University of Southern Queensland

Faculty of Health, Engineering and Sciences

Estimation of peak flow in Queensland using Quantile Regression Technique and ARR 2019.

A dissertation submitted by

Samuel Walker

in fulfilment of the requirements of

ENG4111 and 4112 Research Project

towards the degree of

Bachelor of Engineering (Honours) (Civil)

Submitted October, 2019

Abstract

Design flood estimation for small and medium sized catchments is required for the design of culverts, small to medium sized bridges, causeways and other engineering projects. Currently, the most widely used approaches within Australia include the Rational Method, the Quantile Regression Technique (QRT) and the Regional Flood Frequency Estimation (RFFE). Unfortunately, these techniques have been found to produce unreliable and inconsistent results for various reasons.

Empirical models (QRT, for example) are correlated to a particular set of data, they require updates when more data or improved methods are available. Australian Rainfall and Runoff has published 2019 version superseding the industry standard 1987 version. In addition, 1987 rainfall intensity-frequency-duration tables are also out-of-date and will no longer be available as of June 2020, replaced by 2016 version. The aim of this project is to update and improve overall performance of the QRT method presented by Palmen and Weeks (2011), utilising the latest 2016 rainfall intensities and new ARR 2019 procedures. Catchments are limited to smaller than 1000km^2 in area.

The project involved reviewing the quality of some rating curves with the aid of two-dimensional hydraulic models. Flood frequency analysis at each gauge produced discharge quantiles which were correlated by regression with catchment area and 2016 design rainfall intensities. A variety of standard frequency-durations of rainfall intensity was tested to find the best performing. The final model was compared to common methods in the industry, including those mentioned above, frequency analysis results and gauge data.

The result of this investigation is a set of six two-parameter equations for each design probability; 50%, 20%, 10%, 5%, 2%, and 1% AEP events, requiring only readily accessible catchment areas and design rainfall intensities to quickly and reliably obtain design flood estimates. The model was found to have similar performance to RFFE and so provides and alternative method. It may be used to validate larger and more complex hydrologic models and is particularly suited for planning, preliminary design and use when little information is known or available.

University of Southern Queensland
Faculty of Health, Engineering and Sciences

ENG4111 & ENG4112 Research Project

Limitations of Use

Certification

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.

Samuel Walker
Student Number: [redacted]
Date: 17/10/2019

Acknowledgements

I would like to thank Carlos Gonzalez for providing training and assistance in many aspects of this project; my supervisor, Rezaul Chowdhury, for guidance throughout; and the Queensland Department of Transport and Main Roads for providing support and resources to complete this project. In addition, Queensland Department of Natural Resources, Mines and Energy provided some guidance and data for the project.

Table of Contents

List of Figures

List of Tables

Abbreviations

1D	One-Dimension
2D	Two-Dimensions
3D	Three-Dimensions
AEP	Annual Exceedance Probability
AHD	Australian Height Datum
AM	Annual maximum series
ARI	Annual recurrence interval
ARR	Australian Rainfall and Runoff
BCC	Brisbane City Council
BoM	Bureau of Meteorology
CDF	Cumulative density function
CPU	Central processor unit
DEM	Digital Elevation Model
DNRME	Queensland Department of Natural Resources, Mines and Energy
DTMR	Queensland Department of Transport and Main Roads
ELVIS	Elevation Information System
FFA	Flood frequency analysis
GEV	Generalised Extreme Value probability distribution
GIS	Geographic Information System
GLS	Generalised least squares regression
GPU	Graphical processor unit
HPC	Heavily Parallelised Compute/Computing
IFD	Intensity Frequency Duration table
KS test	Kolmogorov-Smirnov test
LiDAR	Light Detecting and Ranging
LP3	Log-Pearson III probability distribution
MGA94	Map Grid of Australia 1994
MGBT	Multiple Grubbs-Beck test
OLS	Ordinary least squares regression
PDF	Probability density function
POT	Peak over threshold series
PP	Plotting Position
QLD	Queensland
QRT	Quantile Regression Technique; also referring to method developed by Palmen & Weeks (2011)
RFFE	Regional Flood Frequency Estimation online tool
TUFLOW	A hydrodynamic modelling software program
WLS	Weighted least squares regression

1 Introduction and Idea Development

1.1 Project Objectives and Scope

The aim of this project was to update and improve the overall performance of the Quantile Regression Technique (QRT) method presented by Palmen and Weeks (2011), utilising the latest 2016 rainfall intensities and ARR 2019 procedures, and improved gauge rating curves at several locations throughout Queensland. The intended outcome of this investigation was a set of simple equations, requiring only readily accessible catchment areas and design rainfall intensities to quickly and reliably obtain design flood estimates.

There are several of objectives that Palmen and Weeks (2011) targeted when developing their model. This project maintain as the same objectives, as follows;

- Applied routinely,
- Simple,
- Data is easily available in the office, and
- Repeatable.

At-site flood frequency analysis was completed on stream gauge stations, desirably fifty, to obtain standard design quantiles. Matching the requirement for catchment size by Palmen and Weeks (2011), only small to medium stations was selected, less than 1000km^2. Although initially a minimum of fifty years of record will be considered, this may be relaxed for the sake of distribution over Queensland and missing years of data in the record. Historical trend analysis, catchment changes, upstream and downstream influences, and urbanisation effects is out of scope for this project. Stream gauge rating curves, provided with time series data, was assessed with the aid of two-dimensional hydraulics models. The Queensland Department of Transport and Main Roads (DTMR) assisted with this task. Uncertainty analysis of stream data record and flood frequency analysis is not considered.

The regression methodology of all site design quantiles is detailed throughout the report. The same equation structure was used as in Palmen and Weeks (2011); no alternative forms, such as linear exponential coefficients, were considered. No regionalisation analysis for both frequency analysis and regression equations was conducted. However, examination as to whether other rainfall intensity duration/frequencies selection produce more accurate results was investigated. Testing of the final models and comparison to Palmen and Weeks (2011) findings and other common methods was then completed.

