, as the components are all individually proven (control, navigation and sampling/surveying) but there are few commercially available systems. This makes ASVs a key area to invest in, as they could bring new capability into the management of river systems and would allow for new types of outcomes.

The platforms act as a force multiplier, allowing surveys to be undertaken in more places, more often (Figure 5). Autonomous platforms therefore create the potential for continuous surveys of long river systems to be undertaken, allowing for better detection of issues such as low oxygen events, run-off, and changes in river profile with minimal user intervention. The systems increase in scale simply by having more units. Mobile platforms can be linked to static monitoring stations to give a more complete picture of the river system, and platforms can be tasked with responding to particular events, such as floods or changes in river levels. Systems can work at night and in conditions that may be unsuitable for people, such as areas with marine pests. Sonar-based river profile surveys, for example, can be undertaken at night. A series of configurable platforms allows for a range of responses to be quickly deployed in response to events as they unfold.

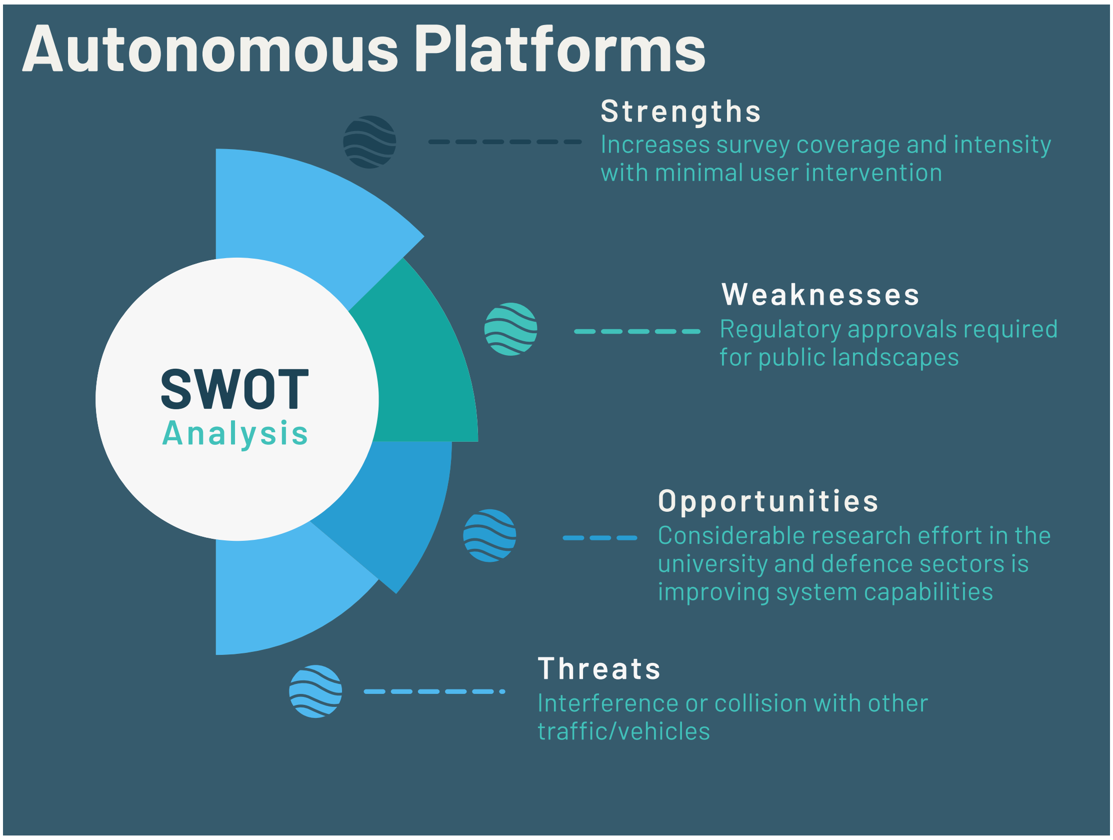


Figure 5 *Summary of the strengths, weaknesses, opportunities and threats of Autonomous Platforms.*

Current Applications

ASVs have been employed for dam monitoring, either for water quality or for the extent and depth of the dam and resulting water levels. Other ASVs have been used to monitor and manage the marine environment (Box 2). Further, a number of studies (e.g., Dunbabin and Grinham 2010; Bin Mat Idris et al. 2016) have demonstrated the utility of automated platforms in undertaking routine surveys over large areas with the potential for results to be analysed in real-time, allowing for reactive management of events.

Potential Applications for the Murray-Darling Basin

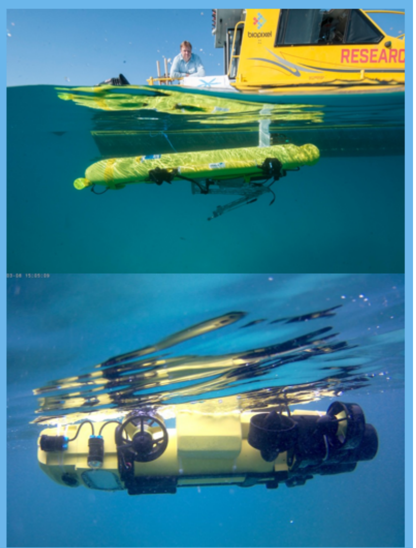
Automated platforms present several opportunities for deployment in the environmental monitoring of the Murray-Darling Basin. An ASV could undertake regular automated surveys of large stretches of rivers where navigable waters exist and survey for basic biochemical parameters such as temperature, turbidity, salinity, dissolved oxygen, and water flow. In areas that are not frequently trafficked, these surveys could be done on a routine and fully automated basis. In areas with regular traffic, more restrictive surveys would need to be done to ensure the safety of other watercraft. It is possible to build fully automated selfcharging platforms that could undertake entire river surveys with multiple recharge points as required. Further, the platforms could be configured so that as real-time data is collected, any unusual or significant results are transmitted to the appropriate authorities, allowing for real-time event detection and subsequent response. A range of sensors can be fitted to autonomous platforms, such as bio-chemical, optical and imaging, including current metres and sonar systems.

Box 3: Application of Autonomous Platforms in aquatic monitoring – COTSbot and RangerBot

The COTSbot and RangerBot Autonomous Underwater Vehicles (AUVs), led by Professor Matthew Dubabin (Queensland University of Technology) and Dr Feras Dayoub (University of Adelaide), have been designed to assist in monitoring the Great Barrier Reef and managing marine pest species such as the crown-of-thorns starfish. Both AUVs use advanced real-time image processing techniques to detect the presence of invasive species, conduct counts of individuals and map their distribution.

The COTSbot is designed specifically to assist in the eradication of crown-of-thorns starfish, an invasive invertebrate that feeds on coral and can have devastating impacts when present in large numbers, resulting in significant coral losses on the Great Barrier Reef. The COTSbot combines robotic vision with classification algorithms to deliver eradication programs at scale, requiring no tethering system and minimal human interaction once deployed. It is able to identify crown-of-thorns starfish in real-time with 99.4% accuracy using automated image-based detection, and autonomously deliver a lethal injection of bile salts into the starfish with no residual impacts on the environment. The COTSbot can search the marine environment for up to eight hours, delivering more than 200 doses of bile salt. The next generation of the COTSbot, the RangerBot, is completely vision-based, and is used to monitor reef health indicators such as coral bleaching and water quality. It can also monitor and control pest species, detect obstacles and complete management tasks, such as capturing and delivering coral larvae to damaged reefs. Its convenient size means the RangerBot can be easily deployed by one person and operated using a tablet-based interface.

The COTSbot and RangerBot have the ability to 'swarm' autonomously, with one operator able to coordinate many robots to work collaboratively on management tasks. These types of AUVs represent the next generation of underwater monitoring and conservation management, and are able to perform the same tasks as divers at lower cost and survey effort, over a larger scale.



COTSbot (top) and RangerBot (bottom)

Images used with permission of Queensland University of Technology Research. Thanks to Dr Feras Dayoub (University of Adelaide) and Professor Matthew Dunbabin (Queensland University of Technology).

Costs and Limitations

A number of commercial autonomous platforms are currently available (e.g., https://www.ysi.com/hycat). However, these systems are a combination of prototype and early commercial systems, and are therefore costly and only partially operational. Opportunities exist for bespoke platforms to be developed with a total product cost of < $50K given the currently available components, and this would allow for multiple units to be deployed at any one time.

The main risk associated with autonomous platforms is that of interference with other river traffic, and the possibility that a platform may become stranded or go off course. The latter can be dealt with by monitoring systems and using existing navigation data to ensure the platform stays within navigable waters. Gaining regulatory approval for autonomous platforms operating around traditional traffic is an additional consideration, but there is ongoing work to facilitate this (see: https://tasdcrc.com.au/).

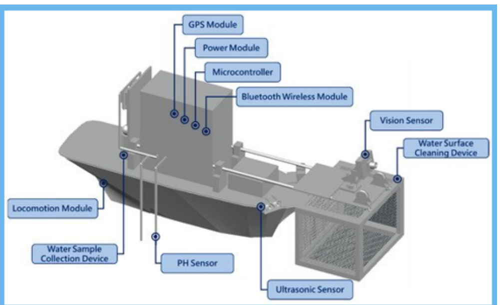
Regulatory compliance of autonomous platforms is managed from a variety of stakeholders from the Civil Aviation Safety Authority to location specific Parks management authorities. The regulatory framework is in a state of evolution and so permits, licensing and compliance requirements must be considered as part of introducing these platforms into operations.

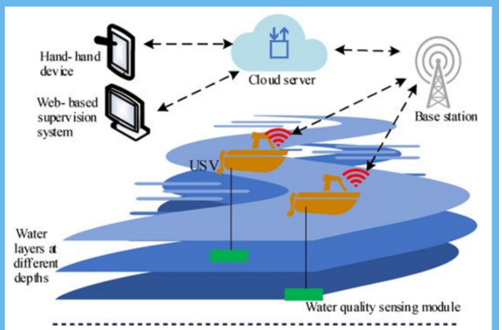
There is a large interest in the adoption of autonomous platforms, and ASVs in particular. The components of these systems exist and are field ready; therefore the path to uptake and use is relatively simple. A number of water authorities are already interested in their applications (e.g., SEQ Water). Currently, uptake is limited to water authorities that manage restricted waterways such as dams, as the regulatory framework for open river use is still under development.

Box 4: Application of Autonomous Platforms in aquatic monitoring – automated surface vehicles for water quality monitoring

Water quality monitoring is a task that is essential for managing Australia's water resources. Currently, there are reliable and accurate monitoring stations throughout many of Australia's aquatic systems, but their fixed positions and expensive cost limit their scale and, thus, their ability to respond to specific events. Autonomous surface vehicles (ASVs) can be rapidly deployed, autonomously travel throughout aquatic environments, sample hard to reach places and capture site variability.

Chang et al. (2021) developed an ASV capable of object avoidance, surface water cleaning, water quality monitoring (pH) and water sample collection.
Madeo et al. (2020) developed a low-cost prototype ASV capable of monitoring water quality (e.g. dissolved oxygen, salinity, flow, pH) under manual control or autonomous point-to-point navigation.
Cao et al. (2020) used a surface vehicle with an automatic cruising system capable of monitoring water quality (pH, total dissolved solids and turbidity) at specific points on a lake. Sampling points can be organised into a grid to capture the full spatial heterogeneity of the system. Additionally, the water quality monitoring device can be lowered into the water by up to two meters to capture the lake's three-dimension variation.





Technological Development

As all of the components (platforms, control systems, navigation systems and instruments) for autonomous platforms currently exist as mature systems, the only real development required is the packaging of these components into a functional system that meets regulatory approval. Most of this development is currently in the research and university sector, but there is also development underway in defence as well as the commercial arena.

The Australian Institute of Marine Science (AIMS) is undertaking a parallel approach for a coral reef ASV and has undertaken an Expression of Interest round under AUSTender. A second round will be undertaken in early 2022, with the delivery of the first platform scheduled for late 2022. This may be a development pathway that the MDBA can follow, as although the needs of the organisations differ, the basic requirements will be similar. Modelling by AIMS has shown that autonomous platforms allow work to be done on reef surveys in parallel – the platforms undertake some tasks, while the human team contributes the value-add work, for example, allowing work to be done in areas with marine pests. Autonomous platforms are a critical pathway to scaling the work required in order to meet the needs of organisations working across large spatial scales.

Key Experts

Key experts in autonomous platforms in Australia include AIMS, Queensland University of Technology (QUT), Defence Science and Technology Group, the Australian Defence Force, CSIRO's Data61 initiative, and the Australian Centre for Field Robotics.

Animal Tracking Technology

What is animal tracking technology?

The process of remotely tracking animals in the wild provides highly valuable information on spatial ecology, and is a key method to inform wildlife management. Recent advances in GPS, radio- and acoustic telemetry technology have allowed wildlife managers and ecologists to collect movement data for a diversity of species (fish, amphibians, reptiles, birds, mammals) at a range of spatial and temporal scales. Tags are becoming smaller and more lightweight while simultaneously increasing in accuracy and performance. These technologies are capable of generating huge amounts of data on the movements of a tracked animal, thereby providing precise and valuable insight into the movement, habitat use and behaviour of individuals across terrestrial and aquatic ecosystems (Figure 6).

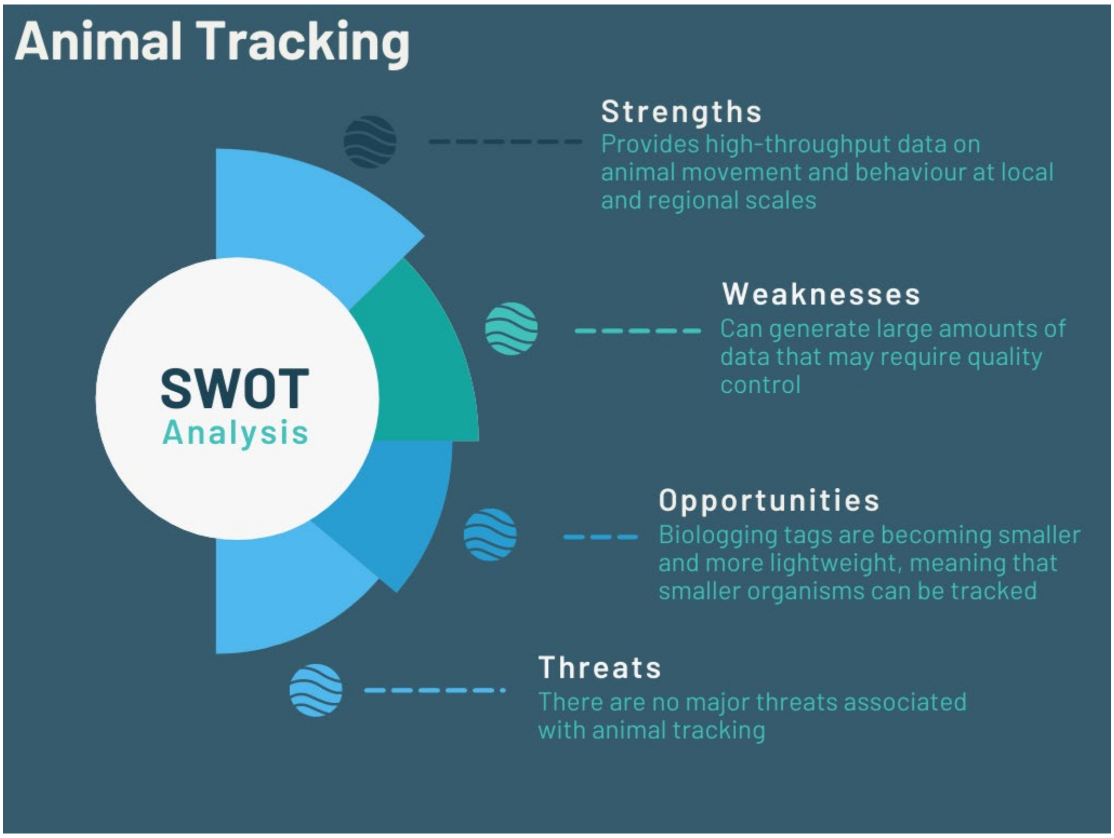


Figure 6 *Summary of the strengths, weaknesses, opportunities and threats of Animal Tracking.*

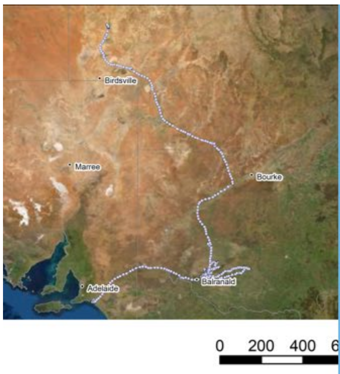
Current Applications

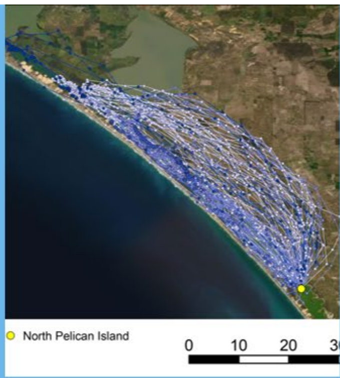
One example of this type of high-throughput animal tracking technology is passive integrated transponder (PIT) tagging. PIT tags are electronic microchips with unique digital identification codes that are typically implanted into animals subcutaneously or in the peritoneal cavity. PIT tags are inactive until activated by a reader (which may be a handheld scanning device or a remote system installed in the environment), at which time the tag transmits its unique identification code and the presence of an individual can be recorded. The technology was designed for fisheries research and management, but its applications have since expanded to terrestrial wildlife, livestock and veterinary purposes, and even the live animal trade. PIT tags can now be produced as small as ~8 mm in length and thus may be used to monitor the movements of small animals, including invertebrates (e.g., Foote et al. 2018).

