University of Southern Queensland

Faculty of Health, Engineering & Sciences

An Investigation of the Predictive Accuracy of Salinity Forecast using the Source IMS for the Murray-Darling River

A dissertation submitted by

Harry James Mccullagh

In fulfilment of the requirements of

ENG4111/ENG4112 RESEARCH PROJECT

Towards the degree of

Bachelor of Engineering (Civil)

Submitted October, 2016

Abstract

The Murray Darling Basin (MDB) is Australia's largest and most important river system. Today, the Murray Darling Basin Authority (MDBA) manages and operates the river system through the oversight of key components such as water storage, quality, markets, trade, sharing and salinity. In order to provide defensible operational decisions and enable effective planning, the MDBA has developed a model of the Lower Murray Darling River using the Source Integrated Modelling System (IMS).

A key functionality of the model is the ability to forecast salinity. The forecasting of salinity enables justification of key water sharing and management decisions in relation to their effects on future salinity levels. In order to predict salinity, the current method is driven by three key inputs being salinity concentration (mg/L), flow (ML) and inflow salt load (Tonnes). Currently, salinity and flow are forecast using trend or average functions while inflow salt load is forecast using the average of the most recent month extrapolated forward.

This research project worked to determine the current accuracy of salinity predictions within a new Source model and investigated methods used to estimate and forecast additional salt loads between the reaches. The project worked to improve the model prediction through investigating a variety of data smoothing methods in order to determine whether monthly averaging is the best representation of including the salt inflow loads within the current model. The project then worked to refine the existing forecast method using two approaches: one being trend extrapolation, and the second being application of an Artificial Neural Network (ANN).

The results of the data smoothing analysis indicate that monthly averaging is the best representation of additional salt inflow used within the model. The results of the forecast analysis indicate that rather than using the average of the most recent month for forecasting, trend methods may provide a more effective option. Finally, the research found that the developed neural network was unable to recognize patterns present in the salt inflow data enabling an effective forecast. However, the research highlighted that the application of artificial neural networks are well suited to the prediction of water resource variables such as salinity and would make an excellent option for future research.

University of Southern Queensland

Faculty of Health, Engineering & Sciences

ENG4111/ENG4112 Research Project

Limitations of Use

Dean

Faculty of Health, Engineering & Sciences

CERTIFICATION OF DISSERTATION

I certify that the ideas, designs and experimental work, results, analyses and conclusions set out in this dissertation are entirely my own effort, except where otherwise indicated and acknowledged. I further certify that the work is original and has not been previously submitted for assessment in any other course or institution, except where specifically stated.

Harry Mccullagh

0061046419

Signature ____________________

Date _____/_____/_____

ACKNOWLEDGMENTS

I would like to thank the many staff from the University of Southern Queensland for the effort and time that they have invested in providing me with an excellent education.

Many thanks is needed for both eWater and the Murray Darling Basin Authority in providing assistance during the course of the project along with the availability of both software licensing and a current Murray Darling Basin Model. In particular, I need to thank Mr Alistair Korn, Dr Matthew Bethune and Mr Jabed Bhuiyan who provided external supervision and assistance.

I wish to thank my family and friends for their support throughout the course of my studies and the understanding when it came to prioritizing study over them.

Finally, I wish to thank my supervisor Dr MD Jahangir Alam who provided a constant source of feedback and knowledge. The ability to quickly and easily schedule a meeting and discuss ideas throughout the project was greatly appreciated, Thank-you.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

Glossary of Key Terms

ANN: Artificial Neural Network

CRC: Cooperative Research Centre

EC: Electrical Conductivity

IMS: Integrated Modelling System

NSE: Nash-Sutcliffe Model Efficiency Coefficient

MDB: Murray Darling Basin

MDBA: Murray Darling Basin Authority

RMSE: Root Mean Square Error

CHAPTER 1: INTRODUCTION

1.0 BACKGROUND

The Murray Darling Basin (MDB) is Australia's largest river system which provides water to 39% of Australian agricultural farms (Australian Bureau of Statistics, 2008). Modelling the MDB represents a highly complex task. Its catchment features five basin states resulting in highly varying landscapes, climates and demands. As a result of Australia's heavy reliance on the system, extensive research and policy reform has been undertaken in order to ensure that the river system is effectively managed.

Today, the Murray Darling Basin Authority (MDBA) manages and controls the Murray Darling River system through the oversight of key components such as water storage, quality, markets, trade, sharing and salinity. The MDBA has recently developed and adopted a daily time step model of the River Murray and Lower Darling which will be used to analyse and support decision making within the river system. The model has been built within the Source Integrated Modelling System (IMS) platform developed by eWater and its Cooperative Research Centre (CRC) partners. eWater is an Australian Government owned, non for profit organisation that has been developed to facilitate the implementation of the National Hydrological Modelling Platform in Australia.

The Source Model provides a diverse range of features enabling the analysis of flow routing, storage, ground water interaction, water management rules, operations, demands and supply, whilst also providing constituent generation and transport models. The flexible nature of the software ensures that new scientific research may be utilised within the system as it becomes available. This aims to ensure that Source remains at the forefront of analysis and decision making within the Australian water resources sector.

Previously, salinity was modelled by the MDBA using the MSM-Bigmod model, which would model water management decisions such as the operation of storages, water accounting, resource assessment, irrigation demand and simulate water flow and salinity routing computations on a monthly time-step as described by Close et al. (2003). This modelling task was increasingly complicated as a result of downstream inputs being dependent on upstream outputs and different basin states using different models that featured varying time steps and inputs.

The Water Act 2007 required the MDBA to build an integrated model running on a daily timestep which linked existing state models based on different modelling platforms into an Integrated River System Modelling Framework as described by Bethune et al. (2015). The Murray Darling Basin Plan 2012 Implementation Agreement (MDBA 2012) identified Source as the standard for future water resource plan accreditation. A National Hydrological Modelling Platform was required in order to ensure optimal management of the river system.

In 2013 the development of Source models by the MDBA began. The current Source models developed by the MDBA are now being tested and compared to the existing MDBA operational and planning models in order to build confidence in both the Source platform and the models themselves. Currently discrepancies exist between the new models and previous models and ongoing work is being undertaken in order to understand the underlying causes of the differences (Bethune et al. 2015).

An important functionality of the Source program is the ability to forecast salinity within the system. The MDBA is currently developing a six-month salinity forecast plugin in order to aid management decisions with particular focus on the scheduling of pumped water extraction. Currently, salinity forecast is driven by three variables being; salinity level (mg/l), flow (ML) and additional salt inflow (Tonnes). The development of accurate forecast methodologies leads to refined and more defendable management decisions which will aid in ensuring the effective long term management of salinity within the Murray Darling River system.

This research project works to examine the current accuracy of salinity modelling within the MDBA model of the Lower Murray, and examines the method of salt inflow calculation and forecast. The project works to calculate the salt loads for a case study reach and examines its effectiveness when compared to having no salt loads. The research then compares the calculated salt load to current salt loads being included to investigate whether the monthly averaging is the best representation of including the salt loads, or whether alternative smoothing methods can improve the predictive capacity of the model. In addition, the project worked to examine the best way to forecast additional salt inflows in providing useful forecasts.

The effective modelling of salinity is critical in ensuring optimal economic growth within the agricultural sector along with providing an accurate indication of the effects of different water usages. Salinity modelling enables defensible predictions, leading to enhanced decision making, optimised salt inception processes and enhanced scheduling of water extraction. Overall, salinity modelling assists in ensuring that the salinity targets for future decades are reached in the most efficient way possible.

1.1 Project Aims and Objectives

The primary objective of the project is to investigate the methodology of additional salt load calculation and forecast by the MDBA using the Source IMS. This will be completed with the aim of improving the existing modelling approach through enhancing the predictability of salt inflow loads, which in turn leads to more accurate salinity predictions. The specific objectives required to complete the project include:

1) Obtain and review the exiting model of the Murray and Lower Darling River.
2) Complete a literature review that examines salinity in the MDB, policy reform leading to the adoption of Source, advances in computational modelling and advances in forecast approaches.
3) Analyse the effectiveness of the current modelling approach used in the Source Program.
4) Hindcast the flow and salinity level in the river for a historical period and evaluate the accuracy of the existing model.
5) Calculate the theoretical salt inflow which will form the input for the data smoothing and forecast analysis.

6) Manipulate the existing method of salt inflow forecast in the aim of developing a greater forecast methodology suitable for adoption in the new MDBA salinity forecast plugin.
7) Investigate the feasibility of applying an artificial neural network to recognise patterns in salt inflow in order to provide predictions.
8) Evaluate the accuracy level of the salinity forecast.

The key aims of this research project were to investigate and refine the existing salinity forecast techniques to develop a more accurate predictive salinity model. This was completed by investigating the best way to estimate the incoming salt loads and determine the most effective way to use this information in order to provide useful forecasts.

1.2 Methodology Overview

In order to complete the project objectives, several steps were required. An initial project preparation phase was required in order to develop the necessary knowledge to use the Source program and understand the modelling methodologies used by the MDBA. This stage of the project involved completing the literature review, gaining access to the software and completing a sensitivity analysis in order to understand the Source program and associated algorithms.

A case study was then required for the analysis, which featured a number of key requirements. It needed a distance great enough that an unaccounted salinity difference could be recorded. In addition, the case study required gauges featuring data of sufficient quality to enable analysis. A Hindcast of the model was then completed in order to gain an understanding of the accuracy and validity of the models salinity modelling capabilities.

Salt inflow was then calculated using the methodology previously applied to the Bigmod model. This salt inflow was then smoothed using a variety of methods in order to determine if monthly averaging was the best technique or if alternative solutions yielded a better result.

The calculated salt load was then forecast using a variety of trend functions in order to determine if trend extrapolation methods could more accurately predict future salt as opposed to the monthly average extrapolated forward. Finally, a neural network was tested on the additional salt load in order to determine if machine learning and pattern recognition could successfully recognise trends that would enable accurate forecasting. The results of these tests have then been discussed and recommendations for future research have been provided.

1.3 Consequences of research/ Ethics

There are a number of possible consequences resulting from the project work:

1. Through the calculation of the additional salt inflow, the project effectively provides a new calibration to the Lower Murray model. If increases in model performance are achieved, this may provide guidance to the MDBA and the wider research community regarding methodology for model calibration and salt inflow calculation.

2. The investigation of forecast methodologies has the potential to increase in the accuracy of additional salt load and thus salinity forecast. If this is achieved, positive ramifications of the work include upgrade of the existing methods used by the MDBA and in turn, greater confidence in water management decisions as a result of more accurate salt load forecasts.

3. The recommendations for future research provided in the conclusions of this report may guide future projects and therefore needs to have adequate justification.

Ethical consideration of the project work includes:

1. The research will provide an assessment on the current accuracy of the model which provides a highly important function in the context of future water management decisions. Consequently, there is an ethical responsibility to provide accurate and unbiased results and discussions, while also providing justified recommendations.

1.4 Limitations Of The Project

A number of limitations were present in the project which ultimately shaped the direction of the research.

- As a result of the complex nature of the software and resources required to develop models within Source, the project will be limited to the model provided by the Murray Darling Basin Authority.
- As a result of not having access to the MDBA hydro database, the MDBA forecast plugins and the use of the MDBA operational models, assessment of the forecast functions were not available for the project. As a result, only the additional salt inflow was forecast, not the upstream flow and salinity as this required application of the MDBA plugins and hydro database. Assessment of forecast accuracy was completed via comparison with data provided by the MDBA. The experiments were based on determining the best ways to use historical additional salt load in order to improve the performance.

1.5 Expected outcomes and Project Justification

It is expected that this research project will provide multiple benefits for future users and the wider research community. Firstly, the project is designed to analyse the Source software which will provide valuable peer assessment regarding the validity of the constituent modelling method adopted in Source. The project will provide both future users of the software and those hoping to conduct further water quality modelling research with an up to date summary of research into modern day salinity modelling practices.

The Murray Darling Basin Authority has developed the Basin Salinity Management 2030 Plan (BSM2030), in which a key focus is stated as:

> "The investment in knowledge to reduce uncertainty and potentially avoid the need for future capital investment in new joint works and measures regarding salinity measurement" (MDBA, 2015).

The project will therefore be providing work that directly fulfils the BSM2030 plan by aiming to increase the accuracy of salinity forecasts. The calculation of additional salt inflow and comparison with current calculations may provide an enhanced inflow timeseries that enables more effective model performance. In addition, it may provide peer review and justification of the existing processes. The analysis of additional salt inflow prediction methods will provide valuable research regarding the best ways to use historical information to provide forecasts into the future. Finally, if a more accurate salinity model and forecasting process can be developed, future users will have more confidence in salinity forecasts. This will have positive repercussions as more accurate results allow for defendable river management decisions.

1.6 Structure of the Thesis

The dissertation is comprised of the following sections:

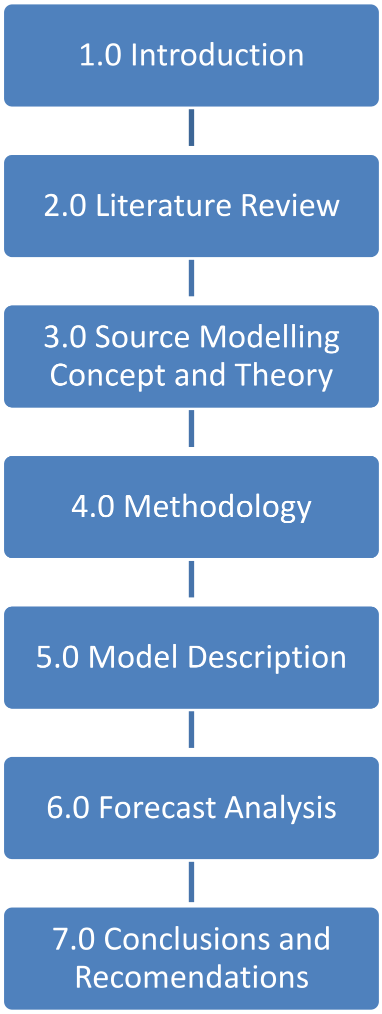


Figure 1.1 Report Structure

CHAPTER 2: LITERATURE REVIEW

2.0 CHAPTER OVERVIEW

Chapter two provides a literature review which examines the current methods used to model flow and salinity within the Murray Darling Basin and examines the demand and subsequent process that has led to the development of the Source Integrated Modelling System. The review then examines the fundamentals of salinity modelling within the program and evaluates the outlook for salinity levels and future of modelling techniques within the Murray Darling Basin. Finally, the literature review examines the advances in forecasting methods, with particular focus of artificial neural networks, and examines the potential of alternative methods as future forecasting methods.

2.1 LITERATURE REVIEW

2.1.1 SALINITY IN THE MURRAY DARLING BASIN

Salinity is defined as the concentration of dissolved salt in water and is expressed in the form of concentration (mg/L) or electrical conductivity (EC) as described by the Murray Darling Basin Authority (2015). The Murray Darling Basin Authority (2015) states that:

> "Salt in large quantities, occurs naturally in the Murray–Darling Basin. In high concentrations, salt can affect ecosystem health, impact on drinking water, and cause economic loss in irrigated agriculture".

Nielsen et al. (2003) states that current research suggests salinity effects ecological health once salinity exceeds 1000 mg/l. The New South Wales Office of Environment & Heritage (2015) categorises the causes of salinity into four steams being dryland, irrigation, urban and industrial salinity. Dryland Salinity is widely regarded as the major cause of salinity in the Murray Darling Basin (Peck et al. 2003). Dryland salinity refers to salinity caused due to the mobilisation of accumulated salts in the soil surface and groundwater. Dryland salinity has been compounded due to Australia's large agriculture and grazing industries.

Traditionally native plants featuring a deep root system would absorb the water before saline ground water could rise to the surface (Department of Sustainability, Environment, Water, Population and Communities, 2012) (Van Dijk et al. 2007). The literature suggests that the spread of dryland salinity in the Murray-Darling Basin has been caused due to the clearing of native vegetation for grazing and agricultural land use (Peck et al. 2003). This has caused water tables to rise and promote saline ground water to the surface in which it can then flow and accumulate in Australia's inland river systems.

Irrigation salinity is a result of improper irrigation techniques exacerbated by factors such as: inefficient water use, poor drainage and/or irrigating on unsuitable soils. Irrigation salinity occurs when the surface strata becomes over saturated resulting in the water table shifting towards the surface in a cone shape. This mobilises salt accumulated in the underlying soil layers (The New South Wales Office of Environment & Heritage, 2015). The effects of irrigation salinity are greatly exacerbated when the water used for irrigation is already saline; a factor often associated with heavy bore water irrigation usage.

Urban salinity occurs, similar to dry land salinity, whereby the ground water table rises due to the clearance of traditional deep rooted vegetation for residential and commercial development. Urban salinity can often increase salinity in coastal areas when salt is transferred through rain and wind to buildings where it dries and is eventually washed into soil (The New South Wales Office of Environment & Heritage, 2015).

Finally, industrial salinity is categorised as salinity resulting from industrial processes that increase the salinity effluent output generated from cities. For example, coal fired power stations boil water to generate steam leaving concentrated levels of salt in the remaining water. This concentrated salinity must then be disposed of resulting in higher levels of downstream salinity if not effectively managed. (The New South Wales Office of Environment & Heritage, 2015).

The Murray Darling Basin Authority (2015) states that the only way in which the salinity generated inland can be disposed of naturally is through mobilisation to the ocean, this is largely dependent on water flow, water use and climatic conditions.

2.1.2 Salinity Management in the Murray Darling Basin

As Australia’s environment has evolved, thus has the way in which it manages its natural resources. In order to ensure that the negative impacts of salinity were minimised, significant policy reform has been completed through collaboration with the Australian Government and industry bodies over previous decades.

Lenblanc et al. (2012) states that the first piece of policy developed to manage the river system was The River Murray Water Agreement which was implemented between 1915–1917, and remained in use for over 70 years. As a result of increasing salt loads associated with growing population and industry, the 1988 Salinity & Drainage (S&D) Strategy was developed as the first interstate salinity management process.

In 1993 the Murray Darling Basin Act was signed by all catchment states and territories which enabled greater transparency and cooperation between States, a factor that had long proved problematic. The 1999 Basin Salinity Audit highlighted that saline groundwater was being mobilised as a result of rising water tables associated with land use changes across the Murray Darling Basin (Murray Darling Basin Authority, 2001). The 2001-2015 Management Strategy developed by the Murray Darling Basin Commission saw enormous gains the management of salinity within the MDB. This strategy was developed with the goal of reducing salinity within the Murray Darling Basin through salt interception schemes, greater planning of land use and new management processes. A key result of this plan saw for the first time, a basin salinity target set at Morgan. The target was to maintain the modelled average daily salinity level of less than 800 EC for at least 95% of the time. The Basin Salinity Target was met for the first time in 2010.

The success of this strategy formed the basis for river management within the Murray Darling Basin. The current basin strategy exists as The Basin Salinity Management 2030 (BSM2030) which builds upon the successful salinity management processes produced by The Basin Salinity Management Strategy 2001-2015.

It has been forecast that under current levels of development and management the Basin Salinity Target at Morgan can be met until approximately 2035 (The Murray Darling Basin Authority, 2015). However, the 2001-2015 Basin management highlights that due to increasing salt mobilisation, continued efforts are required (Murray Darling Basin Authority, 2015). (Murray Darling Basin Authority, 2015) has stated that:

> "'Business as usual' would mean that the reduction in lower River Murray salinity achieved over the last decade would be cancelled out within 20 to 30 years, and median salinity levels would exceed the Australian Drinking Water Guidelines for good quality water within 50 to 100 years".

2.1.3 Modelling methods used in the Murray Darling Basin

The Murray Darling Basin itself is fed via the catchments of five State and Territories. Lenblanc et al. (2012) states that:

> "Maintaining healthy rivers and wetlands has long been a challenge due to a long history of salinisation, intensive water regulation and infrastructure (large reservoirs, dams, weirs), increasing demands from irrigation and urban areas, and the complexities of sharing water allocations between the five States, all governed by different legislation and policies".

In order to ensure that the negative consequences associated with salinity are mitigated, salinity modelling is required to play a critical role. Salinity modelling, enables set targets to be measured based on potential changes in river management. Targets allow policy makers to systematically prioritise public spending and facilitate the introduction of trading in environmental credits (Peck et al. 2003). Water quality models are used to estimate the positive and negative impacts of management actions within the water resources sectors as discussed by Littleboy et al. (2015).

Given the differences in landform, climate and state bodies for which the river system passes, a number of models have been used for river modelling in Australia. The major method previously used included IQQM, Realm and Bigmod.

The Integrated Quantity-Quality Model (IQQM) was initially developed by the New South Wales Department of Infrastructure, Planning and Natural Resources. Its aim was to produce a fully integrated daily model of the River Murray for use in water resource planning as described by Close (2003). IQQM represents the physical world using a series of interconnected nodes and links which can be customised to simulate any river system as highlighted by Simons (2015).

The Resource Allocation Model (REALM) was developed by the Victorian Department of Environment and Primary Industries. The model, similar to IQQM, uses nodes and links to as the primary geographical representation within the model space as highlighted by George (2011). The model uses a combination of water balance models with a network linear program algorithm (RELAX) to transfer water from sources to demand centres (George 2011).

MSM Bigmod was developed by the Murray Darling Basin Commission (MDBC) in order to achieve four primary goals described by Close (1996) including:

1) To replace the existing salt routing models used in planning studies.
2) To make short term flow and salinity forecasts.
3) To analyse historical water quality monitoring data to calculate the solute loads entering the river reaches between water quality monitoring stations.
4) To provide a capability for modelling daily flows.

In order to complete flow routing the Bigmod model divides the larger river system in smaller reaches. The number, length and connection of each reaches are defined by the input file (Close, 1996). The equations used in the modelling of MSM Bigmod can be found in appendix B and are currently used by the Murray Darling Basin Authority. The diagram highlighted in figure 2.1 provided by the Murray Darling Basin Authority (2015) highlights the flow routing models used in each section the Murray Darling River system.