1.2 Justification and Expected Benefits

Design flood estimation for small and medium sized catchments is required for the design of culverts, small to medium sized bridges, causeways and other engineering projects. Currently, the most widely used approaches within Queensland include the Rational Method (RM), the Quantile Regression Technique (QRT; by Palmen & Weeks 2011), and the Regional Flood Frequency Estimation (RFFE). Palmen & Weeks (2011) is an empirical model and was developed using ARR 1987 methodologies and rainfall data. Therefore, it needs to be updated with additional data now available, ARR 2019 guidelines, and 2016 rainfall Intensity-Frequency-Duration tables (IFDs). The update to ARR 2019 has included better techniques and probability distributions for flood frequency analysis (Kuscera & Franks 2019). Gaffney & Babister (2019) found that QRT was superior to RFFE and suggest improvement is needed for both. Despite the trend toward complicated and realistic models, engineers, planners, and politicians need fast methods that are reasonably accurate.

This study will equip engineers with an alternate technique to RFFE based on available local data and will help in reducing the uncertainty in the observed flood data due to factors such as limitations in record length and rating curve extrapolation. It may be used effectively to validate larger hydrologic models and is particularly suited for use when little information is known or available. Other benefits would be to identify potentially problematic results, inform that a model may be incorrect, and allow greater confidence in the results of complex models. Moreover, it is appropriate for planning tasks and requires less time and resources to furnish an answer.

During this project, RFFE was compared with the resultant models and at-site FFA. Consequentially, the accuracy of RFFE can be observed for use over Queensland. Similarly, the most common probability distribution, Log-Pearson III or Generalized Extreme Value, may indicate the best suited to Queensland conditions for flood frequency analysis. These observations are not the focus of this project and was not extensively scrutinized in their own right.

The results and findings of this project will contribute to the scientific, research and engineering community and may prompt or assist further study. This research may broaden hydrological understanding and may influence industry practice.

1.3 Project Structure

The following headings outline the format and basic content of this report.

Literature Review

The literature review briefly discusses the relevant content and theory of the project. The original work by Palmen and Weeks (2011) is summarised along with some concepts involved when building RFFE. The update of ARR 2019 from 1987 edition is described throughout. Hydrology concepts regarding rainfall and how catchments influence stream flow was outlined. Gauging stations, data measurement and errors were discussed, along with an explanation of statistical analysis of floods. Lastly, regression methods for creation of the final equations are explained.

Methodology

All steps undertaken throughout the project are detailed in this section. Data acquisition and processing of the selected sites are presented. Management of missing data in stream flow records is outlined along with the methodology of reviewing rating curves using TUFLOW. Flood frequency analysis, regression methodology, and testing procedures are articulated. A commentary on execution of work tasks and reasoning of decisions is expounded throughout.

Results and Discussion

All results are presented and examined.

Conclusions and Recommendations

Overall findings are presented along with any recommendations. Project success is evaluated. The impact of the findings, and possible future work are also discussed.

1.4 Project Resources

Data

The data used comes from multiple sources. Department of Natural Resources, Mines and Energy (DNRME) reports stream gauge information through their Water Monitoring Information Portal (WMIP) service. The WMIP provides station data such as location, stream conditions and more, as well as water level records. Some stations also record rainfall and water quality measurements. Manual recorded gaugings are also recorded and most stations have a rating curve derived by these gaugings, interpolation and extrapolation techniques.

Rainfall design intensity tables are provided by the Australian Bureau of Meteorology (BoM). The BoM phased out 1987 IFDs by June 2019, replacing them with the new 2016 IFDs compliant with ARR 2019 update. All stations' design 2016 IFDs can be obtained using BoM online enquiry tool, which caters for multiple points. Conversely, 1987 IFD tables need to be obtained individually through another online tool supported by BoM.

Elevation data for 2D hydrodynamic modelling of rating curves was provided by DTMR and through the Elevation Information System (ELVIS) website. ELVIS is a website where governments and organisations throughout Australia provide lidar, aerial survey, DEMs and bathymetry for the public to source using an interactive map.

Software

Three mapping software were used during this project. Queensland Globe (Queensland Government 2019), an online interactive map supported by the Queensland Government, has Queensland specific data sets readily available and is easy and quick to use. MapInfo Professional version 15.0 (Pitney Bowes 2016) was the primary software in creating geographical information system (GIS) files and layers for 2D hydraulic modelling using TUFLOW (WBM Pty Ltd 2018). DTMR has provided access to this software. QGIS (QGIS Brighton 2017) was also used to review some of the 2D modelling results as it is freely available and has useful functionality not available in MapInfo.

TUFLOW was the 2D hydraulics modelling software used in determining rating curves. TUFLOW has no user interface, so Notepad++ (Don Ho 2019) and MapInfo are used to feed commands and files to the TUFLOW engine. Flike program (WBM Pty Ltd 2011) is recommended by ARR 2019 for flood frequency analysis. DTMR provided access to both Flike and TUFLOW licences. Various tasks were completed using MATLAB, Student Version (MathWorks 2019). The Curve fitting toolbox was also used for rating curves. Scripts were developed for extraction of data from WMIP (DNRME 2019) and BoM rainfall files. Additional software utilized includes the Microsoft Office suite, especially Excel.

1.5 Project Tasks

Summary of the principal tasks completed as part of this project is provided below;

- *Literature Review:* Review relevant material and current practices in order to inform the project methodology, risks and findings.
- *Data Preparation:* Gather data and transform into required formats.
- *Site Selection:* Filter gauges based on project criteria. Further remove sites as necessary throughout project.
- *Revise Rating Curves:* Assess provided DNRME rating curve to identify suspect gauges. Complete 2D hydraulic modelling to assess/revise gauge rating curves.
- *Flood Frequency Analysis (FFA):* Assemble Annual Maximum series and complete FFA to calculate quantiles for each site at 50%, 20%, 10%, 5%, 2%, 1% AEPs.
- *Regression:* Use quantiles to find parameters for equations that best fit real data at each AEP.
- *Validation:* Test the derived models for accuracy against observed data and other methods. Assess and investigate results.
- *Dissertation:* Write final report and present findings.

2 Literature Review

2.1 Background

Australia is a land of extreme weather. Being the driest inhabited country on earth (Bradshaw 2012), water storage is essential to Australian society's survival. However, when the droughts break, infrastructure and community are strained by large storms and flooding. Engineers must understand these natural extremes to design structures appropriately and manage risks. The field of hydrology – analysing rainfall to runoff – is always changing, increasing knowledge and improving practices. Therefore, methods need to be constantly reviewed and updated as necessary.

