

Hawwa Raufath Nizar

SUPERvisor:

S4755382

TiTLE

Subtitle

# Background

There is a complex interplay between spatial and temporal variability in fish migratory behaviour, and their influence on fish morphology (Chapman et al., 2015) and growth (Clausen et al., 2015; Tattam et al., 2017). In lotic habitats such as riverine ecosystems, these influences are further compounded by the inherent environmental variability, including factors such as streamflow variability and shifting thermal regimes. Research indicates that for some species, stable low-flow periods that follow periods of sudden and increased flow rates tend to yield higher growth rates for some species, especially coupled with warmer temperatures (Haworth & Bestgen, 2016; Tonkin et al., 2017). These findings have significant implications on the monitoring and management of riverine ecosystems. For instance, the positive correlation between flow variability and growth of fish highlights how regulatory measures that suppress the flow of water in lotic ecosystems could be ultimately detrimental to the growth of certain species. These effects are also subject to much temporal variability, with higher growth rates not only being associated with periods of warmer temperature in general, but also higher hydrological variability during phases and seasons critically important to the species' lifecycle. (Tonkin et al., 2017). Therefore, any management interventions must carefully consider not only the environmental factors that serve as predictor variables, but their spatio-temporal scale, as well.

Within this context, the distinctive hydrological characteristics of South and West Queensland's dryland rivers serve as a compelling case study for the profound impact of environmental variability on fish migration, growth and survival. The Bureau of Meteorology classifies large swathes of this region of Queensland as arid and semi-arid areas (Bureau of Meteorology, 2007), where habitats are typically characterised by either low rainfall throughout the year, or by wet summers and low rainfall winters. Dryland rivers, which are a critical aquatic habitat in this region, experience cyclic conditions of “booms” in productivity due to episodic floods and “bust” periods after the waters recede, leaving behind perennial and semi-perennial waterholes (Bureau of Meteorology, 2007). These waterhole refugia play a central role in shaping ecological dynamics of these intermittent riverine ecosystems, governing critical mechanisms and processes such as growth and resilience in species that inhabit such environments (Marshall et al., 2016). Alongside more common environmental factors, natural phenomena such as droughts and flooding also create key disturbance events in these ecosystems, that have been shown to significantly impact the magnitude of the role such refugia play.

This project aims to look at the impact of various environmental and hydrological factors on migratory patterns and ultimately, growth rate of golden perch (*Macquaria ambigua*), Bony bream (*Nematalosa erebi*) and Common carp (*Cyprinus carpio*) populations in local riverine systems that experience flow intermittency. Otolith biochronology will be used as a means of examining the impact of these factors, including their spatiotemporal variability, on incremental growth rate in these three species across 11 rivers in the Northern Murray-Darling Basin. This project not only aims to answer specific research questions regarding the magnitude and effect of various predictive factors on growth, but also to contribute to the broader discourse on sustainable water resource management in the face of increasing climate vulnerability and anthropogenic disturbance. Thus, in adhering to the principles of reproducible research, this project will ensure that all research outputs are transparent and accessible via its GitHub repository, facilitating further research and application.

# Methods and Approach

## 2.1 Data Sources

A range of ecological and environmental datasets was utilized for this project, including data on otolith-derived incremental growth rates for the three fish species, river flow metrics, and annual average temperature readings. These datasets were accessed and extracted via a Power BI Solution developed by La Trobe University, with input from the Department of Environment and Science (DES), Queensland. The streamflow data and temperature data contained therein were sourced from stream gauges installed within the study area. The raw datasets were collated and organized via R, to create a consolidated dataset that was used for further data cleaning and analysis.

## 2.2 Predictor Variables

As outlined above, the primary focus of the analysis will be to evaluate the impact of various environmental and hydrological factors on the annual growth rates of three lotic fish species. The predictor variables, derived from the data sourced through the Power BI dashboard, will include:

* Flow Volume: The mean, minimum, maximum levels, to reflect the dynamics of water flow within the habitat.
* Flow Duration: The length of time for which water flow is sustained at various levels, affecting connectivity and movement.
* Bankfull Flow Conditions: Indicates the maximum carrying capacity of the river, without overflowing, and consequently disrupting sediment transportation and habitat structure.
* Flow Days: The number of days with significant flow, potentially affecting feeding opportunities and other interactions within the food web.
* Water Temperature (Annual Average): Indicating the thermal conditions experienced at the study sites annually.
* Temperature Accumulation (Degree Days): A cumulative measure of heat exposure over time, which influences metabolic rates and growth cycles.
* Drought and Flood Events: Instances of extreme low and high volume, affecting habitat quality and food availability.