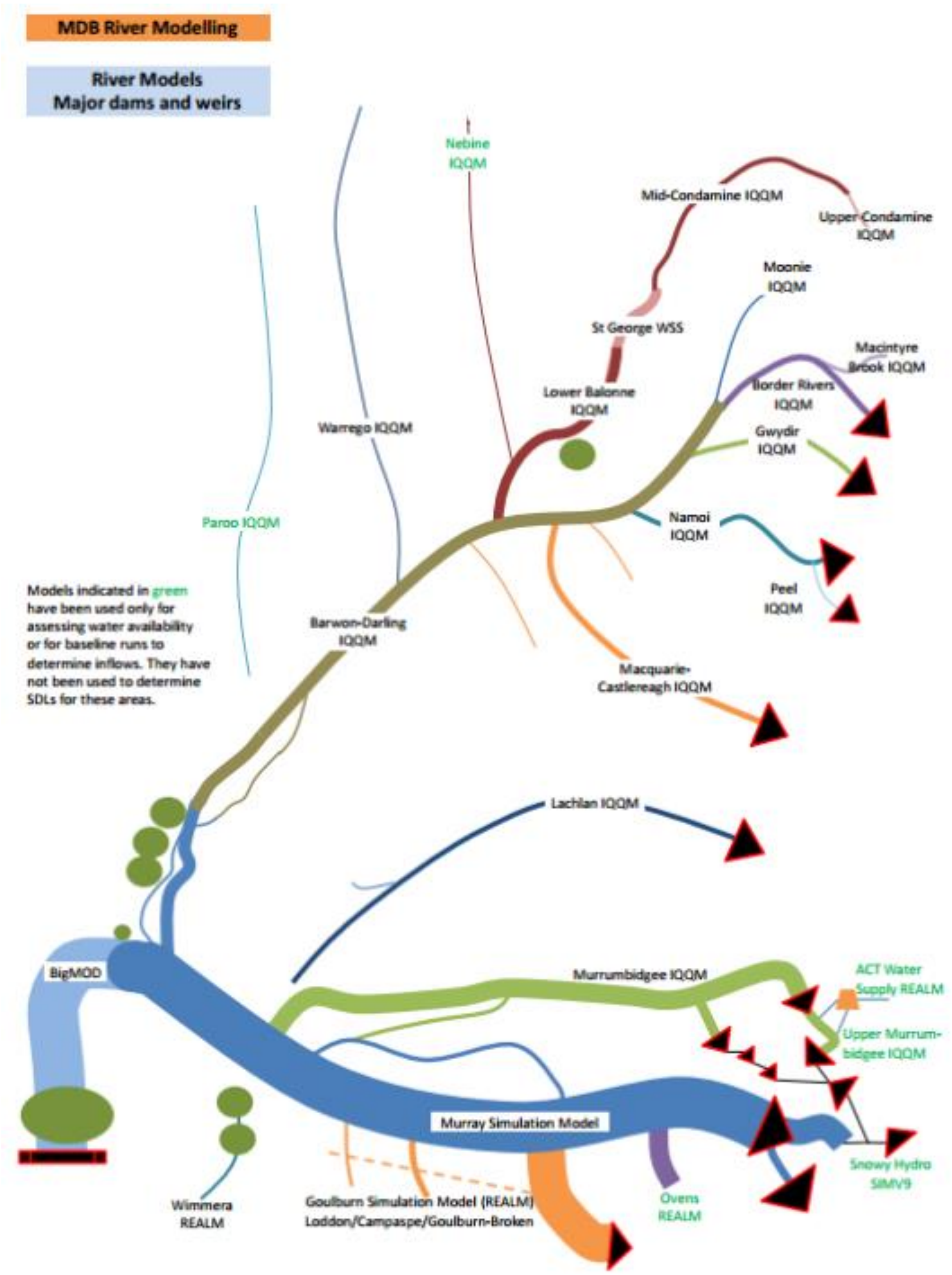


Figure 2.1 Murray Darling Basin Modelling Methods (MDBA, 2012)

While significant gains have been achieved in accurately modelling flow and salinity, the Murray Darling Basin Authority has identified multiple areas of the river systems management that possess capacity for improvement. These are outlined in The Murray Darling Basin Authority 2014 and 2015 annual reports as:

> “The need to invest in new knowledge particularly in relation to reducing the uncertainty in the catchment salinity projections”.

2.1.4 Salinity Outlook

As previously discussed, climatic factors and land use changes have a substantial effect on salinity levels. In order to ensure that salinity levels stay within the targets set by the MDBA continuous adaptation of management, policy, forecasting and modelling techniques are required. Smarter and more efficient salinity measurement techniques are required in order to mitigate the detrimental effects that salinity has on ecosystem health and agricultural yields.

Vaze et al. (2003) examines the effects of land use climate, topography soils and geology on a water and salt balance. The paper then examines the effects of land use change and models salinity using CATSALT and FLOWTUBE. Many authors have examined ways to quantify the effects associated with climate change and land use and have discussed their effects on salinity levels. Connor (2012) estimates the impacts of climate change on supply variability and salinity using a mild, moderate, and severe global warming scenario. The findings state that challenges arising from climate-induced water scarcity are likely to be compounded and these challenges are likely to be under-represented with models that do not include the effects of salinity.

Wei et al. (2011) discusses the findings of (Potter et al. 2010 and Chiew et al. 2008) being that based on the median of 15 global climate models, rainfall in 2030 over the Murray Darling catchment will decrease by an estimated 2% in the north of the basin, and by approximately 5% in the south.

Climatic factors are a critical input needed in the assessment of salinity. Rainfall is a complex variable due to its ability to initially mobilise salt and eventually flush it out the system. Dry conditions cause evaporation which has been found by numerous sources to exacerbate the effects of salinity. Quinn (2011) states that:

> “Land salination and salinity impacts on rivers are exacerbated by drought” and “certain salt sensitive agricultural crops experience progressive yield declines when the salt concentration of applied irrigation water exceeds a certain threshold resulting in economic losses to the agricultural sector”.

The effects of climate change and potential increases in salinity will clearly have an impact on those who depend on the Murray Darling River. As a result, ongoing salinity management is clearly required to face and overcome potential future challenges associated with changing climatic conditions.

2.1.5 Source Development

Extensive research driven by a need for effective management of river systems has resulted in a vast array of models available to hydrologic modelers to utilise when conducting flow and water quality modelling. The Source IMS has been developed due to demand from water users for a single piece of software that utilises and expands on existing methods.

The demand for Source has arisen from a need for an integrated modelling system that can be used as the basis for the Australian National Hydrological Modelling Platform. In addition to this, there has been demand for a contemporary, well documented software solution that enables modelling based on a daily timestep as opposed to monthly as highlighted by Welsh (2013).

Previously the MDBA used 24 hydrological models based on a number of different software packages in order to model the different reaches of the Murray Darling Basin (MDBA, 2012). An example of the different methods used is highlighted in figure 2.1. These models were linked together using an Integrated River System Modelling Framework (IRSMF) developed by CSIRO (MDBA, 2012). Welsh and Podger (2008) highlight the difficulties associated with this process stating that:

> “This process made combining individual models for the whole basin highly cumbersome, as downstream models require the outputs of upstream models as inputs, and these models are often run at different time steps”

Bethune et al. (2015) states that the current Source models developed by the MDBA are now being tested and compared to the existing MDBA operational and planning models in order to build confidence in both the Source platform and the models themselves. In addition to this, the MDBA is working with the basin state Government bodies to replace their existing models with the Source platform as discussed by Bethune et al. (2015).

2.1.6 ADVANCEMENT OF COMPUTATIONAL MODELS

Over recent decades the management of river systems such as the MDB has become increasingly complicated due to rapid socio economic growth as highlighted by Welsh (2013). The Australian Government has invested extensively in organisations such as The Murray Darling Basin Authority, CSIRO, Department of Infrastructure, Department of Natural Resources and Mines, Planning and Natural Resources and Ewater. This research investment has led to the development of a number of salinity modelling programs.

Littleboy et al. (2016) highlights that URS for the National Action Plan for Salinity and Water Quality found that there are over 100 models developed or undergoing development across Australia that will be used to mode salinity management options.

Littleboy et al. (2016) further describes these modelling techniques and approaches used to model dryland salinity by dividing the models into four primary categories, these include:

1. Salinity Hazard
2. Trend Model
3. Scenario
4. River Basin

Salinity hazard models are used to indicate sites at risk though analysis of factors that lead to a predisposition towards high salinity. These factors include cropping, irrigation and/or pasture. Trend models use statistically derived relationships to extrapolate into the future in order to determine future salinity levels. Scenario modelling works to determine the impacts of salinity management actions. Finally, river basin modelling describes the relationship for which salt moves through a system.

Major river routing models exist in the form of; IQQM, MSM-Bigmod and REALM (Littleboy et al. 2016). Welsh et al. (2013) brings to light commonly used models such as Mike SHE, TOPMODEL, HECHMS, WEAP, HSPF, MODSIM, WRAP and describes that while useful, they are not sufficient to model complex policies and water sharing rules.

Littleboy et al. (2016) categorizes scenario models into a number of groups being unsaturated, groundwater, catchment hydrology and salt balance models. Littleboy et al. (2016) states that:

> "Unsaturated models use water balances to predicted how variation in climate, vegetation, soils and land management influence the water balance as part a wider salinity modelling activity, these more complex models have more demanding data requirements that often prevent widespread application".

Examples of unsaturated models include Hydrus as part of Catsalt, PERFECT, APSIM and GRASSGRO (Littleboy et al. 2016). Major ground water models exist as Flowtube, Modflow and Perfect Wlag (Ewater, 2012) which assess long term trends in ground water levels in order to predict when mobilisation of salinity will occur. Catchment models exist in the form of BC2C, CATSALT, Catcher and 2CSalt (Stenson et al. 2011) and are used to model wider catchment areas.

Qureshi (2013) describes the use of Positive Mathematical Programming (PMP) combined with the BIGMOD model developed by the Murray Darling Basin Authority as a means to determine the effects of salinity in terms an economical cost to the agricultural industry. Quinn (2011) highlights that this approach provides a common technical basis for developing numerical load limits that have application in water trading schemes. The use of a salt balance is further echoed by Biggs et al. (2013) as an effective technique to monitor salinity levels.

Biggs et al. (2013) describes the use of a salt balance as an indicator of the stage of hydrologic change after changes in land use and salinity development. Yihdego (2012) applies a water balance model using a bucket model and the Darcy equation to analyze factors such as evaporation and surface inflow/outflow in order to determine the variation of salinity during different seasonal trends.

Quantifying the factors causing salinity has long proved to be the major drawback in developing salinity models. Historically, salinity studies have often been confined to only utilising few factors or processes in the landscape as inputs into salinity models. Biggs et al. (2013) describes the analysis of a wide array of salt inputs as too big of a task to be considered feasible.

Examples of research into salinity resulting from the physical environment include Doble (2004) who uses the factors such as land usage, elevation, and soil type to model ground water salinity using MODFLOW 96. The paper provides a case study analysis which models the effects of groundwater on salinity levels through land use data whilst also highlighting the effects of elevation on groundwater seepage. These inputs are echoed by Amerasinghe et al. (2011) who use similar inputs in their groundwater salinity investigation. The paper utilises groundwater conductivity measurements, water table height, the analysis of local geology, and finally land use in the development of their models. While much work has been conducted in the analysis of groundwater as a salinity input, this will be considered beyond the scope of the project.

A critical component of any effective model exists in its calibration. Zhang et al. (2016) states that when working to calibrate a model featuring both water quality and quantity, the calibration of each factor often conflicts. The paper provides a case study of multi-objective optimisation as a suitable calibration process in order to mitigate the effects discussed above, along with the subjectiveness associated with model calibration.

2.1.1 Source Salinity Forecasting

The MDBA has developed a salinity forecast plugin for use within the Source platform in order to provide forecasts for up to 6 months into the future. Currently room for improvement exists in the models forecast capabilities. For instance, the result of 1-month salinity forecasts at Morgan resulted in a coefficient of determination of 0.11 between observed and predicted salinity Bhuiyan (2016). The salinity forecast itself is driven by three unknown forecast parameters: inflow (ML), inflow salinity (mg/L) and addition salt inflow (Tonnes). In order to forecast these factors, salinity and inflow are predicted using trend and average forecasts while additional salt inflow is predicted using the average salt load for the previous month extrapolated forward as described by Bhuiyan (2016).

Bhuiyan (2016) states that the additional salt load within a reach is estimated using the unaccounted salinity at a downstream reach location. The salt load is then extrapolated forward in order to provide the prediction.

The report produced by Bhuiyan (2016) highlights that the inflow is determined using known flow and an inflow trend forecast which is provided in Source, specifies a target flow along with a recession rate. In order to forecast salinity, two different types of functions are used being trend and average. In addition, climate, diversion, salinity and salt inception scheme processes are required and utilised within the plugin.

2.1.2 Advances in Forecast Approaches

While trend and average models may provide useful short term forecasts, more complex methods exist for forecasting. The application of artificial neural networks (ANN) have been used more recently in order to predict water resource variables as highlighted by Mairer and Dandy (2000). ANN models are well suited to the forecast of water resource variables as the models do not require prior knowledge of the complex physical relationships present in the system. The ANN models work by recognizing often overlooked patterns within the input data. These patterns are then applied as forward feedbacks to new inputs in order to predict appropriate values. Nielsen et al. (2003) provides a simplified relationship between salinity and water level in the following figure:

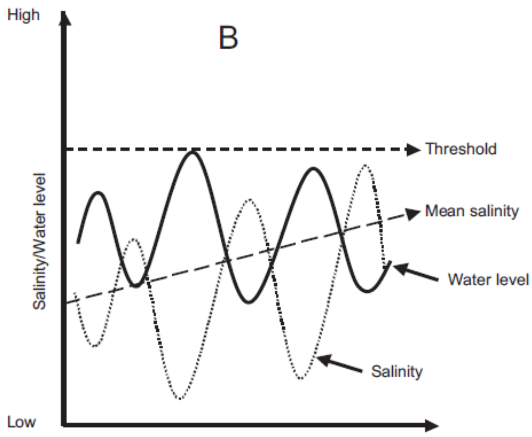


Figure 2.2 Simplified Relationship Water Level Vs Salinity (Nielsen et al. 2003)

Salinity forecast is well suited to the application of ANN's due to a large database of generally high quality data provided by government organisations such as the Bureau of Meteorology. In addition, the data is often correlated and is able to represent the full physical system through data such as flow, salinity, evaporation, rainfall and land use.

The general representation of an ANN is structured by an initial input layer; this layer then moves into what is referred to as a hidden layer where the inputs are multiplied by weightings that are determined through a learning/calibration run of the model. These altered values are then transformed using a summation function. Finally, these values are moved into the next layer for which they are converted into output values through application of an activation function. This generalised description can be highlighted in the following diagram provided by Palani et al. (2008).

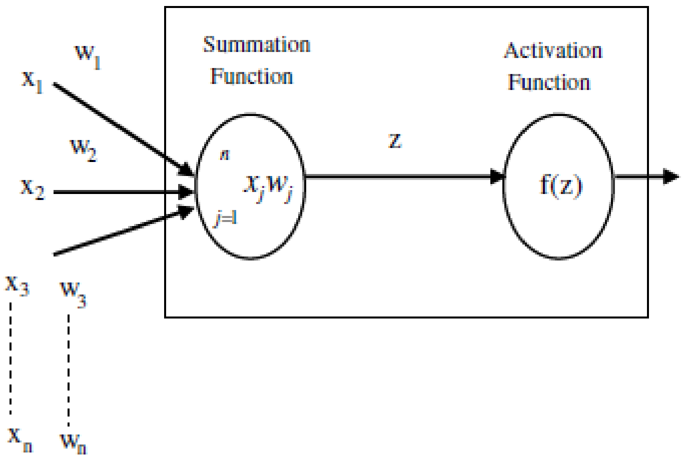


Figure 2.3 Generalized Artificial Neural Network Model (Palani et al. 2008).

In order to develop an ANN a number of requirements must be met, such as a normalization, data pre-processing, data partition, initial weight values, stopping criteria and selection of an appropriate activation function.

A number of studies exist highlighting the application of ANN in the context of forecasting water resource variables. Palani et al. (2008) applies an ANN in order to predict the salinity in sea water among other variables. The study used the Ward Net model architecture associated with Neuroshell software. Chen et al. (2015) utilised the continuity equation and fuzzy pattern recognition combined with an ANN in order to forecast discharge in the Yellow River, Georgia, USA. Here the NSE and RMSE are used as an evaluation criterion in order to determine accuracy of forecast. The ANN accounted for river flow and seasonal trends though the input of daily discharge and rainfall. The result of the model was a NSE of 0.8789 for training and 0.6329 for testing at a 3-day prediction.

A key piece of research that is highly applicable to this project is the work by Bowden et al. (2005) which examines the background and methodology for determining neural network models in water resource application based on previous work done by co-authors Dandy and Maier (2000). This work highlights that neural network learning becomes difficult and cumbersome when too many inputs are used. The article then provides a methodology for selecting the appropriate number of inputs.

Mechanisms proposed for selection of inputs include relying on prior knowledge of the system being modelled, cross correlation, heuristic approach, sensitivity analysis or a combination of the mentioned methods. The report examines the use of a partial mutual information algorithm and a genetic algorithm general regression neural network to assess model inputs for which both were recommended in the papers conclusion. The findings are then applied to a case study for forecasting salinity within the Murray Darling River.

The case study used a feed forward model, trained using the back propagation algorithm. This model type and training methodology is a consistent theme used in developed ANN forecast models particularly within water resource related applications. The ANN was developed and trained using a commercially available software; NeuralWorks Professional II/Plus. Data was divided based on the GA data division method in order to ensure that training, validation and testing were statistically representative of the sample. Ratios of data division were 80% for training and 20% for testing. Linear and hyperbolic functions were used as the transfer and activation functions. The number of nodes and layer architecture were determined by using a trial and error approach. In order to determine a starting point, the following formula was used:

$$N^{H} \leq \frac{N^{TR}}{N^{I}+1}$$

Where:

N^{H} = Number of hidden nodes

N^{TR} = Number of training samples

N^{I} = Number of inputs

The partial mutual information algorithm (method 1) and genetic algorithm general regression neural network (method 2) were used to assess model inputs and determine the lag between variables such as flow and salinity. The results of the case study ranged from a RMSE using EC units between the values of 29.3 and 46.2.

Maier et al. (2000) examines the effectiveness of neural network models highlighting a number of ways in which they can be enhanced. Having analysed 43 papers that use neural networks to forecast water resource variables, all but two of the papers used feed forward networks while the majority use a back propagation algorithm. Maier et al. (2000) highlights the issues effecting successful neural network application in predicting water resource variables as:

1. Effective data pre-processing
2. Data division
3. Stopping criteria
4. Optimized network geometry
5. Internal network parameters

Maier et al. (2000) highlights that an important component of building an artificial neural network is in selecting an appropriate performance criterion. The paper states that the criteria choice will have a significant impact on the model architecture and optimal weight optimisation algorithm.

Banerjee et al. (2011) highlights that the optimal number of hidden nodes and layers are dependent on the complexity of the modelling problem and the goals of the modeler. In addition, a relationship is described for which more hidden neurons and layers are associated with a more complex model.

In addition to ANN models, a variety of alternative methods existed that may assist in improving the Source model and salinity forecast. Tratar et al. (2016) states that exponential smoothing methods provide powerful tools for de-noising time series and predicting future demands. Exponential smoothing itself is based on the assumption that time series are built on components such as level, trend and seasonal effects.

Another widely used approach is the application of regression analysis in order to determine linear, polynomial, exponential and power relationships. Adamowski (2008) compares wavelet forecasting method (WT) to multiple linear regression analysis (MLR), autoregressive integrated moving average analysis (ARIMA), and artificial neural network analysis (ANN) for forecasting daily stream flows at 1, 3 and 6-day lead times. In order to develop the linear and nonlinear relationships regression models were developed using Microsoft Excel. The results found that Wavelet transformations combined with ANN yielded the best forecasting followed by ANN, nonlinear regression, linear regression and auto regression integrated moving average (ARIMA) respectively.

In addition, Adamowski et al. (2012) compares the same variables for water demand in summer and winter in order to determine seasonal trends. The report found that while over the short term of 2 days the wavelet transformation was better at forecasting when compared to the above methods, ANN was superior in 6-day forecasting.

Autoregressive integrated moving average models (ARIMA) provide a means for forecasting. Kim et al. (2016) describes that ARIMA models place more importance on newer values as opposed to older values in generating forecasts. This occurs as the forecast constantly updates as new values are predicted.

Using the information learnt in the literature review, this project will work enhance forecasting approaches using a nonlinear auto regression neural network. This will work to utilise the pattern recognition capabilities of neural networks in order to provide salinity predictions. In addition to this, linear and nonlinear regression will be tested in order to investigate the effects of using long term trends as opposed to the use of the single monthly average in forecasting additional salt inflow.

2.2 SUMMARY

As Australia continues to develop and grow its economy, the role of the water quality modeler will be ever more important. The combination of reduced rainfall and a change of climate resulting in less consistent rainfall, increased agricultural use and deforestation all combine to place stress on the Murray Darling System. In order to manage these changes, forecasts and predictions will be critical in providing justification and defence of key policy and water management reform. It is only through the continued refinement and expansion of this software and scientific methods that will ensure that Australia features a safe and sustainable water supply for generations to come.

Over previous decades the MDB as seen considerable reform from a policy standpoint while also having been forced to adapt in an environment filled with considerable options for water quality and quality modelling. In order to model the system in the most efficient way possible, the MDB has selected the Source integrated modelling system. This highly adaptable, robust piece of software will enable the MDBA to successfully model the arising challenges of the system for the immediate future. Over recent decades' significant research has been completed regarding forecasting techniques in a wide range of sectors. Today, the application of ANN's appears well documented as an effective avenue for prediction of water resource variables.

CHAPTER 3: SOURCE MODELLING CONCEPT AND THEORY

3.0 CHAPTER OVERVIEW

This chapter provides an overview of the modelling concepts and theory which form the foundation of the Source software, as relevant to the research project. The objective of this chapter is to provide the necessary knowledge and understanding of the mathematical and scientific principals being utilised to model salinity within the Source program. Additionally, this chapter aims to provide a description of the hydrologic flow method adopted which will form the basis of the analysis.

3.1 SOURCE OVERVIEW

The Source software features a standard Microsoft user interface featuring a range of tabs and toolbars that enable model development, manipulation and display. Scenarios or operations can be viewed and edited using a number of methods. These include:

- A geographic editor which provides a geographic representation of the model.
- A schematic display which highlights the models components such as nodes and links.
- A tabular editor can be used in operational scenarios to display data.