In aquatic environments, acoustic telemetry is commonly used to investigate the spatial ecology of fishes and marine mammals. Tags come in various sizes and battery lives and are either implanted into animals or tethered externally. The tags emit a sub-audible ping with a unique identification code that is recorded if within the range of an acoustic receiver deployed in the environment. Individual projects typically deploy 'arrays' of receivers in a strategic manner to address specific hypotheses related to animal movement. The rapid uptake of this technology has led to the deployment of large communal arrays of receivers in certain locations to bolster individual projects and streamline the sharing of acoustic data. In recent years, acoustic telemetry has developed beyond a means of investigating movement, to facilitating investigations of physiology and behaviour. Tags can now be coupled with sensors for monitoring environmental conditions experienced by animals (e.g., pressure (depth), temperature, dissolved oxygen) and accelerometers to measure the speed of acceleration during certain activities (e.g., foraging).

Radio-telemetry is a further technology commonly employed for tracking animal movements in terrestrial and freshwater environments. As per acoustic tags, radio tags come in various sizes and battery lives and are either implanted into animals or tethered externally. The tags emit radio waves that communicate information to nearby reader devices, which can be fixed remotely in the environment. Alternatively, animals can be tracked manually by researchers using a radio receiver and antenna. In terrestrial environments, signals can propagate over large distances (kms), making the technology suitable for tracking wide-ranging animals.

On a larger scale, satellite telemetry, which marks an animal with a unique biologging tag equipped with a satellite geolocation sensor, can be used to accurately track the movements of animals across large distances using GPS. This technology is particularly useful for monitoring highly mobile species across vast land- and ocean-scapes (e.g., migratory birds, pelagic sharks and marine mammals) and understanding long-term patterns of habitat use (Box 3).





Box 5: Application of animal tracking technology

Satellite trackers fixed to animals can assist in determining where animals are moving and what habitat they are utilizing. Highly mobile animals such as birds and fish can move many hundreds of kilometres. Managing habitat for these organisms requires an understanding of when and where they are using the various habitat.

Satellite and radio tracking of individuals is used in the Murray Darling Basin to monitor fish and bird movements.

Dr Rowan Mott and Dr Micha Jackson from the University of Adelaide are tracking the movement of birds in the Coorong. Both migratory and ‘permanently” resident bird species are being tracked.

Satellite trackers on Pelicans (lower panel) have revealed that some individuals move to central Australia when rainfall in the north creates suitable habitat in that region (top left panel). Other breeding individuals are observed to feed along the North Lagoon of the Coorong returning to the rookery at Pelican Island (top right panel).

Potential Applications for the Murray-Darling Basin

Animal tracking using high-throughput technologies presents a range of opportunities for application in the MDBA’s business and, indeed, is already being applied. Due to the wide range of spatial and temporal scales at which these technologies can operate and, indeed, the small size and lightweight tracking devices currently available, most species of interest can be tracked. To date, radio- and acoustic telemetry have been used to investigate the movement of a range of fish species and generated significant new knowledge, including identifying key migration periods and relationships with hydrology and the impact of flow regulating infrastructure on movement. This information has directly influenced site, state and basin-scale management, notably the delivery of environmental water and remediation of barriers to movement. Additionally, the MDBA is the custodian of a ‘Basin-scale’ array of acoustic receivers that supports various fish-related investigations across the basin.

PIT technology is also currently used in parts of the MDB and has assisted with MDBA-funded research and river management. All fishways on the main channel weirs of the River Murray have PIT readers installed that have been used to monitor fish movement and passage through fishways.

For radio-telemetry, acoustic telemetry and PIT technology, the majority of studies have predominantly focused on adults of large-bodied fish species (e.g., adult length typically >300 mm). The development of small acoustic and PIT tags now presents an opportunity to investigate the movements of juvenile fish and potentially adults of small-bodied species (adult length typically <100 mm).

Despite the large spatial scale of the MDBA's operation, the movement of bird and mammal species could also be effectively tracked using satellite telemetry, with some applications already occurring in some ecosystems in the MDB, such as the Coorong.

Costs and Limitations

The cost of PIT monitoring lies predominantly in the monitoring system; PIT tags themselves are typically low cost, around $4-5 each, while station monitoring systems can cost ~$25-200k depending on the complexity. Acoustic tracking tags, meanwhile, can cost $350-450 each, while the receivers are priced at around $1700-2000. Similarly, radio tracking tags also cost $300-400 each, but the readers can reach $25-35k.

Although the large amount of data generated by high-throughput animal tracking technologies has many benefits, it can also present challenges to wildlife managers. Short tracking intervals and high temporal resolutions can increase location error, which may introduce uncertainty. This may be overcome by "cleaning" the data - inspecting the data and removing outliers. Performing this task manually on large datasets is difficult and time-consuming, but the application of an automated process using protocols and models can assist here.

Box 6: Application of animal tracking technology – acoustic micro-transmitter for monitoring small fish

Monitoring and conservation of aquatic animals requires an understanding of their movements and behaviours and their interactions with hydraulic structures such as dams, locks and barrages. Acoustic telemetry has been the primary method for achieving this in aquatic environments due to its long detection range and high accuracy and performance in aquatic environments. However, the size of these transmitters has typically limited these techniques to animals weighing at least 11 grams.

Deng et al. (2021) developed an acoustic micro-transmitter (AMT) that can remotely track the three-dimensional movements of aquatic animals that weigh as little as 4 grams. The AMT has a dry weight of 0.08 grams and can function for 30 days at a 5-second transmission rate with no adverse effects on fish mortality or swimming ability expected.
In freshwater, the detection range of the AMT is approximately 80–140 m. Receivers could be placed throughout riverine and estuarine environments to gain crucial information on animal migration and dispersion and shine a light on juvenile movement and behaviours in a detail that was previously unfeasible. Receivers could also be placed at fish passageways and hydraulic structures to determine their impact on fish movements.

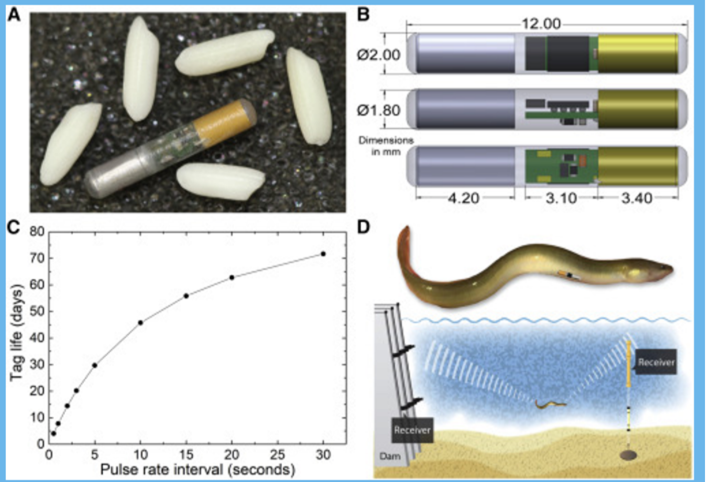


Image from Deng et al. (2021), licensed under CC-BY-NC-ND 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0/

Technological Development

Tracking and bio-logging devices are becoming smaller, lighter, more powerful and more efficient every year, whilst improving in capacity to gather environmental and physiological data in association with movement data. This has broadened the potential applications of these technologies. It is now possible to track the movement of small species, such as invertebrates. Progress is also anticipated in the field with the combination of data-rich movement and behavioural studies, and analysis of this information in response to environmental factors and climatic change.

Key Experts

Key experts in Australia on advanced animal tracking methods include Dr Rowan Mott and Dr Misha Jackson, University of Adelaide, Chris Bice and Dr Leigh Thwaites, South Australian Research and Development Institute (SARDI), Dr Jason Thiem, NSW Department of Primary Industries (Fisheries), Wayne Koster, Arthur Rylah Institute, Heather McGuiness, CSIRO, and Karl Pomorin, KarlTek.

Category 2: Data Analysis

An Introduction to Data Analysis

Analysis of large and complex datasets has become far more achievable in recent years with the advent of increasingly accessible high-powered computing technology. Wildlife researchers and managers can now harness bioinformatics platforms to collect and analyse many types of data, including imagery, genomics, audio and movement. Excitingly, artificial intelligence and machine learning technologies present opportunities to process massive amounts of data in a highly sophisticated manner, at a lower cost and effort than manual analysis previously allowed. With the large datasets now being generated by remote sensing technologies, the uptake of these analysis tools is a natural progression for environmental managers.

Commercial platforms for machine learning and artificial intelligence are becoming more readily available and more affordable. In most cases, there remains a need for skilled users to further develop and "train" this software to fit the desired purpose. However, once harnessed, these data analysis platforms present many promising avenues for environmental monitoring at a landscape scale through tools such as object identification, pattern recognition and visual question answering.

Given the rapid advance of data analysis technologies in recent years, and the predicted development in speed, scope and accuracy in the future, it is an ideal time to integrate machine learning, artificial intelligence and other advanced bioinformatics pipelines into the MDBA's business. Potential applications of this technology at the landscape scale include monitoring the extent of vegetation or water bodies, monitoring change in vegetation over time, estimating the biomass of organisms, and identifying organisms to species level.

Machine Learning Based Analysis

What is Machine Learning Based Analysis?

Machine learning is a branch of artificial intelligence that can 'learn' the characteristics of the input data to build models that can be used to predict the content of unseen data or images. This is mostly applied to imagery, but can also be applied to time-series data, sound data and other media. For images, it can be used in 2 ways; object identification and scene classification.

Machine learning models are trained to recognise an object or pattern by exposing them to thousands or even millions of images to'learn' the characteristics of the object. The accuracy of the models to correctly identify objects in previously unseen images is directly related to the number and variety of images used to teach the model. The application of machine learning is, therefore, directly related to the number, quality and type of images used to generate the model, and so image libraries are the new 'gold' in developing and applying.

As it gets easier to collect data from cameras, automated instruments and data capture systems, the problem then becomes how to analyse this. Scaling the data collection only moves the problem to the analysis side. Machine learning is currently the only analysis method we have that will also scale and so it is the key technology piece in scaling for far-reaching analyses (Figure 7).

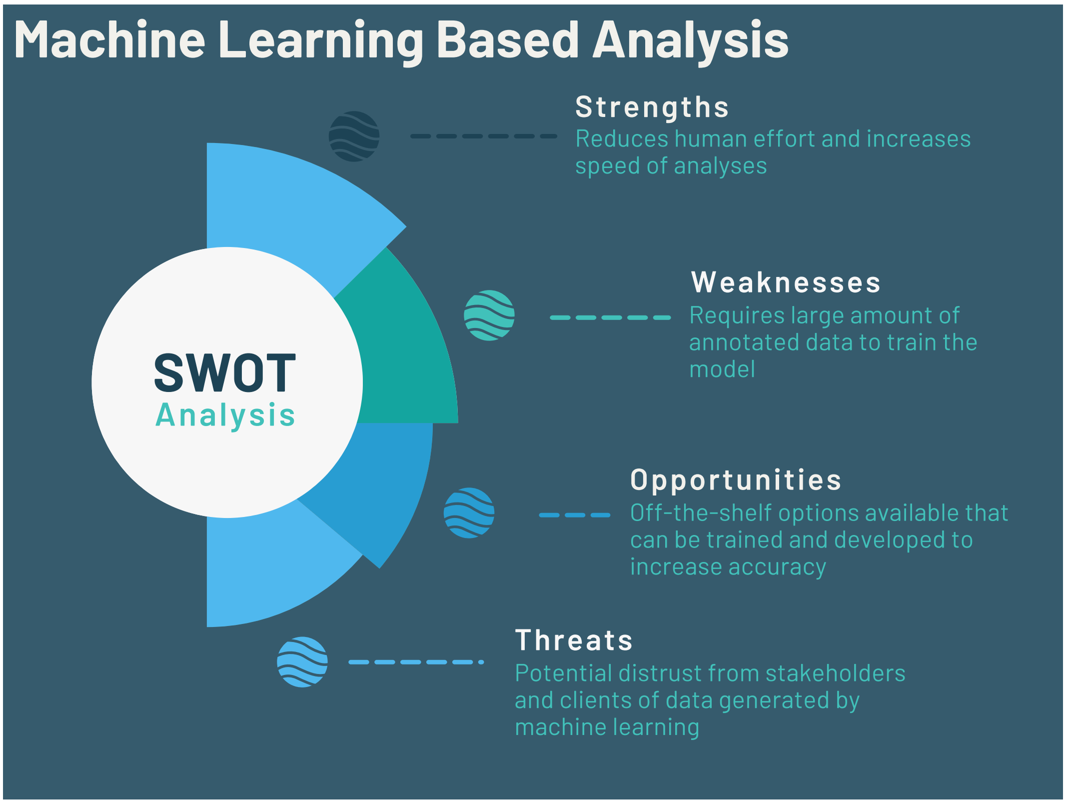


Figure 7 *Summary of the strengths, weaknesses, opportunities and threats of Machine Learning Based Analysis.*

Current Applications

There is little historical application of machine learning given that this is a relatively new area (although building on many decades of work). It is currently used in a range of day-to-day applications, including internet search engines and mobile apps and functions such as voice recognition.

Potential Applications for the Murray-Darling Basin

There are 3 potential applications of machine learning to environmental monitoring:

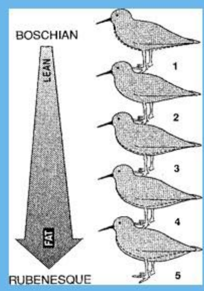
- Object Identification – automatic identification of objects of interest including pest species, economically important species, disease, invasive species and other environmental indicators, as well as vessels and other river traffic and illegal activities such as trespass and fishing (see Box 4).
- Scene classification – automated classification of imagery for per cent cover of the major benthic forms, biomass estimates of aquatic vegetation, automated length and size estimates of fish and so on.
- Pattern recognition – using time series data, such as monitoring data or other data, to recognise patterns in the data that represent particular events, such as linking individual station rainfall patterns to potential flood events or environmental data to spawning events.

Of these, the best developed are the object identification and image classification applications; the development of complex pattern recognition is still an area of active research.

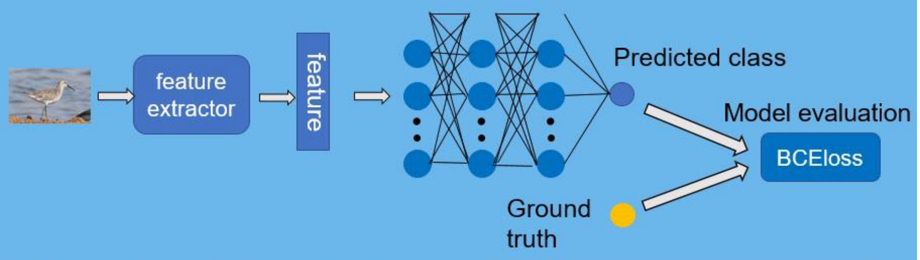
Machine learning can be applied as an adjunct to traditional methods. One current use in marine monitoring is to classify benthic images for per cent cover of the main benthic forms. Initially, machine learning was used to filter out certain image classes (mostly abiotic such as sand) to reduce the number of images analysed by hand. Now that the models are more robust, the computer-generated identifications are subject to human quality control (QC) with a focus on the areas where the models get confused or where the identification accuracy is low. This way, the human simply does a QC pass over the computer-generated values, reducing the human work to around 30% or less than a traditional workflow. As the models improve, this will reduce to the point where only a cursory QC analysis is required. At this point, the analysis can scale in sync with the scaling of the collected data.

Box 7: Application of machine learning to ecological monitoring – determining body condition score

Wiersma & Piersma (1995) developed an approach to assessing shorebird body condition based on scoring a bird's accumulated fat reserves based on its side-on abdominal profile. Migratory shorebirds should display noticeable fat accumulation (i.e. an increase in body condition score) prior to departure on their northward migration in the austral autumn.



A team of researchers* are developing a two-step machine-learning approach that first determines image suitability for body condition scoring and then classifies body condition for suitable images based on feature extraction. This could be used as a tool to assess habitat condition and compare the health of shorebirds at different sites.