2.1.1 Hydrology Modelling

"A model represents the physical/chemical/biological characteristics of the catchment and simulates the natural hydrological processes," (Lohani n.d.). Lohani (n.d.) goes on to explain that the model is a tool to help make decisions. Devi (2015) suggests that the best model tries to reduce complexity while maintaining accuracy to the real world. Despite the general trend toward more sophisticated models (Rui et al. 2013), "there is no such thing as a 'perfect model'" (Teng et al. 2017); there will always be simplifications and assumptions. Furthermore, models are made for purpose. For example, large complex models that strain time and resources are no good for early planning or optioneering where the design may change significantly or even be cancelled. Although we have the ability of large 2D and more physically representative modelling, there is still a place and need for fast simple-to-use methods for a variety of reasons.

Hydrological Modelling

Models may be classified several different ways depending on what is being considered (Lohani n.d.; Devi et al. 2015; Teng et al. 2017). The most common classification is based on the description of hydrological process; empirical, conceptual and physical-based, all under the umbrella class of deterministic models. Deterministic means that the same input parameters will always produce the same results. Alternatively, statistical and stochastic modelling is governed by randomness producing varying results with each simulation. Monte Carlo process runs a large number of simulations and produces the most likely or significant result.

Empirical models only look at input data and make correlation to output data. A hydrological example would be deriving a relationship between rainfall data and catchment discharge at a recording station on a stream (the method adopted by Palmen and Weeks 2011). There is no consideration for physical processes or interactions. Often parameter units are not conserved or balanced. This method has minimal ability to explain the system being modelled and is limited to the specific conditions under which an empirical model is made. Departing from the purpose and scope of the model will produce uncertainty.

Empirical models are designed within a range of conditions and can fall out-of-date. As an example, Rational Method is a very popular method but is limited to < 25 km2, thus generally suited to smaller catchments, and was calibrated to 1987 IFDs and therefore out-of-date. Palmen and Weeks (2011) is an empirical model and was developed using ARR 1987 methodologies and rainfall data. Therefore, it needs to be updated with additional data now available, ARR 2019 guidelines and 2016 IFDs.

Increasing in complexity and resemblance to reality is the conceptual model. Parameters are employed to make the model more versatile, able to be applied to a number of similar catchments. These parameters usually represent a characteristic of the catchment or hydrological system but do require calibration. Rui et al. (2013) mentions a danger of losing physical meaning of parameters and that models are being misused, out of their 'niche'. The roughness parameter 'n' may not be the same for different models. Conceptual models are very useful; they require less computational effort and so are less time-consuming, they are flexible so one can study changes to a catchment by adjusting parameters and some diagnoses of cause and effect can be assessed. Conceptual models may only consider one or two hydrological processes, such as evaporation loss or infiltration into the soil, or many processes.

On the continuum, physical-based models most aptly represent reality; though some simplifications and assumptions are still made. Often coupled with the classification of distributed, as opposed to lumped, these models vary in space and time. Although physical-based models can be used for a wide range of applications, because they aim to simulate the complex principles behind the physical processes, they unfortunately necessitate very high computational demand.

Uncertainty & Errors

Typically, any data that has to be recorded or varies in space and time, such as rainfall, evaporation, and temperature, will have error associated with it (Lohani n.d.). While it is the aim of the hydrologist to minimise errors as much as possible, they come from many sources. Errors that are recognised can usually be treated/accommodated or at least understood. Yen (2002, as cited in Kidson & Richards 2005) suggests five sources of errors or uncertainty (Table 2.1). Assumptions are made all the time, mainly due to inadequate information and data. Errors can be accumulated simply by rounding individual steps in a lengthy process.

Table 2.1. Sources of model uncertainty (Yen 2002 as cited in Kidson & Richards 2005)

	Uncertainty Type	**Sensitive to**
1	Natural uncertainty	Nonstationary conditions
2	Model uncertainty	Choice of model
3	Parameter uncertainty	Fitting technique; goodness-of-fit test
4	Data uncertainty	Data choice; accuracy of observed/gauged data
5	Operational uncertainty	Human errors/decisions

2.1.2 Australian Rainfall and Runoff

Australian Rainfall and Runoff (ARR) – A Guide to Flood Estimation is Australia's principle guideline for hydrology and hydraulics practice. The first edition was published in 1958 and is now in its 4th revision (Geoscience Australia 2019a). In 2016, ARR was published as an advanced draft for most books of the multi-volume publication. The final copy was released in 2019 and can be found at http://arr.ga.gov.au/. It consists of nine books covering various aspects of hydrology and hydraulics. In conjunction with ARRs update, rainfall Intensity-Frequency-Duration data was updated in 2016, and online Regional Flood Frequency Estimation tool was developed. Both are discussed later.

Prior to 2016 the last edition of ARR was 29 years old, published in 1987. Since 1987, computers have improved largely and changed much of the way hydrology is done. Additional research and mathematical and statistical techniques have been developed. Significantly more data is also available. The update was broken up into 21 projects to review current industry practice, complete research and generate content (Geoscience Australia 2019b). Almost all tasks of this dissertation utilise ARR 2019, much of which differs notably from the previous 1987 version.

2.1.3 Palmen and Weeks (2011) Method

Palmen and Weeks' (2011) objective was to develop a simple and easy-to-use method for peak flow estimation to a reasonable accuracy. The study encompassed 289 Queensland catchments, all less than 1,000km^2. Nine catchment characteristics, with hydrological significance, were considered and tested, to analyse their degree of influence. Those with little influence or high difficulty of calculation were not included in the equations.

The result of Palmen and Weeks' (2011) research was a set of six equations for the most common design probabilities as follows;

$$Q_{2y} = 0.122 A^{0\ 757} I_{72h,50y}^{1\ 588} \tag{2.1}$$

$$Q_{5y} = 0.664 A^{0\ 709} I_{72h,50y}^{1\ 301} \tag{2.2}$$

$$Q_{10y} = 1.419 A^{0\ 682} I_{72h,50y}^{1\ 174} \tag{2.3}$$

$$Q_{20y} = 2.547 A^{0\ 673} I_{72h,50y}^{1\ 074} \tag{2.4}$$

$$Q_{50y} = 4.731 A^{0\ 656} I_{72h,50y}^{0\ 968} \tag{2.5}$$

$$Q_{100y} = 7.031 A^{0\ 644} I_{72h,50y}^{0\ 899} \tag{2.6}$$

where Q_n is the estimated peak discharge for n years ARI (m^3/s), A is the catchment area (km^2), and $I_{72h,50y}$ is the design rainfall intensity for a 72 hour, 50 year ARI (mm/h). Intensity was determined by the 1987 IFDs.