In order to represent the physical environment within the model space, nodes and links are used as the primary interface. Ewater (2013) states that:

> "Nodes represent places where actions or measurements occur in a river system, where water can be added, extracted, stored, recorded, or change ownership in a model"

And describes links as:

> "Connections between the nodes in which they link, store and route water passing between nodes".

A schematic diagram highlighting this process can be seen in figure 3.1 which is was provided by Welsh et al. (2013).

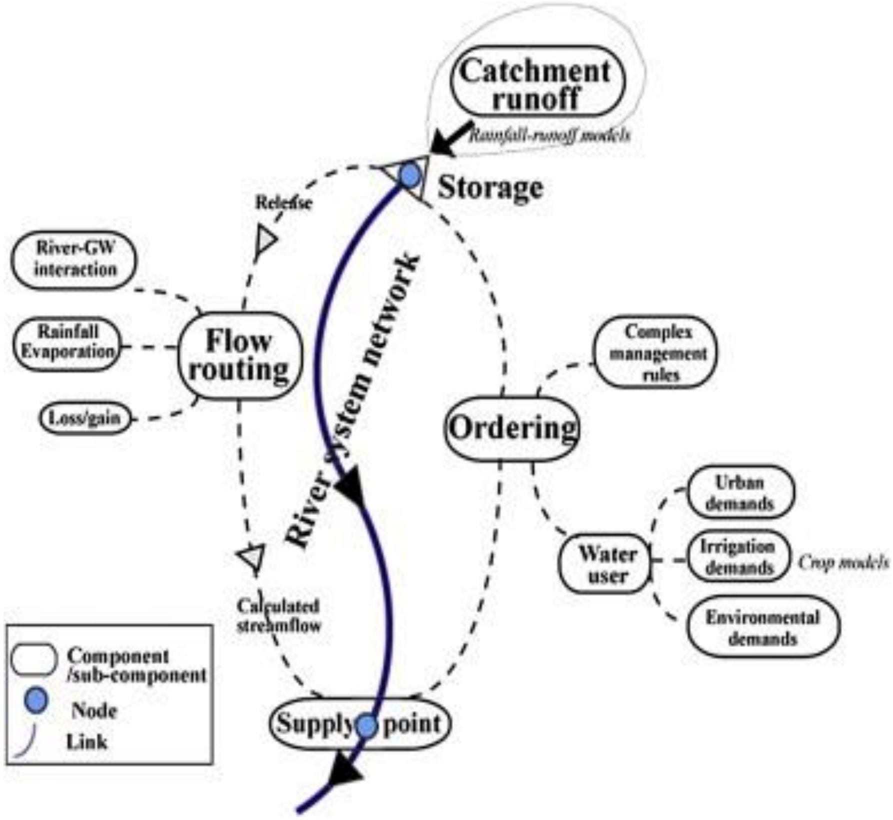


Figure 3.1 Node and Link Representation (Welsh et al. 2013)

In figure 3.1 it can be seen that an upstream node is present for which the boundary condition of the model can be defined. This is generally present in the form of either catchment runoff or as the upstream boundary of a river system. A storage routing link is then used to rout the water downstream. This link is used to model the various interactions the water may have with the physical environment such as ground water interaction, rainfall, evaporation, losses and gains along with changes in ordering and allocation as a result of different demands and water management rules. This highly flexible interface, along with a large suite of settings and available options, ensures that any physical environment can be effectively modelled within the system.

The source software works to encompass the three major modelling methods used in Australia being IQQM, Bigmod and Realm whilst also incorporating new scientific research (Welsh et al. 2013). Source enables users to select from six conceptual daily rainfall runoff models being: AWBM, IHACRES, Sacramento, SIMHYD, SMARG and GR4J (Welsh et al. 2013 and eWater 2012). The Source program itself can be used in either operational or planning applications. This involves use of either the river operations or river management scenario modes.

3.2 Hydrologic Flow Simulation

In order to model the Murray Darling River system, the schematic editor was used in order to model the river flow and salinity interactions. This model utilises nodes and links to represent physical factors and to provide the necessary input into the model space. The nodes and links used can be seen in figure 5.3 which highlights the schematic model of the Lower Murray which was used in the project work. The following section discusses the purpose and characteristics of the nodes and links featured in this case study site used for the analysis. In addition, this section provides an overview of some of the key functionalities of the Source program that are relevant to the project work.

3.2.1 Nodes

Nodes have been used throughout the model in order to provide a mechanism for data input and output interactions within the model. Source features a large array of node types which feature different settings and input functions which are configured in such a way that is representative of the real world. The nodes used within the model can be summarized as:

Confluence Node

The Confluence node is used to represent the section of the river in which two reaches converge and become one. The characteristics of the confluence node are largely dependent on whether the upstream branches are regulated or unregulated.

Controlled Splitter Node

The Controlled Splitter node is used to represent the process in which a single stream flow divides into two. Ewater (2012) states that the controlled splitter has one inflow which is the upstream channel, and two outflows; the main channel and the effluent. Ewater (2012) describes the controlled splitter node as having three distinctive functionalities:

The Controlled Splitter: The controlled splitter is characterized by the presence of some form of structure on the effluent that can be adjusted in response to the flows coming from upstream.

The Uncontrolled Splitter: The uncontrolled splitter has no structure on its effluent and as a result the flows leaving the splitter downstream are a fixed function of the upstream flow rate.

The Loss: If the effluent branch is not connected to anything, the water flowing out the effluent branch will leave the model and appear to be a flow rate related loss. Ewater, (2012) highlights that the controlled splitter node utilises the same scientific theory applied in IQQM, REALM and MSM-BigMod.

Gauge

The gauge node is used to represent a point within the river system that enables the measurement of flow or constituents with observed data or as a location that requires modelled outputs to be recorded for comparison. The gauge point is a location that enables time series data to be loaded and rating curves can be applied to determine water height in meters.

The gauge node can be converted to an unaccounted differences node by enabling the set flow check box. This overrides the modelled flow and sets it as observed flow for the gauge; a technique often used to increase the accuracy of the model or define new boundary conditions.

Inflow

The inflow node is used as a point within the system to allow for additional flow input through catchments or headwaters. The inflow can be loaded in four ways:

1) Through a file as a time series data set
2) Imported from a scenario
3) Mathematical Expression
4) Forecasted (Note: This is highlighted in section 3.3)

Storage

The storage node represents a point within the river system that holds water for long periods of time. The storages operate through maintaining a water mass balance. In order for the storage node to function it must feature dimensions, inflows, outflows and losses within the node feature editor.

Supply Point

The supply point node represents a point in which water may be extracted in order to meet demands.

Water User

The water user node represents a node that is used to model water demand. The node features three components being demand, distribution and storage. The demand accounts for losses in the system as a result of urban, industrial or environmental usages. Distribution enables the modeler to specify how water is supplied to the end user. Storage enables the management of water storages.

Unaccounted Differences

The unaccounted difference node calculates the difference between modelled flow and observed flow in order to ensure that a mass balance is present within the system.

Minimum Flow Requirements

The minimum flow requirement represents the minimum flow required for particular activities. This includes factors such as irrigation, dilution of waste waters, water for environmental flows and urban water supply eWater (2012). In order to enable this function, minimum daily flow requirements are required as an input data set, expression or rule curve.

3.2.2 LINKS

In order to enable the movement of water between nodes, Source utilises three types of links as highlighted by eWater, these include:

1. Straight Through links
2. Lagged Flow links
3. Storage Routing links

Straight through links do not feature many of the available configurations of alternate link options as they simply enable the user to transfer reach inflow to outflow in a single time step. This enables modelers to transfer flow when no reach modelling is needed. Lagged Flow routing enables the modeler to transfer flow with a delay in time steps which represents water delay as it moves downstream.

Storage routing links enable the modeler to rout the movement of water through the use of a range of hydraulic routing methods. It enables configuration of a range of factors such as reach shape and size, ground water interaction, evaporation and surface water input. In order to rout the water through the storage links Source utilises the Muskingum method which is based on mass conservation (eWater, 2013) for which the storage in a link can be described by:

$$Flow\ in - Flow\ out = \frac{ds}{dt}$$

Where:

$$s = Reach\ storage\ m^3$$

$$t = Time\ step\ of\ model\ (s)$$

The Muskingum storage function used in Source is represented by:

$$S = K \cdot (x \cdot I + (1 - x) \cdot O)$$

Where:

$$S = Reach\ Storage\ (m^3)$$

$$I = Inflow\ (m^3/s)$$

$$O = Outflow\ (m^3/s)$$

$$K = Constant\ (s)$$

$$x = Weighted\ factor\ denoting\ the\ importance\ of\ inflow\ relative\ to\ outflow$$

As source utilises average flow, this formula can be written as:

$$\bar{q} = x \cdot I + (1 - x) \cdot O$$

$$\bar{q} = Index\ flow\ rate$$

$$I = Inflow\ (m^3/s)$$

$$O = Outflow\ (m^3/s)$$

$$x = Weighted\ factor\ denoting\ the\ importance\ of\ inflow\ relative\ to\ outflow$$

Source enables the use of variable parameter Muskingum routing for which the x is constant but the k is able to vary with flow. The model used in this research project utilises a lookup table in the form of a piecewise linear editor. The Piecewise linear function describes the relationship between index flow and travel time though enabling the user to manipulate the K variable.

In order to track the constituent particles as they move through the system, the model marker routing or particle tracking as opposed to lumped (fully mixed) routing. Maker routing has been incorporated into Source in order to meet the legislative requirements that exist in The MDBA’s Basin Salinity Management Strategy (BSMS).

3.3 Forecasting

Forecasting describes the process by which water demands, stream flow losses and gains and constituents can be predicted for future periods. Forecasting is a critical component of the Source functionality as it enables users to assess the consequences of unique water use, land management and climatic scenarios. Source evaluates functions for each timestep and returns a single value. Forecasting occurs in two phases, the first occurs through a warm-up phase which involves historical data being loaded into the model. This must be completed before the forecasting phase can be completed. A range of forecasting methods exist within the Source program and are highlighted by eWater (2012) as:

Table 3.1 Forecast Methods (eWater, 2012)

Average	Average over the last specified time-steps
Function	User defined arithmetic expressions/functions
Monthly Average	Daily average for the month in megalitres per day.
Time Series	Supports the inclusion of forecast data using data sources.
Trend	A single target value (either positive or negative) plus a recession rate.

Currently the trend and average functions are used to predict flow (ML) and salinity (mg/l) while the average of the most previous month is used to forecast additional salt load (Tonnes) within the developed salinity plugin by the MDBA.

3.4 Salinity Simulation

3.4.1 Constituent Transport

Constituents are described by eWater, (2012) as:

> "Materials that are generated, transported and transformed within a catchment and affect water quality."

These constituents commonly include sediments, nutrients, contaminants (e.g. pesticides, heavy metals), pathogens and other water quality properties (Ewater 2012). In order to rout the water constituents, a number of approaches can be used within the Source system. When attempting to rout constituents through a reach, two methods are currently available these include Lumped (fully mixed) and Marker Routing (Particle Tracking). Ewater (2012) highlights that the Lumped (fully mixed) approach is more effective when the user wants to measure monthly or annual loads, while the Marker Routing (particle tracking) technique is more accurate and is suitable in measuring concentrations at smaller scales. In order to model the changes that may occur to salinity levels when routing through a reach two models can be used, these include the decay and flux models.

These nutrients and sediments can be modelled in source using three different processes, these include: constituent generation, routing and filtering models. Constituent generation models work to describe how constituents are generated and delivered to the nodes. Constituent routing models describe the movement of constituents along links. Finally, constituent filtering models represent any transformations that may occur between generation and link at an upstream node.

Current constituent models used within Source include: Nil Constituent, Event Mean Concentration (EMC)/ Dry Weather Condition (DWC), Export Rate Model and Power Function.

Nil Constituent

eWater (2012) states that:

> "The Nil Constituent model is used as a substitute constituent generation model where no constituent load needs to be modelled for a given constituent from a given functional unit (FU)."

Event Mean Concentration (EMC)/ Dry Weather Condition (DWC)

The Event Mean Concentration (EMC)/ Dry Weather Condition (DWC) is expressed through:

$$C_{Load} = (SF \times DWC) + (QF \times EMC)$$

Where:

C_{Load} = Constituent load.

SF = Slow flow.

QF = Quick flow.

DWC = Constituent concentration measured during dry weather.

EMC = The flow-weighted average constituent concentration over a storm event.

Export Rate Model

The export rate model is expressed as:

$$C_{Load} = Export\ Rate \times area\ FU$$

Where:

$Export\ Rate$ = Area based scaling factor used to represent the physical amount of constituent exported per unit area per year.

$area\ FU$ = Area of the 'functional unit' taken from the model.

Power Function

The power function is expressed as:

$$Cocentration = A \times (flow)^B + C$$

Where:

A = The coefficient, represents the slope of the curve if plotted on semi-log axes;

B = Represents the "curvature" of the curve; B less than 1 means the curve is convex upwards; and B greater than 1 means the curve is concave upwards.

C = Y Intercept.

3.5 Source Summary

The Source IMS features a highly flexible interface that has been developed based on well-reviewed scientific research and existing modelling methods such as IQQM, REALM, MSM-Bigmod. The model used in the case study makes use of nodes and links in order to represent the real world. While Source enables a vast array of rain fall runoff and constituent generation methods, these components have been defined by the upstream boundary of the river system. The model utilises the physical interactions that occur within the system primarily in the form of evaporation, groundwater, rainfall, and users demands in order to rout the flow downstream. It is the accuracy of this downstream salinity forecast that will be tested and analysed in the project.

CHAPTER 4: METHODOLOGY

4.0 CHAPTER OVERVIEW

Chapter four details the methodology used to investigate the salinity modelling within the current MDBA Source model, along with forecast approaches. The chapter begins with an overview of the project which summarizes the project tasks. The methods and procedures required in each stage of analysis are then discussed.

4.1 PROJECT OVERVIEW

Figure 4.1 highlights the key steps required to complete the research project:

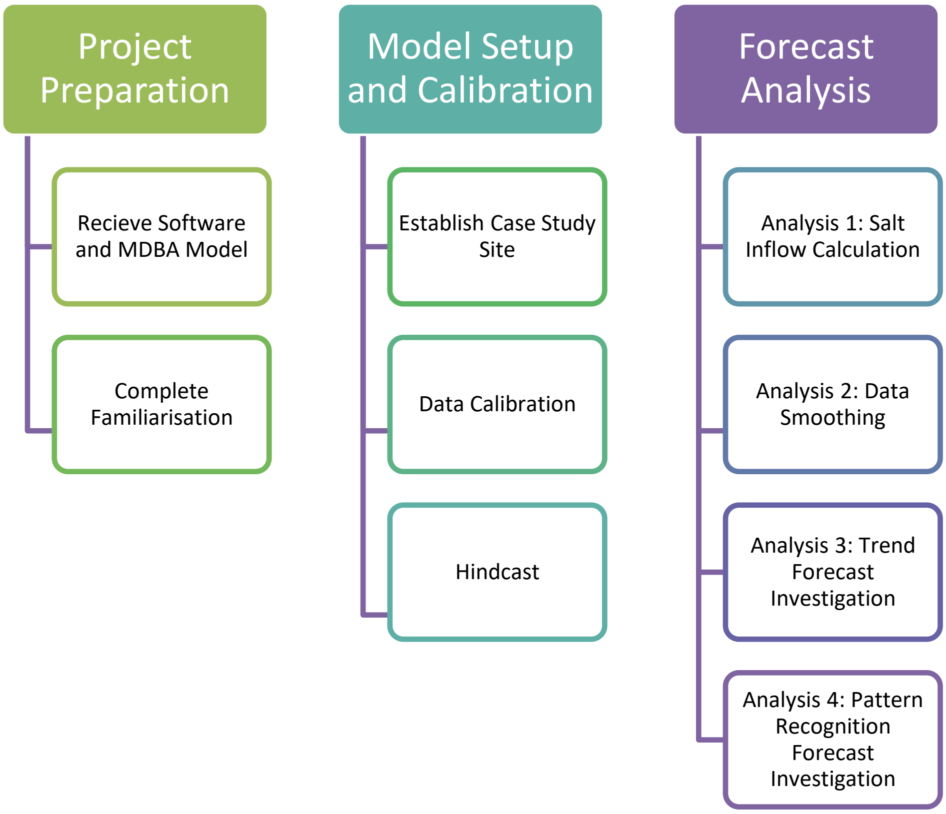


Figure 4.1 Project Methodology Overview

4.2 Project Preparation

Stage one of the project began with obtaining and reviewing the existing model of the River Murray and Lower Darling. After receiving the schematic model provided by the MDBA, a process of familiarization was then completed. Initially, research was conducted in order to establish the necessary knowledge to use the software effectively. This research involved study of the Source user guide and scientific reference guide. A sensitivity analysis was then completed in order to achieve two goals:

1. Understand the importance of different model components and algorithms.
2. Understand the user interface and key functionalities of the Source program.

The sensitivity analysis was completed by adjusting key variables within river reaches such as evaporation, flow volume, reach length and additional salt inflow in order to examine the effect on downstream salinity. The primary objective of the initial preparation phase of the project was to build the necessary foundation of knowledge required to complete the research objectives.

4.3 Model Setup and Calibration

The model was developed by the MDBA, for which the case study site of Lock 5 to Morgan was selected for analysis due to the following factors:

1. Availability of stream gauge flow and salinity data.
2. Length of reach.
3. Simple reach configuration.
4. Importance of reach in monitoring salinity.

In order to increase the accuracy of model predictions, calibrations are completed in which key variables and settings are altered in order to minimize the difference between predicted and observed values. In order to determine whether the current model required further calibration, a hindcast was completed as highlighted in section 5.1.3. Hindcasting was completed at the case study site in order to determine the baseline accuracy of the model in predicting salinity. If a Nash Sutcliffe Model Efficiency Coefficient (NSE) of 0.9 or greater was achieved, no further calibration would be deemed required.

The hindcast was completed through simulating flow and salinity in the river for historical periods based on data availability. The observed and predicted salinity (mg/L) and flow (ML) was then compared with observed values through the application of a statistical analysis. As a result of the highly accurate performance of the model being a NSE=0.959 for salinity and a NSE=0.964 for flow, it was deemed that no further calibration was required.

4.4 Testing Methodology Overview

The testing phase of the project was divided into four sections, which together work to investigate the optimal ways to utilise the historical salt load data in providing salinity predictions. The testing sections include:

a) Calculation of the additional salt load (Tonnes/day) within the case study.
b) Analysis of the effects of different data smoothing techniques on model performance.
c) Forecast of additional salt inflow using trend extrapolation.
d) Forecast of additional salt inflow using an ANN.

The theory behind the testing methodology emerged due to a number of factors that were found upon project initiation. Calculation of the additional salt load was initially required as the salt load provided by the MDBA was already smoothed on a monthly time step, this would not enable proper investigation of the smoothing techniques or forecasts.

The analysis of smoothing techniques was an initial research goal of the project found through consultation with the MDBA. The MDBA was interested in research which would determine if the calculated additional salt inflow used within the model was most effective when smoothed on a monthly time step or if another method resulted in more accurate model performance.

The analysis of the additional salt inflow forecast was a key objective of the project due the identification of the possibility to expand on, and improve, the existing method of salt inflow forecast. The use of trend extrapolation was proposed as an option and would be tested and compared to the existing monthly average method.

Finally, the potential for using an artificial neural network to forecast additional salt inflow within the Source model was tested. This was found to be an interesting concept developed during the literature review, as a number of authors had suggested that neural networks demonstrated the ability to forecast water resource variables. A neural network model was then investigated in regards to predicting salt loads and the results were discussed and recommendations for future avenues of this research were made.

4.4.1 Salt within a reach calculation

Before the future research phases of the project could be completed, the salt inflow needed to be calculated. The first step of the calculation would require summating the total length of the reach which can be summarized in the table below:

Table 4.1 Length of Reaches

Reach	Length (m)
R112	36100
R162	9800
R117	84800
R118	20400
R163	28000
R164	20900
R119	47100
Total	247100

The lengths of each reach were provided in the model schematic from the storage routing links which represented the river sections. The procedure outlined in appendix B and C was then followed for completing a salt balance and determining the required salt inflow in order to minimize the difference between observed and predicted salinity. This process required a number of changes to be made to the model these included:

1. Forcing upstream flow and salinity to equal the recorded flow and salinity.
    a. The Lock 5 salinity was set to the gauged salinity effectively converting the gauge node to an unaccounted differences node.
    b. The flow at Lock 5 was found to be of insufficient quality and thus, flow was set to -9999 at Lock 5, effectively using the SA recorded flow.

2. Set additional salt inflow within the reach zero.
    a. This was completed by setting the salt inflow at all storage routing links to zero.

Through following the above steps, the difference between observed and predicted salinity at the downstream Morgan gauge when no additional salt flow is present could be determined. The next phase involved adding an additional salt load of 1 Tonne/km in order to determine the effect of a unit value additional salt inflow. 1 Tonne of salt was then added for each Km of reach length. The output at the downstream Morgan gauge for predicted salinity could then be recorded.

The following formula was then applied in order to determine the additional salt load per km of reach:

$$Salt\ load\ (Tonnes) = \frac{C - A}{B - A}$$

Where:

A = Salinity at Morgan assuming no salt.

B = Salinity at Morgan assuming the addition of 1 Tonne per km.

C = Gauged salinity at Morgan.

This resulted in a time series that could be multiplied by the reach length in order to calculate the additional salt inflow in Tonnes per day for the entire reach from Lock 5 to Morgan.

4.4.2 Data Smoothing

The goal of the data smoothing analysis was to determine whether monthly smoothing is the best representation of including additional salt loads within the model. The research aimed to determine whether alternative methods exist for smoothing the calculated daily salt loads in order to improve the predictive capacity of the model. A number salt loads were compared in the analysis, these include: no salt, original, daily, monthly, weekly, seasonal, exponential, weekly moving, monthly moving and weighted moving average which are further discussed in table 4.2 below.