Additional variables might be considered depending on their availability and relevance to the growth patterns observed in the otolith data. These could encompass environmental features such as habitat composition, water quality parameters, and anthropogenic influences. The inclusion of these variables will be adaptive, contingent upon their statistical significance to the models, the insights they provide into the growth rates of the species being studied and the convergence properties of the models, themselves.

## 2.3 Modeling

The primary final product will be a comprehensive analytical script prepared with the programming language R that tidies, prepares for analysis and thoroughly explores the dataset to examine the impact of the various environmental and hydrological factors outlined above, on the annual growth of Golden perch, Bony bream, and Common carp. The initial steps will involve data visualisation and generation of descriptive statistics to guide the modeling process. Following this, a methodologically iterative approach will be taken to explore the relationships. This means starting with simpler models to understand basic relationships, then progressively incorporating more complex models to capture nuanced patterns. This phased approach allows for a thorough exploration of the data, ensuring that the final model(s) provide insightful and reliable predictions about the impact of hydrological factors on fish growth rates. The aim is not only to identify significant predictors but also to understand the magnitude of their effects, contributing to informed management strategies for riverine ecosystems. Some of the model types that will be considered include the following.

**Multiple Linear Regression:**

* Overview: A foundational method to quantify and model the relationship between fish growth and one or more of the independent variables.
* Implementation: The lm() function will be used to estimate model parameters, starting with a base model and iteratively adding potential predictors.
* Evaluation: Model diagnostics such as residual plots, QQ-plots, and the variance inflation factor (VIF) will be employed to check the assumptions and fit of the model. The R-squared value will be used to provide insight into the explanatory power of the model.

**Mixed Effects Models:**

* Overview: These models will consider both fixed and random effects of the predictor variables on fish growth and are predicted to be especially useful where there is spatial or temporal nesting within the data.
* Implementation: Implementation will be using the lmer() function from the ‘lme4’ package in R, fixed effects of predictors will be analysed while accounting for the random variations across catchments and sites.
* Evaluation: Model fit will be evaluated using likelihood ratio tests, and AIC and BIC criteria. As with multiple linear regression, model assumptions will be checked via residual plots.

**Generalized Linear Models (GLMs):**

* Overview: These models will extend linear regression to allow for response variables that have error distributions other than a normal distribution.
* Implementation: The glm() function will be employed, and Gaussian GLMs will be the primary focus, but the project will explore variations, adjusting fixed and random effects, and trying different link functions based on data distribution.
* Evaluation: Deviance and residuals will be the key evaluation metrics. The goodness of fit will be evaluated using the AIC and BIC.

**Advanced Techniques (Tentative):**

Random Forest and Artificial Neural Networks (ANNs) may be explored for their ability to capture non-linear relationships and complex interactions within the data. The choice to use these techniques will depend on initial findings from simpler models and the complexity of the data.

## 2.4 Data Splitting and Model Validation

Prior to modeling, the dataset will be divided into training and testing sets to validate model performance on unseen data.

**Training Set:** This subset will include 80% of the original dataset, and will be used for model development and training, enabling algorithms to learn the relationship between the dependant and independent variables.

**Testing Set:** This subset will comprise the remaining 20% of data, it will be used to evaluate the performance of each model on ‘unseen’ data, to prevent overfitting.

An additional validation set will be used specifically for Artificial Neural Networks, to fine-tune the model parameters without impacting the test set, ensuring it remains an unbiased measure of model performance.