The method used for deriving the equations was quantile regression technique (QRT). First, flood frequency analysis was completed for all the sites. This provided quantiles of discharge for regression with the catchment's characteristics. Ordinary Least Squares approach was deemed appropriate for regression.

Palmen and Weeks (2011) mentioned the possibility of large uncertainty in the stream flow data. Gauged level was reported up to 100% of maximum recorded stage, with a median of 18%. Hence most rating curves required large extrapolation leading to uncertainty of discharge values.

2.1.4 Regional Flood Frequency Estimation Model

Regional Flood Frequency Estimation model (RFFE) was developed by Dr Ataur Rahman and Dr Khaled Haddad in conjunction with ARRs update and is recommended by ARR 2019 for peak flow estimation of ungauged catchments (Rahman & Haddad 2019a, Rahman et al. 2019). The objective of RFFE tool is to be quick, simple and require only readily accessible information (Rahman et al. 2019). It was developed as Project 5 – Regional Flood Methods in three stages. Since 2015, when RFFE application was first released, there has been one minor update in August 2016 where results for Arid Region were disabled due to anomalies, a station was removed from zone 1, and some graphs were added to the output (Rahman & Haddad 2019c). ARR 2019 Book 3 Chapter 3 describes RFFE functionality and fundamental theory (Rahman et al. 2019). The tool uses different methodologies based on the regions shown in Figure 2.1. Queensland covers Region 1 and 7, Fringe areas 14, 4, and 1.

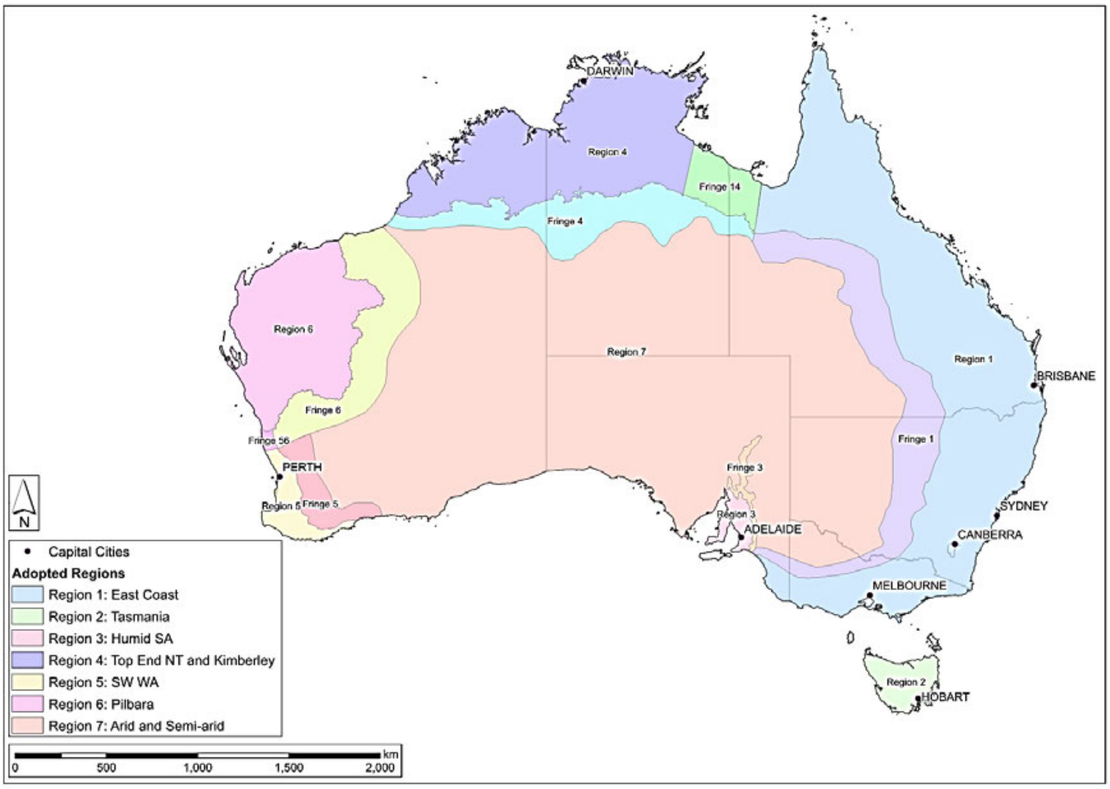


Figure 2.1. Regional Flood Frequency Estimation Model regions. (Geoscience Australia 2019c).

Catchment information and at-site flood frequency analysis (FAA) data from 853 gauged catchments across Australia, approximately 1 gauge per 8,800 km^2, are kept in a database used by RFFE. The application "transfers flood frequency characteristics from a group of gauged catchments to the location of interest" (Rahman et al. 2019). Therefore, it is greatly dependent on the number, likeness, and distance of nearby gauges.

RFFE gives a number of warnings and limitations for the model. For any catchments with more than 10% urbanisation, with dams or weirs that effect flow, that have had large scale clearing, or are significantly affected by agricultural activities, irrigation infrastructure, or mining activities, RFFE cannot be applied (Rahman & Haddad 2019b). RFFE is not well suited for catchments smaller than 0.5 km^2 or greater than 1,000 km^2, or when the nearest gauge is greater than 300km away. Arid areas have very few gauges, hence, accuracy may be low. In addition, catchments with distinctly abnormal shape, storage, or other characteristics may not be catered for by RFFE methodology.

2.2 Hydrology Concepts

Hydrology is specifically the study of the earth's hydrosphere (also called water cycle), however, it may need to draw on geology, meteorology and other sciences. Hydrology is an old science with ancient civilisations manipulating rivers, such as the Tigris and Euphrates in Mesopotamia, Indus in Pakistan and the Hwang Ho in China, primarily for agriculture (Jamal 2017). As early as 3000 BC gauges were installed on the Nile, Egypt (Jamal 2017). Over time, understanding of water mechanics has grown steadily. During the 1930s and 1940s, engineering hydrology moved forward swiftly (Rui et al. 2013). The development of computers has enabled major advances in this field, especially adding complexity and rigor to calculation methods.