Table 4.2 Data Smoothing Methods

Data Smoothing Method	Description of Method
No Salt	The no salt test represents the scenario for which no salt has been added within the reach, this will essentially set a baseline performance of the model.
Original Salt Load	The original salt load represents the calibrated original salt inflow time series provided by the MDBA.
Daily Salt Load	The daily time series represents that calculated daily salt inflow determined in section 4.4.1.
Monthly Average	The monthly average represents the calculated daily salt inflow smoothed on a monthly average based on a sample width of 30 days.
Weekly Average	The weekly average represents the data smoothed based on weekly average using a sample width of 7 days.
Seasonal Average	The weekly average represents the data smoothed based on the basis of its season. This used a sample width of 3 months.
Exponential Smoothing	The exponential smoothing technique is an example of a moving average method, which smooths the data based on the equation: $$s_t = \alpha x_t + (1-\alpha)s_{t-1}, t > 0$$ Where: $$\alpha = \frac{2}{width + 1}$$ $$s_t = smoothed\ observation \text{ at } timestep\ t$$ $$x_t = original\ observation$$ For the analysis a value of $\alpha = 0.333$ resulting from a width size of 5.

Weekly Moving Average	The weekly smoothing works by taking the average of 7 samples, the next value in the sample is then generated by dropping the last sample number and including the sample piece for the current day. This process is repeated throughout the data in order to smooth it.
Monthly Moving Average	This same method is applied to the monthly smoothing but rather a sample size of 30 is used. The longer width moving averages, account for a lag within the data and it's therefore slow to change. For example, the seven day moving average has the capacity to change far quicker than the monthly moving average.
Weighted Moving Average	The final smoothing method to be tested is the weighted moving average which uses the same method applied to the weekly moving average but places more value on newer samples as opposed to older samples. The weighting basis used for the weighted moving average can be highlighted by the following: $x_7 \; = 0.01\; x_6 = 0.02\; x_5 = 0.04\; x_4 = 0.06\; x_3 = 0.9$ $x_2 = 0.12\; x_1 = 0.16$

In order to process the data and perform the smoothing, a combination of Microsoft Excel and Matlab were used. The smoothed time series when then divided based on reach length and applied as additional salt inflow to each storage routing link. The model was then run within Source which resulted in a corresponding predicted salinity time series for each data smoothing method. This predicted salinity value could then be compared with the observed salinity. The NSE value was used as the performance criteria in order to determine the effectiveness of each smoothing method.

4.4.3 Trend Forecast

Currently, when the MDBA forecasts future salt loads the salt load calculation determined in section 4.4.1 is averaged for the previously month (Note: The current month cannot be averaged as a full month of data is not available). The salt load for the previous month is then extrapolated forward in order to complete salinity predictions.

This is a very basic method and does not consider long term trends or patterns that may result in more accurate forecasts. This project works to examine the effectiveness of using trend equations to extrapolate forward in order to predict future additional salt loads. This method is not a current feature in Source and ultimately provides a new approach to forecasting salt inflows. The method assumes that long term trends will provide better salt load forecasts as opposed to only using the last month. In order to determine whether the monthly average is the most effective means of extrapolating future salt inflow, a series of experiments were conducted.

To complete this analysis, the calculated additional salt inflow data will be loaded within Matlab in order to develop the trend equations and to perform the forward extrapolation. The new methods of data extrapolation will include linear, exponential, power and polynomial trend equations. The data will be extrapolated for each day up to a 6-month lead time from three arbitrary dates. These periods will be compared with the calculated additional salt load time series in order to determine the accuracy effectiveness of forecast methods when compared to that of the monthly average.

The current method used to forecast additional salt inflow is through the use of the average of the most recent month carried forward. This will be used as the baseline of analysis and is calculated using:

$$\text{Monthly average } = \frac{\sum x}{N}$$

Where:

$$x = monthly\ values$$

$$N = number\ of\ values$$

In order to complete a polynomial regression of the historical salt inflow data the polyfit code within Matlab was used. The algorithm is presented by the following:

$$y = p_1 x^n + p_2 x^{n-1} + ..+ p_n x + p_{n+1}$$

Where:

$$p = coefficient\ of\ the\ polynomial;$$

$$x = query\ points;$$

$$y = fitted\ values;$$

$$n = degree\ of\ polyniomical\ fit.$$

The Matlab input is completed through the use of:

$$p = polyfit(x, y, n)$$

And;

$$y = polyval(p, x)$$

The polyfit equation fits the data generating coefficients which could be utilised as in equations to extrapolate the timeseries forward. The polyfit equation is based on the least squares method in order to fit the data while polyval evaluates the polynomial at all x values. In order to complete the linear regression, the polyfit 'n' was set to 1. In order to evaluate for 2nd 3rd and 4th order polynomials the polyfit 'n' was set to 2, 3 and 4 respectively.

In order to calculate the power function, the polyfit 'n' was set to 1 however the 'x' and 'y' variables were converted to log values. For the exponential polyfit, the polyfit 'n' was set to 1 while the y values were transferred to log values.

The resulting values generated from the polyfit function could then act as variables in the following equations in order to be extrapolated for the forward lead times:

Linear:

$$y = p(1)x + p(2)$$

Polynomial 2:

$$y = p(1)x^2 + p(2)x + p(3)$$

Polynomial 3:

$$y = p(1)x^3 + p(2)x^2 + p(3)$$

Polynomial 4:

$$y = p(1)x^4 + p(2)x^3 + p(3)x^2 + p(4)$$

Power function:

$$y(x) = p(1) \times x^{p(2)}$$

Exponential function:

$$y(x) = p(1) \times 10^{p(2)x}$$

It should be noted that moving average methods were not applied to the forecast analysis as they were found to quickly convene to a value that was very similar to the original monthly average used.

The forecast methods will be compared using the NSE as the primary performance criteria. Analysis of the dates will be completed at lead times of 3-days, 1-week, 1-month, 3-months and 6-months in order to determine the effectiveness of each method at different stages of forecast. The dates used to extrapolate from have been chosen to represent three scenarios being when the additional salt load appears to be declining, flat and ascending being the 1/1/1993, 1/1/2008 and 1/1/2012 respectively.

The results of the forecast salt inflow were then run though Source in order to examine their effects on salinity performance. By assuming that flow and salinity is known, the experiments will work to examine the effectiveness of using trend methods as opposed to monthly average data. Upon completion of the analysis the results will be discussed regarding their relevance for incorporation and into the current MDBA forecast plugin. In addition to this, recommendations for future research will be discussed.

4.4.4 Artificial Neural Network Forecast

The aim of the ANN forecast analysis was to gain an understanding of the ways in which the historical salt load provided within the Source MDBA model, could be used as an input for a neural network as a potential method for enhanced forecasting. As a result of time constraints, a working model was not achievable, however the research was aimed at gaining refined idea of how an ANN can be developed and to provide recommendations for future research.

As highlighted in the literature review, neural networks provide an effective means of modelling water resource variables as a result of the ability to recognize patterns and relationships between variables based on machine learning. The general neural network design process was followed in the use of the model; the process was provided by Beale et al. (2016) as:

1. Collect Data
2. Create the network
3. Configure the network
4. Initialize the weights and biases
5. Train the network
6. Validate the network (Post training analysis)
7. Use the network

In order to develop the Neural Network, the above process was utilised. Initially, an appropriate software solution needed to be selected. The literature review highlighted the use of commercial software platforms such as Neuroware and Neuroshell to develop the network architecture and train the networks. A desktop study of available options, highlighted Matlab as an effective means for neural network development and application. The Matlab Neural Network Toolbox enables the user to solve a number of forecast problems such as nonlinear autoregressive problems with or without external inputs.

For the project, a nonlinear auto regressive network (NARN) was selected as a mechanism to predict salt inflow. The NARN uses a single input, in this case additional salt load, and applies weights based on the timestep and target data in the attempt to recognize patterns in the data. The selection of a NARN model an attempt to use the patterns within the historical salt inflow fluctuations to generate usable forecasts that would be superior than that of the monthly average.

The neural network was configured using the methods highlighted in the literature review. This included use of a feed-forward back propagation network and use the mean square error as the performance criteria. The tutorial provided by Heath (2015) was used to form the basis of the neural network. All other variables were set to Matlab default settings. A trial and error approach was used to determine an appropriate feedback delay and hidden neuron configuration.

The neural network used to forecast the parameters used in the experiment was based upon a two-layer architecture which is highlighted in figure 4.2. The transfer function used in the neural network was sigmoid function while a linear function was used as the activation function. The sigmoid function is used to normalize the data and present it as a value within 0 and 1. The linear function is used in the final output layer of the neural network in order to approximate the data and convert it back to a usable value. The Matlab script used to initiate the Neural Network is provided in appendix G.

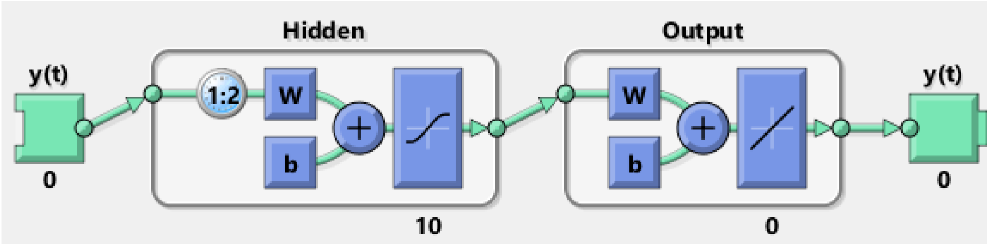


Figure 4.2 ANN Model (Open loop)

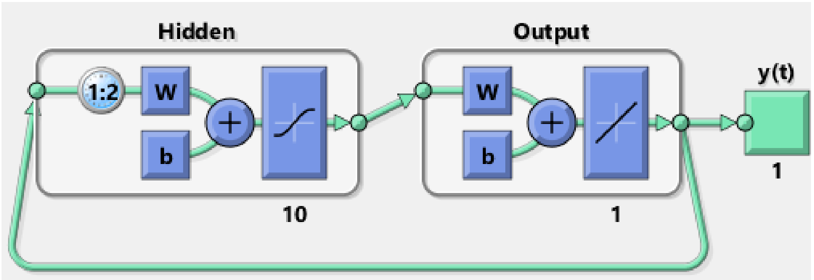


Figure 4.3 ANN Model (Closed Loop)

The methodology for performing the predictions works by first training the neural network based on an open loop. A closed loop is then created which enables multistep prediction to occur. The closed loop makes predictions by continuing to predict based on the timeseries value but without any external feedback, the feedback used in the closed loop is internal feedback which is gained during the open loop training. In order to monitor the performance, the root mean square error was used in both the open and closed loops. In order to reduce the noise of the data, smoothing was utilised through a monthly moving average process. This was completed in order to smooth the data to reduce the risk of over prediction while still allowing for the important patterns to be recognized. A range of different delay and hidden neuron configurations would be tested in order to determine whether a NARN was feasible as a potential forecasting option.

4.5 Statistical Analysis

A number of statistical methods will be required throughout the project in order to analyse the correlation between observed and predicted data, and to justify the conclusions drawn. In order to compare the observed values with the predicted values a number of statistical methods will be used these include:

1) The Nash Sutcliff Model Efficiency Coefficient (NSE).
2) Root mean square error (RMSE)
3) Coefficient of Determination (R^2).
4) Probability of Exceedance.
5) Percentage Bias

4.5.1 Nash Sutcliffe Model Efficiency Coefficient

The Nash-Sutcliffe Model Efficiency Coefficient (NSE) can be highlighted by the following formula:

$$E = \frac{\sum(x_i - \bar{x})^2 - \sum(y_i - x_i)^2}{\sum(x_i - \bar{x})^2}$$

Where:

x_i = Observed data values [mg/l];

y_i = Predicted data values [mg/l];

$\bar{x}$= Mean of observed data values [mg/l].

The NSE value measures of the scatter around a 1:1 line of the observed data vs predicted data in which a value of 1 represents a perfect result. This method is commonly used to determine the predictive power of hydrological models as described by Nash et al. (1970).

4.5.2 Root Mean Square Error

The RMSE is used to provide an indication of the error between two sets of data, for which a value of zero represents a perfect result. The root mean square error is calculated by:

$$RSME = \frac{\sqrt{\sum(y_i - x_i)^2}}{n}$$

x_i = Observed data values [mg/l];

y_i = Predicted data values [mg/l];

n= Number of values.

4.5.3 Coefficient of Determination

The coefficient of determination can be calculated by:

$$R^2 = \left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2\sqrt{\Sigma(y_i - \bar{y})^2}}}\right)^2$$

Where:

x_i = Observed data values [mg/l];

y_i = Predicted data values [mg/l];

$\bar{x}$= Mean of observed data values [mg/l];

$\bar{y}$= Mean of predicted data values [mg/l].

4.5.4 Probability of Exceedance

The percentage bias can be calculated using the following equation:

$$P = 100 \times \left(\frac{m}{n+1}\right)^2$$

Where:

P = Probability of exceedance;

m = Rank from highest to lowest of all daily mean flows;

n = Total number of daily mean flows.

4.5.5 Percentage Bias

The percentage bias was calculated using the formula:

$$Percentage\ Bias = 100 \times \left(\frac{\sum(mod - obs)}{\sum obs}\right)$$

Where:

mod = modelled values.

obs = observed values.

Chapter 5: Model Description

5.0 Chapter Overview

This chapter examines the case study model selected for the analysis. The chapter discusses the rationale for the selection of the case study site as the focus of the dissertation. It also provides a summary of the model's performance in terms of its predicative accuracy while also describing the how the model provides a representation of the physical world.

5.1 Case Study – Lock 5 to Morgan

5.1.1 Case Study Overview

The Morgan stream gauge is the site of The Murray Darling Basin Authorities key salinity target: to achieve less than 800EC for at least 95% of the time. This gauge therefore represents a critical monitoring site for the river system. As Morgan is positioned as one of the lowest gauge sites, it features the accumulation of a larger portion of salt that has been mobilised from upstream reaches.

The model itself forms part of the Lower Murray Catchment which is highlighted in figure 5.1 below. The model begins at Lock 9 in NSW and ends at the Southern Ocean outlet.

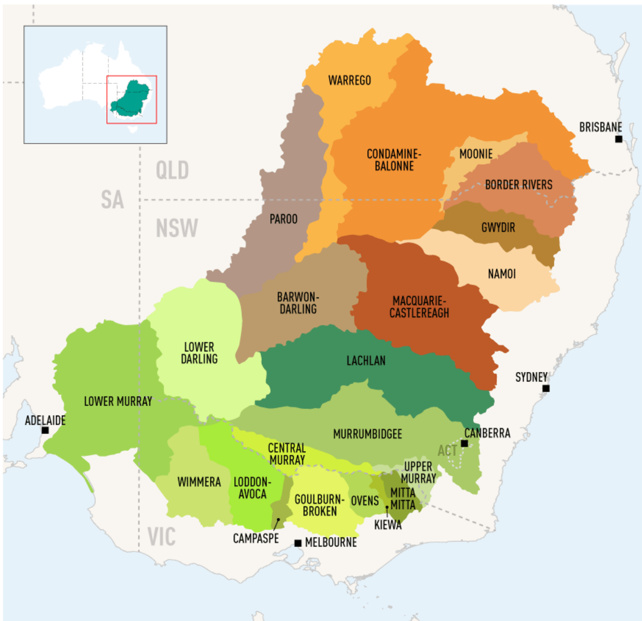


Figure 5.1 MDB Catchments (MDBA, 2016)

A map representing the physical location of key model sites is located in figure 5.2 below. Additionally, the schematic model used for the case study is present in figure 5.3. The case study begins at Lock 5 which is located at Renmark and is represented by an unaccounted differences node, this enables salinity to be set to gauged salinity. The water is then routed into the Berri Gauge, Lock 4, Lock 3, Woolpunda, Waikaerie, Lock 2 and finally Morgan, all of which are represented by gauge nodes. The routing is completed through the use of both straight through routing and storage routing links.

The storage routing links feature climatic data in the form of rainfall and evaporation. In addition to this the storage routing links feature initial conditions and physical descriptions such as reach length, initial flow, and additional salt inflow. A piecewise lookup table is also featured in order to describe the relationship between travel time and flow. Water demands have been presented throughout the model to represent extraction from the system. The final Morgan gauge node provides the end point for the Morgan case study section. This node represents a key stream gauge for which a range of water quality and flow conditions are monitored. This is the point for which all model outputs will be collected and compared with observed data.

5.1.1 Data Preparation

In order to generate the model, a substantial amount of input data was required in the model. This input data is highly critical in enabling the effective development and calibration of the model, fortunately data this was provided by the MDBA.

The climate data features a time series data sets featuring daily rainfall and evaporation (mm) for a range of sites. The flow data present is presented in the form of ML/day. The salinity data contains salinity readings for the gauge points in the form of mg/L, and also provides additional salt inflow data in the form of Tonnes/day.

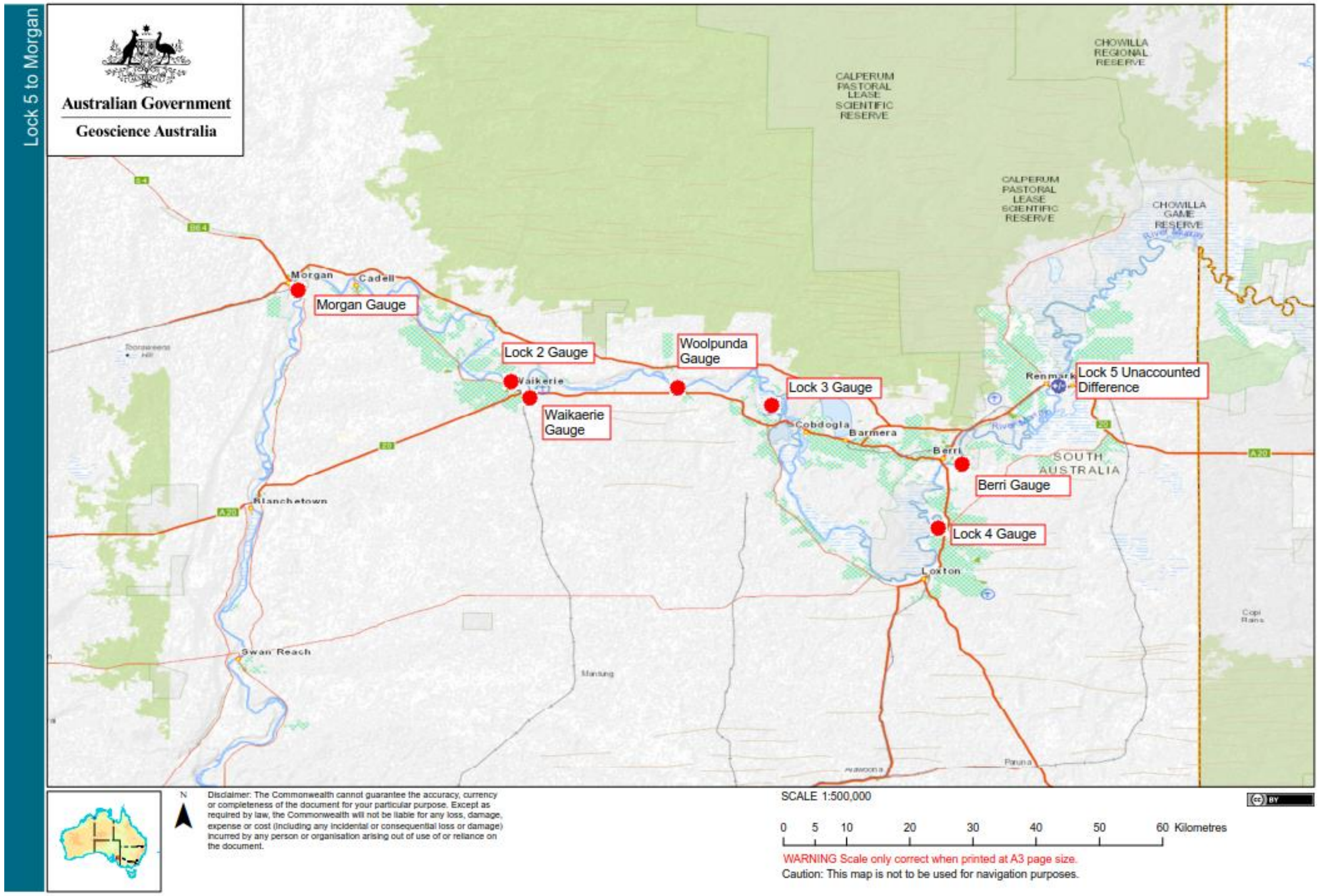


Figure 5.2 Lock 5 to Morgan Map (Geoscience Australia, 2016)

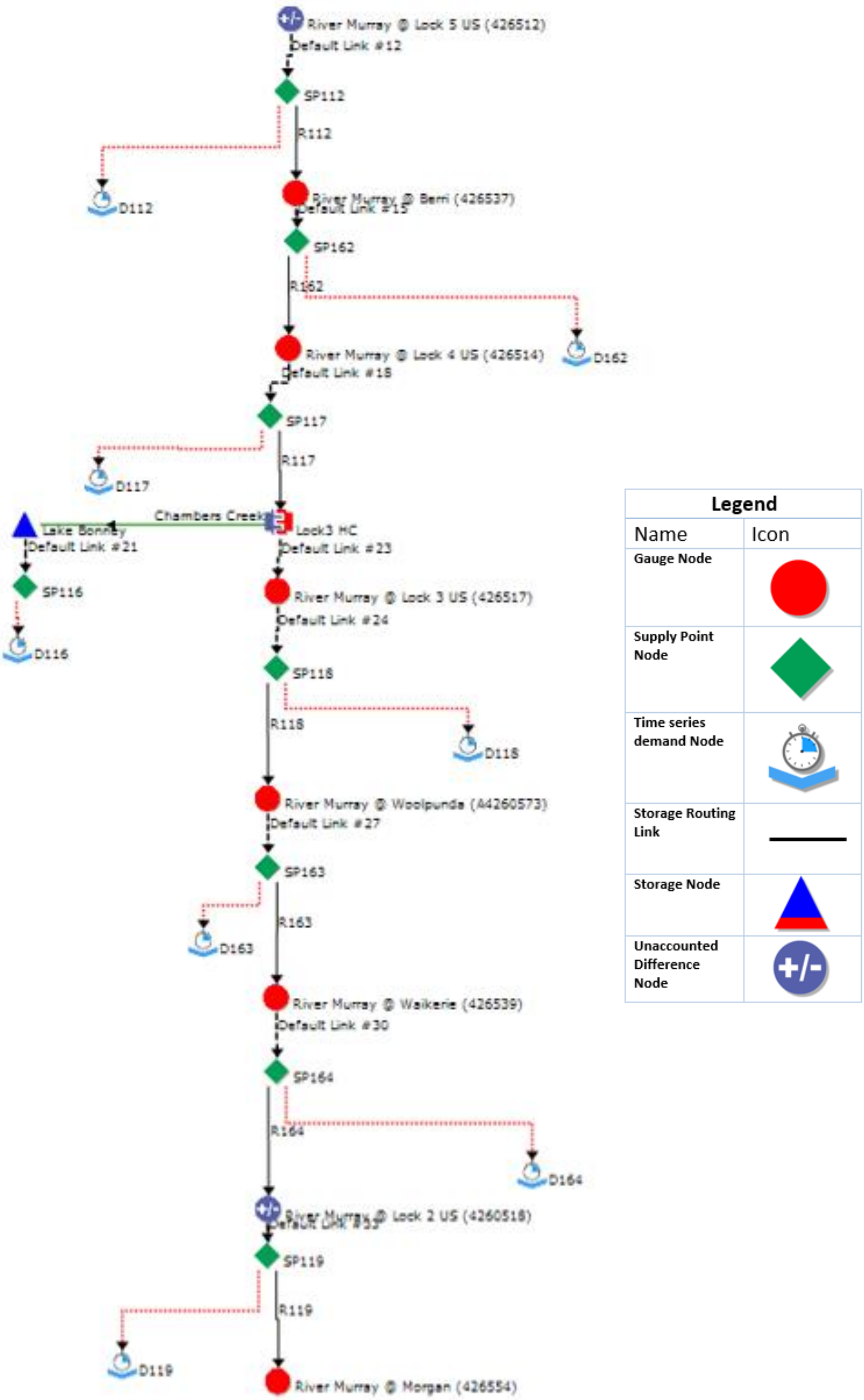


Figure 5.3 Morgan Case Study Schematic Diagram

5.1.1 Boundary Conditions

As the case study section is a portion of the larger Murray Darling Model, the upstream input is governed by the output of the upstream model. An unaccounted difference node at Lock 5 defines the boundary condition of the model. Here salinity is set to the gauged salinity. This represents the values measured at the Lock 5 stream gauge. As a result of missing data, the flow was set to the recorded flow at node SA flow rather than Lock 5. Flow at Lock 5 was set to -9999 indicating missing data. This was required as a result of poor model performance when Lock 5 flow was used as indicated in figure 5.3 and 5.4 below:

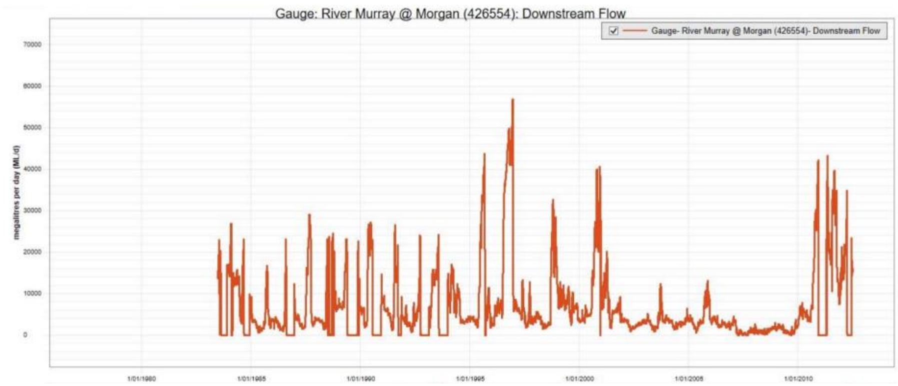


Figure 5.4 Predicted flow featuring Lock 5 data.