*Dr Micha Jackson (University of Queensland), Dr Thomas Prowse (University of Adelaide), Dr Rowan Mott (University of Adelaide), Dr Yangyang Shu (University of Adelaide), and Dr Lingqiao Liu (University of Adelaide).

$$L_{bce}(\hat{y}, y) = -y \log \hat{y} - (1 - y) \log (1 - \hat{y})$$

Costs and Limitations

Key to the application of machine learning is the collection and curation of image libraries of the objects of interest. Strategically, it is worth investing in developing image libraries now in the recognition that any machine learning work in the future will need these. This can be done by harvesting existing imagery, using citizen science to collect new imagery or sponsoring collection programs. Images need to be annotated (the objects of interest identified in each image including the location within the image) and curated so that there is a level of assurance in the images. Modelling and simulation technology can be used to support extending the value from existing image libraries such as gaming engines (e.g., Unreal Engine or Unity). The importance of quality curated image libraries cannot be overstated and typically is the main limiting factor in developing machine learning models that deliver the required level of accuracy.

Even though there are many off-the-shelf machine learning services available (e.g., Microsoft Azure Custom Vision, Google Vertex AI, Amazon Web Services, IBM Watson, etc.) and most of the software for implementing machine learning is free or open-source, there is still a need to fine-tune the models to deliver the required outcomes. This means that there is a need to internally upskill, or partner with an external agency, to ensure the machine learning is applied correctly to the problem. While machine learning is sold as a 'black-box' service, the reality is that some in-house understanding, or access to this, is required to implement it.

The application of machine learning via the main providers is relatively inexpensive but does typically involve the storage and processing of large numbers of images. The expectation is that the machine learning will lead to efficiencies in existing manual methods and so may lead to cost savings. Applications of machine learning by the Australian Institute of Marine Science, for example, have reduced the time for manual image analysis by 30%, and this is expected to increase as model performance increases. Machine learning should therefore deliver a net cost saving.

Machine learning is not a magic solution to scaling data analysis. The methods work as well, in general, as an inexperienced person and only with a lot of work and test images can it get close to what an expert can achieve. Typically, machine learning can correctly identify an object in an image given enough test images to a confidence of 85-90%. For many applications this level of data confidence is enough, but in others, such as compliance, this may not be enough. In these situations, supplementary human checking can be used with the automated analysis to deliver a more accurate result.

Box 8: Application of machine learning to ecological monitoring – detection of Koala's

Corcoran et al. (2019) used machine learning techniques to automate the identification of Koala (*Phascolarctos cinereus*) heat signatures in tree canopies using aerial thermal imagery. The authors combined the outputs of two well-established neural networks (Faster RCNN & YOLO). The networks achieved a greater overall probability of detection (78–100% per mission) than both ground-based observation studies (60–75%) and manual detection using thermal imagery (50–72% per mission).

This approach to monitoring animal populations can be applied to other low-density, cryptic mammals that are sensitive to human disturbances or located in complex, hard to reach environments.

The image to the right displays examples of machine-generated bounding boxes identifying heat signatures for (A) Koala, (B) Kangaroo, (C) Car and (D) Human.

Image from Corcoran et al. (2019), licensed under CC-BY 4.0
http://creativecommons.org/licenses/by/4.0/

The main issues with adoption are two-fold. The first is changing existing manual analysis procedures to semi or fully automated ones, and the subsequent implications for the quality of the data and the resources currently allocated. The second is the degree to which custodians, clients and stakeholders will trust, use and value data that is generated in full, or in part, by computer models versus human-generated data. How the users of the data deal with computer analysis versus human analysis will vary from domain to domain, but the potential impact of this should not be understated.

Technological Development

At a superficial level, a basic machine learning system can be set up with a set of training images and a provider such as Microsoft or Google. With a few thousand images this will generally give a result in the 70-75% accuracy range, depending on the problem. Achieving better accuracy will require tuning the models for the particular problem, which needs either in-house expertise or partner expertise.

A suggested development methodology is to do a proof of concept with either a generic provider or a technical partner, such as a university or AI provider. Once the level of accuracy is understood and any issues identified, then it is recommended that a close partnership with an AI provider be developed. This partnership will further develop the model and then develop software and systems to implement it as a workflow in close consultation with the custodians of that workflow. Machine learning is a rapidly changing area, and so it is recommended that a continued relationship to domain expertise be developed either via in-housing of resources or via those of an external partner.

Key Experts

Key experts in machine learning in Australia include the Australian Institute of Machine Learning (AIML) and CSIRO’s Data61.

Visual Question Answering and Visual Language Navigation

What is Visual Question Answering and Visual Language Navigation?

Visual Question Answering (VQA) and Vision-Language Navigation (VLN) are 2 kinds of artificial intelligence techniques that allow humans to communicate with and command robots/machines in natural language. VQA enables a computer/robot to answer a natural language question regarding a given image, e.g., "How many people are there in the image?" and "Is there a red car in the image?". The questions can be unique questions that the computer has not been trained on. The answer is usually presented in the form of a few words, a short phrase, such as "4 people" and "Yes". From the data perspective, VQA methods fall into image-based and video-based methods. As the term suggests, image-based VQA handles questions regarding an image and video-based VQA handles questions regarding a video. From the question type view, VQA methods can handle (1) general questions without providing candidate answers, such as "how many …"; (2) knowledge questions that can only be answered using common sense, such as "how many *girls* are there in the image?" (3) embodied questions that require a robot to explore an environment in order to give the answer.

VLN techniques equip robots or embodied devices (e.g., drones) to navigate to a target location according to human instructions. Existing VLN algorithms can handle 4 types of instructions:

- Detailed indoor instructions, such as "Go along the hallway and turn left at the round table, then stop in front of the TV".
- High-level indoor instructions, such as "Bring me the blue cushion in the living room".
- Dialogue instructions that allow embodied devices to ask questions when they are uncertain about decisions.
- Detailed outdoor instructions, such as "Go with the flow of traffic, at the first traffic light turn left …", using fine-grained reference objects rather than an exact address input as required by GPS.

Overall, these cutting-edge techniques can play important roles in our intelligent daily life.

VQA is a more flexible approach to using machine learning to analyse images and related data compared to traditional machine learning approaches that require all of the questions to be added up front for training. The VQA approach gives managers the ability to ask the questions they thought of before training a computer vision system, but also enables additional questions to be answered without additional training. This technology would give the MDBA a highly flexible 'situational awareness' tool to quickly ask questions about large datasets without the need for specialist input or re-training machine learning systems (Figure 8).

VLN reduces the need to train humans in how to manually control drones, and enables drones to be able to operate in dangerous situations such as bad weather.

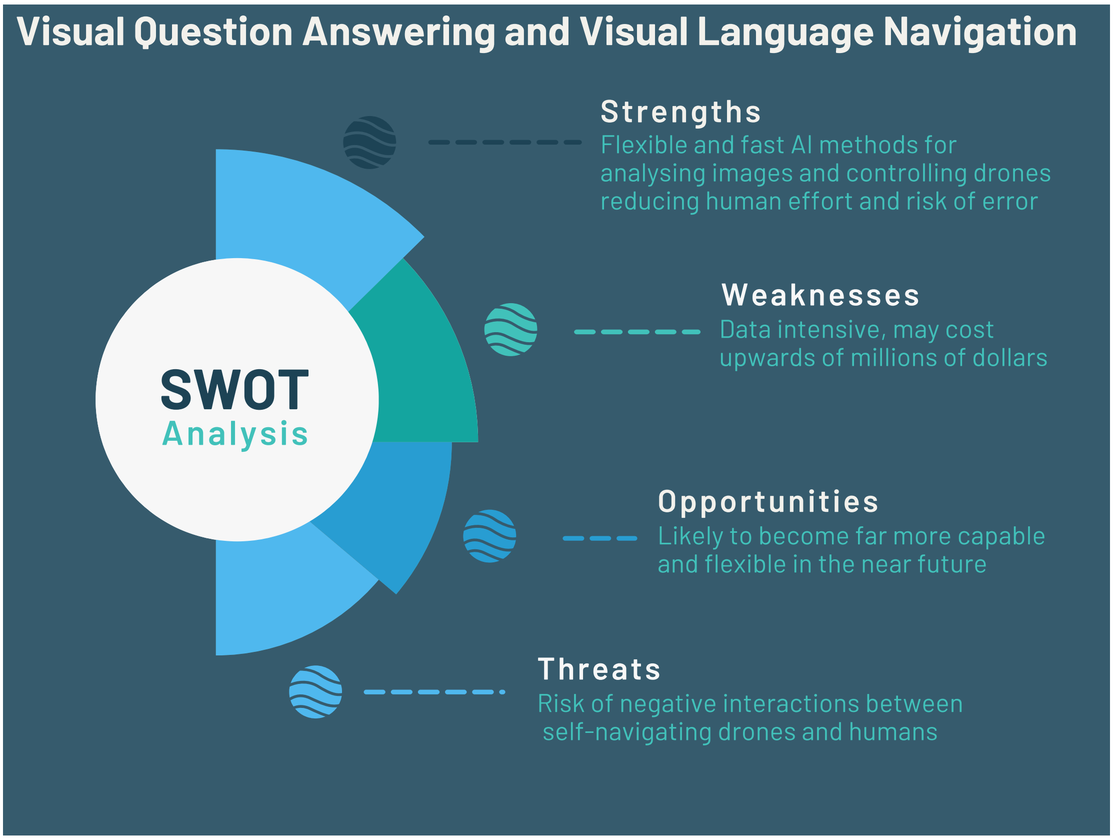


Figure 8 *Summary of the strengths, weaknesses, opportunities and threats of Visual Question Answering and Visual Language Navigation.*

Current Applications

VQA has direct applications to assist visually-impaired users of media, and to improve image retrieval systems (Box 5). VLN has recently been proposed as a method for new and improved approaches to automatic wayfinding in cities.

Box 9: Application of Visual Question Answering and Visual Language Navigation – land change in aerial imagery

Vision and language methods research brings together technologies that understand human spoken and written languages, also referred to as natural language and vision, in the form of images or video. The ultimate goal of this research is to be able to pose questions to a digital device about visual information and receive meaningful and useful answers.

For example, Yuan et al. (2021) applied an integrated change detection and visual question and answering (CDVQA) technique on multi-temporal aerial images. The dataset used contains 122,000 question-answer pairs and 2,968 pairs of multi-temporal images that have had land-cover classified at the pixel level. Examples of the possible question-answer pairs include:

- Q: Have the areas of water decreased? A: No.
- Q: What is the change ratio of low-vegetation in the pre-event image? A: 40–50%.

Image from Yuan et al. (2021), licensed under CC-BY 4.0 http://creativecommons.org/licenses/by/4.0/

For years, change detection using remote sensing data has been an important tool in monitoring Earth's changing surface, though these specialised techniques have typically been limited to experts. By incorporating visual question and answering techniques, the outputs of change detection methods become considerably more accessible for non-specialists.

Potential Applications for the Murray-Darling Basin

VQA can be used for visually and flexibly interrogating databases of images (camera traps, satellite or drone images, photopoints), without needing to finalise all the questions in advance. The sorts of questions that can be asked of images or image libraries include:

- "show me all of the images with birds in them",
- "label each of these images with how many pelicans there are in them"
- "estimate the proportion of wetlands covered with growing vegetation"
- "identify all of the dams in this image/image library"
- "would we expect a fish spawning event in the Lower Murray in the next 3 months?"
- "list all of the wetland birds that can be identified in this image library"

As well as images, other related datasets can be included in the training, which increases the capability of a VQA system – allowing questions to be asked that require both images and other explicit knowledge.

VLN can be used to direct drones to collect images via voice control, for example:

- "fly along the river and take 10 photos every second"
- "fly over that floodplain and record video of the whole area"

Costs and Limitations

For a VQA system, the main costs are collecting the data (satellite images, drone images, camera traps) and developing a good set of training questions. Assuming this is in place, it may cost in the order of $100,000s to develop a prototype solution. There is a cost-benefit trade-off in collecting more images and training the system to become more accurate which cannot be answered here. The most a system might cost to be able to accurately answer management questions about the ecology of the Murray–Darling Basin (without image acquisition costs) would be millions of dollars and no more than $5M. Once established, the only additional costs for a VQA system would be to collect more data and retain the system to become more accurate. These costs cannot be estimated in advance and could range between $10,000s and millions of dollars, depending largely on data acquisition costs. There are no significant risks other than the degree of accuracy.

For VLN, the costs of developing relatively simple visual navigation tools will be in the order of $100,000s. The ongoing costs relate to the drone itself and taking a drone out into the MDB environment. The risks of this technology include interactions with humans, wildlife, infrastructure and livestock when allowing drones to self-navigate.

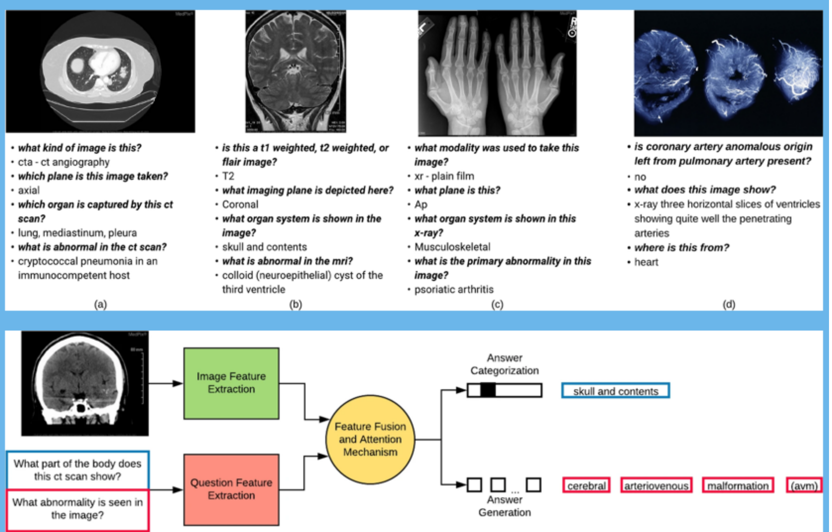
Box 10: Application of Visual Question Answering and Visual Language Navigation – medical imagery

Medical images are challenging to interpret for non-specialists, and experts in any field are prone to human error. Visual question and answering (VQA) allows experts to rapidly gain a 'second opinion' to reduce human error and provide confidence for diagnoses. Sharma, Purushotham and Reddy (2021) developed MedFuseNet, a deep learning model for VQA tasks tailored for medical use.

MedFuseNet aims to maximise the learning while minimising complexity by breaking the task into simpler tasks and predicting the answer. The model is comprised of four main components: (1) image feature extraction, (2) question feature extraction, (3) feature fusion, and (4) answer prediction. Examples of MedFuseNet's real-world question-answer pairs include:

- Q: What organ system is shown in the image? A: Skull and contents.
- Q: What is abnormal in the MRI? A: Colloid (neuroepithelial) cyst of the third ventricle.

The authors assessed the performance of the VQA model by conducting several experiments on real-world medical datasets and comparing the performance of MedFuseNet to five other state-of-the-art and popular VQA approaches. The experiments show that MedFuseNet generally outperforms all other state-of-the-art VQA approaches and showcases the interpretability of the model's predictions.



Technological Development

This is a fast-moving area of technical development, and VQA and VLN are likely to become orders of magnitude more flexible and capable over the next 5 years.

Key Experts

There are several prestigious experts in the related techniques within Australia. Specifically, the experts include: Qi Wu (The University of Adelaide), Stephen Gould (Australian National University), Yuankai Qi (The University of Adelaide) and Xiaojun Chang (RMIT University).

Data Efficient Learning for Vision and Audio Recognition

What is Data Efficient Learning for Vision and Audio Recognition?

Vision and audio recognition aim to extract meaningful information from images, videos, and audio. For example, vision recognition can be used to identify objects at the image level (i.e., image classification), at the bounding box level (i.e., object detection), or at the pixel level (i.e., segmenting objects). Similarly, we could also use the audio signal as the input and output information like the species of a particular animal. Although the applications and purposes of various vision and audio recognition models are very different, they follow a general pipeline called supervised learning. In standard supervised learning, a model is "trained" from a set of training data that comprises input data (e.g., images, audio recordings) and the ground-truth output (e.g., class label of the images or bounding box of objects). Usually, a large amount of training data, ranging from a few thousand to a million samples, are needed to build a recognition model with reasonable performance. However, this requirement is not always affordable, and for some applications collecting annotated data can be quite expensive.

Data efficient learning provides a potential solution to the above issue, and to similar limitations on systems training identified in the previous section on VQA and LVN. The concept of data efficient learning encompasses several techniques in machine learning. Among them, transfer learning, semi-supervised learning, and unsupervised representation learning are the most promising for environmental management. With transfer learning techniques, it is possible to use the rich annotation data from a related problem to help build a machine learning model for the target problem, reducing the number of annotated target images required to fine-tune the resulting model (Box 6). For example, a publicly available image database of one bird species could be used to build a base model for a related species, thus requiring only hundreds of images of the target species to build an effective model.

Semi-supervised learning is based on the fact that the unannotated samples, such as images, video, or audio recordings, are relatively easy to obtain, but their annotations are expensive to collect. Thus, semi-supervised learning will use a few annotated samples and a large number of unannotated samples to jointly teach or develop a machine learning model. The state-of-the-art semi-supervised learning methods show that using only 10% of the annotated training samples can achieve similar performance as using 100% annotated training samples in supervised learning. Unsupervised learning tries to learn the main features from the unannotated data without knowing the required features. The learned representation can then be applied to a wide range of tasks, called downstream tasks. With a good feature representation, the need for many annotations is reduced. A summary of those 3 paradigms is presented in Table 3.