In essence, rainfall to discharge can be thought of as a mathematical function of flow rates. An input quantity of rainfall over a period of time, then passes through a catchment (the operation) and outputs as a flow rate at the point of interest. Statistical analysis of rainfall has been completed throughout Australia and is generally well understood. This will be discussed in Section 2.2.12.2.1. The difficulty is understanding the operation and how the catchment affects the water flowing through it. Many assumptions and approximations are required.

The scope and methodology of this project, as in Palmen and Weeks' (2011) work, does not seek to approximate hydrological processes but, attribute a proportion of flow to different parameters/characteristics of the catchment and rainfall. These parameters must be hydrologically significant, that is increase or reduce peak discharge. Some will be discussed briefly in further sections.

2.2.1 Rainfall

Storms move across catchments, changing position, covering a fluctuating proportion of the catchment area, and varying in intensity over time. Therefore, hydrologists use design storms and patterns. Design storms are typical storm profiles with certain probability of occurrence. ARR 2019 Book 2 details how the design storms are determined and guidelines on applications for use.

Intensity-frequency-duration (IFD) tables, provided by the Bureau of Meteorology, specify the design intensity for a given rainfall event of frequency and duration. Rainfall intensity tends to decrease the longer duration and the rarer the event. This raw design intensity may be modified to cater for spatial or temporal variability or for climate change.

Depending on the catchment size, it is unlikely that a storm will cover the entire catchment area. This areal reduction factor should be automatically incorporated by regression of rainfall and catchment area parameters. Therefore, no modifications to rainfall intensity will be considered in determination of the empirical models.

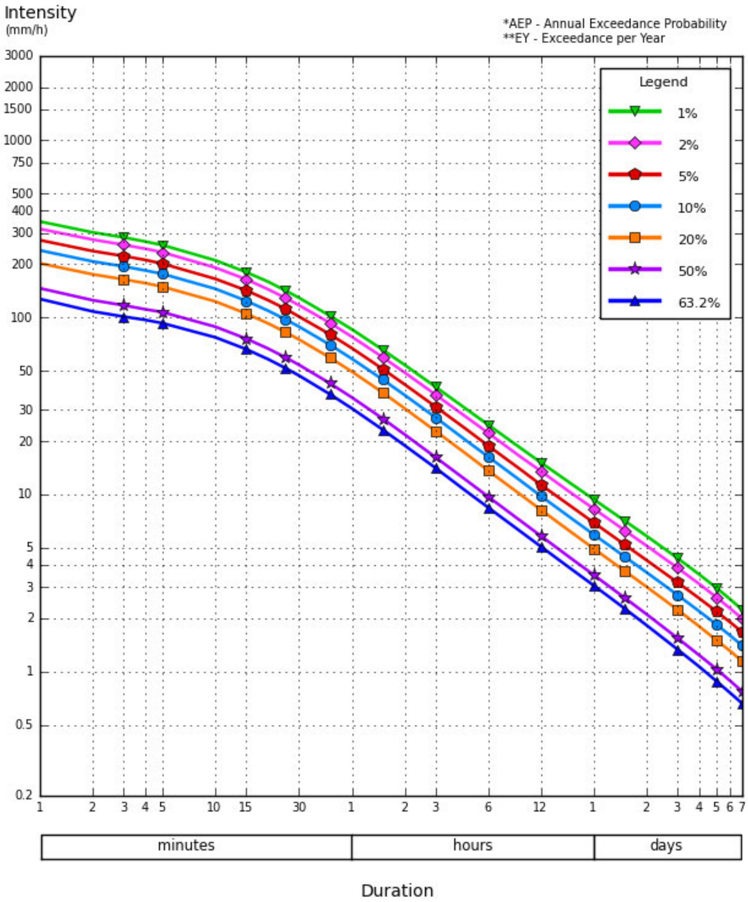


Figure 2.2 Example of Intensity-Frequency-Duration graph.

Probability and Recurrence

The most often used measure of probability of a flood event is Annual Exceedance Probability (AEP), the chance of a flood being equalled or exceeded within a year, or Exceedances per Year (EY), the number of equal or larger floods expected to occur within a year (Ball et al. 2019a). Previously, 'recurrence interval' or 'return period' were also common terms, but tend to be misleading or confusing. Palmen and Weeks' (2011) paper uses Average Recurrence Interval (ARI). The use of any of these terms to express probability may depend on context and derivation or method of calculations. A conversion can be made between AEP and ARI by equation 2.7 below.

$$AEP = 1 - e^{\frac{-1}{ARI}} \tag{2.7}$$

Hence, the events used in Palmen and Weeks (2011) of 2, 5, 10, 20, 50, 100-year ARI relates to approximately 39.35%, 18.13%, 9.51%, 4.87%, 1.98%, and 1% AEP, respectively. For practical purposes 50%, 20%, 10%, 5%, 2%, and 1% AEP events will be presented as final equations for this project.

There is an inherent assumption that a rainfall event will produce a flood at the same frequency. This is not precisely true. A combination of different climatic (spatial and temporal variance) and catchment characteristics (infiltration, storage, and more) may produce the same discharge. Examination of the differences in rainfall to discharge AEP would be rather difficult and time consuming with little expected effect. Moreover, independence, that is determining if one event effects another, is difficult to determine. Many rainfall events may contribute to one flood. Local rainfall independence may not translate to larger area independence.

1987 and 2016 Intensity-Frequency-Duration Tables

As part of ARRs revision, the 1987 IFDs were also updated. This update has had a significant impact on the industry, the most prominent being that most models have been determined from the 1987 IFDs and if there is a change in intensity, coefficients and calibration terms may be incorrect and produce erroneous results. Palmen and Weeks (2011) and the Rational method are affected by this update.

With almost thirty more years of data, contemporary statistical methods, and the advancement of computers, 1987 IFDs are outdated. Computers have allowed analysts to conduct more rigorous quality testing of data, modelling, and advanced, computational heavy statistical methods. A minimum of 500 station years was adopted to ensure accuracy and minimise sampling error. As there is no station in Australia with 500 years of data, a circular region would be increased until this criterion was achieved and regionalisation principles applied. Table 2.2 sets out eight important differences between 1987 and the new IFDs.