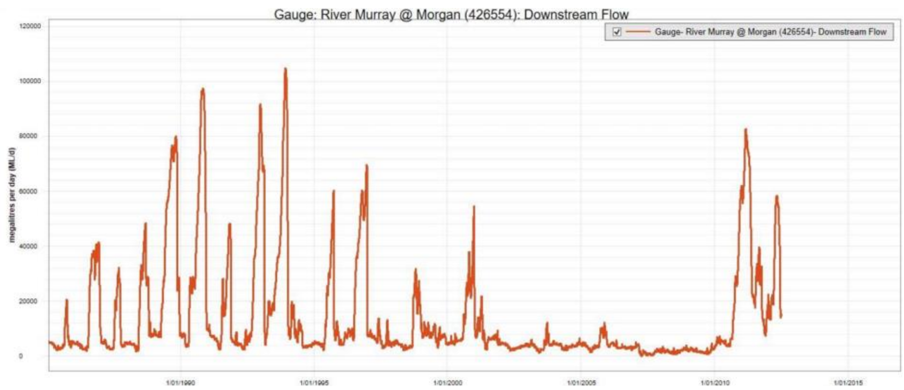


Figure 5.5 Predicted flow featuring SA flow.

Figure 5.4 highlights that as a result poor quality data featured at Lock 5 for flow, significant periods of zero flow can be seen. This is not representative of the real world conditions but rather has occurred as a result of insufficient data quality. Therefore, another method of defining the upstream flow was required.

This was completed by selecting the flow from the SA flow node. A comparison of the input flow timeseries can be seen in figures 5.6 and 5.7 below:

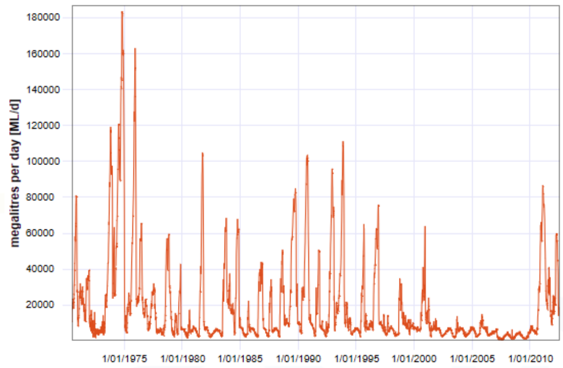


Figure 5.6 SA flow.

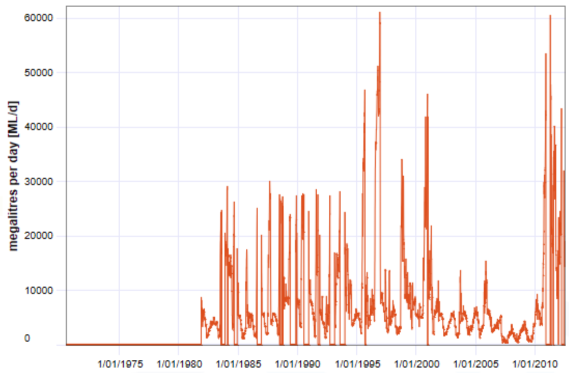


Figure 5.7 Lock 5.

It should be noted that although the SA flow is represented further upstream, when it is routed downstream, the losses and gains in the system for the historical periods have been accounted for through calibration. Therefore, this has not impacted on model performance as highlighted in section 5.1.3.

5.1.2 Simulation Methodology

In order to evaluate the models performance a hindcast was conducted to assess the baseline accuracy of the case study section. This involved running the model as a single analysis over the time period from 1/7/1983 to 30/6/2012. The salinity output generated though running the model was compared to the observed salinity at the Morgan gauge using mg/l. In addition to this, the predicted flow was compared with modelled flow in ML.

5.1.3 Model Performance

The results of the hindcast are represented in figure 5.8, 5.9, 5.10, 5.11 which highlights the comparison of the datasets.

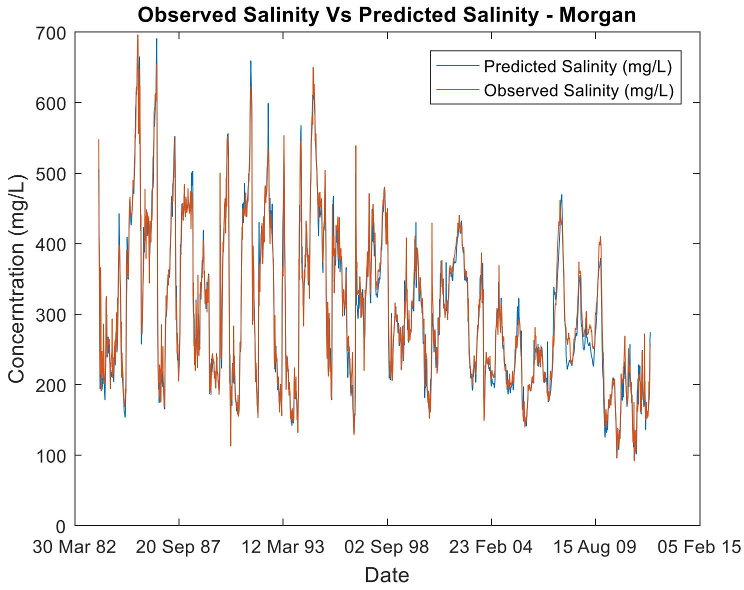


Figure 5.8 Observed Salinity Vs Predicted Salinity - Morgan

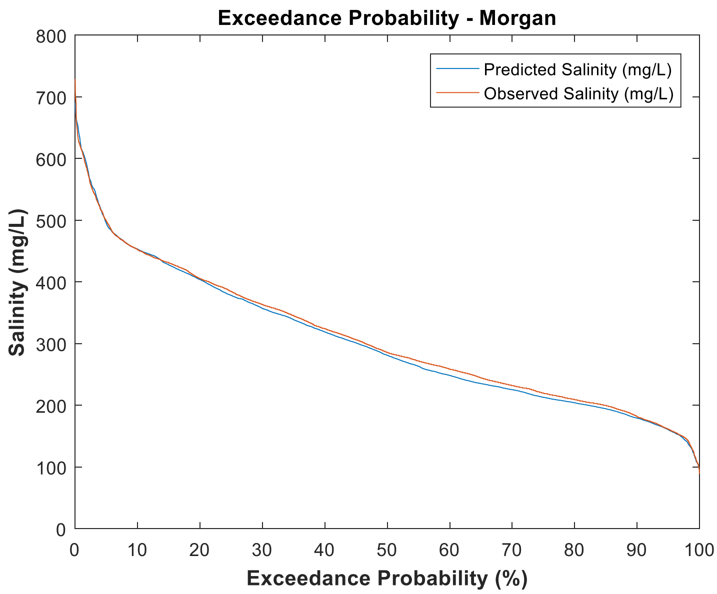


Figure 5.9 Exceedance Probability - Salinity - Morgan

Table 5.1 Statistical Analysis – Salinity - Morgan

Observed Salinity Vs Predicted Salinity - Morgan	
NSE	0.956913
R^2	0.958738
RSME	23.05288

Flow data was not present within the Source model. Consequently, data was acquired from the Murray Darling Basin Authority for the Morgan Gauge. The historical data featured a small time series of available data, hence only a small portion of the modelled downstream flow could be compared.

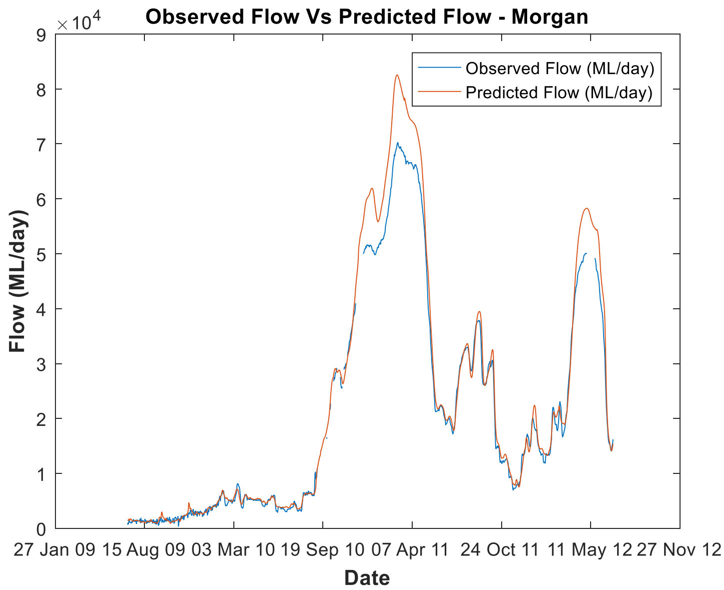


Figure 5.10 Observed Salinity Vs Predicted Salinity - Morgan

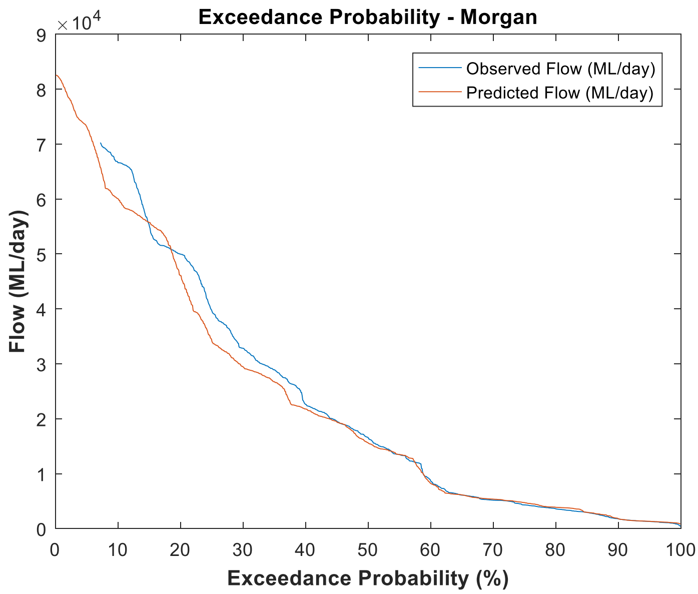


Figure 5.11 Exceedance Probability - Flow - Morgan

The available observed Morgan flow data featured a total of 133 missing data points. As a result, these values were omitted from the statistical analysis. The result of the statistical analysis for flow can be highlighted in table 5.2.

Table 5.2 Statistical Analysis - Flow - Morgan

Observed Salinity Vs Predicted Salinity - Morgan	
NSE	0.964177
R^2	0.993197
RSME	3816.89

The results of the two analysis examples provide a number of insights into the performance of the current model in its calibrated condition. Firstly, the salinity prediction is excellent with consistent correlation between observed and predicted salinity throughout, along with a NSE value of 0.957. This indicates that that a highly effective calibration of the model has been completed.

The flow analysis highlighted excellent performance in low flow conditions. However, during large flood events, the model appeared to over predict flow. Although this over prediction occurred, the majority of the output is extremely accurate and has resulted in a NSE of 0.964.

Although the salinity graphs appear to demonstrate a much closer correlation between observed and predicted flow, the calculated statistical values did not follow the trend. This has occurred as a result of a much larger range of data available for salinity analysis and the over and under predictions are proportionally are larger than that of the flow predictions. In addition, the flow prediction was almost perfect except for the instances in which major flow occurred. Together the NSE values of 0.957 for salinity and 0.964 for flow represent highly accurate model performance.

CHAPTER 6: FORECAST ANALYSIS

6.0 CHAPTER OVERVIEW

Chapter six highlights the results of the four testing components of the project. The chapter begins with the results of applying the additional salt calculation methodology to determine an estimated salt inflow within the river system. The results of the data smoothing analysis are then provided for which the optimal methods are analysed. The results from the trend forecast and artificial neural network are then examined in terms of their potential as future forecasting techniques.

6.1 SALT INFLOW CALCULATION

The initial salt inflow time series can be highlighted in figure 6.1 (Note: this is a summation of the inflow for every storage routing link):

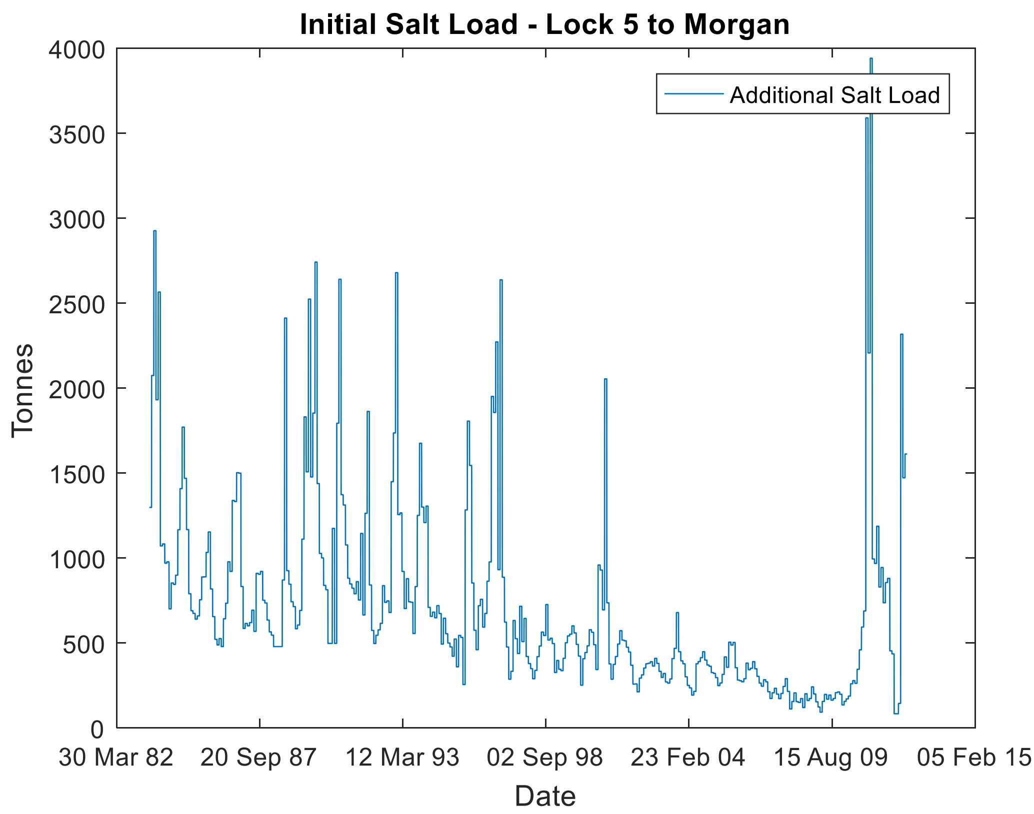


Figure 6.1 Initial Salt Inflow Timeseries

In order to calculate the salt inflow on a daily timestep, the methodology for salt inflow was applied. This resulted in the following time series for the entire reach Lock 5 to Morgan, where the yellow time series represents the daily unsmoothed additional salt inflow in Tonnes/day:

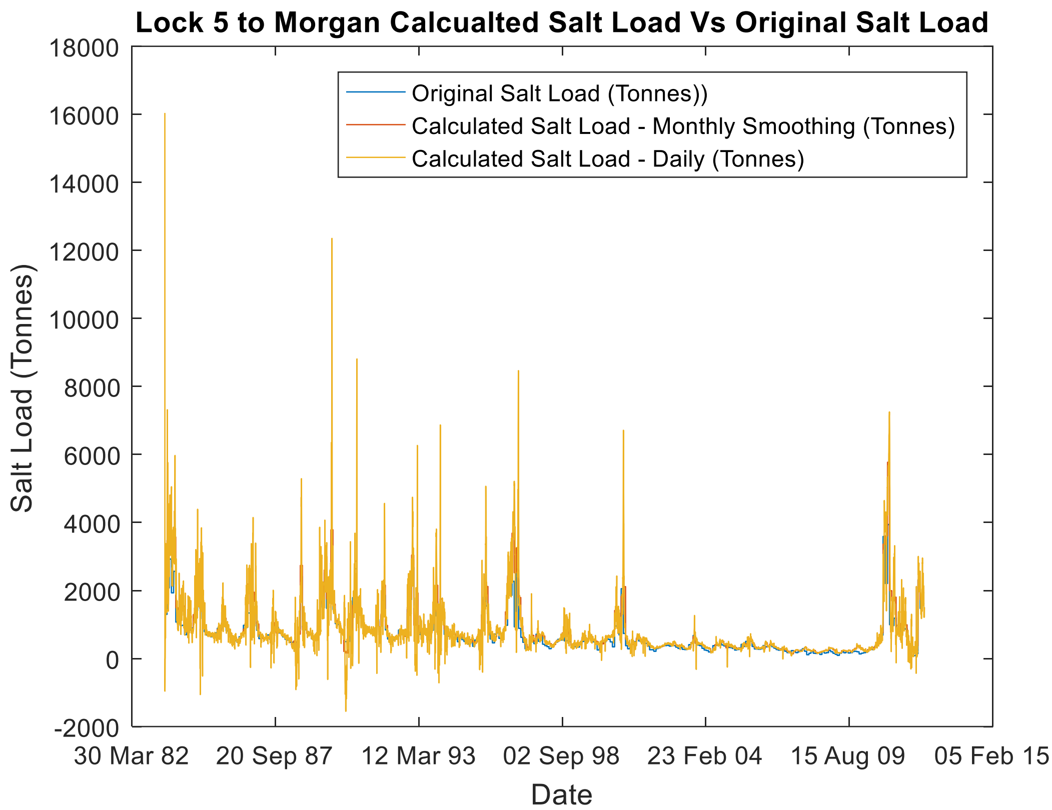


Figure 6.2 Calculated Salt Inflow

The results feature a similar pattern to that of the initial time series provided by the MDBA for salt inflow (Tonnes). Correlation can be seen for major peaks in salt inflow such as the 2012 peak event. It should be noted that much larger fluctuations in the salt load are present in the yellow timeseries present in figure 6.2 compared to the blue time series in figure 6.1. This has occurred largely as a result of the data being present without any smoothing being applied. For instance, the yellow timeseries uses a daily timestep while the blue timeseries features a monthly average smoothed timeseries, meaning that no data smoothing has occurred.