Table 3 *Summary of data efficient machine learning paradigms.*

Method	**Learned representation requirements**	**Downstream task requirements**
Transfer learning	Annotated training data from other applications, called source data	A few annotated target problem data A lot of source data samples
Semi-supervised learning	Unlabelled target problem data	A few annotated target problem data A lot of unannotated target problem data
Unsupervised learning	Unlabelled target problem data	A few annotated target problem data A lot of unannotated data (not necessarily from the target problem)

The benefit of using data efficient machine learning is to have a more efficient way of leveraging the data (Figure 9). It can significantly reduce the number of annotations needed for building a machine learning model for environmental monitoring. More specifically,

- It can reduce the cost of collecting annotated data.
- It can accelerate the development of a prototype system since data annotation can be time-consuming.
- Given the existing data annotation, data-efficient learning can potentially boost the performance of the current model by further leveraging other data sources or unlabelled data.

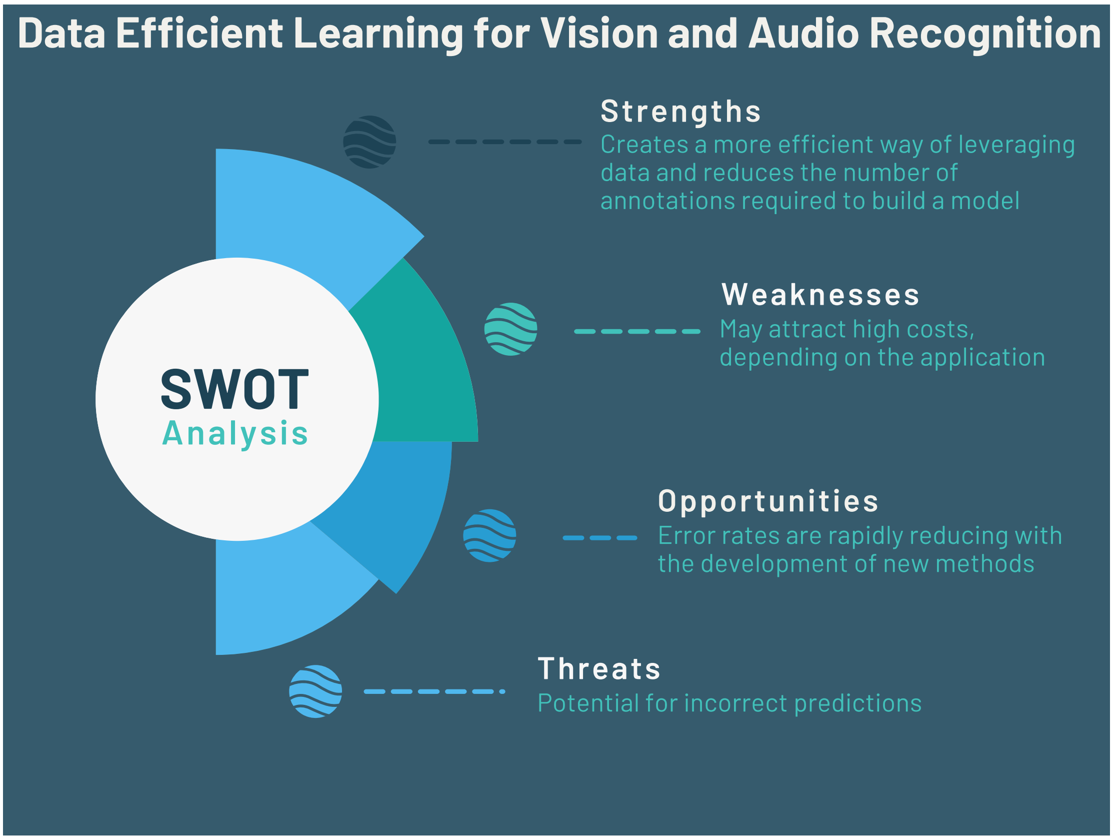


Figure 9 *Summary of the strengths, weaknesses, opportunities and threats of Data Efficient Learning for Vision and Audio Recognition.*

Current Applications

An example of the usage of data-efficient learning is the application of semi-supervised learning to conduct crowd counting using images or videos. Additionally, unsupervised machine learning was used to classify a pelican's movement as either migrating or foraging without the use of labelled training data (Wang 2019).

Potential Applications for the Murray-Darling Basin

Vision and audio data-efficient learning can be used as an alternative strategy for building environmental monitoring machine learning systems to the traditional supervised learning strategy. Based on the existing literature, the data-efficient learning approach has been successfully applied to problems like image classification, object detection, image segmentation, object counting, and audio recognition. Here follows a list of examples relevant to environmental monitoring:

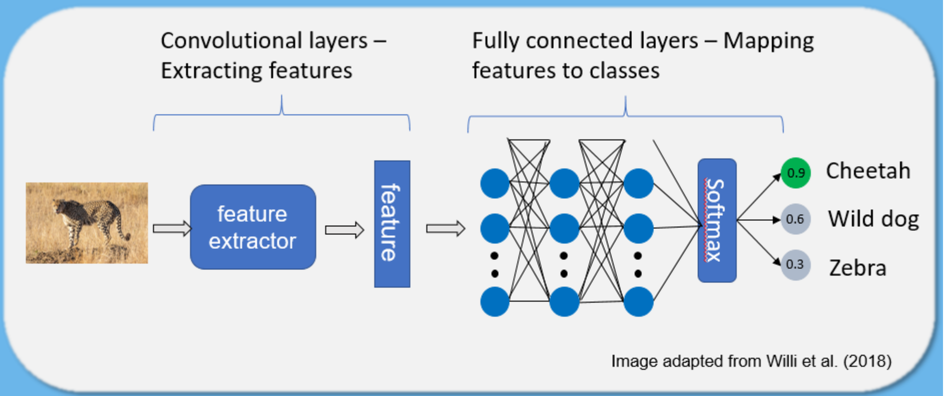
- A system that can identify birds/fish species from audio recordings
- A system that can identify birds/fish species from images from a camera trap (Box 6)
- A system that can use satellite images to identify the region of interest

- A system that can search images with similar content for animals/persons/objects with the same identity

The use of data efficient learning is an important tool in reducing the cost and resources required for the application of Machine Learning and so can be seen as a supporting technology for many of the other proposed analysis methods.

Box 11: Application of Data Efficient Learning for Vision and Audio Recognition – transfer learning for training computer vision

Willi et al. (2018) used a transfer-learning approach to leverage a neural network trained on a large dataset of camera trap images containing many species and applied it to a different, much smaller dataset. To achieve this, the weights in the convolutional layers (responsible for extracting features) were copied from the larger model and preserved. Next, the fully connected layers (responsible for mapping features to classes) were retrained on the new, smaller dataset.
The resulting image classifier was able to discern what species was pictured in an unlabeled image with an accuracy of 85.8% after being trained on 17,671 labelled images. In contrast, a model developed without transfer learning achieved an accuracy of 75.5% after being trained with the same data.



Transfer learning is a powerful tool for training machine learning algorithms, particularly when training datasets are limited. This technique can extend the monitoring capabilities of camera traps, audio moths, satellite imagery and other visual or auditory sensors by reducing the time, resources and human labour needed to train an accurate model for recognition and classification algorithms.

Costs and Limitations

The cost of developing data efficient learning prototypes for one or several related applications is dependent on the problem, but is likely to cost $10,000s to $100,000s depending on the complexity, underpinning data and required accuracy. The major costs may come in collecting data to train the model, and this cannot be estimated without knowing what data already exists and how costly it is to collect additional data to train a required task. If the analysis is offline, then a standard GPU-powered server will be needed to execute the algorithm for analysing the video, image and audio captured from various sensory data sources. The cost of a standard GPU-powered server will be less than $10,000. The only risk is the potential problem of an incorrect prediction.

Box 12: Application of Data Efficient Learning for Vision and Audio Recognition – semi-supervised learning for monitoring seagrass habitat

Semi-supervised learning aims to address the data-hungry nature of training machine learning models for vision and audio recognition. Semi-supervised learning leverages a 'few' annotated samples and many unannotated samples to jointly teach a machine learning model.

For example, Hobley et al. (2021) used the Teacher-Student method to train a semi-supervised model capable of classifying and segmenting aerial images of seagrass habitats. This process involves a 'teacher' model that is pre-trained with annotated images that guides the 'student' model to make the same predictions when given unannotated images.
The average accuracy of the semi-supervised model was greater than the supervised model. When classifying pixels from RGB imagery, the semi-supervised and supervised models had an average pixel accuracy of 83.3% and 76.79%, respectively. The semi-supervised model also had greater accuracy in classifying pixels from multispectral imagery than the supervised model, averaging 88.44% and 85.27%, respectively.

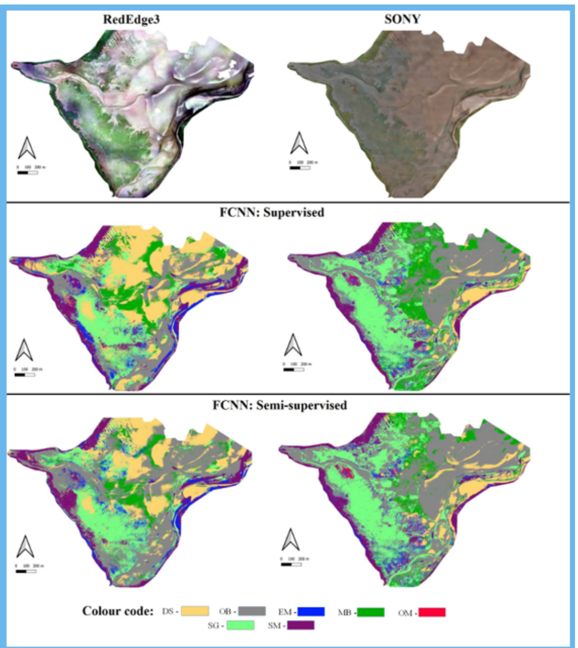


Image adapted from Hobley et al. (2021). Image shows the results image segmentation by supervised and semi-supervised models. Images in the left column are RedEdge3 multispectral images and images in the right column are SONY RGB images).

Technological Development

Data efficient learning is a fast-developing area in machine learning, computer vision, and audio-processing. For example, in 2019, the prediction error rate of semi-supervised learning algorithms on a standard benchmark was almost halved with the development of new methods in that year. We are expecting to see similarly rapid development in the coming years.

Key Experts

There are several prestigious experts in the related techniques. Specifically, the experts include: Lingqiao Liu (The University of Adelaide), Piotr Koniusz (Data 61), and Lei Wang (The University of Wollongong).

Semantic Change Detection in Images

What is Semantic Change Detection in Images?

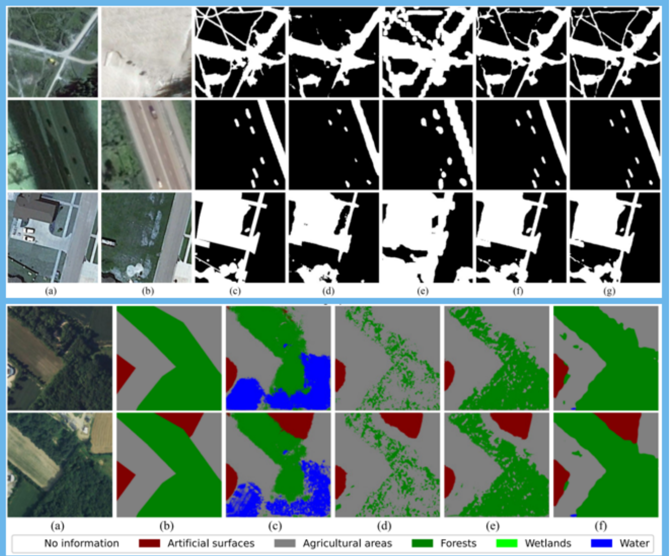
Semantic change detection in images involves comparing images of a scene or object over time to look for changes of interest while ignoring incidental changes. A typical example would be comparing 2 satellite images of the same location but from different times. Appearance changes due to variations in lighting, pose and viewpoint should generally be ignored, whereas significant changes such as new/missing objects or changes in object appearance should be detected. It is common for the pixel brightness difference of incidental changes to be much larger than for significant changes. For example, brightness changes due to a shadow cast on a face are much more significant numerically than changes due to ageing or expression of the face. In early approaches, images were aligned and then pixel brightness was compared either directly or via hand-crafted image features. Modern machine learning and deep learning improve on this by automatically learning features that are invariant to incidental changes and sensitive to changes of interest.

Generally, when comparing 2 or more images for change, they need to be aligned at the pixel level. There are standard methods for this in image processing, computer vision and remote sensing. These methods typically use hierarchical warping and optical flow estimation to minimise photometric error, or a mutual information criterion in the case of imagery from different kinds of sensors. The temporal spacing and frequency of images impacts on results. For example, if there is a long time gap between images it will be challenging to detect small-scale, incremental patterns of change. Conversely, comparing many images over the same period will show smaller changes between image pairs and may make it more difficult to discern long-term changes. To avoid such collection biases, images should be collected at the same locations at regular intervals.

Box 13: Application of Semantic Change Detection In Images – very high resolution remote sensing images

Using a pair of bi-temporal aerial images, semantic change detection is able to identify pixel-level transitions of land-cover between image pairs. With the continual advancement in remote-sensing sensors and model optimisations, research in semantic change detection for environmental monitoring, urban planning and damage assessment has been growing steadily.

Zhao et al. (2022) proposed a novel end-to-end network for state-of-the-art semantic change detection called spatially and semantically enhanced Siamese network (SSESN).
To compare the performance of SSESN, its performance in binary change detection (top image) and semantic change detection (bottom image) was compared to several alternative state-of-the-art change detection methods. SSESN outperformed all other approaches for both tasks and achieved an overall accuracy of 0.919 and 0.890, depending on the dataset used. Future work may include the fusion of spectral data to develop a hyperspectral semantic change detection model.



Images from Zhao et al. (2022). Top: Visualisation results of binary change detection using the CDD dataset. (a) and (b) the original bi-temporal image pair, (c) ground truth, (d-f) result from three alternative models, (g) result from the authors novel model (SSESN).
Bottom: Visualisation results of semantic change detection using the HRSCD dataset. (a) original bi-temporal image pair, (b) ground truth, (c-e) result from three alternative models, (f) results from the authors novel model (SSESN).

The definition of what constitutes a significant change is task-dependent. For instance, if the task is to look for new man-made structures in satellite images, then vegetation change would be largely irrelevant and should be ignored. However, in a study of vegetation health, these are exactly the changes that should be detected. Consequently, generic change detectors are often useless, giving a high false detection rate. There are 3 main approaches to semantic change detection:

1. Object-based: image object detectors are applied to both images, and the detections are compared between images. This is useful for well-defined compact objects, but not for regions such as fields in satellite images.
2. Segmentation-based: each image is semantically segmented (also referred to as pixel-based classification) so that every pixel is given a discrete label (e.g., grass, building, tree). Then the labels in the 2 images are compared for differences, and the different pixels are grouped into change regions. While this is more general than object-based methods, it requires fine alignment of the images, and is limited in that each image is first labelled separately.
3. Supervised learning: this would require a dataset of labelled changes of interest, used to train a deep neural network to detect such changes given the 2 input images (see Box 7). This is the most general case; it does not require fine alignment of images and allows the algorithm to exploit information from both images when detecting the change. It does, however, require a very large labelled dataset.

Related technologies include:

- Anomaly detection: detect unusual objects or regions in images. The machine learning model learns what is "normal" from data, and uses this as a comparison for abnormal occurrences.
- Background modelling in video: a video stream is processed to model the background and detect "foreground" objects. This is a kind of change detection in video over different time scales.

Change detection can also be applied to other kinds of data, such as acoustic signals.

The 2 main benefits of this technology are automation and scalability of environmental monitoring for change detection (Figure 10). Machine learning can automate laborious repetitive tasks that are currently performed by humans at great cost. It can also be applied at scale, so that greater area and more time instants can be handled by the automated algorithms than the capacity of the human analysts. Questions can be asked of the data at a space-time scale that is really beyond the ability of human analysts.