Table 2.2. Comparison of 1987 & 2016 IFDs (Bureau of Meteorology 2019d)

Method	ARR87 IFDs	New IFDs
Number of rainfall stations	Daily read - 7500 Continuous - 600	Daily read - 8074 Continuous - 2280
Period of record	All available records to up ~ 1983	All available records up to 2012
Length of record used in analyses	Daily read >= 30 years Continuous > 6 years	Daily read >= 30 years Continuous > 8 years
Source of data	Primarily Bureau of Meteorology	Bureau of Meteorology & other organisations collecting rainfall data
Extreme value series	Annual Maximum Series (AMS)	Annual Maximum Series (AMS)
Frequency analysis	Log-Pearson Type III (LP3) distribution fitted using method of moments	Generalised Extreme Value (GEV) distribution fitted using L-moments
Extension of sub-daily rainfall statistics to daily read stations	Principal Component Analysis	Bayesian Generalised Least Squares Regression (BGLSR)
Gridding	Maps hand-drawn to at-site distribution parameters, digitised and gridded using an early version of ANUSPLIN	Regionalised at-site distribution parameters gridded using ANUSPLIN

It is expected that there will be differences between the old and new. Figure 2.3 presents the differences between 1987 and 2016 IFDs in the Brisbane area as an example. Blue displays an increase and red reduction of rainfall intensity, and the darker the shade, the larger magnitude of change. The left image shows an increase greater than 50% at one location, which is substantial. The two images show how the differences are not proportional but vary in duration as well as location. 72hour event shows a reduction up to 50%, while at the same location (just inland from the Gold Coast), a general increase of intensity for a 12hour event. Variation in differences are apparent with regard to frequency of events also.

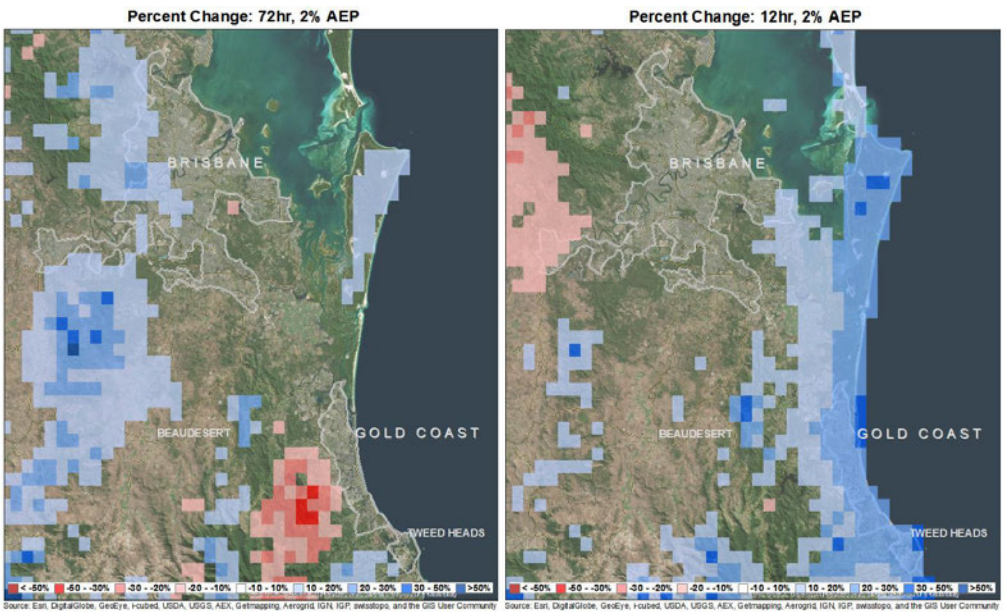


Figure 2.3. Percentage change from 1987 IFD to 2016 IFD in Brisbane area (Bureau of Meteorology 2019b)

Climate Change & Stationarity

Climate change is usually implemented by modifying the design storms as input into hydrologic models. All aspects of rainfall design are theorised to be affected by climate change; however, only minimal research has quantified these effects (Bates et al. 2019). Rare events, long duration, spatial and temporal behaviour with respect to climate change is not well understood (Ball et al. 2019b). However, Westra et al. 2013 (as cited in Ball et al. 2019b) detected statistically significant increases in rainfall intensity for short duration events. Temperature scaling is more accurate than assessment of rainfall for climate change risk projections, as outlined in ARR Book 1 Chapter 6 (Bates et al. 2019).

Climate change can be easily dealt with as an input to models, although stationarity of the data used for development and calibration of a model, which may present in the time series similarly, may require further investigation. The common assumption that a time series is stationary (i.e. it has no long-term trends) is important for many statistical and regression methods and even the model structure. Non-stationarity of stream flow may be caused by a number of things; modification to the catchment vegetation and land use, stream morphology, siltation and erosion are a couple.

It is difficult to test for non-stationarity, especially when dealing with rare events, highly variable and periodic data. A 1% AEP flood requires about 500 years of data to get a reasonably accurate result (Bureau of Meteorology 2019a). If you only look at 50 years of data, you may see half of a 100year cycle, resulting in a false upward or downward trend. Currently, due to Australia's highly variable rainfall in space and time, no trend could be identified in rainfall records to indicate non-stationarity (Bureau of Meteorology 2019c & Bureau of Meteorology 2019a). The difference between 1987 and 2016 IFDs is likely to stem from sampling errors, use of contemporary techniques, and more rigorous and complex modelling and quality testing. It is also expected that once the precipitation volume passes through a catchment, further noise will occur in mass rate. Hence it is even more difficult to identify a trend in stream gauge record. Therefore, all stream gauge data will be considered stationary for this project.