Figure 6.3 highlights a snapshot of the salt inflow timeseries for the 2012 salt inflow spike. This figure provides a comparison between three different timeseries being:

1. Calculated salt load inflow (daily timestep)
2. Calculated salt load inflow (monthly timestep)
3. Original salt load inflow (monthly timestep)

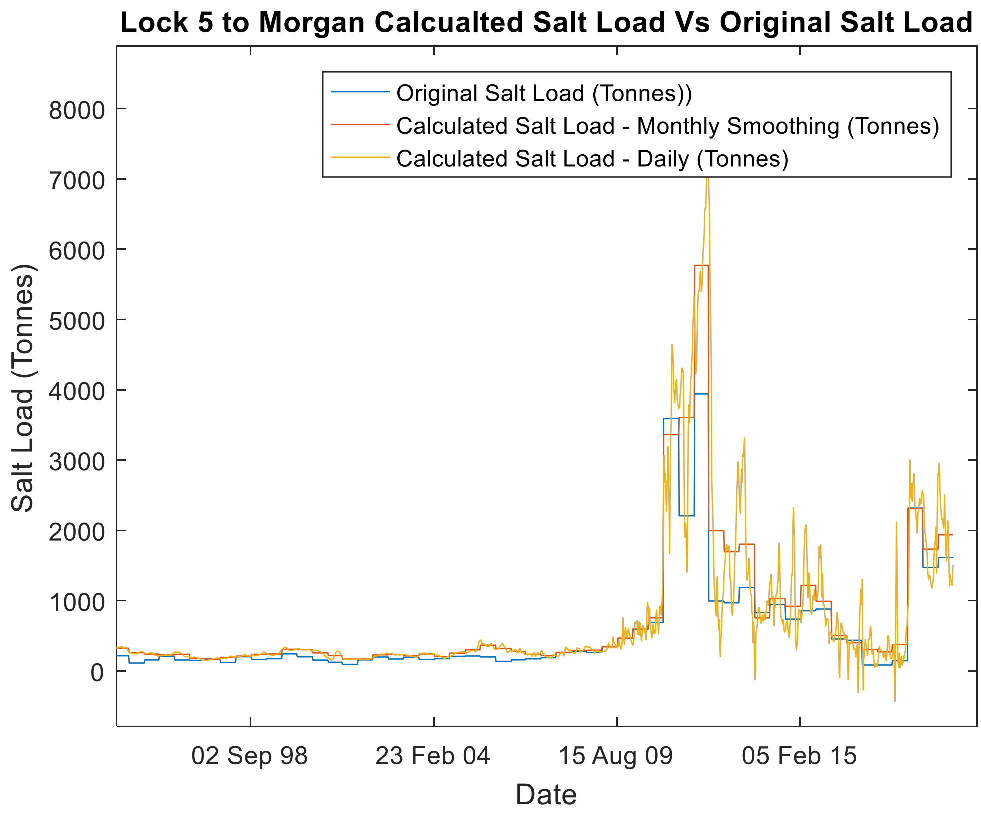


Figure 6.3 Calculated Salt Inflow – 2012 Spike

Figure 6.3 also highlights that the calculated salt load generally predicted higher salt loads than the original timeseries. This may be due to a number of factors including:

1) Different model configurations present when the salt balance was completed.
2) Further calibration methods which reduced the amount of salt inflow.

It needs to be noted that the original salt inflow timeseries featured a complex additional calibration which has likely reduced the salt inflow compared to that for which was calculated in the project. The analysis of this additional calibration methodology was considered beyond the scope of the project work.

6.2 Data Smoothing

In order to test the effects of different data smoothing techniques on the overall model performance, a number of experiments were completed. The results of which are highlighted in table 6.1:

Table 6.1 Data Smoothing Results

	Original	Daily	No Salt	Monthly	Weekly	Seasonal	Exponential	Weekly Moving	Monthly Moving	Weighted
R^2	0.959	0.945	0.795	0.952	0.950	0.937	0.930	0.924	0.904	0.929
NSE	0.956	0.943	0.060	0.951	0.940	0.933	0.926	0.920	0.895	0.926

Table 6.1 highlights in the top row, the data smoothing technique which was applied to the additional salt inflow data. The smoothed salt load for each method was then applied as additional salt inflow to the storage routing links present in the Lock 5 to Morgan case study in order to determine a salinity level. The column on the left indicates the statistical measure used for the analysis. The values present in the table represent the statistical measurement between predicted and observed salinity based on the entire run of data from 1/07/1983 – 30/6/2012.

The results of the data smoothing analysis can be summarized in order of their effectiveness via the following list: (Note: The NSE value was used as the primary means of evaluating performance)

1) Original
2) Monthly
3) Daily
4) Weekly
5) Seasonal
6) Exponential
7) Weighted Moving
8) Weekly Moving
9) Monthly Moving
10) No Salt

A number of conclusions can be drawn from the results highlighted in the above table and list. Firstly, the additional salt load within the reaches clearly has an impact on the model performance. This can be indicated by the no salt performance of E=0.06 representing the worst performance of the model. All model outputs which featured a salt inflow significantly outperformed the no salt inflow results. This indicates that substantial salt in entering the system between Lock 5 and Morgan via methods such as groundwater or surface inflow which is consistent with expectations as highlighted in the literature review.

No smoothing techniques were able to increase the performance resulting from the use of the original salt inflow timeseries provided by the MDBA which registered a NSE=0.956. The performance of the model could however be increased from the initial salt inflow calculation through the use of data smoothing techniques. The monthly smoothed data featured the highest performance with an NSE=0.951 which is greater than the daily salt inflow value which resulted in E=0.943. This indicates that smoothing does in fact have a positive effect on model performance.

The monthly smoothing has likely to have increased the performance of the daily data by reducing the peaks in the timeseries, while also accounting for the travel time of the salt inflow. Interestingly, the remainder of the smoothing methods did not have a positive effect on model performance as the weekly and seasonal smoothing effects worsened the performance of the model when compared to the daily salt inflow. The results in table 6.1 highlighted that the moving averages had a negative effect on model performance. All smoothing techniques based on moving average methods resulted in reduced correlation between observed and predicted salinity.

It can be noted that the initial large spike in the salt inflow calculation is likely the result of a model warmup. Here the marker tracking system which assigns an age to the particles at the upstream node and another at the downstream node has the same initial age at both locations being 0.5. As a result of this, the travel time from upstream to downstream in considered 0 or extremely small for the first 7 days, this results in large quantities of salt being required in the system to account for the high flow rates which are associated with the low travel time. The data is however, quickly normalized and after approximately 2 weeks the estimation was considered reasonable.

6.3 TREND FORECAST

In order to complete the trend forecast analysis the methodology discussed in chapter 4 was utilised. Table 6.2 highlights the forecasting results along with the trend extrapolation that featured the best performance based on the NSE as the primary performance criteria. The top row indicates the lead time of the forecast being 3-days, 7-days, 1-month, 3-months and 6-months. The columns on the left hand side indicate the amount of data provided for data fitting as 1, 3 or 9 years. The second column indicates the dates which was used as the point of extrapolation.

Table 6.2 Trend Results of Best Forecast Methods

Optimal Forecast Method Analysis						
		3 Day	7 Day	1 Month	3 Months	6 Months
1 Year	1/01/2008	Linear	Linear	Linear	Linear	Power
	1/01/2012	Exponential	Exponential	Exponential	Exponential	poly 2
	1/03/1993	Exponential	Exponential	Exponential	Exponential	Exponential
3 Years	1/01/2008	Monthly	Monthly	Monthly	Monthly	Poly 2
	1/01/2012	Monthly	Monthly	Monthly	Monthly	Exponential
	1/03/1993	poly 3	poly 2	Power	Linear	Exponential
9 Years	1/01/2008	Exponential	Exponential	Exponential	Linear	Poly 2
	1/01/2012	Power	Power	Power	Power	Linear
	1/03/1993	Poly 2	Linear	Poly 3	Poly 2	Poly 3

Figure 6.4 provides an example of an extract from table 6.2. This figure represents the trend analysis which was completed using 1 year of data to fit the trend equation. For this example, the forecast was completed using 1/1/2008 as the point of extrapolation. It can be seen that the monthly average over estimates the salt load by using such a small sample to extrapolate from, while the use of 1-year trend data extrapolated forward using exponential, power and linear methods are much more effective as they are more effective in recognizing the long term data trends as opposed to being susceptible to over or under estimating based on short term spikes.

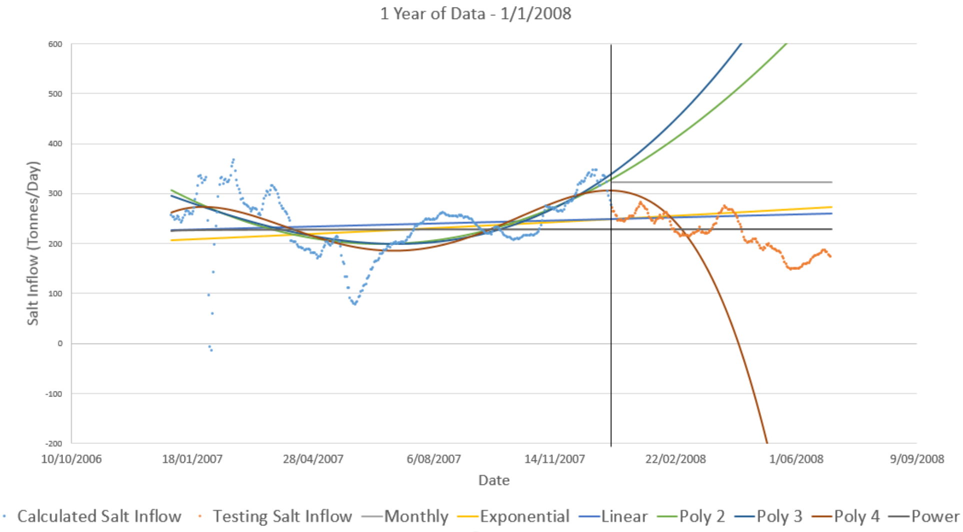


Figure 6.4 Trend Example

Figure 6.4 provides an example of what was found throughout the majority of the forecast analysis; being that the monthly average was likely to grossly under or overestimate the salt load within the reach while the trends of linear, exponential and power were much more effective in accurately predicating the future salt load over the longer term period of six months.

Table 6.3 summarizes the results for the best forecasting method in representing future salt inflow. The table highlights that the exponential method is the most effective method of forecast. The linear and monthly method also proved to be effective both representing the best method 8 times.

Table 6.3 Trend Summary Results

Monthly	Exponential	Linear	Poly 2	Poly 3	Poly 4	Power
8	14	8	6	3	0	6

In order to verify that the method was effective in increasing the salinity performance an exponential case was run back though Source as an additional salt inflow. Figures 6.5, 6.6 and 6.7 highlight the exponential timeseries extrapolated from 1/1/2012 using 1 year, 3 years and 9 years to develop the trend, representing 3 different scenarios.

These figures highlight the timeseries data that was used to fit the data, along with the exponential extrapolation which is extrapolated from the 1/1/2012.

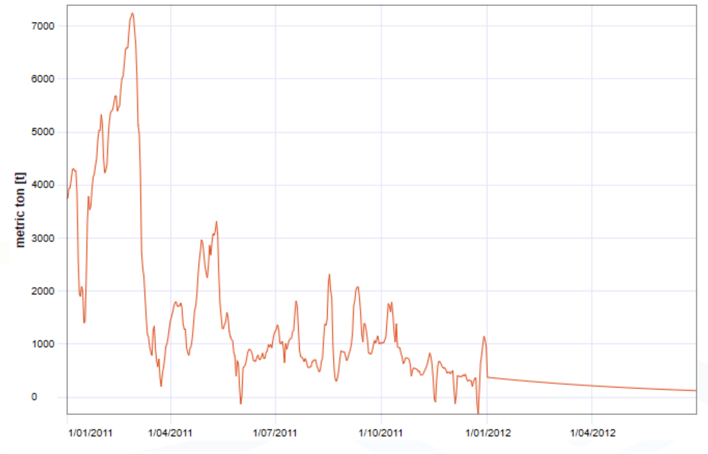


Figure 6.5 Exponential Salinity Forecast - 1 year - 1.1.12

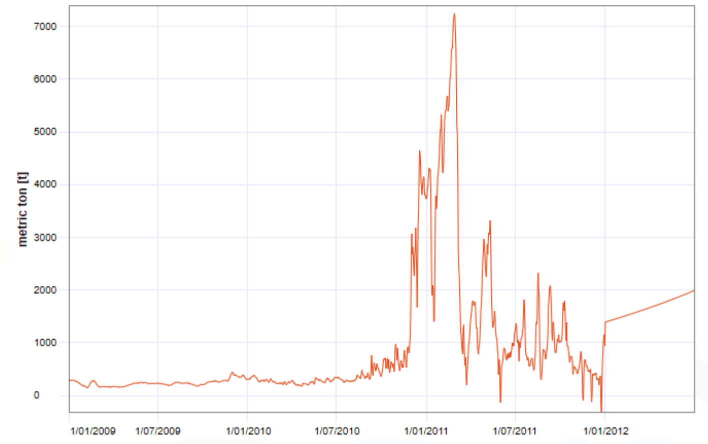


Figure 6.6 Exponential Salinity Forecast - 3 year - 1.1.12

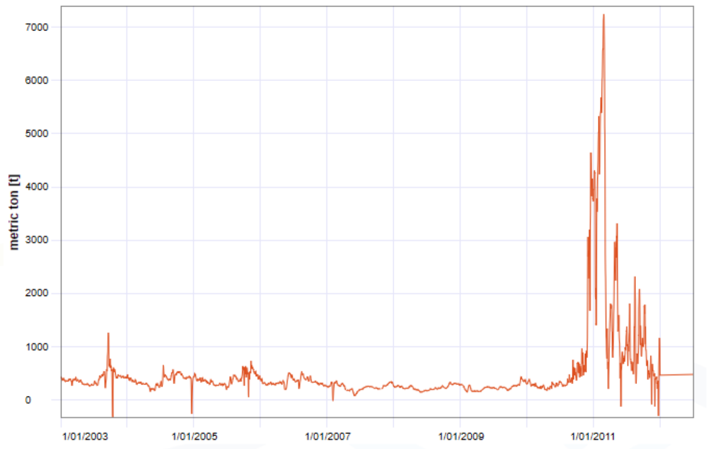


Figure 6.7 Exponential Salinity Forecast - 9 year - 1.1.12

These timeseries were then divided proportionally as salt inflow based on the length of each reach. This was applied as additional salt inflow to the storage routing links present in Source. After running the models, 3 outputs were generated and are highlighted in figure 6.8.

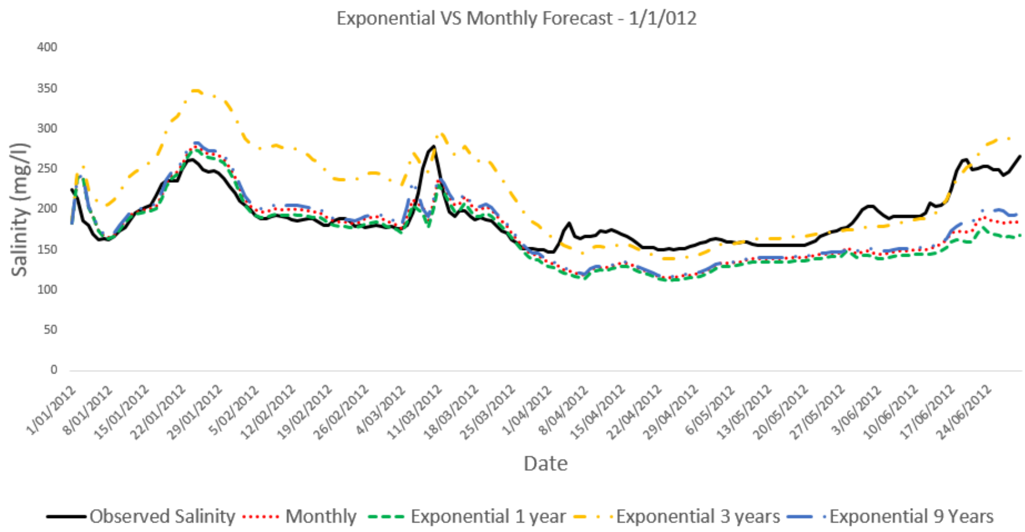


Figure 6.8 Exponential Effect on Salinity Forecast

Figure 6.8 highlights salinity output generated from Source using four different inputs. The graph provides the output for the forecast period which is from the 1/1/2012 – 30/6/2012 which represents a 6-month period.

A number of conclusions can be drawn from the results seen in figure 6.8. Firstly, the NSE value which represented the correlation between forecasted additional salt inflow and calculated additional salt inflow in table 6.2 was reflected in the accuracy of salinity predictions. Tables 6.4 and 6.5 highlight the correlation between salt load forecast and salinity prediction:

Table 6.4 Salt Inflow Performance

Salt Inflow Performance					
Statistical Measure	3-day	7-day	1-month	3-month	6-month
Monthly 1 year (NSE)	-0.0005	-0.0078	-0.1408	-0.0635	-0.6345
Monthly 1 year (Percentage Bias)	-3.4829	10.6672	32.2654	26.3489	-65.3004
Exponential 1 year (NSE)	0.0016	0.0058	0.0235	-0.0364	-1.0737
Exponential 1 year (Percentage Bias)	-9.8874	2.0827	13.5898	-8.5436	-80.2149
Exponential 3 year (NSE)	-2.2455	-5.6672	-19.1032	-13.3659	-0.0766
Exponential 3 year (Percentage Bias)	237.2188	288.1792	375.0542	381.8667	45.2633
Exponential 9 year (NSE)	-0.0065	-0.0582	-0.4062	-0.2346	-0.5103
Exponential 9 year (Percentage Bias)	487.1042	102.0021	76.7593	50.5888	-58.8847

Table 6.5 Salinity Output Performance

Salinity Output Performance					
Statistical Measure	3-day	7-day	1-month	3-month	6-month
Monthly 1 year (NSE)	-5.7896	-0.8248	0.6627	0.6125	-0.0335
Monthly 1 year (Percentage Bias)	6.4517	12.9472	4.2825	2.6813	-9.0004
Exponential 1 year (NSE)	-5.7320	-0.7885	0.7097	0.6308	-0.3251
Exponential 1 year (Percentage Bias)	6.3510	12.5707	3.0052	0.0244	-11.9914
Exponential 3 year (NSE)	-8.3673	-3.3207	-3.2805	-4.0495	-1.2309
Exponential 3 year (Percentage Bias)	10.3890	26.1891	28.9194	31.7101	15.8922
Exponential 9 year (NSE)	-5.9449	-0.9220	0.5883	0.5298	0.0724
Exponential 9 year (Percentage Bias)	6.7071	13.8285	5.8963	4.6936	-7.1761

For example, in Table 6.4 a poor correlation between forecast salt load and real salt load is present using a 1-month lead time for the 3-year exponential data fit, being a NSE value of -19.1. Using this forecast salt load data as the additional salt inflow in the reach, a poor correlation between predicted and observed salinity is present indicated by a NSE value of -3.28. This relationship is present through all data points run back through Source highlighting that ineffective salt load forecast results in a correlated ineffective salinity forecast.

This graph and associated statistical tables highlight that the forecasted additional salt load within a reach directly impacts the performance salinity predictions. Consequently, a conclusion can be drawn in that whatever results in the most accurate estimation of true salt load extrapolated forward, will result in the most accurate salinity prediction. This is a result of the salt inflow value being a value that works to minimize the difference between measured and observed salinity. The more accurate this prediction is, the more accurate the overall salinity prediction will be.

6.4 ARTIFICIAL NEURAL NETWORK FORECAST

The purpose of the Neural Network analysis was to assess the potential for an ANN to perform forecasts of additional salt inflow based on historical data. In addition, the analysis was competed in order to gain an enhanced understanding of the ways in which a neural network could be utilised in future research.

The neural network developed used only the historical salt inflow as input. This method was found to rely extensively on pattern recognition in order to generate useful forecasts. The literature review highlighted that neural networks could be effectively used to forecast variables when a range of external inputs where utilised. These however required known forward estimates of additional variables, in order to predict the target variable.

Given that the research highlighted scope for improvement of the salt inflow forecast, the aim of the neural network forecast was to examine whether salt inflow could be forecast without external input. This method would rely on seasonal patterns or trends in the salt inflow in order to produce forecast effectively using the weights developed in the open loop scenario.

Using the methodology highlighted in section 4.4.4 a neural network was develop using both the Matlab Neural Network Toolbox and the tutorial provided by Heath (2014). Each network was run and trained based on an open network architecture until the performance criteria being the mean square error, the Matlab default, was met. The loop for each case study was then closed and an extended timeseries of 6-months was used. The network would use the weights generated in the open loop to provide feedback for the predictions.

The results of this analysis are provided in appendix H which highlights the range of network architectures that were selected in attempt to successfully train a network that was able to recognize patterns in the historical salt inflow.

The results indicated an inability for the nonlinear autoregressive network (NARN) to recognize patterns. In all cases the open loop performance was excellent, featuring relatively low RMSE values. However, when the loop was closed the model was unable to replicate the open loop performance.

The models appeared to be able to follow the targets for a short period; this could be increased through increasing the amount of feedback delays in the network architecture. The ratio between feedback delays and hidden neurons was however, something that proved to be highly temperamental in the development of the models. Too few hidden neurons would not allow for a sufficiently complex architecture to be developed in order for the weightings to represent the relationships. Too many hidden neurons caused confusion and over-fitting of the data. This over-fitting and confusion of the model can be highlighted in figure 6.9 which provides an example of the attempted fit in the close loop scenario based on 24 feedback delays and 10 hidden neurons:

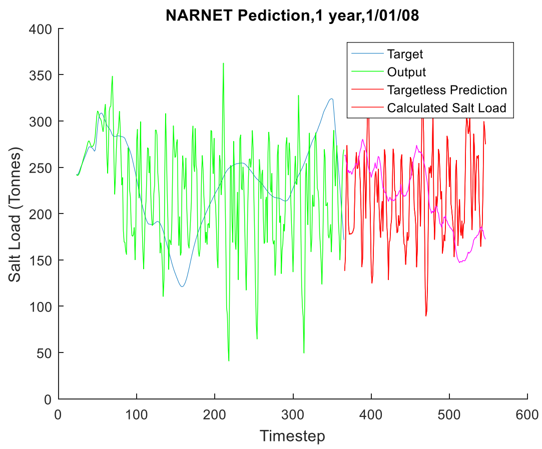


Figure 6.9 ANN Example – Overfitting

The network is able to match the target for the first rise however once irregular patterns begin to occur, the network becomes confused. Once targets are no longer present, being the salt load provided for training, the model is unable to determine suitable forecasts as indicated by the random fluctuations of the red prediction line.

During the trials of different network architectures, a much better result was developed using a configuration of 30 feedback delays and 10 hidden neurons as highlighted in figure 6.10 below:

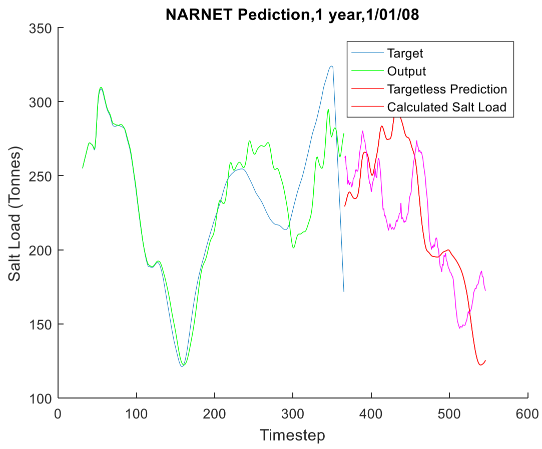


Figure 6.10 ANN Example – Best Result

It can be seen that the output when a target is present is very accurate up until timestep 200. The forecast period in red is able to somewhat predict the calculated additional salt load in magenta highlighted by an initial rise, followed by a decline. While promising, reaching this configuration was highly cumbersome given the trial and error configurations that were required.