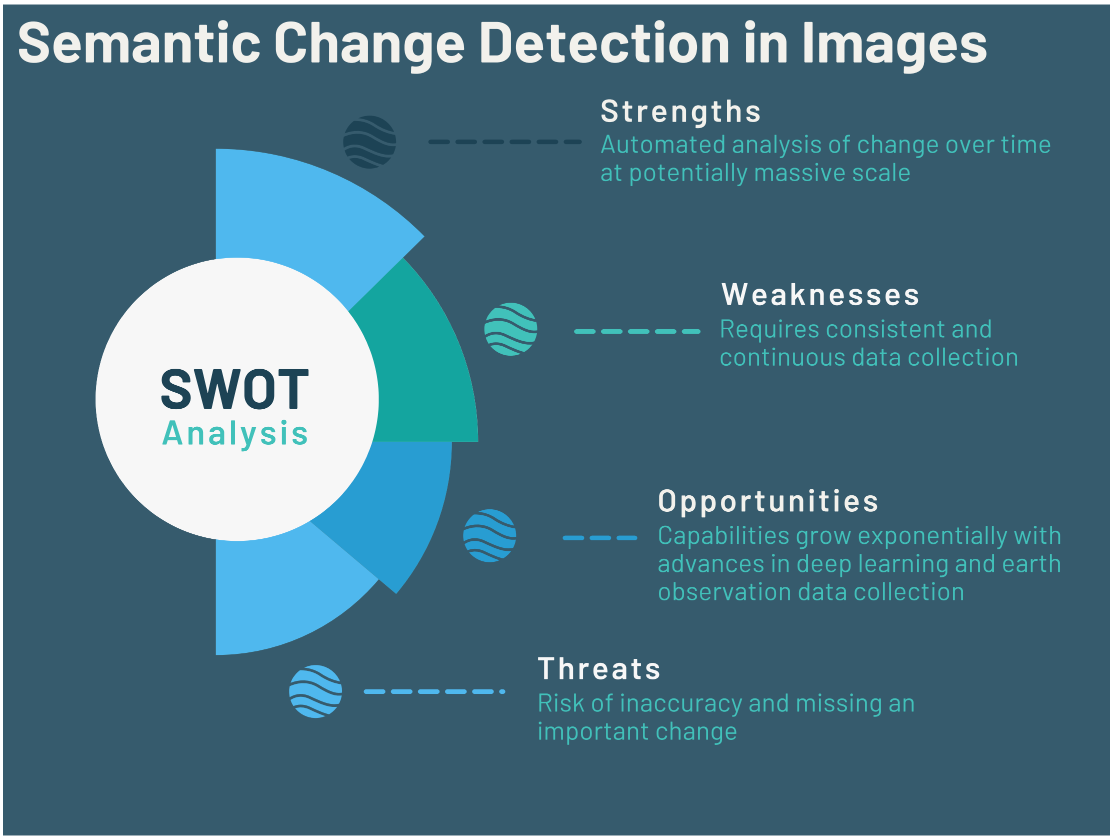


Figure 10 *Summary of the strengths, weaknesses, opportunities and threats of Semantic Change Detection in Images.*

Current Applications

An example of current applications of semantic change detection in images is the Xview2 challenge. Established by the US Defence Innovation Unit, the Xview2 challenge was designed to detect damage to buildings due to floods. The dataset consists of registered satellite image pairs before and after flooding.

Potential Applications for the Murray-Darling Basin

The analysis of changes over time is arguably the most important task in environmental monitoring. Overhead imagery collected from satellites or aircrafts can be stored and processed to analyse a variety of changes. Change detection or anomaly detection can also be applied to ground-based imagery or underwater imagery. Change detection in audio data could also be useful. For example, if populations of insects are increasing or decreasing, this would change the average signal intensity for their sounds. Below are some applications for overhead imagery:

Object-based change detection:

- Count populations of animals and to study change over time. Wetland waterbirds, flood management structures on floodplains and dams can be measured using aerial imagery. A similar technique is used by the company Orbital Insight which counts objects such as cars and oil tanks from satellite images and provides high-level reports of trends over time.
- Detect recreational vehicles on rivers in satellite images to monitor recreational use of the river systems.
- Detect arrival and departure of animals from ground-based cameras.

Segmentation-based change detection:

- Delineate and measure wetland flooding extent.
- Delineate and measure vegetation responses to flooding.
- Delineate and measure regions of tree death.
- Measure growth rates of vegetation over time.

Supervised learning:

- Intensive monitoring of an area such as a wetland for changes (e.g., Box 7). This would involve iteratively creating a labelled dataset marking false alarms to improve future performance. It can be used for discovery and anomaly detection by looking for any kind of change, including changes that couldn't have been anticipated by the user.

- Detection and delineation of areas affected by water management, such as floodplains and irrigation areas.

Change detection could thus be applied in a large variety of ways, and at a massive scale (Box 7). For example, the entire Murray–Darling Basin could be monitored regularly for changes by flying regular image collection missions over the same area. Regularly collecting the imagery for an extended period would create a highly valuable database that can be used to answer a variety of questions both now and for decades to come.

Box 14: Application of Semantic Change Detection In Images – Monitoring land change in a coastal marsh

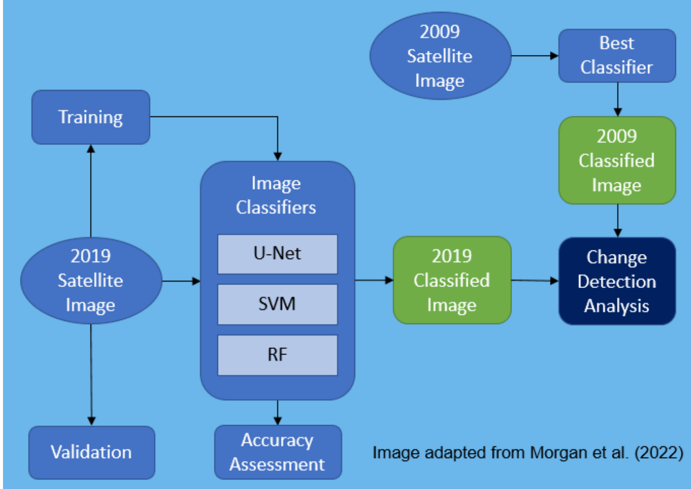


Image adapted from Morgan et al. (2022)

Morgan et al. (2022) trained deep learning algorithms to classify a coastal marsh's land cover (marsh, urban, water, etc.) from satellite imagery taken in 2019. The accuracies of three image classifiers (U-Net, SVM & RF) were compared, and the best performing classifier was used on satellite imagery of the same marsh taken in 2009. Change detection analysis was carried out on the two classified images to determine land change over the 10-year period.

U-Net was found to be the best performing classifier for this data, achieving an overall accuracy of 85.3% and 92.4% for the 2009 and 2019 images, respectively. In addition to monitoring the direct modification of marsh environments, this technique could be used to detect an upstream land change, such as pollution and development. Similar approaches could be applied to rapidly detect changes in floodplain inundation, tree cover and tree cover. Data efficient learning techniques such as transfer learning are recommended to overcome the limited labelled training data for these tasks.

Costs and Limitations

A simple prototype or study would cost in the $100,000s. An operational system has software development, maintenance, and data acquisition costs. Software development costs for a large-scale system would be in the order of a few million dollars, and ongoing cloud costs, maintenance and improvement could cost $50,000 to $500,000 per year, depending on the scale. Data costs depend on the application – the cost and frequency of airborne or satellite collections, for example, could be free or cost $100,000s or millions of dollars depending on the chosen platform and the number of data collection events and scale required to answer questions. The main risks involve the accuracy of the system and the cost of missing an important change.

Technological Development

This is a technology involving large amounts of data, including earth observation data, and advances in deep learning technology. The capability and scalability of semantic change detection is expected to grow at the same exponential rate as both of these factors.

Key Experts

Experts on semantic change detection in images in Australia include Jamie Sherrah (University of Adelaide), Anton van den Hengel (University of Adelaide/Amazon), Dong Gong (University of Adelaide), and Peter Kovesi (University of Western Australia).

Category 3: Models

An Introduction to Models

Previous sections of this report have addressed the automated collection of big data using remote sensing, and the rapid analysis of these large and complex datasets using artificial intelligence and machine learning. Recent technological developments have extended the capabilities of environmental managers and researchers even further, however; the application of models now enables the prediction of complex environmental interactions and patterns using trained artificial intelligence platforms. Managers can populate these models with existing data and a set of simple rules to forecast ecological responses to management strategies, environmental change or climatic events. This enables decision-makers to make informed choices when designing and implementing management programs.

These models are advancing in sophistication and capability at a rapid rate, and are likely to become orders of magnitude more accessible in the near future. It is an ideal time for the MDBA to incorporate modelling into its management strategies and prepare for future improvements in the area. Modelling technology will soon allow managers to input data directly into models for real-time forecasting, enhancing understanding of the complex interactions between wildlife, natural resources and the environment and refining management strategies.

With the uptake of big data, analysis and modelling in an organisation operating at a landscape scale such as the MDBA, it is of high importance that efficient data storage and sharing platforms are established and that members of the workforce use common data language and management practices.

Sequential Decision-Making and Reinforcement Learning

What is Sequential Decision-Making and Reinforcement Learning?

Sequential decision problems are those that try to anticipate the compounding future results of decisions made over a period of time. Predicting the outcome of sequential decisions requires modelling the delayed consequences of policies that are made today. These problems are particularly challenging to model accurately. Consider, for example, setting a policy for both braking and accelerating in an autonomous car. Policies developed now must anticipate how external independent events may interact with the car in the near future (i.e., the pedestrian jumping in the street, a car braking ahead, changing traffic lights). The car's control mechanism needs to consider a trade-off between throttling and braking to avoid undesired consequences. When the number of variables is large and the environment is complex, developing policies that can predict all possible outcomes to achieve an optimal outcome becomes impractical for humans.

An emerging and powerful tool for solving complex sequential decision problems is reinforcement learning. The use of this tool requires reasonably realistic simulators of the environment in which the policies must operate. Reinforcement learning requires a deep learning agent to 'play' the simulator multiple times and learn by example which mix of policies delivers the best outcome. In doing so, it can then 'play' the simulator optimally, demonstrating how to 'win' in a particular scenario, or it can print a series of 'policies' on how to operate under various circumstances and scenarios to optimise outcomes (see Box 8).

The benefit of using these artificial intelligence (AI) technologies is that it allows us to automate the processes that are otherwise laborious, time-consuming and, in many cases, require human modelling that could be prone to mistakes and only can handle a few variables at the time (Figure 11). Using AI to tackle such problems is one of the most promising ways forward that Australia could significantly benefit from being one of the pioneers.

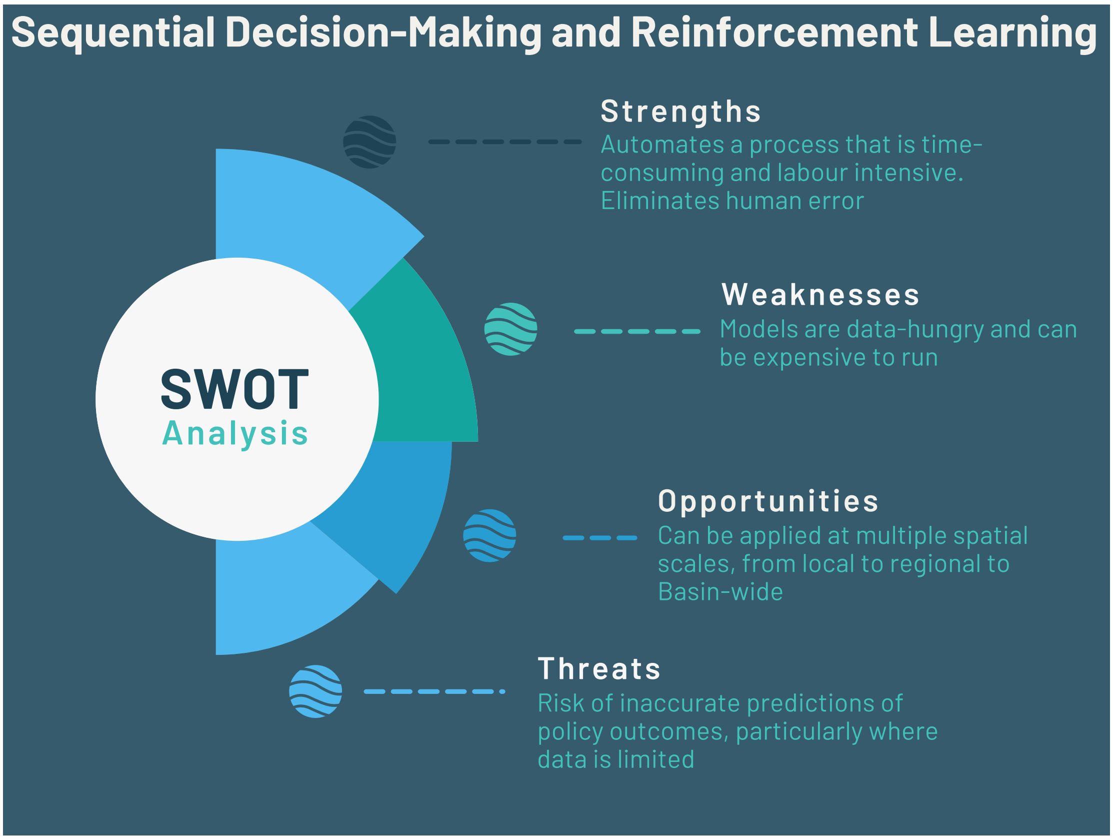


Figure 11 *Summary of the strengths, weaknesses, opportunities and threats of Sequential Decision-Making and Reinforcement Learning.*

Current Applications

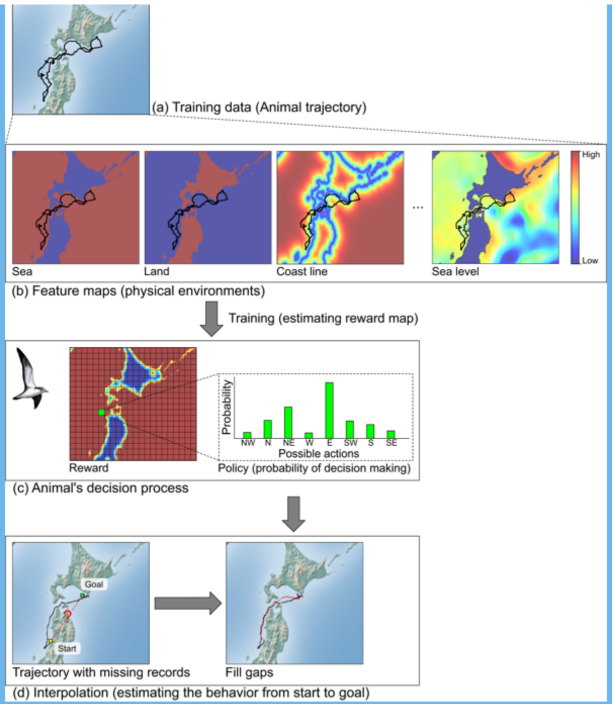
Some examples of how reinforcement learning can outperform humans in developing policies for managing complex environments include:

- 'Go' is a strategy game similar to chess. In 2015 AlphaGo (a model) defeated the human Go champion, demonstrating that sequential decision-making models in general, and deep reinforcement learning in particular, are capable of outperforming humans in highly complex, strategic tasks that require an anticipation of the far future
- Reinforcement learning is employed for weather prediction and has dramatically improved short term rain prediction
- Bushfire prediction (see Box 8)
- Autonomous driving and robotic navigation
- Dialogue between machine and humans
- Sports strategy and coaching

Box 15: *Application of sequential decision-making and reinforcement learning – filling gaps in animal tracking data*

Instruments that monitor the movements of animals (e.g. GPS, VHF triangulation, Argos satellite tracking) can provide crucial information on animal behaviour such as habitat selection, foraging strategies, predator avoidance and dispersion.
However, these tracking techniques sometimes experience data gaps due to the animal losing a direct line of sight with a satellite or receiver. Additionally, GPS systems often sample coarsely at even intervals to conserve battery. The limitations of these technologies may result in data gaps ranging from seconds to hours. A straight line between data points has been one of the most common methods of path interpolation, but this underestimates the tortuosity and length of real-world paths.

Hirakawa et al. (2018) trained an inverse reinforcement learning (IRL) model to interpolate the movement of streaked shearwaters (*Calonectris leucomelas*) in response to the physical environment (e.g. sea, land). Compared to linear interpolation, the resulting model more accurately predicted the flying behaviour of *C. leucomelas*, such as avoiding flying over land. Ogawa et al. (2019) used a similar IRL approach to achieve even greater accuracy that is less computationally costly. In addition to filling data gaps, these techniques could extend the capability of GPS systems by allowing for more coarse sampling intervals, thereby extending battery life.
To further improve interpolation accuracies, additional environmental features known to be associated with animal behaviour could be incorporated into feature maps. For example, sea surface temperature and chlorophyll-a influence seabird foraging behaviour and their inclusion may further improve model accuracy.