2.2.2 Catchment

Overland there are different flow regimes depending on volume and behaviour of the water. Initial runoff is spread out and relatively shallow, sheet flow, which then starts carving a drainage network of rills, channels, creeks and rivers as volume increases. Water flowing through this network is affected by many things. Total volume of water may decrease due to losses such as infiltration into the soil or storage in reservoirs. As the accumulated wave travels, it is resisted by vegetation, slope, terrain, structures and obstacles. When a rain event occurs, this retention will reduce (attenuate) and delay (lag time) the peak flow discharge of the system. All these effects, and more, are further variable between seasons and year to year due to climatic influences (droughts) and modifications to the catchment. Kuczera & Franks (2019) cites Kiem et al. (2003) comparing forty New South Wales gauges as having a 1 in 100 AEP flood under positive Interdecadal Pacific Oscillation (IPO, a common climate indicator) index is equivalent to a 1 in 6 AEP when IPO drops to negative.

Even though rainfall data may be stationary this does not imply stream flow record is because of historical changes in the catchment. However, this is also impossible to cater for. ARR 2019 Book 3 Section 2.2.3 states that;

> "It may be difficult or impossible to adjust the data if the catchment conditions under which the flood data were obtained have changed during the period of record, or are different to those applying to the future economic life of a structure or works being designed." (Kuczera & Franks 2019)

A great deal of historical information is required and is simply not available. Therefore, it is almost always assumed that stream record is stationary for the purposes of mathematical and statistical techniques.

Different catchment characteristics will have a different influence on the volume of discharge from a catchment. Palmen and Weeks (2011) assessed a number of these. The catchment area and rainfall intensity were found to have the most influence out of stream length, catchment slope, average annual rainfall, potential evapotranspiration, stream density, quaternary sediment area, and forest area. There are other methods available that use different assortment of these. Having a scalar catchment area will quantify the total volume of water caught from rainfall, however, slope, stream length, vegetation etc. can delay and reduce the peak discharge. Catchment shape is another parameter that is widely used and is part of RFFE tool. Zhang et al. (1999) reviewed literature and analysed the effect reducing vegetation has on the water cycle. "A clear conclusion was that reduction of forest cover increases water yield" (Zhang et al. 1999). The consequential effects of land use and vegetation change is complex.

2.2.3 Manning's Roughness Coefficient

A roughness coefficient, especially Manning's, is used in many applications in hydrology and hydraulics. Surface roughness is not clearly defined in hydrology and is generally described as resistance to flow (Arcement & Schneider 1989). It may encompass micro and macro variations in soil surface, vegetation including branches and tree trunks, debris and so on. The value of roughness is also dependent on water depth, flow dynamics and even the relative scale of what is being assessed. Deep water on clayey soil will have a much lower roughness then gravel bed with shallow water, because the water has to push over and around the aggregates. In free surface conditions, that is water is not pressurised, roughness is largely correlated to water depth. Resistance to flow will reduce velocity causing the cross-sectional area to increase to conserve discharge.

Selecting a roughness value is complex and rather subjective. There are some equations available like Limerinos and Strickler equations (BCC 2003) which are based on aggregate size, but restricted to certain conditions. Different methodologies have been proposed over the years, however, there is no ultimate solution.

Cowan developed a procedure in 1956 for estimating roughness in channels (BCC 2003, Arcement & Schneider 1989, and Philips & Tadayon 2006). It looks at six factors that are attributed to roughness. Cowan's equation is presented below;

$$n = \left(n_b + n_1 + n_2 + n_3 + n_4\right)m \tag{2.8}$$

where each parameter is defined as; n_b for channel material, n_1 for degree of irregularity, n_2 for variation in cross section, n_3 for effect of obstructions, n_4 for amount of vegetation, and m for degree of channel meandering. Parameter values are selected from a table with description of the channel conditions. The table is large and has not been included here but can be found in various sources including Appendix C of BCC (2003) Natural Channel Design Guidelines. Cowan's method is further adapted, by adjusting roughness values, to floodplains, with the same parameters except for cross section and meandering.

It is important to treat channels and floodplains independently (BCC 2003), as descriptions are not necessarily transferable. Figure 2.4 below shows an example of how one might apply roughness to areas for different conditions. Typically, roughness on the banks of a channel will be significantly different to the main bed, then bank full and overflow will change again. Most commonly, Manning's roughness values can be obtained from tables with a description of the conditions. Photos are sometimes provided with calculated or suggested roughness values to assist visualisation and use. Sometimes roughness needs to be changed based on height; once water force is great enough to overcome the strength of vegetation it will lie flat, reducing the roughness, or the water depth reaches overhanging branches. Furthermore, roughness does change over time and may not reflect historical conditions.

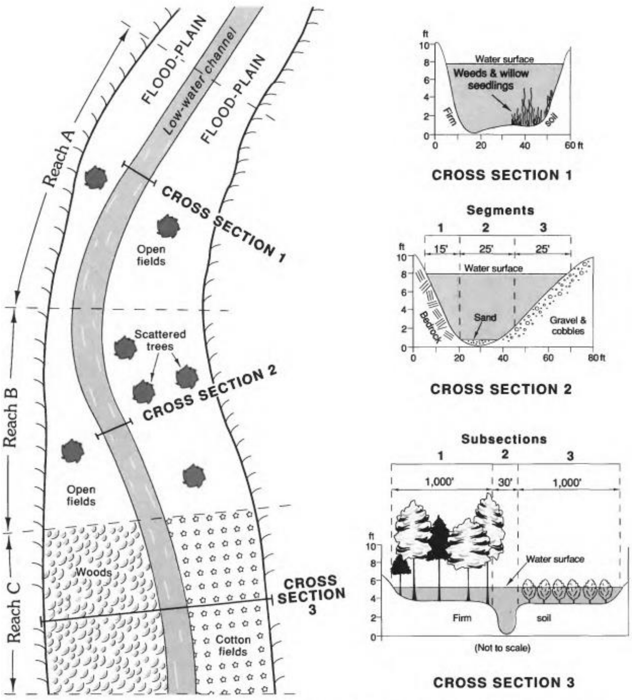


Figure 2.4. Schematic of assigning roughness values to a hypothetical stream. (Arcement & Schneider 1989)

2.3 Stream Gauge

2.3.1 Queensland Gauges

The Department of Natural Resources, Mines and Energy (DNRME) maintains most of Queensland's stream gauge stations. They provide stream flow records and metadata about the gauges to the public through Water Monitoring Information Portal (WMIP) website. Gauge records are used for water allocation and enforcement, flood monitoring, stream and floodplain management and other such activities.