Future research needs to be undertaken in this respect in order to further understand an effective means for determining the network architecture. The trial and error process is not feasible as given different training samples, different network architectures will be required.

The one year forecast period appeared to be the most promising in terms of generating correlated forecasts as opposed to the 3 and 9 year training periods. It appeared that as a result of the large quantity of values present in the 3 and 9 year training periods, the network featured far too many data points for the network to be able to recognize patterns effectively.

Without external inputs it appears as though the neural network using only historical salt load does not provide a feasible avenue for future research based on the highly irregular data series. Future research should therefore be aimed in two primary directions, being network architecture selection for NARN models and the use of a NARN to predict salinity using external inputs of flow (ML), salinity (mg/L) and additional salt load (Tonnes).

Chapter 7: Conclusions and Recommendations

7.0 Chapter Overview

This chapter concludes the discussion and analysis present in the project work before recommending future avenues of research that can expand on the project's findings.

7.1 Conclusions

This research project has provided an investigation of the predictive accuracy of salinity forecasting using the Source integrated modelling system for a case study of the Murray Darling Basin. The project worked to initially determine the current performance of the newly developed Source model and then worked to examine the ways in which the additional salt loads could be more effectively utilised. This was completed through four stages of analysis including:

1. Calculation of the additional salt inflow present within the case study.
2. Investigation of the effects of different data smoothing techniques on model performance.
3. Trend forecast analysis.
4. Neural network forecast analysis.

The research found that the current Source model of the Lower Murray, developed by the MDBA, performs effectively in its current calibrated state. This was indicated by a predicted salinity vs observed salinity NSE value of 0.956.

In order to begin the testing analysis, the additional salt inflow was required for a case study area. This was calculated as per the method previously applied in Bigmod. The effect of different data smoothing techniques were examined to determine how each method effected the models ability predict salinity. This process found that the monthly average smoothing was the most effective. Additionally, it was found that moving average techniques decreased the models performance when compared to the unsmoothed output. The results also highlighted that significant salt inflow does occur within the reach and that accounting for this has a significantly positive effect on the model's salinity performance.

The trend forecast analysis highlighted that more effective methods exist in forecasting salt inflow as opposed to simply using the monthly average of the most recent month prior to forecast. The results highlighted that exponential extrapolation appeared to more accurately predict the future salt load as opposed to the monthly value. This was indicated by the exponential extrapolation resulting in the highest correlation in 14 of the cases as opposed to 8 for the monthly average. Although this was promising, more case studies would be needed in order to verify the results. Polynomial extrapolation yielded a poor correlation between predicted and estimated additional salt load while linear and power extrapolation resulted in strong results scoring the highest correlation 8 and 6 times respectively.

The forecasted salt inflow timeseries were then applied in Source as salt inflow within each reach. It was found that the timeseries that had a close correlation with forecasted additional salt inflow and calculated salt inflow provided superior salinity predictions as opposed to the forecasts that featured poor correlation between forecasted and calculated additional salt inflow.

The application of the Artificial Neural Network proved to be highly problematic in terms of selecting an effective network architecture. It appeared that the lack of patterns present within the salt inflow timeseries meant that the neural network was unable to provide adequate forecasts based on the NARN model.

7.2 Recommendations

This research project examined a wide array of components that play a critical role in salinity modelling and forecasting. As a result of this, a variety of recommendations have been made regarding avenues for future research.

Firstly, research into the effects of different seasonal trends which effect river salinity would be of value. When forecasting, a methodology that defines an appropriate salinity value or range for a particular date may provide a valuable option for research. This type of pattern recognition and use of historical data is essentially a simplified version of an artificial neural network that utilises information gained from calendar dates and flow conditions to determine the likely future salinity value or range.

The research of salt inflow based on the types of cropping, land use and macro environmental conditions present would also be of high value. A key assumption that was overlooked in the trend forecasting was whether or not different storages were being released as this would have a significant effect on future salinity levels. If the effects of the flows and demands associated with land use and storage operation types could be accounted for accurately in the future, then the salinity prediction may be enhanced.

The analysis of using Neural Networks for forecasting salinity appears to be a form of forecasting that features high potential. As computing power continues to increase and more research and development is invested into artificial intelligence and machine learning, the feasibility of using pattern recognition and machine learning techniques will only grow. Neural Networks are clearly suited to using multiple inputs for training and forecast and this should therefore be the focus of future research in this area.

Finally, a key finding that has resulted from this research is that the NARN Matlab model is unlikely to be able to predict salt inflow based on a single timeseries alone. A research path that would more likely increase the overall salinity prediction would be through the use of a NARNX model as opposed to a NARN model. The NARX uses additional external inputs in order to assist in model training and prediction. Three key inputs being flow (ML), salinity (mg/L) and salt load (Tonnes) could potentially be utilised in order effectively to forecast salinity.

REFERENCES

Adamowski, JF 2008, 'Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis', *Journal of Hydrology*, vol. 353, no. 3–4, pp. 247-66.

Adamowski, J, Chan, HF, Prasher, SO, Ozga-Zielinski, B & Sliusarieva, A 2012, 'Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada', *AN AGU Journal*, vol. 48, no. 1.

Amerasinghe, Slatter, Hammond, Mullen 2011, 'Calperum Station Groundwater and Salinity Investigation', *Australian Bureau of Agricultural and Resource Economics and Sciences*, <http://data.daff.gov.au/data/warehouse/9aaw_001/csgssig9aaws001/CalperumStation_GwSalinity2011.pdf>

Australian Bureau of Statistics, Water and the Murray-Darling Basin - A Statistical Profile, viewed 30 October 2015, < http://www.abs.gov.au/ausstats/abs@.nsf/mf/4610.0.55.007cument >.

Banerjee, P, Singh, VS, Chatttopadhyay, K, Chandra, PC & Singh, B 2011, 'Artificial neural network model as a potential alternative for groundwater salinity forecasting', *Journal of Hydrology*, vol. 398, no. 3–4, pp. 212-20.

Bethune, M, Hardy, MJ, Teng, J, Lerat, J 2015, 'Implementation of source, the national hydrologic modelling platform, by the MDBA', *36th Hydrology and Water Resources Symposium 2015: The art and science of water*, Barton, ACT, pp.1123-1130.

Bhuiyan, J 2016, 'Source Salinity Forecasting in River Murray', Technical Report No. 2016/05.

Biggs, AJW, Silburn, DM & Power, RE 2013, 'Catchment salt balances in the Queensland Murray–Darling Basin, Australia', *Journal of Hydrology*, vol. 500, pp. 104-113, <http://www.sciencedirect.com/science/article/pii/S0022169413005441>.

Bowden, GJ, Dandy, GC & Maier, HR 2005, 'Input determination for neural network models in water resources applications. Part 1—background and methodology', *Journal of Hydrology*, vol. 301, no. 1–4, pp. 75-92.

Bowden, GJ, Maier, HR & Dandy, GC 2005, 'Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river', *Journal of Hydrology,* vol. 301, no. 1–4, pp. 93-107.

Chen, XY, Chau, KW & Busari, AO 2015, 'A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model', *Engineering Applications of Artificial Intelligence*, vol. 46, Part A, pp. 258-68.

Close, Andrew and Sharma, Pradeep, 2003. Hydrologic Models Used for Water Resource and Salinity Management in the River Murray System, 28th International Hydrology and *Water Resources Symposium: About Water; Symposium Proceedings*. Barton, A.C, 2003.

Close, Andrew. A New Daily Model of Flow and Solute Transport in the River Murray, Hydrology and Water Resources Symposium 1996: Water and the Environment; Preprints of Papers. Barton, ACT: Institution of Engineers, Australia, 1996: 173-178. *National conference publication (Institution of Engineers, Australia)* ; no. 96/05. <http://search.informit.com.au/documentSummary;dn=360100209749138;res=IELENG>

Connor, JD, Schwabe, K, King, D & Knapp, K 2012, 'Irrigated agriculture and climate change: The influence of water supply variability and salinity on adaptation', *Ecological Economics*, vol. 77, pp. 149-157, < http://www.sciencedirect.com/science/article/pii/S0921800912000857>.

Doble, Jolly, Walker, 2004, 'Spatial modelling of groundwater discharge patterns to predict floodplain salinisation and impacts on vegetation health ', *CSIRO Land and Water Technical Report No. 1/04*, <http://www.researchgate.net/profile/Rebecca_Doble/publication/265066805_Spatial_modelling_of_groundwater_discharge_patterns_to_predict_floodplain_salinisation_and_impacts_on_vegetation_health/links/546539b10cf2052b509f2e18.pdf>.

Engineering Statistics Handbook, Singe exponential smoothing, viewed 18 August 2016, http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm>.

Ewater, 2012, Source Scientific Reference Guide, eWater Cooperative Research Centre, Canberra, Australia.

Ferbar Tratar, L, Mojškerc, B & Toman 2016, A 'Demand forecasting with four-parameter exponential smoothing', *International Journal of Production Economics*.

George, B, Malano, H, Davidson, B, Hellegers, P, Bharati, L & Massuel, S 2011, 'An integrated hydro-economic modelling framework to evaluate water allocation strategies I: Model development', *Agricultural Water Management*, vol. 98, no. 5, pp. 733-46.

Greg Heath 2014, Narnet Tutorial On Multistep Ahead Predictions, Matlab News Group, Mathworks, viewed 24 August 2016, < https://au.mathworks.com/matlabcentral/newsreader/view_thread/338508>

Kim, S & Shin, DH 2016, 'Forecasting short-term air passenger demand using big data from search engine queries', *Automation in Construction*, vol. 70, pp. 98-108.

Leblanc, M, Tweed, S, Van Dijk, A & Timbal, B 2012, 'A review of historic and future hydrological changes in the Murray-Darling Basin', *Global and Planetary Change*, vol. 80–81, pp. 226-46.

Littleboy et al, An Overview of Modelling Techniques and Decision Support Systems and Their Application for Managing Salinity in Australia, CRC for Catchment Hydrology, Canberra, ACT.

Littleboy et al, New South Wales Government Office of Environment and Heritage, An overview of modelling techniques and decision support systems and their application for managing salinity in Australia, viewed 12 March 2016, <http://www.environment.nsw.gov.au/resources/salinity/pursllittleboyetal.pdf>.

Maier, HR & Dandy, GC 2000, 'Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications', *Environmental Modelling & Software,* vol. 15, no. 1, pp. 101-24.

Murray Darling Basin Authority, 2012, Hydrologic modelling to inform the proposed basin plan: Method and Results, Murray– Darling Basin Authority, Canberra, viewed 12 March 2016, < http://www.mdba.gov.au/sites/default/files/archived/proposed/Hydro_Modelling_Report.pdf?>.

Murray Darling Basin Authority, 2015, Basin Salinity Management Strategy, viewed 30 October 2015, <http://www.mdba.gov.au/sites/default/files/pubs/BSMS-2013-14-air.pdf>.

Murray Darling Basin Authority, 2015, Hydrological modelling using Source, Salinity, viewed 30 October 2015, < http://www.mdba.gov.au/managing-water/hydrological-modelling/hydrological-modelling-using-source>.

Murray Darling Basin Authority, 2016, River data and info, Salinity, viewed 30 October 2015, < http://www.mdba.gov.au/river-data/water-quality/salinity>.

Murray Darling Basin Authority, 2010, Basin Plan, Fact Sheet 5 water quality and salinity management plan, Murray Darling Basin Authority, Canberra, viewed 12 March 2016, < http://www.mdba.gov.au/sites/default/files/pubs/BP-2010-WQSMP-fact-sheet-v4.pdf>.

Murray Darling Basin Authority, 2013, Approach for estimating salt export from the River Murray System to the Southern Ocean, Murray Darling Basin Authority, Canberra, viewed 12 March 2016, < http://www.mdba.gov.au/sites/default/files/pubs/Approach-for-estimating-salt-export-from-the-River-Murray-System-to-the-Southern-Ocean.pdf>.

Murray Darling Basin Commission, 2001, Basin Salinity Management Strategy 2001–2015 Murray-Darling Basin Commission, Canberra, viewed 12 March 2016, < http://www.mdba.gov.au/sites/default/files/pubs/BSMS-full.pdf>.

Nash, JE & Sutcliffe, JV 1970, 'River flow forecasting through conceptual models part I — A discussion of principles', *Journal of Hydrology*, vol. 10, no. 3, pp. 282-90.

Nielsen et al, 2003, Effects of increasing salinity on freshwater ecosystems in Australia, viewed 12 March 2016, < http://www.environment.nsw.gov.au/resources/salinity/abjnielsenbrocketal.pdf>.

NSW Department of Water and Energy, 2008, Instream salinity models of NSW tributaries in the Murray-Darling Basin, Volume 6 – Murrumbidgee River Salinity Integrated Quantity and Quality Model, NSW Department of Water and Energy, Sydney, viewed 12 March 2016, <http://www.water.nsw.gov.au/__data/assets/pdf_file/0004/547870/quality_iqqm_mbd_murrum.pdf>.

Palani, S, Liong, S-Y & Tkalich, P 2008, 'An ANN application for water quality forecasting', *Marine Pollution Bulletin,* vol. 56, no. 9, pp. 1586-97.

Peck, AJ & Hatton, T 2003, 'Salinity and the discharge of salts from catchments in Australia', *Journal of Hydrology*, vol. 272, no. 1–4, pp. 191-202, http://www.sciencedirect.com/science/article/pii/S0022169402002640

Potter, NJ, Chiew, FHS & Frost, AJ 2010, 'An assessment of the severity of recent reductions in rainfall and runoff in the Murray–Darling Basin', *Journal of Hydrology*, vol. 381, no. 1–2, pp. 52-64.

Queensland Government (Department of Environment and Resource Management). 2011, Salinity management handbook, second edition, Department of Environment and Resource Management, Brisbane, viewed 12 March 2016, < https://publications.qld.gov.au/storage/f/2013-12-19T04%3A10%3A23.754Z/salinity-management-handbook.pdf>.

Quinn, NWT 2011, 'Adaptive implementation of information technology for real-time, basin-scale salinity management in the San Joaquin Basin, USA and Hunter River Basin, Australia', *Agricultural Water Management*, vol. 98, no. 6, pp. 930-940, <http://www.sciencedirect.com/science/article/pii/S0378377410003793>.

Qureshi, ME, Whitten, SM & Franklin, B 2013, 'Impacts of climate variability on the irrigation sector in the southern Murray-Darling Basin, Australia', *Water Resources and Economics*, vol. 4, pp. 52-68, <http://www.sciencedirect.com/science/article/pii/S2212428413000303>.

Simons, M, Podger, G & Cooke, R 1996, 'IQQM—A hydrologic modelling tool for water resource and salinity management', *Environmental Software*, vol. 11, no. 1–3, pp. 185-92 , < https://www.researchgate.net/publication/222921034_IQQM-A_hydrologic_modelling_tool_for_water_resource_and_salinity_management>.

South Australia State Government, Managing Hazards, Risk Assessment Matrix, viewed 30 October 2015, <http://www.safework.sa.gov.au/uploaded_files/RiskAssessmentMatrix.pdf>.

Stenson, MP, Littleboy, M & Gilfedder, M 2011, 'Estimation of water and salt generation from unregulated upland catchments', *Environmental Modelling & Software*, vol. 26, no. 11, pp. 1268-78.

Van Dijk, AIJM, Hairsine, PB, Arancibia, JP & Dowling, TI 2007, 'Reforestation, water availability and stream salinity: A multi-scale analysis in the Murray-Darling Basin, Australia', *Forest Ecology and Management*, vol. 251, no. 1–2, pp. 94-109.

Vaze et al, 2003, Modeling the effects of land use change on water and salt delivery from a catchment affected by dryland salinity, viewed 12 March 2016, < http://www.environment.nsw.gov.au/resources/salinity/modelingtheeffectoflandusechange.pdf>.

Victorian State Government, What are the risks?, viewed 30 October 2015, <http://www.worksafe.vic.gov.au/safety-and-prevention/health-and-safety-topics/office-work/about-the-problem/what-are-the-risks>.

Victorian State Government Department Land Water and Planning, Resource Allocation Model (REALM), Victorian State Government Department Land Water and Planning,, Melbourne, viewed 12 March 2016, < http://www.depi.vic.gov.au/water/water-resource-reporting/surface-water-modelling/resource-allocation-model-realm>.

Wei, Y, Langford, J, Willett, IR, Barlow, S & Lyle, C 2011, 'Is irrigated agriculture in the Murray Darling Basin well prepared to deal with reductions in water availability?', *Global Environmental Change*, vol. 21, no. 3, pp. 906-16.

Welsh, WD, Vaze, J, Dutta, D, Rassam, D, Rahman, JM, Jolly, ID, Wallbrink, P, Podger, GM, Bethune, M, Hardy, MJ, Teng, J & Lerat, J 2013, 'An integrated modelling framework for regulated river systems', *Environmental Modelling & Software*, vol. 39, pp. 81-102, <http://www.sciencedirect.com/science/article/pii/S1364815212000813>.

Welsh, W., Dutta, D., Vaze, J., Kim, S. and Nicholls, D., 2011, December. Improvement in short-term streamflow forecasting using an integrated modelling framework. *In Proceedings of the 19th International Congress on Modelling and Simulation*. Perth, Australia (pp. 3405-3411).

Welsh, W.D., Dutta, D., Vaze, J., Kim, S.S. and Nicholls, D., 2012. A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in eWater source. *Water resources management*, 26(15), pp.4397-4415.

Yihdego, Y & Webb, J 2012, 'Modelling of seasonal and long-term trends in lake salinity in south-western Victoria, Australia', *Journal of Environmental Management*, vol. 112, pp. 149-159, <http://www.sciencedirect.com/science/article/pii/S0301479712003635>.

Zhang, Y, Shao, Q & Taylor, JA 2016, 'A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model', *Journal of Hydrology*, vol. 538, pp. 802-16.

APPENDIX A - PROJECT SPECIFICATIONS

University of Southern Queensland
Faculty of Health, Engineering and Sciences

FOR:	HARRY MCCULLAGH.
TOPIC:	AN INVESTIGATION OF THE PREDICTIVE ACCURACY OF SALINITY FORECAST USING THE SOURCE PROGRAM FOR THE MURRAY-DARLING RIVER
SUPERVISORS:	MD JAHANGIR ALAM.
EXTERNAL:	DR MATHEW BETHUNE and Mr. ALISTAIR KORN.
ENROLMENT:	ENG4111/ENG4112
PROJECT AIM:	THE PRIMARY AIM OF THE PROJECT IS TO DEVELOP AN EHANCED PREDICTIVE SALINITY MODEL OF THE RIVER MURRAY AND LOWER DARLING SYSTEM.
SPONSOR:	MURRAY DARLING BASIN AUTHORITY (MDBA), CANBERRA

Program: ISSUE A, 8/02/2016

1) Literature review of the salinity issues, management and modelling practice in relation to salinity forecast.
2) Obtain and review the exiting model of the River Murray and Lower Darling River.
3) Analyse the effectiveness of the current modelling approach used in the Source Program.
4) Model set up for a case study section of the Murray - Darling River.
5) Hindcast the flow and salinity level in the river for a historical period and evaluate the accuracy of the existing model.
6) Manipulate the existing method in the ambition of developing an enhanced predictive salinity model.

As Time Permits

1) Examine the relationship between different land use management practice and salinity exports.
2) Examine the use of artificial neural network method in improving the salinity forecast.
3) Provide confidence level in salinity forecast.

APPENDIX B - RISK ASSESSMENT

In order to conduct a risk assessment for the project the following risk matrix provided by Safe Work South Australia (2015) has been used.

	RISK SEVERITY / CONSEQUENCE				
LIKELIHOOD	**FATALITY** *(may cause one or more fatalities)*	**CRITICAL** *(may cause severe injury – more than two weeks lost time)*	**MAJOR** *(injury resulting in at least one day lost time)*	**MINOR** *(medical treatment injury – return to work)*	**NEGLIGIBLE** *(first aid treatment – no lost time)*
VERY LIKELY *(exposure happens frequently)*	Extreme	Extreme	High	Medium	Medium
LIKELY *(exposure but not frequently)*	Extreme	High	Medium	Low	Low
UNLIKELY *(exposure could happen but only rarely)*	High	Medium	Low	Low	Low
VERY UNLIKELY *(exposure could happen but probably never will)*	Medium	Medium	Low	Very low	Very low

Figure 1 - Risk Matrix - Safework South Australia, 2015.

As no experimentation is required in the process, the entire project can be completed using a desktop computer. This task is deemed as a very low risk task. The major safety hazards can be summarized by that of a conventional office environment, that of which are highlighted online by WorkSafe Victoria (2015) as:

- **physical** (e.g. tripping or slipping hazards, glare or reflections from computer screens, hot components of photocopiers, poorly designed chairs that do not provide adequate back support, and tasks that demand prolonged work in a fixed posture)
- **psychological** (e.g. the need to complete excessive workloads under pressure, stress, inadequate recognition for work performed, lack of job satisfaction, or repetitive work)
- **mechanical** (e.g. getting 'caught' by equipment and filing cabinets that tend to tip when heavily laden top drawers are open)
- **chemical** (e.g. vapors or fumes from paint, solvents and photocopier toner)
- **electrical** (e.g. damaged electrical cords or overloaded power points that may cause electric shock)

In order to counteract these risks, the following steps will be carried out:

- Take regular breaks at 1 hour intervals.
- Ensure that good posture is maintained.
- Tripping hazards are removed.
- Electrical cables are safe.
- The room is well ventilated.