Potential Applications for the Murray-Darling Basin

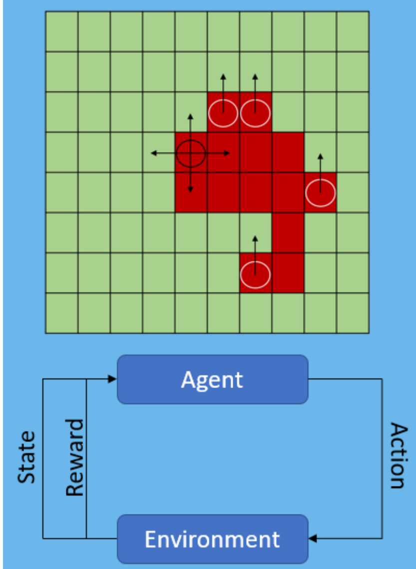
Environmental management is a classic sequential decision-making problem. It is challenging because it requires decisions to be made about complex landscapes, where the outcome may not be able to be measured for many years or decades. Reinforcement learning can support decision-makers in understanding what policies will lead to optimal outcomes and, therefore, what monitoring may be needed. Reinforcement learning can be used to predict the outcome of individual threats or policies across multi-decades on a wide range of measures. Consider the following:

- **Impact of environmental watering policies**: Using existing models of water flow in the MDBA, reinforcement learning could be used to propose environmental watering policies that would optimise for a range of agreed outcomes. This would require a relatively accurate understanding of how watering river channels and floodplains affects biological organisms. Such a system could generate heat maps of where the best and worst impacts of particular policies would be. The technology could be applied for an individual or many species, and to a single site, a region, or to the Basin as a whole.
- **Short-term and long-term precipitation forecasting**: It is important to be able to accurately predict the weather as well as rain conditions. Using deep sequential decision-making approaches, the prediction accuracy could be significantly improved. This leads to more accurate predictions

that lead to better policies (see https://tinyurl.com/y6q3mqgn for an example of using these technologies for precipitation prediction).

- **Animal migration modelling:** Using the existing patterns and AI technologies, we can predict where certain species are more likely to move to. It could also assist in predicting in which environments certain species can thrive.

Box 16: Application of sequential decision-making and reinforcement learning – Predicting forest fires



Images adapted from Jindal et al. (2020) & Hooshyar et al. (2020)

Reinforcement learning (RL) is a machine learning technique that computes the optimal behaviour of an 'agent' to reach a specific goal in reaction to its 'environment'. With this technique, Jindal et al. (2020) predicted the spread dynamics of a forest fire by positively rewarding the agent (the fire) for spreading to a likely tile (according to land-cover type, wind speed and direction, etc.) and negatively rewarding the agent for spreading to an unlikely tile, such as water. This cell-based fire spread policy can be applied to satellite imagery and used to predict the future spread of fire faster and more accurately (up to 82% accuracy) than previous methods. These predictions could assist with fire management and identify animal populations at greater risk of fire.

Additional uses for reinforcement learning include creating optimised sampling and patrolling schemes or the optimal creation and operation of dams.

Costs and Limitations

Depending on the scale and objectives of employing the AI technologies discussed here, the models could cost upwards of $100,000. For some applications, if realistic simulators are needed to be designed, that could add a substantial additional cost.

One risk of this technology is that the model may inaccurately predict the outcomes of policies. The more inaccurate the models and the more limited the data upon which to train, the greater this risk is. There is also a risk that the results of the modelling, if seen as counter-intuitive by an interest group, will be dismissed. These same risks occur for human-generated policies. More investment in better data and the incorporation of these models into new applications could lead to innovations and new insights.

Technological Development

This is a fast-moving area of technical development, and it is likely to become orders of magnitude more capable over the next few years.

Key Experts

Key experts in this field in Australia include Ehsan Abbasnejad (University of Adelaide), Reza Haffari (Monash University), and Javen Shi (University of Adelaide).

Advanced Hydrological Models

What are Advanced Hydrological Models?

In the last decade, environmental water allocations have been used to deliver in-channel pulses or have been paired with pumping and engineering solutions to artificially inundate wetlands and anabranches. To date, these actions have had a primary focus on flow rates and inundation extent, with limited attention to in-channel hydraulics and water quality. However, consideration to the changes in physical attributes due to changes in flow is fundamental to the rehabilitation of riverine ecosystems and should be integrated with hydrological restoration.

Hydrological models are essential tools that have been used to underpin MDB policy and operational decisions for decades. The models have been developed to simulate discharge, as well as diversions and other components of the water balance. Significant investment in the Source Murray Model has resulted in this model replacing MSM-Bigmod for almost all of MDBA's modelling requirements in the Murray and Lower Darling Rivers. These models are required to extrapolate monitoring at specific locations across the basin, as well as simulate scenarios of interest, such as the conditions without environmental water or different water delivery strategies.

However, hydrological models do not produce outputs that are most relevant for predicting the ecological response to changes in water availability and delivery. At best, local information is used to develop relationships between discharge and particular environmental responses, and due to interactions with geomorphology and operational structures, these empirical relationships between discharge and an outcome of interest cannot be extrapolated beyond the site scale. However, with the development of advanced modelling capabilities, these hydrological models can now incorporate factors other than flow rates to more accurately evaluate and predict environmental responses to water management at a landscape scale, and can express ecological flow requirements in hydrological terms (Box 9).

The benefit of converting modelling inputs from flow rates to physical attributes experienced by the biota is that 1) outputs are expected to correlate to ecological responses more strongly, and 2) these correlations are expected to be applicable across river reaches (Figure 12). However, relationships still need to be developed between the physical variables and ecological responses. In some cases, the relationship is obvious, such as dissolved oxygen thresholds for impacts on behaviour or mortality. Velocity thresholds have been developed in South Australia for predicting Murray Cod larvae and macroinvertebrate abundance (Gibbs et al., 2020), and to determine the risk of algal blooms in rivers by setting velocity thresholds which need to be maintained to allow for sufficient mechanical mixing to suppress BGA bloom formation (e.g., Mitrovic et al. 2003). Not reaching such velocities under low flow situations was a main cause of the 2018/19 Menindee fish kills, as water was allowed to stratify for a longer than usual period. This significantly depleted dissolved oxygen in the hypolimnion, with fatal consequences after mixing events. However, further work is required to extend these reach-specific studies of velocity thresholds to a basin-scale application and pinpoint critical situations in space and time.

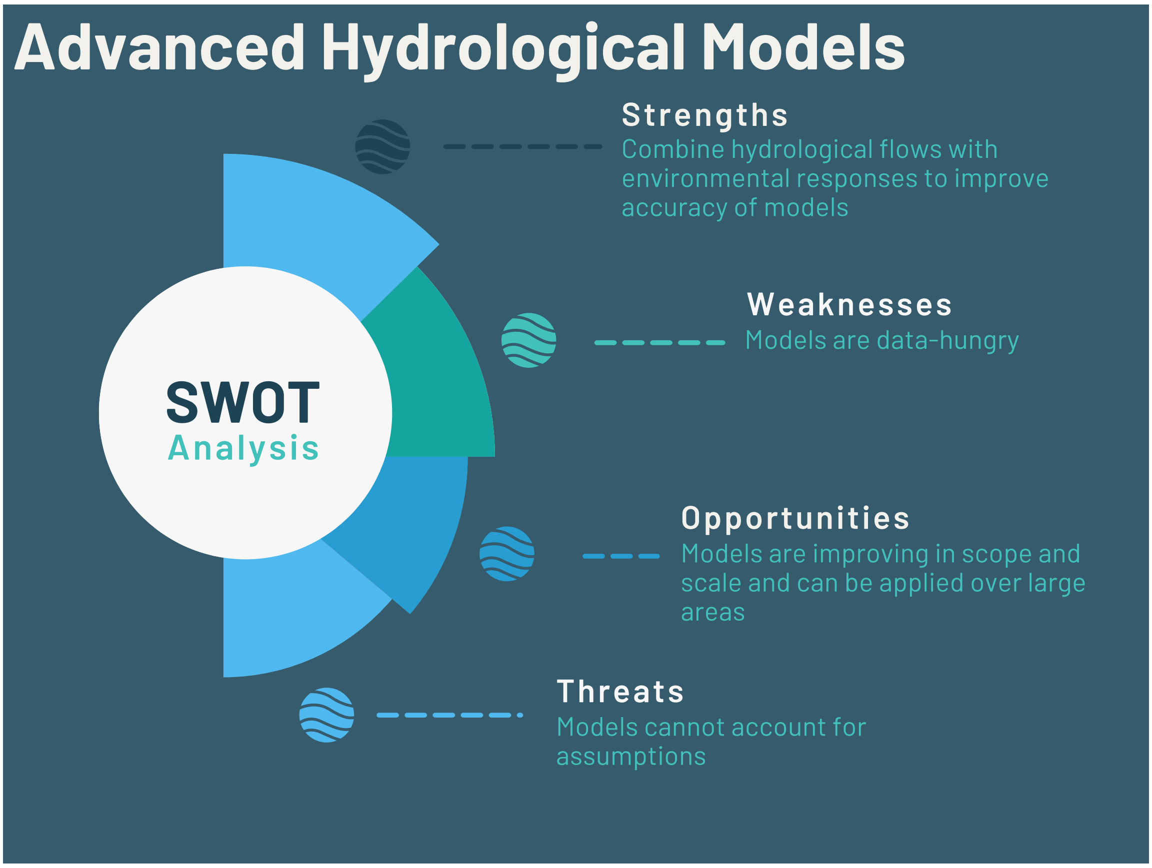


Figure 12 *Summary of the strengths, weaknesses, opportunities and threats of Advanced Hydrological Models.*

Current Applications

There are reach scale examples of hydrological models being extended to provide outputs of physical variables that have a more direct relationship with ecological responses, such as:

- Three of the selected areas for the Long-Term Intervention Monitoring Project/Monitoring Evaluation and Research project use relationships between discharge and velocity statistics to evaluate and interpret the ecological monitoring undertaken at the Lower River Murray (Ye et al. 2020), Goulburn River (Webb et al. 2015) and Edward-Wakool River (Watts et al. 2015). Methods to move from the reach to basin-scale are being evaluated as part of the WERP and Ecosystem Functions project.
- Water quality models that make use of the hydrological models have been developed, including the Dissolved Oxygen and Dissolved Organic Carbon (DODOC) plugin for Source models (Mosley et al. 2021). This extension is regularly used by the South Australian Department of Environment and Water to assess the risk of hypoxic conditions when operating floodplain infrastructure (e.g., Gibbs et al. 2020). The New South Wales Department of Planning and Environment is currently using the tool to assess changes in the risk of hypoxic blackwater from different Constraints Management Strategy options. Water temperature is a key input to this.

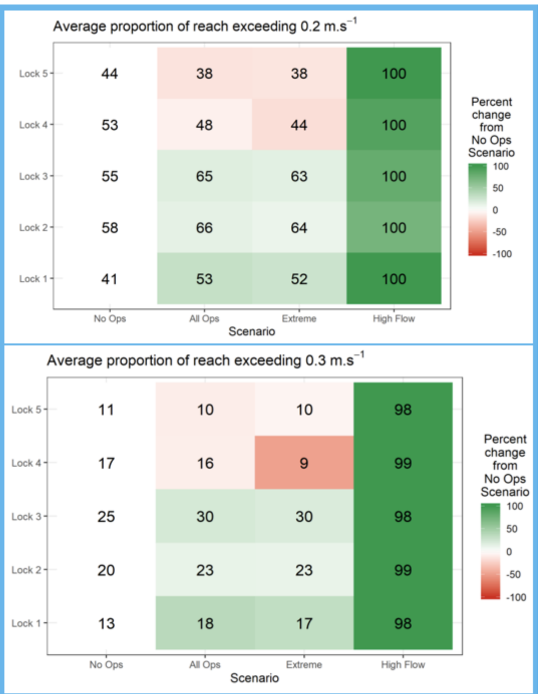
There are potential flow-on benefits from the current applications of the technology not yet tested. For example, the ability to simulate water temperature and dissolved organic carbon, as well as nutrients and productivity from other water quality models, provide the fundamental inputs necessary to relate changes in water delivery and water management to the amount of energy available to different trophic levels of food webs.

Box 17: Application of advanced hydrological modelling – modelling critical velocity thresholds for zooplankton Murray cod

Gibbs et al. (2020) used data gathered from Australia's Lower River Murray to estimate critical velocity thresholds required to entrain zooplankton (*Trichocerca spp.)* and Murray cod (*Maccullochella peelii*) larvae in the river drift (>0.2 $m.s^{-1}$ & >0.3 $m.s^{-1}$, respectively). These criteria were incorporated into the 'Source hydrological model of the SA River Murray' (Beh, Montzeri & Gibbs 2019) to determine what proportion of the Lower River Murray exceeded these thresholds under four operational scenarios (1) no Operations (No Ops), (2) all operations (All Ops), (3) extreme operations (Extreme) and (4) a high flow event (High Flow).

Relative to 'No Ops', all operation scenarios saw an overall increase in conditions suitable for the creation of suitable larvae retention habitat and zooplankton entrainment, though small decreases were estimated for Locks 5 and 4.

The addition of propagule transport with the River Murray allows the Source hydrological model to more comprehensively assess the efficacy of water operations and manage potential water risks following high flow events.



Images from Gibbs et al. (2020) used with permission of the Department for Environment and Water (DEW)

Potential Applications for the Murray-Darling Basin

There are a number of potential applications of advanced hydrological modelling for the MDBA's business that have the potential to enhance understanding of the relationship between hydrological flows and ecological requirements. For example, inundation models to convert flow rates into dynamic inundation patterns including the depth of water, have been developed (Teng et al. 2019) and are being extended. These models can be coupled with remotely sensed inundation extents to determine water depth, interpolate between available images (which can be taken weeks apart), and enable scenarios to be assessed. These tools are beginning to be used in the FLOW-MER project for the vegetation Basin-scale indicator, and can now be applied at spatial scales beyond the capabilities of traditional hydraulic models. Such models could be applied at scale across the MDB.

Costs and Limitations

The cost of extending hydrological models to provide more relevant outputs to predict ecological responses is likely to require further research and application to develop, test and improve the methods. However, the costs involved in deploying the technology are low; once the models are set up, running additional scenarios for the next year of data of a given scenario is a relatively small-time commitment, especially if the hydrological models are already required to be run. Data collated for these models (such as bathymetry) is likely to have multiple uses and provides a valuable foundational dataset.

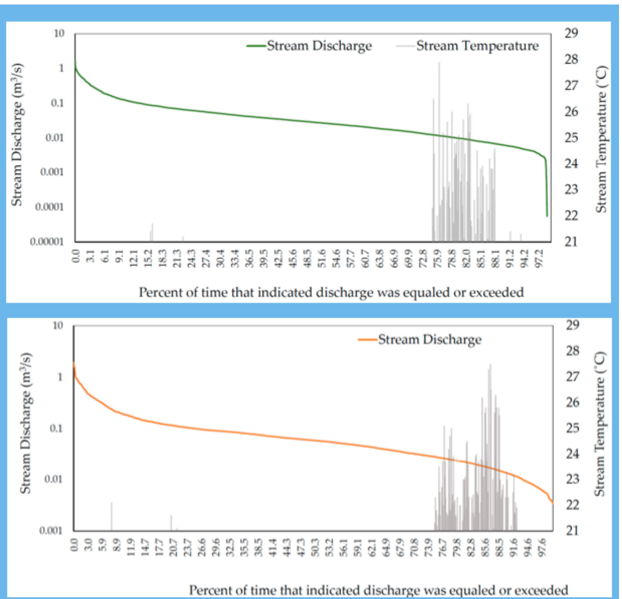
Box 18: Application of advanced hydrological modelling – identifying stressful events in fish habitat

Chambers, Pradhanang and Gold (2017) used the Soil and Water Assessment Tool (SWAT) with an incorporated hydroclimatological component to model the hydrology and stream temperatures in the Cork Brook watershed of New England. This study aimed to identify historical water quality trends so that the future impact of climate change on stream hydrology may be better understood.

The modelling period was from 1980–2009 and aimed to measure the frequency of 'stressful events' for brook trout (*Salvelinus fontinalis*). To account for both flow regime and temperature, 'stressful events' were defined as days where stream temperatures exceeding 21 °C co-occur with days of high or low flow (25^{th} and 75^{th} percentiles).

Modelling revealed that the number of stressful events increased from 84 during 1980–1989, to 131 during 2000–2009, representing a 56% increase in frequency. Additionally, of the 338 stressful events simulated over the three decades, only seven occurred during high flow conditions, demonstrating the increased susceptibility of temperature stress during low flow conditions.

The water quality trends simulated in this study provide an important baseline for future projections, and the identification of stressful events can assist in the management of cold-water species.



Simulated flow duration curves generated by SWAT with the hydroclimatological component. Top: Flow duration curve for the period 1980 – 1989. Bottom: Flow duration curve for the period 2000 – 2009. Note the different scales in the left y-axis. Images adapted from Chambers, Pradhanang and Gold (2017), licensed under CC-BY 4.0, https://creativecommons.org/licenses/by/4.0/

Technological Development

Learnings from the application of the DODOC model have found that current Source models, developed with a focus on the water balance, have in some areas made assumptions that do not provide sufficient information to apply the models directly. For example, a wetland is represented as a bulk water loss, rather than a storage that must fill and has evaporation and seepage losses associated with it, or transmission losses are represented as an unaccounted loss rather than attributed to a process. Hence, some hydrological model development may be required to enable the extensions of interest. Typically, these assumptions have had to be made due to a lack of data available (for example, bathymetry and outputs from hydraulic models to configure hydrological model nodes), or sufficient data to separate the effects of different processes (such as seepage loss, evapotranspiration, surface water groundwater interaction).