# 3.0 Schedule

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | April | | May | | | |
| Milestone description | Start | Days |  | 3 | 4 | 1 | 2 | 3 | 4 |
| **Literature Review and Finalisation** | | | | | | | | | |
| Review background literature. | 15/04/24 | 2 |  |  |  |  |  |  |  |
| Finalize objectives and expected outcomes of the project. | 17/04/24 | 1 |  |  |  |  |  |  |  |
| **Accessing and Preparing Data** | | | | | | | | | |
| Combine datasets to create a comprehensive and consolidated dataset for modeling | 18/04/24 | 1 |  |  |  |  |  |  |  |
| Identify missing data points and inconsistencies | 19/04/24 | 1 |  |  |  |  |  |  |  |
| Preliminary data exploration and creating a tidy dataset | 20/04/24 | 3 |  |  |  |  |  |  |  |
| Data preprocessing: Normalize, standardize, or transform data as necessary. | 23/04/24 | 2 |  |  |  |  |  |  |  |
| Split data into training and testing sets (and validation set for ANNs). | 25/04/24 | 1 |  |  |  |  |  |  |  |
| **Modeling** | | | | | | | | | |
| Initiate modeling process with Multiple Linear Regression Models and Polynomial Regression Models. | 26/04/24 | 2 |  |  |  |  |  |  |  |
| Fit models, test assumptions, and evaluate initial results. | 28/04/24 | 2 |  |  |  |  |  |  |  |
| Continue modelling process with Generalized Linear Models (GLMs). | 30/04/24 | 2 |  |  |  |  |  |  |  |
| Evaluate the GLMs using AIC and BIC. | 02/05/24 | 2 |  |  |  |  |  |  |  |
| Start Random Forest modeling. | 04/05/24 | 2 |  |  |  |  |  |  |  |
| Evaluate predictor importance and model performance. | 06/05/24 | 2 |  |  |  |  |  |  |  |
| Start developing Artificial Neural Networks (ANNs). Adjust architectures and activation functions as needed | 08/05/24 | 2 |  |  |  |  |  |  |  |
| Validate models using the dedicated validation set to prevent overfitting | 10/05/24 | 1 |  |  |  |  |  |  |  |
| Evaluate the ANN models using MAE and RMSE | 11/05/24 | 1 |  |  |  |  |  |  |  |
| Compare all models side-by-side and identify the most accurate and efficient model(s) for predicting growth. | 12/05/24 | 1 |  |  |  |  |  |  |  |
| **Report Drafting** | | | | | | | | | |
| Draft a comprehensive report detailing methodologies, results, findings, and recommendations. | 13/05/24 | 5 |  |  |  |  |  |  |  |
| Review, edit, and finalize the report | 18/05/24 | 5 |  |  |  |  |  |  |  |
| **Seminar Preparation** |  |  |  |  |  |  |  |  |  |
| Synthesize key points from report | 18/05/24 | 2 |  |  |  |  |  |  |  |
| Develop visual aids and prepare PowerPoint presentation | 20/05/24 | 3 |  |  |  |  |  |  |  |

# 4.0 Project Deliverables

The following deliverable will be produced at the conclusion of the project and the associated modelling process.

**Predictive Models and Relevant Code:**

The outcomes will include Multiple Linear Regression Model, Polynomial Regression Models, Generalized Linear Models (GLMs), Random Forest, and Artificial Neural Networks (ANNs), all tailored for predicting water temperature in lotic ecosystems based on relevant parameters. In the interest of ensuring transparency and reproducibility, the full R code for each of these modeling processes will be shared via GitHub and other agreed means.

**Visual Representations:**

Graphical representations and plots showing the predictive abilities of each model. Where Artificial Neural Networks are utilised, architecture diagrams will also be included, depicting layers and activation functions.

**Model Evaluation:**

Detailed statistical evaluations of each model's performance, including R-squared values, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) will be presented.

Comparative visual charts elucidating the strengths and potential shortcomings of each model will be generated.

**Final Report:**

A comprehensive project report detailing the background, methodologies employed, datasets used, modeling processes, findings, and recommendations will be generated. The report will delve into informed recommendations based on the findings, and provide advise on the feasibility and work required for a state-wide roll out. This will include information on potential future refinements and model development, including model types, factors such as higher variability over larger spatial scales, the differences in stime frames and other nuances inherent in such extrapolation. This report is envisioned to serve as a reference point for future endeavours in lotic ecosystem temperature predictions and a guide for management and conservation efforts.

**Seminar Presentation:**

A PowerPoint presentation designed to convey the project key objectives, methodology, results, and implications for future research and conservation issues will be given. This seminar will not only focus on our findings but will also engage the audience in a discourse on the potential applications and future directions of our work.

These deliverables aim to provide a holistic understanding of the relationship between water temperature and a myriad of environmental factors in fast-moving aquatic ecosystems. Through this undertaking, the aim is to equip future researchers and other stakeholders with the knowledge and resources necessary for effective and accurate water temperature modelling, and consequently, contribute to efforts towards aquatic ecosystem management and conservation.

# References