Appendix C - Salinity Calculation within a reach

The current method used by The Murray Darling Basin Authority to calculate salt within a reach is completed through:

1) Calculate daily salt load
2) Smooth the daily salt load

<u>Calculation of daily salt load</u>

1) Force upstream flow and salinity to equal recorded salinity
2) Set salt inflow in reach to equal zero
3) Run Model
4) Output per day of run
    a. = difference mean observed (EC_{Obs}) and modelled salinity (EC_{mod}) at downstream gauge
    b. TT_{daily} Average daily marker travel time for reach
5) Run the model assuming 1Tonne/km of salt inflow into reach
6) Output per day of model run
    a. Difference between mean and observed salinity at downstream gauge
7) Calculate the daily salt load required to minimize the difference between observed and modelled salinity

Where:

$$\Delta EC_{No\ Salt} = EC_{obs} - EC_{mod}$$

$$TT_{Daily} = Age_{DS} - Age_{US}$$

$$\Delta EC_{1t} = EC_{obs} - EC_{mod}$$

$$SL_{Daily} = \frac{EC_{No\ Salt}}{\Delta EC_{1t}}$$

Where:

$\Delta EC_{No\ Salt}$ = the daily difference between modelled and observed salinity at downstream gauge in a reach.

EC_{obs} = Daily observed salinity at downstream gauge.

EC_{mod} = Daily modelled salinity at downstream gauge.

Age_{DS} = Average daily age of markers at downstream gauge.

Age_{US} = Average daily marker age at upstream gauge.

TT_{Daily} = Daily average time taken for a marker to travel through a reach.

ΔEC_{1t} = The daily difference between modelled and observed salinity at downstream gauge in a reach when 1 Tonne of salt is added per km of reach.

SL_{Daily} = daily amount of salt added per km to minimize difference between observed and modelled salinity.

<u>Smoothing of daily salt load</u>

The approach in BIGMOD involves averaging the daily salt load to a monthly timestep for the amount of time it takes for salt to move through a reach in a month.

For each daily salt load:

1) Determine date when salt load was added to reach
   a) $Date_{Saltload}$ = Current date - TT_{Daily}
2) Calculate average salt load for month. This is the average of all daily salt loads added to the reach that contribute to salt leaving the reach in the current month.
   a) If $Date_{Saltload} < 0$ – then this daily salt load is contributing to salt leaving in the current month
   b) If $Date_{Saltload} > 0$ – this daily salt load contributes to salt leaving in the previous month. It is possible that salt added may take several months to move through a reach.
   c) SL_{Daily} = Average (SL_{Daily} where month = month($Date_{Saltload}$))

APPENDIX D - BIGMOD

In order to rout the flow of the water the model utilises the following expression stated by Close, 1996:

$$q_{out} = q_{in} + S_0 - S_1 - d - E \times \frac{A}{100} - L_{hf} - L_{em}$$

$$q_{out} = flow\ out\ of\ reach\ (ML/Day)$$

$$q_{in} = flow\ into\ reach\ (ML/Day)$$

$$S_0 = reach\ storage\ start\ of\ day\ (ML)$$

$$S_1 = reach\ storage\ end\ of\ day\ (ML)$$

$$d = distributed\ diversion\ (ML)$$

$$E = net\ evaprtation\ rate\ (mm/Day)$$

$$A = surface\ area\ of\ reach\ (ha)$$

$$L_{hf} = high\ flow\ losses\ of\ reach\ (ML/Day)$$

$$L_{em} = constant\ monthly\ losses\ (ML/Day)$$

In order to model salinity within the system the model using the following equation to determine the solute inflow between water monitoring sites Close, 1996:

$$Sol = (C_{ds} - C_{ni}) \times \frac{K_{us}}{C_{lu}}$$

$$Sol = calculaated\ solute\ inflow\ (units/day)$$

$$C_{ds} = concertration\ at\ downstream\ site$$

$$C_{ni} = concerntration\ at\ upstream\ site\ routed\ downstream$$
$$assumiong\ no\ additional\ solute\ inflows$$

$$K_{us} = kilometers\ from\ upstream\ site$$

$$C_{lu} = increase\ in\ concerntration\ that\ would\ result\ from\ a\ inflow$$
$$of\ 1\ unit\ of\ solute\ per\ day\ per\ km$$

This equation is used to model salinity and nutrient trends for the River Murray, provides data on the dynamics of between groundwater and the river and the interaction between nutrient content and sediments.

APPENDIX E - MATLAB TREND EXTRAPOLATION

```
%% Clearing Old Data
clc
clear all

%% Loading Data
fileid = fopen('9year_forecast1.1.12_csv.csv');
if fileid>0
data = textscan(fileid,'%s %f %s
%f','Delimiter',',','HeaderLines',1);
fclose(fileid);
end

%% Loading Predicted Salt Inflow
date =[datenum(data{1},'dd/mm/yyyy')];
saltload =data{1,2}';
datevalue = 1:length(date);
tp = 1:length(date)+183;

% power fit
power_1=polyfit(log10(datevalue),log10(saltload),1);
m_power_1=power_1(1);
b_power_1=10^power_1(2);
fx_power_1=b_power_1.*datevalue.^m_power_1;
fx_power_1_g=b_power_1.*tp.^m_power_1;

% Exponential fit
exponetial_1=polyfit(datevalue,log10(saltload),1);
m_exponetial_1=exponetial_1(1);
b_exponetial_1=10^exponetial_1(2);
fx_exponential_1=b_exponetial_1*10.^(m_exponetial_1*tp);
fx_exponential_1_g=b_exponetial_1*10.^(m_exponetial_1*tp);

% Linear fits
coeff1 = polyfit(datevalue,saltload,1);
fx_linear_1 = polyval(coeff1,datevalue);
fx_linear_1_g = polyval(coeff1,tp);

% Polynomials 2,3,4 fits
coeff2 = polyfit(datevalue,saltload,2);
coeff3 = polyfit(datevalue,saltload,3);
coeff4 = polyfit(datevalue,saltload,4);


% Evaluate polynomials
poly2_day1_g = polyval(coeff2,tp);
poly3_day1_g = polyval(coeff3,tp);
poly4_day1_g = polyval(coeff4,tp);
```

```
%% Excel output
% exponential
zexpoutput = abs(fx_exponential_1_g');
% power
zpoweroutput=abs(fx_power_1_g');
% linear
zlinearoutput = fx_linear_1_g';
% poly 2
zpoly2output = poly2_day1_g';
% poly 3
zpoly3output = poly3_day1_g';
% poly 4
zpoly4output = poly4_day1_g';

%% End script
```

APPENDIX F - MATLAB DATA SMOOTHING

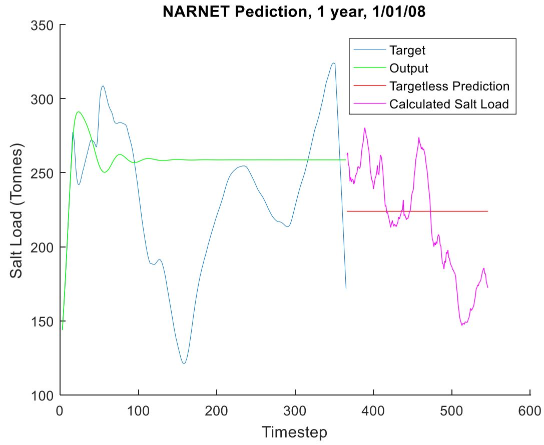
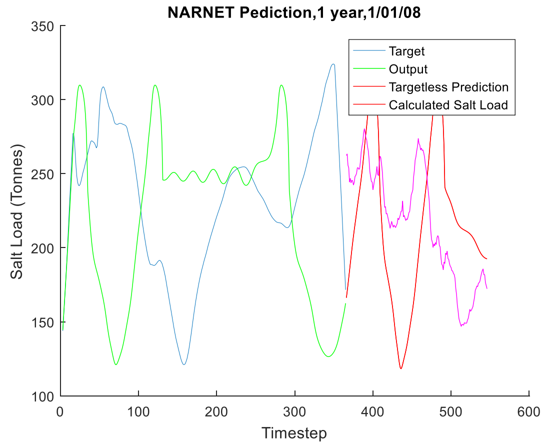
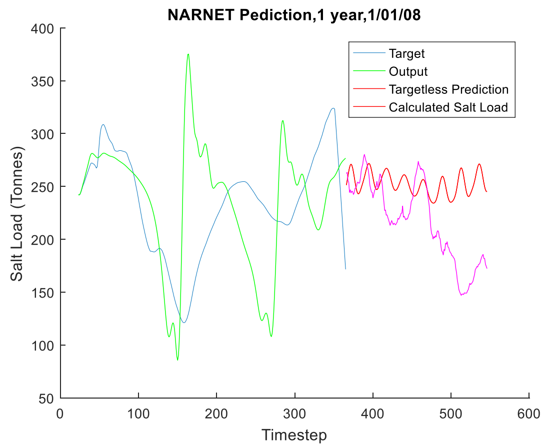
```
%% Clearing old data
clear all
close all
clc

fileid = fopen('calcualtedsaltload_csv.csv');
if fileid>0
data = textscan(fileid,'%s %f','Delimiter',',','HeaderLines',1);
fclose(fileid);
end

tsobj = fints(datenum(data{1},'dd/mm/yyyy'), data{1,2});
salt = fts2mat(tsobj.series1);

%% Exponential Data Smoothing
salt_ecp_smooth = tsmovavg(salt,'e',30,1);

%% Weekly Rolling Average
salt_simplerollingLag_7days = tsmovavg(salt,'s',30,1);

%% Monthly Rolling Average
salt_simplerollingLag_30days = tsmovavg(salt,'s',30,1);

%% Weighted Moving Average
weights = [0.01 0.02 0.04 0.06 0.9 0.12 0.16];
salt_weightedmovingavg = tsmovavg(salt,'w',weights,1);

%% End script
```

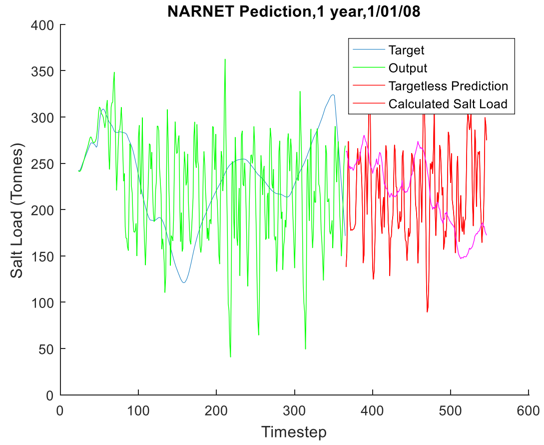
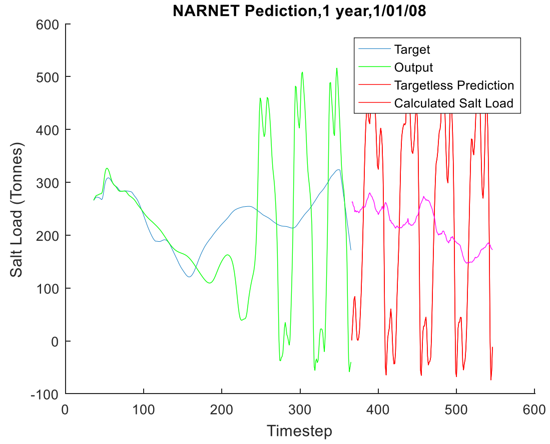
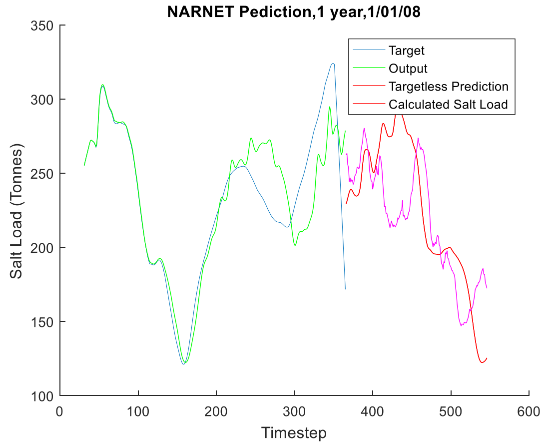
APPENDIX G - MATLAB ANN

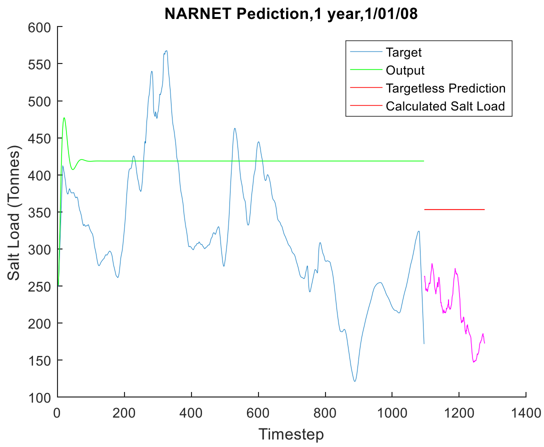
```
%% Script initially generated by Neural Time Series app
% Reference needs to be given to the members of the Matlab
Newsgroup
% in particular Heath (2014) who has provided a vast range of
resources
% and whose tutorial formed the basis of the network architecture.

%% Clearing old data
close all
clear all
clc

%% Loading Data
fileid = fopen('1year_forecast1.1.08_csv.csv');
if fileid>0
data = textscan(fileid,'%s %f %s
%f','Delimiter',',','HeaderLines',1);
fclose(fileid);
end

FDvalue = data{1,2};
t = FDvalue';

fileid = fopen('1.1.08_onwards_csv.csv');
if fileid>0
data = textscan(fileid,'%s %f %s
%f','Delimiter',',','HeaderLines',1);
fclose(fileid);
end

d2 = data{1,2};
saltonwards2 = d2';

t=smoothts(t,'b',30);

T=con2seq(t);
%% Network configuration

[I,N]=size(T)
FDvalue=30
FeedbackDelays=1:FDvalue;
HiddenNeurons=10;

neto=narnet(FeedbackDelays,HiddenNeurons);

%% Network settings
neto.divideFcn='divideblock';
net.divideMode = 'time';
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
net.performFcn = 'mse';  % Mean Squared Error
rng('default')
%% Prepare for training
```

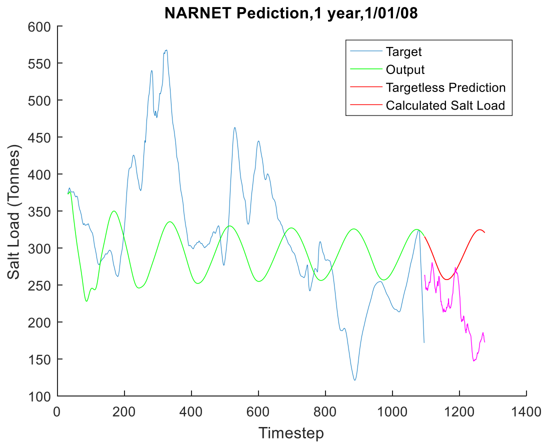
```
[x,xi,ai,To]=preparets(neto,{},{},T);

%% Training open network
[neto,tro,Yo,Eo,Aof,Xof]=train(neto,x,To,xi,ai);
[Yo,Xof,Aof]=neto(x,xi,ai);

Eo = gsubtract(To,Yo);
RMSEo=sqrt(mse(Eo))

%% Closing loop and training network
[netc,Xci,Aci]=closeloop(neto,xi,ai);
[Xc,Xci,Aci,Tsc]=preparets(netc,{},{},T);
[Yc,Xcf,Acf]=netc(Xc,Xci,Aci);
[netc,trainingrec] = train(netc,Xc,Tsc,Xci,Aci);

tsc = cell2mat(Tsc);
yc = cell2mat(Yc);
Ec = gsubtract(tsc,Yc);
RMSEc = sqrt(mse(Ec))

M=181;

%% Performing multistep prediction
[Yc,Xcf,Acf] = netc(Xc,Xci,Aci);
Xc2 = cell(1,M);
[Yc2,Xcf2,Acf2] = netc(Xc2,Xcf,Acf);
yc2 = cell2mat(Yc2);
yc2L=yc2';

%% plotting results
figure(1)
hold on
plot(FDvalue+1:N,tsc,'LineWidth',0.2 )
plot(FDvalue+1:N,yc,'g','LineWidth',0.5 )
plot(N+1:N+M, yc2,'r','LineWidth',0.5 )
plot(N+1:N+M,yc2,'r','LineWidth',0.5 )
plot(N+1:N+M,saltonwards2,'m','LineWidth',0.5 )
legend('Target','Output','Targetless Prediction','Calculated Salt
Load')
title('NARNET Pediction,1 year,1/01/08')
xlabel('Timestep')
ylabel('Salt Load (Tonnes)')

%% End script
```

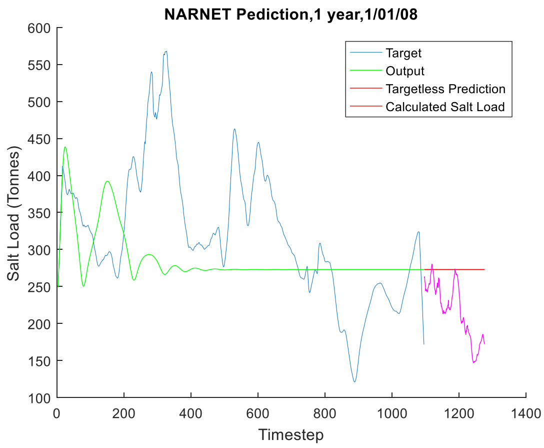
APPENDIX H - ANN OUTPUT NARNET

Narnet Predictions, 1 year, 1/1/08		
Delay, Neurons	Graph	RMSE
2,10		Open = 1.15 Closed = 56.6
2,20		Open = 1.76 Closed = 74.15
22,4		Open = 0.66 Closed = 75.3

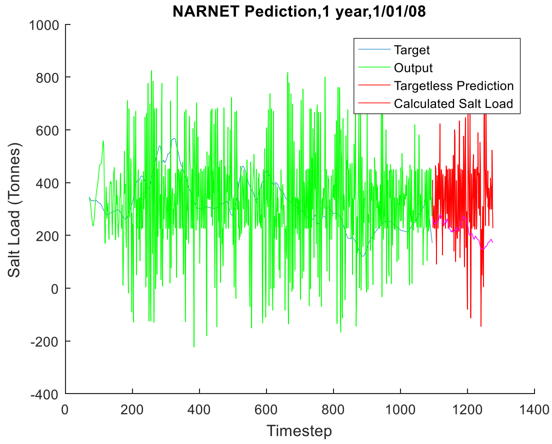
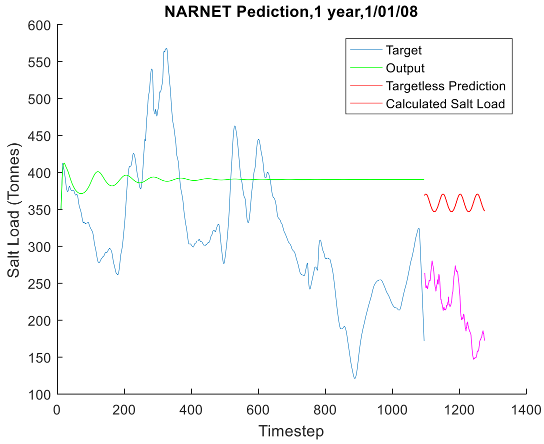
24,10		Open = 1.51 Closed = 74.8
35,5		Open = 0.592 Closed = 144.53
30,10		Open = 1.88 Closed = 67.51

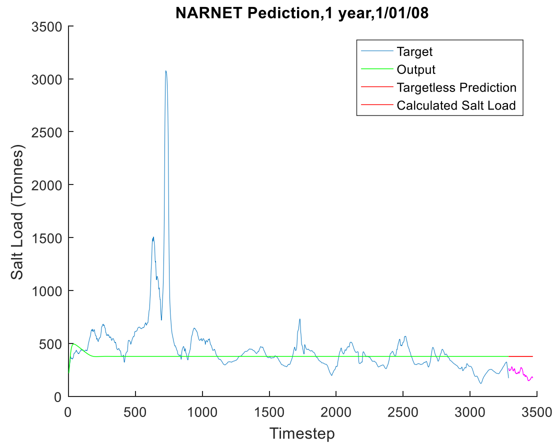
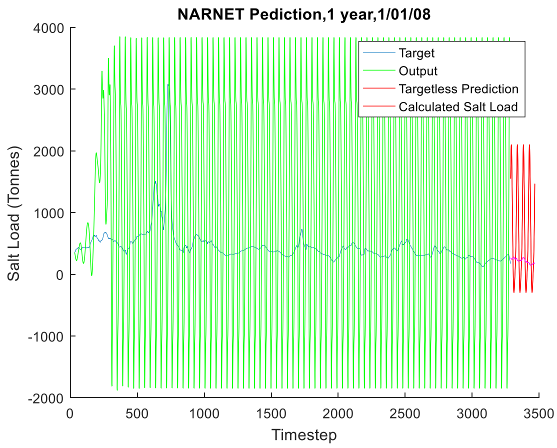
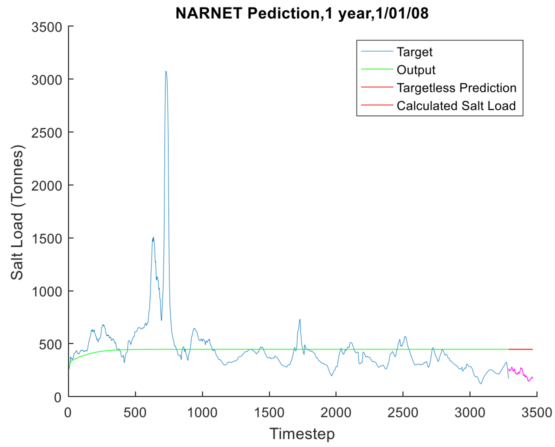
<table>
<tr><th colspan="3">Narnet Predictions, 3 year, 1/1/08</th><th></th></tr>
<tr><th>Delay, Neurons</th><th colspan="2">Graph</th><th>RMSE</th></tr>
<tr><td>2,10</td><td colspan="2">
</td><td>Open = 3.91

Closed = 132.5</td></tr>
<tr><td>30,3</td><td colspan="2">
</td><td>Open = 1.24

Closed = 98.89</td></tr>
<tr><td>2,15</td><td colspan="2">
</td><td>Open = 3.76

Closed = 101.81</td></tr>
</table>

70,5		Open = 1.34 Closed = 198.12
10,20		Open = 2.19 Closed = 122.78

Narnet Predictions, 9 year, 1/1/08		
Delay, Neurons	Graph	RMSE
2,10		Open = 3.04 Closed = 306.33
30,5		Open = 2.87 Closed = 1916.6
5,30		Open = 2.3 Closed = 298.2