This technology is developing quickly, however often in a disjointed or disconnected way. A coordinated improvement program is likely to realise the major benefits. The risks involved in further model development incudes the complexity in managing the many models across the basin, as well as the inputs and outputs from the model. The importance of this process has been highlighted recently, and the MDBA Model Uplift Project aims to provide technical solutions to support model and output management.

Key Experts

CSIRO have a long history of basin scale modelling for environmental outcomes. This includes developing hydrologic models and hydraulic relationships (Dr Matt Gibbs, University of Adelaide, and Dr Ashmita Sengupta, CSIRO), basin scale inundation modelling (Dr Jin Teng, CSIRO) and water quality models (Dr Matt Gibbs, University of Adelaide, Dr Klaus Joehnk, CSIRO).

The Aquatic Ecodynamics group (Dr Matt Hipsey, University of Western Australia) have experience developing detailed, reach scale, hydraulic-water quality models for reaches in the River Murray, Coorong, and estuaries in WA and NSW.

Streamology (Dr Jeff Vietz, University of Melbourne) has developed hydraulic relationships to be applied for evaluation of environmental water benefits, in particular the Goulburn River.

Collecting and curating large ecological datasets - tips from TERN

Collecting and curating large datasets presents a number of challenges for ecologists and natural resource managers. To ensure consistency and accessibility across large networks such as the MDBA, there are a number of key considerations, including consistent data curation from the point of collection, database integration and shared repositories using a standard language, and model harmonisation. The Terrestrial Ecosystem Research Network (TERN) is a continental-wide open data network encompassing a variety of sensor streams, data formats, and scales. Data repositories and monitoring systems of this size and scale require advanced data infrastructure and standardised protocols, which could be adopted by the MDBA as new technologies are incorporated into the business.

Monitoring types

The first step for environmental monitoring should be to ensure that you are planning the correct types of environmental monitoring to provide the information types needed for effective decision-making. Key environmental questions can be broadly summarised in 6 types, with different forms of environmental monitoring excelling at answering different question types. The 6 key types of questions can be summarised as follows:

- Where is change occurring?
- When is that change occurring?
- What components of the environment are changing?
- What is the magnitude and direction of change?
- Why is change occurring?
- How can we manipulate change in a way we are happy with?

(Sparrow et al. 2020a)

It is important to choose monitoring activities that are well suited to addressing the questions of interest, and in many cases a mix of monitoring activities will need to be conducted to provide the information required for decision-making (Figure 13).

TYPES OF QUESTIONS

MONITORING TYPE	Where is environmental change occurring in the landscape?	When is environmental change occurring? and Is speed of change increasing or decreasing?	What elements within the environment are changing?	What is the direction and magnitude of that change?	What is the cause of the environmental change we are observing?	How can we take action to ameliorate deleterious change and encourage change that is positive?
Landscape	✓ Continuous data - Regionally and Globally	✓ Daily/ Weekly/Monthly for variables it can inform on	Only correlative with field data based on spectral response - not detailed information	Good within sensing limitations	Only able to suggest cause of change by correlation	Only able to correlate observed change with management interventions
Surveillance	Broadscale information at point locations across regions or continents	Useful data on trend over longer time periods - limited by re-visit frequency	✓ Good quality broad scale information on a wide range of variables if designed correctly	✓ Good across large areas	Only able to determine cause of change by correlation	Only able to correlate observed change with management interventions
Targeted	Only within the local study area	Usually good data for the study area / investigation species	Most often known prior to study - usually testing one interaction	Great for study area	✓ Given focus on a specific question understands cause of that change well	✓ Given accuracy at determining cause also good at determining how to manage unwanted change, single drivers are known.

__Figure 13__ Environmental monitoring decision-making guidelines, from Sparrow et al. (2020a) (used with permission).

Standardising monitoring

Once monitoring activities are determined it is then important to collect that information in the most standardised manner possible, in a way that maximises objectivity and the ability to provide the information required. Standardisation can apply to field sensors, field observations and measurement and remote assessment techniques. It is also best practice to collect objective data wherever possible, with measurements and continuous data providing far more utility than the collection of subjective, categorical data. The latter has much greater inaccuracies in collection, and the collection of categorical values makes it harder to detect change, particularly subtle change. If data is collected consistently across the entire area of interest, then the ability to inform on change is maximised. Collecting information in a standardised manner across all sites enables the intercomparison between sites (spatially), the comparison of an individual site through time (temporally) and the ability to assess how different sites are changing in comparison to each other (spatially and temporally). This task is much more difficult, if not impossible, without the application of standardised methods. Collecting standardised, objective data also maximises the potential for data reuse, leading to greater efficiencies (both time and financial) and improved value from monitoring activities. It also enables the data to be utilised for purposes that were not necessarily considered at the time of collection, enabling the greatest chance of data still being useful when questions change as a result of adopting adaptive management principles.

Ensuring compliance with monitoring methods is best achieved through a combination of well-considered and well-documented monitoring protocols. These protocols should include a rationale for data collection so that observers understand why they are collecting that data, and how it is likely to inform decision-making. Practical information on equipment and time needed to conduct the protocol, along with tips and tricks to help streamline data collection for data collectors are also important. For a protocol to be of maximum effectiveness, it is essential to include a step-by-step guide for the core method, written in such a way to avoid any ambiguity or ability to interpret the protocols in differing ways. Whilst this often takes time to get right, it is worth persevering with. Prior to developing protocols, it is important to determine if any already exist that are fit for purpose and can be easily implemented. Implementing pre-existing protocols is likely to save significant effort on method development and data management (see below) and maximise the compatibility of different monitoring programs.

Even with these protocols published and made widely available, it is also important to develop and deliver training activities in a variety of ways to assist with widespread implementation. Recently hybrid in-person and online events have become common, but where possible in-person training should be prioritised so that trainees can interact with trainers, ask questions and see methods used practically. TERN have found that training conducted in areas where there are a diversity of environments occurring (analogous to those the trainees are likely to need to sample) provides the best experience for trainees. A site that provides the opportunity for more formalised presentations (such as a lecture theatre style) for theory and rationale components, along with practical demonstrations of field data collection, provides the best experience for trainees. Once data collection activities are underway, it is sensible to run regular calibration activities within and between field teams (at least once per field season) to ensure that methods are still being implemented consistently by all data collectors (McCallum in Prep; White et al., 2012; Sparrow et al., 2020b).

Box 19: Collecting and curating large ecological datasets – The Terrestrial Ecosystem research Network (TERN)

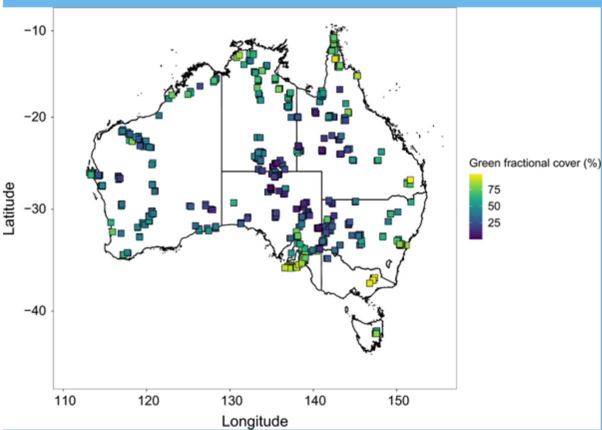


Image from Munro et al. (2021). Image depicts TERN AusPlots monitoring locations generated within the ausplotsR package. Licensed under CC-BY 4.0, https://creativecommons.org/licenses/by/4.0/

Australia's Terrestrial Ecosystem Research Network (TERN) is a 'network of networks', with 800 ecosystem surveillance sites, 13 landscape calibration and validation sites and 16 'SuperSites' that monitor biodiversity and ecosystem processes such as energy balance, micro-meteorology and water balance (TERN 2022).

To effectively manage and share the vast amount of data gathered by this network, TERN uses nationally consistent and standard ecosystem measures to monitor changes in ecosystems. For example, TERN's 'AusPlots' has developed survey protocol manuals for rangelands (AusPlots-Rangelands) and forests (AusPlots-Forests). These manuals outline a standardised plot-based sampling procedure that allows the gathering of consistent information that can be seamlessly integrated with other data sources. A companion application has been developed to streamline the data collection process for field scientists. Data can then be immediately uploaded to a cloud-based server for storage and curation.

TERN also developed a manual for 'Effective calibration and validation practices' (TERN 2018), a handbook that provides methods for collecting and using ground measurements to calibrate and validate remotely sensed data. Additionally, the manual also describes the best management practices for earth observation data.

Sample collection

It is strongly recommended that standardised monitoring methods are combined with robust sample collection protocols for a variety of reasons. A simple "chain of custody" system can be set up, where samples have barcodes attached in the field and scanned into a field data collection app (see below) to ensure a physical sample can always be associated with other data collected at the location. Samples enable accurate taxonomic identification to combine with other information collected at the monitoring site. This circumvents accuracy issues associated with field-based identifications and enables samples to be identified by the relevant taxonomic expert. Samples are also able to be placed in long-term storage, which enables them to be re-identified as taxonomy changes to ensure data is compatible with modern taxonomy. This ensures greater longevity of the data collected. These samples can also be stored to provide access to scientists at a future time, enabling their use for purposes that were not conceived at the time of collection. The samples themselves can be used for a range of analyses that enhance the initial data collection (e.g., taxonomic revisions, eDNA approaches, biological, chemical and physical analysis). There is also immense value in having all of these activities conducted on samples collected at the one field site at the one collection time, enabling a range of cross-disciplinary analyses that would otherwise not be possible. Managing these samples as ongoing research infrastructure provides ongoing benefits to both the research and management communities.

Box 20: Collecting and curating large ecological datasets – The U.S. National Ecological Observatory Network (NEON)

The U.S. National Ecological Observatory Network (NEON) gathers and provides standardised ecological data from across the country. NEON began full operation in 2019 and, as of 2022, has 47 terrestrial and 34 freshwater monitoring sites spanning 20 ecoclimatic domains (NEON 2022). Within each of these domains, NEON uses sensor and field measurements to gather data on plants, animals, soils, nutrients, freshwater and atmosphere, capturing the country's full range of ecological and climatic diversity.

NEON uses standardised sampling techniques, consistent data formatting and quality control to manage the thousands of sensors, billions of data points, and terabytes of data. For example, for processes requiring people to sample data in the field, NEON has used the 'Fulcrum platform' to develop tailor-made applications for that specific sampling protocol. Field scientists can then access these applications while in the field to ensure real-time quality assurance while gathering data. To undertake quality control for sensor-derived data, NEON uses algorithms to flag data that are out of the normal range or implausible. Additionally, random recounts and crosschecking is performed by NEON to further assess the quality of collected data.

Whether from automated sensors, airborne instruments or field staff, NEON must ensure that the data gathered can be seamlessly ingested for further processing, quality control and, eventually, publication. To this end, NEON has developed standardised documentation for organising and naming data to improve interpretability and readability for both humans and machines. Where possible, existing data formats, vocabularies or ontologies have been incorporated to enhance interoperability between NEON datasets with datasets from other research programs.

Image from Thorpe et al (2016), licensed under CC-BY 3.0 https://creativecommons.org/licenses/by/3.0/

Field data collection and centralised data management

To assist with standardisation and to simplify the curation, storage and delivery of data, it is recommended that consideration is given to electronic field data collection. Collecting information into a purpose-written app at the point of collection has a wide range of benefits, namely;

- *Ensures compliance with standards* by collecting information in a way that ensures that data is entered in the way it is intended; by setting criteria for fields (pre-defined categories, checks for sensible values, enforcement of particular collection accuracy, and checking of data dependencies) and checking the completeness of data prior to online submission, along with the ability to check and highlight likely errors when the data are being collected (and hence easily re-collected).

- *Prevents data transcription errors* by removing the need to collect, translate and enter data from poorly collected and messy handwriting from field data sheets.

- *Removes interpretation* by providing both structure and categories when necessary for data collection, along with the ability to provide definitions and guidance for data collection at the point of collection.

- *Enhances data security* by backing up data as it is collected, both in several places locally (onboard memory and SD card), and where there is a capacity for internet connectivity, data can be backed up to a cloud database as it is collected, minimising the chance of data loss.

- *Ensures easy data federation* by ensuring that data from multiple sites is collected in the same manner and is able to be uploaded to a centralised database as per the FAIR data principles (Wilkinson et al., 2016)

- *Ensures efficient data management* by automating the submission of data to a centralised database, and enabling data curation activities to be centralised to a small and specialised team rather than having a significant data management component in each field team. It also enables the submission of subsequent data (i.e. herbarium determinations), interoperability with other data sources, and the efficient delivery of data.

- *Enables a chain of custody to be easily established for collected samples* by utilising the camera of the device to scan pre-prepared, unique barcodes (or QR codes) that are physically stuck to samples collected in the field, so that they are always able to be easily associated to the site and other data from where they were collected (Tokmakoff et al., 2016).

- *Enables the rapid and easy collection of contextual data* by collecting photos of the site or specimens that are collected with, and linked to, other data collected.

- *Enables unique data collection methods to be integrated,* such as the collection of three-dimensional photopoints, or the automated assessment of groundcover or canopy cover using data collected by the device camera.

- *Enables the collection of some data automatically* where the device being used can inform (without intervention from the user) fields such as date, time, location, and direction that is output by the device, decreasing the likelihood of further transcription errors.

- *Can provide advice* by the inclusion of app pop-up reminders, tips and tricks, links to definitions, and links to electronic protocol manuals, ensuring that the data collector has all of the necessary information needed to collect the data appropriately.

- *Can provide some pre-analysis* to ensure data is in a form that makes it easier for ecologists to subsequently analyse. These apps also make analysis easier by being able to connect the live database to analysis software, ensuring researchers have the most up-to-date data and tooling to conduct analysis.

- *Centralised data management* can simplify data management, curation and delivery through the use of specialists in data management and the creation of specialist tools to enable the widest dissemination of data possible.

- *Can facilitate data access restrictions for sensitive information* through both authentication and obfuscation of sensitive data (Lowe et al., 2017).

Data Sharing Platforms

The sheer avalanche of information now gathered by survey and monitoring technologies poses obvious issues for researchers and practitioners of storage infrastructure and processing and data sharing. Furthermore, the integration of multiple data sources collected at different spatial and temporal scales is critical.

Data sharing simply means making data available for use by others. Perhaps the key part of that definition is "for use". Data sharing policies should consider both the development and adoption of data and metadata standards, as well as identify the availability and best-practice for federated data repositories and software tools to access and manage those repositories. Such practices will facilitate the use of open and reproducible data curation by the MDBA.

Centralised repositories for data are essential, but these no longer have to be in the form of fixed single databases. Instead, federated data systems make use of cloud storage and comprehensive metadata to allow access to any user and consequently facilitate data access. A key consideration for the uptake of shared data systems is determining how structured the repositories should be, or whether simple 'data lakes' of raw format data are stored without any transformation. Such an approach then requires more substantive data processing prior to analysis and visualisation but at least ensures all data sources are centrally accessible whilst limiting the costs of data management. When combined with metadata standards, this approach can produce significant efficiency gains with limited setup costs. More broadly, organisations should focus on data management and sharing, and how they might consider automating the data pipeline of handling, storing, and processing the massive amounts of data that can be generated by these sensor networks, particularly how the process can be operationalised at the large spatial and temporal scales of interest. As mentioned above, this component also should consider the artificial intelligence pipelines that can process these large and complex data streams into usable data for subsequent analysis. There is a clear need for these technologies to utilise cloud computing for data storage, and potentially programming and processing techniques that split data into manageable chunks and process them in parallel.

The concept of “data sharing readiness” refers to reliable and consistently curated data that are stored in accessible repositories for use by a broad user community. In more technical terms, this corresponds to standardised data sources shared in repositories that are recognised and accessible to the scientific communities. The review should consider how to integrate databases or data repositories into the technology pipeline such that recorded data are accessible online, but also that this remote access should consider both how data can be submitted, potentially automatically by connected sensor networks, and also retrieved at any time by end-users. In terms of data retrieval, “data sharing readiness” should mean that the shared repositories of data are consistent with FAIR (Findable, Accessible, Interoperable and Reusable; Wilkinson et al. 2019) and TRUST (Transparency, Responsibility, User focus, Sustainability and Technology; Lin et al. 2020) data principles. Data repositories such as that of the United States Geological Survey (USGS) can be extremely valuable in providing readily accessible and well-catalogued environmental data.

Skilling A Workforce

All of the technologies described in this Innovation Sweep require some degree of bioinformatics and data processing skills to be executed. While it is possible to outsource tasks requiring these capabilities to external contractors, it is clear that the next phase of ecology will be far more computerised than in previous years, and the skillsets of ecologists must evolve in parallel. While bioinformatics is increasingly taught as part of undergraduate coursework, there is still a mismatch between the available training and the rapidly increasing demand for data scientists. Further, few established ecologists have the capacity to undertake additional tertiary studies once in the workforce.

Studies have found that environmental professionals with established careers show a preference for short, self-paced online courses when upskilling (Attwood et al., 2019). Such training modules could be incorporated regularly into the MDBA's operations to improve bioinformatics proficiency in the workforce. While the MDBA may choose to develop their own training program targeted towards areas of interest and need, there are several online platforms currently available that provide a large body of training materials and courses on a variety of bioinformatics tools. An example is GOBLET (https://www.mygoblet.org), a repository of high-quality bioinformatics learning materials established to be a network of trainers and end-users (Corpas et al., 2015). Such resources could be readily used by current staff seeking to improve their bioinformatics skills – meanwhile, a stronger focus on data expertise when recruiting new team members would improve the capability of the MDBA workforce overall to work with new technologies requiring large, complex data sets.

Next Steps

Given the rapid development predicted for the near future of the technologies discussed in this report, the MDBA would benefit from regular revisions and conference-style workshops with experts in the field. This would allow the MDBA to keep up to date with progressions and advancements of technologies already in use, as well as maintain awareness of other innovations that may emerge. It is recommended that the MDBA conduct reviews of technologies of interest annually, as rapid growth is anticipated in many areas of innovation, including data analysis and modelling platforms that utilise artificial intelligence. These audits may be conducted within the MDBA by connecting and communicating with the network of experts identified within this report, or by outsourcing the review to an external organisation. This will ensure that the MDBA remains at the forefront of ecological management techniques in Australia, and will streamline transitions towards the application of next-generation technologies.

Liaising directly with commercial providers of data sensing and sharing platforms may also be a cost-effective pathway towards accessing cutting-edge products and training. It is recommended that in-house training or embedded placements with technology companies and providers be made available to staff within the MDBA, as this will not only facilitate skill-building within the workforce, but encourage critical and innovative thinking. Building teams of staff within the organisation that can confidently and effectively implement these technologies will be highly beneficial to the MDBA's business.

Conclusion

This Innovation Sweep provides a summary of emerging technologies in ecology and natural resource management that have the potential to improve environmental outcomes throughout the Murray-Darling Basin. A range of advanced sensing technologies and machine-learning assisted data analysis techniques have been described, with the capacity to increase the accuracy and efficiency of the MDBA's monitoring efforts. A summary of the potential on-ground areas of application for these technologies can be found in Table 4. Further, modelling approaches enhanced by artificial intelligence have been detailed that can inform decision-making for long-term management projects such as environmental watering. Using such models, managers can predict the future outcomes of various strategies with a high degree of accuracy. Although each of these technologies are powerful tools alone, they become exponentially more useful when partnered and used together. For example, autonomous platforms may collect long-term data on water levels that can be quickly and efficiently analysed using machine learning-based analysis. The results of this analysis could then be fed into sequential decision-making and reinforcement learning models to assist managers in identifying best-practice environmental watering strategies.

Based on the innovations explored in this report, a number of skills and capabilities were identified that emerged as high priority within the workforce in the move towards adopting next-generation ecological technologies. It is apparent that bioinformatics and coding skills are becoming increasingly relevant and valuable for ecologists, and often have the added benefit of improving data management and curation practices. A stronger focus on training and recruiting individuals skilled in these areas would be beneficial for the MDBA in the future.

The technologies described in this report have undergone rapid development in recent years, and many are predicted to improve in speed and accuracy by orders of magnitude in the near future. It is therefore timely that the MDBA consider incorporating these tools into management planning and practice and preparing their workforce to uptake the required skills.

__Table 4__ Summary of sensor and data analysis technologies discussed in this Innovation Sweep and their potential to inform on areas of interest to the MDBA. ✓= directly, ✦ = indirectly.

	Species/Individual Presence				**Species/Individual Abundance**				**Soil Chemistry**	**Water Chemistry**	**Water Level**	**Water Temperature**	**Turbidity**	**Flow**	**Vegetation Cover**
	Birds	**Fish**	**Invertebrates**	**Micro-organisms**	**Birds**	**Fish**	**Invertebrates**	**Micro-organisms**							
eDNA	✓	✓	✓	✓	✓	✓	✓	✓	✦	✦					
Autonomous platforms									✓	✓	✓	✓	✓	✓	
Animal tracking	✓	✓	✓												
Machine Learning Based Analysis	✓	✓	✓		✓	✓	✓				✓				✓
Visual Question Answering and Visual Language Navigation	✓	✓	✓		✓	✓	✓				✓				✓

Data Efficient Learning for Vision and Audio Recognition	✓	✓	✓		✓	✓	✓				✓				✓
Semantic Change Detection in Images	✓	✓	✓		✓	✓	✓				✓				✓

References

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A. and Schneider, M. V., 2019. A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*, 20 (2), 398–404.

Beh, E., Montzeri, M. and Gibbs, M. S., 2019. Refinements to the River Murray Source model in South Australia, DfEaW Government of South Australia, Adelaide.

Bin Mat Idris, M. H., Al Azhari Bin Che Kamarudin, M., Sahalan, M. I., Bin Zainal Abidin, Z. and Rashid, M. M., 2016. Design and Development of an Autonomous Surface Vessel for Inland Water Depth Monitoring. *In*: *Design and Development of an Autonomous Surface Vessel for Inland Water Depth Monitoring*. Presented at the 2016 International Conference on Computer and Communication Engineering, ICCCE), 177–182.

Bracken, F. S. A., Rooney, S. M., Kelly-Quinn, M., King, J. J. and Carlsson, J., 2019. Identifying spawning sites and other critical habitat in lotic systems using eDNA "snapshots": A case study using the sea lamprey Petromyzon marinus L. *Ecology and Evolution*, 9 (1), 553–567.

Brookes, J. D., Huang, P., Zhai, S. Y., Gibbs, M. S., Ye, Q., Aldridge, K. T., et al., 2022. Environmental flows to estuaries and coastal lagoons shape the salinity gradient and generate suitable fish habitat: Predictions from the Coorong, Australia. *Frontiers in Environmental Science*, 10, 796623.

Brunt, T., Griffiths, J. and Bino, G., 2018. Detecting a Paradox. *Wildlife Australia*, 55 (4), 4–8.

Bylemans, J., Furlan, E. M., Hardy, C. M., McGuffie, P., Lintermans, M. and Gleeson, D. M., 2017. An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (Macquaria australasica). *Methods in Ecology and Evolution*, 8 (5), 646–655.

Chang, H. C., Hsu, Y. L., Hung, S. S., Ou, G. R., Wu, J. R. and Hsu, C., 2021. Autonomous water quality monitoring and water surface cleaning for unmanned surface vehicle. *Sensors*, 21 (4), 1102.

Cao. H., Guo, Z., Wang, S., Cheng, G. and Zhan, C., 2020. Intelligent wide-area water quality monitoring and analysis system exploiting unmanned surface vehicles and ensemble learning. *Water*, 12 (3), 681.

Chambers, C., Pradhanang, S. M. and Gold, A. J., 2017. Assessing thermally stressful events in a Rhode Island coldwater fish habitat using the SWAT model. *Water*, 9 (9), 667.

Chang, H.-C., Hsu, Y.-L., Hung, S.-S., Ou, G.-R., Wu, J.-R. and Hsu, C., 2021. Autonomous water quality monitoring and water surface cleaning for unmanned surface surface vehicle. *Sensors*, 21 (4), 1102.

Corcoran, E., Denman, S., Hanger, J., Wilson, B. and Hamilton, G., 2019. Automated detection of koalas using low-level aerial surveillance and machine learning. *Scientific Reports*, 9 (1), 3208.

Corpas, M., Jimenez, R. C., Bongcam-Rudloff, E., Budd, A., Brazas, M. D., Fernandes, P. L., Gaeta, B., van Gelder, C., Korpelainen, E. and Lewitter, F., 2015. The GOBLET training portal: a global repository of bioinformatics training materials, courses and trainers. *Bioinformatics*, 31 (1), 140–142.

Deng, Z. D., Li, H., Lu, J., Xiao, J., Myjak, M. J., Martinez, J. J., Wang, Y. and Zhang, J., 2021. An acoustic micro-transmitter enabling tracking of sensitive aquatic species in riverine and estuarine environments. *Cell Reports Physical Science*, 2 (5), 100411.

Dunbabin, M. and Grinham, A., 2010. Experimental evaluation of an autonomous surface vehicle for water quality and greenhouse gas emission monitoring. Presented at the 2010 IEEE International Conference on Robotics and Automation, IEEE, 5268–5274.

Foote, A. R., Stratford, C. N. and Coman, G. J., 2018. Passive integrated transponder (PIT) tagging black tiger shrimp, Penaeus monodon: Applications for breeding programs. *Aquaculture*, 491, 321–324.

Furlan, E. M., Gleeson, D., Wisniewski, C., Yick, J. and Duncan, R. P., 2019. eDNA surveys to detect species at very low densities: A case study of European carp eradication in Tasmania, Australia. *Journal of Applied Ecology*, 56 (11), 2505–2517.

Gibbs, M. S., Bice, C., Brookes, J., Furst, D., Gao, L., Joehnk, K. Marklund, M. Nicol, J., Pethybridge, H., Szarvas, S., Wallace, T., Welti, N. and Zampatti, B., 2020. Ecological connectivity of theRiver Murray: Managing ecological outcomes and water quality risks through integrated river management. Goyder Institute for Water Research Technical Report Series No. 20/03.

Hirakawa, T., Yamashita, T., Tamaki, T., Fujiyoshi, H., Umezu, Y., Takeuchi, T., et al., 2018. Can AI predict animal movements? Filling gaps in animal trajectories using inverse reinforcement learning. *Ecosphere*, 9 (10), e02447.

Hobley, B., Arosio, R., French, G., Bremner, J., Dolphin, T. and Mackiewicz, M., 2021. Semi-supervised segmentation for coastal monitoring seagrass using RPA imagery. *Remote Sensing*, 13, (9), 1741.

Hooshyar, M., Mousavi, S. J., Mahootchi, M. and Ponnambalam, K., 2020. Aggregation-decomposition-based multi-agent reinforcement learning for multi-reservoir operations optimization. *Water*, 12 (10), 2688.

Jindal, R., Kunwar, A. K., Kaur, A. and Jakhar, B. S., 2020. Predicting the dynamics of forest fire spread from satellite imaging using deep learning. Presented at the 2020 International Conference on Electronics and Sustainable Communication Systems, ICESC, 344–350.

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M. and Westbrook, J., 2020. The TRUST Principles for digital repositories. *Scientific Data*, 7 (1), 144.

Lowe, A. J., Smyth, A. K., Atkins, K., Avery, R., Belbin, L., Brown, N., et al., 2017. Publish openly but responsibly. *Science* 357 (6347), 141.

Lugg, W. H., Griffiths, J., van Rooyen, A. R., Weeks, A. R. and Tingley, R., 2018. Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, 9 (4), 1049–1059.

Madeo, D., Pozzebon, A., Mocenni, C. and Bertoni, D., 2020. A low-cost unmanned surface vehicle for pervasive water quality monitoring. *IEE Transactions on Instrumentation and Measurement*, 69 (4), 1433–1444.

Mitrovic, S. M., Oliver, R. L., Rees, C., Bowling, L. C. and Buckney, R. T., 2003. Critical flow velocities for the growth and dominance of *Anabaena circinalis* in some turbid freshwater rivers: *Critical flow velocities for the growth of* Anabaena circinalis. *Freshwater Biology*, 48 (1), 164–174.

Mosley, L. M., Wallace, T., Rahman, J., Roberts, T. and Gibbs, M., 2021. An integrated model to predict and prevent hypoxia in floodplain-river systems. *Journal of Environmental Management*, 286, 112213.

Munroe, S., Guerin, G., Saleeba, T., Martín-Forés, I., Blanco-Martin, B., Sparrow, B. and Tokmakoff, A., 2021. ausplotsR: An R package for rapid extraction and analysis of vegetation and soil data collected by Australia's Terrestrial Ecosystem Research Network. *Journal of Vegetation Science*, 32 (3), e13046.

NEON (National Ecological Observatory Network), 2022. National Ecological Observatory Network, accessed 13 May 2022, <https://www.neonscience.org/>

Ogawa, D., Tamaki, T., Hirakawa, T., Raytchev, B., Kaneda, K. and Yoda, K., 2019. Improved activity forecast for generating trajectories. *arXiv Preprint arXiv:1912.05729*.

OceanAlpha, 2022. ESM30: Autonomous water sampling and monitoring USV, accessed 4 April 2022, <https://www.oceanalpha.com/product-item/esm30/>

Rojahn, J., Gleeson, D., Furlan, E. M., Rojahn, J., Gleeson, D. and Furlan, E. M., 2018. Monitoring post-release survival of the northern corroboree frog, Pseudophryne pengilleyi, using environmental DNA. *Wildlife Research*, 45 (7), 620–626.

Sharma, D., Purushotham, S. and Reddy, C. K., 2021. MedFuseNet: An attention-based multimodal deep learning model for visual question answering in the medical domain. *Scientific Reports*, 11 (1), 19826.

Shokralla, S., Spall, J. L., Gibson, J. F. and Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular ecology*, 21 (8), 1794–1805.

Smart, A. S., Weeks, A. R., van Rooyen, A. R., Moore, A., McCarthy, M. A. and Tingley, R., 2016. Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, 7 (11), 1291–1298.

Sparrow, B. D., Edwards, W., Munroe, S. E. M., Wardle, G. M., Guerin, G. R., Bastin, J. F., et al., 2020a. Effective ecosystem monitoring requires a multi-scaled approach. *Biol Rev Camb Philos Soc* 95 (6), 1706-1719.

Sparrow, B. D., Foulkes, J. N., Wardle, G. M., Leitch, E. J., Caddy-Retalic, S., van Leeuwen, S. J., et al., 2020b. A Vegetation and Soil Survey Method for Surveillance Monitoring of Rangeland Environments. *Frontiers in Ecology and Evolution* 8 (157).

Teng, J., Vaze, J., Kim, S., Dutta, D., Jakeman, A. J. and Croke, B. F. W., 2019. Enhancing the Capability of a Simple, Computationally Efficient, Conceptual

Flood Inundation Model in Hydrologically Complex Terrain. *Water Resources Management*, 33 (2), 831–845.

TERN (Terrestrial Ecosystem Research Network), 2018. Effective field calibration and validation practices: A practical handbook for calibration and validation of satellite and model-derived terrestrial environmental variables for research and management, accessed 13 May 2022, <https://www.tern.org.au/wp-content/uploads/TERN_CalVal_Handbook_web.pdf>

TERN (Terrestrial Ecosystem Research Network), 2022. TERN Ecosystem Research Infrastructure, accessed 13 May 2022, <https://www.tern.org.au/>

Thorpe, A. S., Barnett, D. T., Elmendorf, S. C., Hinckley, E.-L. S., Hoekman, D., Jones, K. D., LeVan, K. E., Meier, C. L., Stanish, L. F. and Thibault, K. M., 2016. Introduction to the sampling designs of the National Ecological Observatory Network Terrestrial Observation System. *Ecosphere*, 7 (12), e01627.

Tokmakoff, A., Sparrow, B., Turner, D., and Lowe, A., 2016. AusPlots Rangelands field data collection and publication: Infrastructure for ecological monitoring. *Future Generation Computer Systems-the International Journal of Escience* 56, 537-549.

Tosa, M. I., Dziedzic, E. H., Appel, C. L., Urbina, J., Massey, A., Ruprecht, J., Eriksson, C. E., Dolliver, J. E., Lesmeister, D. B., Betts, M. G., Peres, C. A. and Levi, T., 2021. The Rapid Rise of Next-Generation Natural History. *Frontiers in Ecology and Evolution*, 9, 698131.

Wang, G., 2019. Machine learning for inferring animal behavior from location and movement data. *Ecological Informatics*, 49 (1), 69–76.

White, A., Sparrow, B., Leitch, E., Foulkes, J., Flitton, R., Lowe, A. J., et al., 2012. AusPlots Rangelands Survey Protocols Manual. Adelaide. University of Adelaide Press.

Wiersma, P. and Piersma, T., 1995. Scoring abdominal profiles to characterize migratory cohorts of shorebirds: An example with Red Knots. *Journal of Field Ornithology*, 66 (1), 88–98.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018.

Wilkinson, M. D., Dumontier, M., Sansone, S. A., da Silva Santos, L. O. B., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M. and Schultes, E., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific data*, 6 (1), 1–12.

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M. and Fortson, L., 2018. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10 (1), 80–91.

Yang, L., Hou, J., Cheng, L., Wang, P., Pan, Z., Wang, T., et al., 2021. Application of habitat suitability model coupling with high-precision hydrodynamic processes. *Ecological Modelling*, 462 (5), 109792.

YSI, 2022. HYCAT Autonomous surface vehicle, accessed 4 April 2022, <https://www.ysi.com/hycat>

Yuan, Z., Mou, L., Xiong, Z. and Zhu, X., 2021. Change detection meets visual question answering. *arXiv Preprint arXiv:2112.06343*.

Zhao, M., Zhao, Z., Gong, Z., Liu, Y., Yang, J., Xiong, X. and Li, S., 2022. Spatially and semantically enhanced Siamese network for semantic change detection in high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2563–2573.