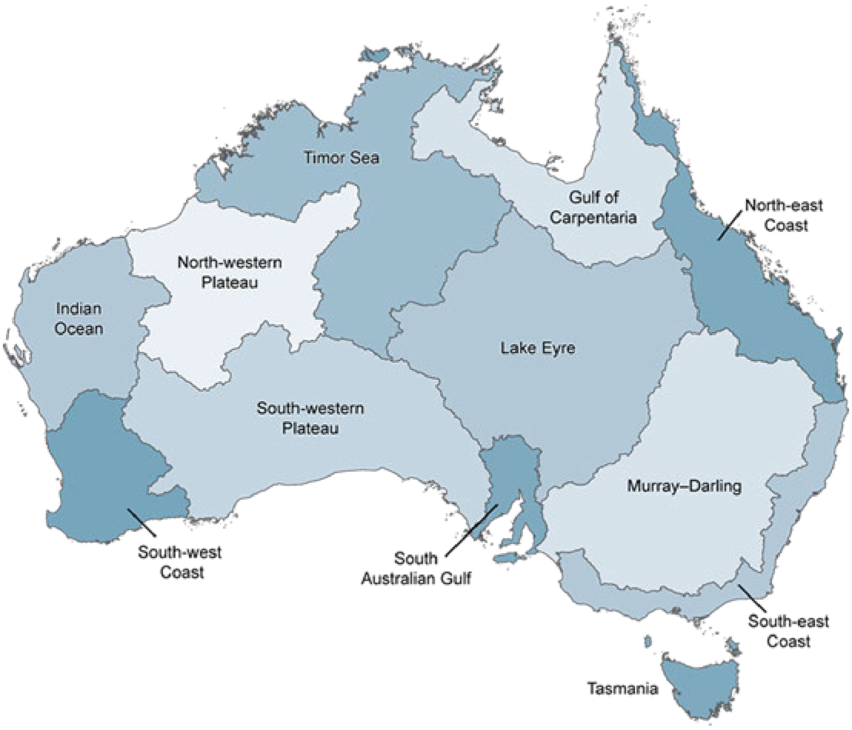


Figure 2.5. Australian drainage divisions (Hatton 2011).

Queensland covers all or part of four drainage divisions; Gulf of Carpentaria, Lake Eyre, Murray-Darling and North-east Coast. Basins are grouped into these divisions based on where the basin outlets to: Cooper basin flows to Lake Eyre. A total of seventy-seven designated basins cover Queensland. Some larger basins are further subdivided into river regions according to the major rivers. Gauges are numbered with reference to these divisions, basins, and river regions. Figure 2.6 breaks down the numbering system for Nebo Creek at Nebo gauge, 130407A. Stations that have been rebuilt or moved are identified with a different terminal letter in its gauge number; A, B, C etc. Gauge records that have been joined for this project will list the revisions used.

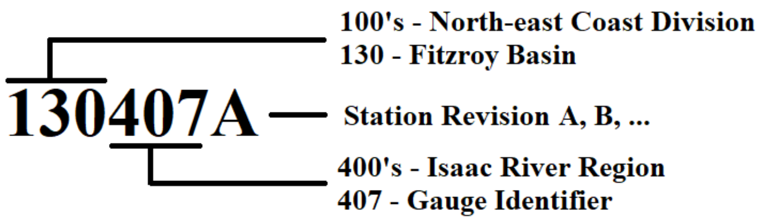


Figure 2.6. Gauge station number ID break down.

There are 1205 gauge stations across Queensland, with 413 remaining open. Most of Queensland's gauges are along the coast with a density of 6.23 per 10,000km^2 for North-East Coast division compared to 0.26-2.77 per 10,000km^2 for the other divisions. The highest density basin is Barron on North-East Coast with 65 per 10,000km^2.

Table 2.3. Summary of gauges in each division.

Drainage Divisions	**Open Gauges**	**Density (no. /10,000km^2)**	**Closed Gauges**	**Density (no. /10,000km^2)**
Lake Eyre (000's)	14	0.26	14	0.26
North-East Coast (100's)	279	6.23	574	12.82
Murray-Darling (400's)	72	2.77	95	3.66
Gulf of Carpentaria and Islands (900's)	48	1.06	109	2.40
Total	***413***	***2.43***	***792***	***4.67***

2.3.2 Measurement

Gauge stations generally comprise of a hut where instruments are housed, and may have depth markers along the rated cross section, a formed weir, and other miscellaneous equipment. A gauge record is measured manually or by computer system with sensors. Other information about the gauge site is recorded and constantly updated to retain currency. The rating curve (refer next section 2.3.3) is of particular interest and requires measurements of discharge and depth for real floods of various sizes, these are called gaugings.

Measurements are costly and time consuming, and impractical at times (Braca 2008). Large floods are hazardous, although these are the measurements most favoured. Sometimes gaugings are measured at a different location within the reach than at the exact gauge site for a variety of reasons including safety and access (Maynard 2014). Adjustments are then made to correct the results for travel time and/or storage effects.

The discharge of a flood is not measured directly. In the cross-section velocities and depths vary, a great deal of work is required to estimate the flow rate. Vegetation further complicates measurements and calculations. Often a boat is required to measure velocities across the channel presenting access problems and practical issues in taking the measures in a flood with debris and often fast flowing water. Depending on the nature of the flood event, the time needed to complete measurements may have an influence due to the flood rising, or falling.

Dealing with incomplete, missing, or poor-quality data is difficult in any field. Hydrology is not lacking in this area. The probability of malfunction of recording instruments is increased during major floods, which is of particularly importance and concern in flood frequency analysis that aims to use all of the major flood peaks (Kuczera & Franks 2019). Problems in the record occur due to instrument failure, uncertainty with the quality of measurements, changes in the environment, local hydraulic effects, human error, and the list goes on. Maynard (2014) even identifies wetting and drying of the soil, especially swelling clays, will cause gauge markers to move, making it difficult to keep accurate height datums.

2.3.3 Rating Curves

Rating curves are extensively used in surface hydrology (Braca 2008) to convert a measured stage to flow rate through the cross section. They are often presented as a graph, table, or set of equations. Preparing a rating curve is a complex task and mostly empirical (Braca 2008). Streams and catchments change over time which will change the stage-discharge relationship. Therefore, it is important to have a current and accurate rating curve for use in research, flood monitoring and water resources management.

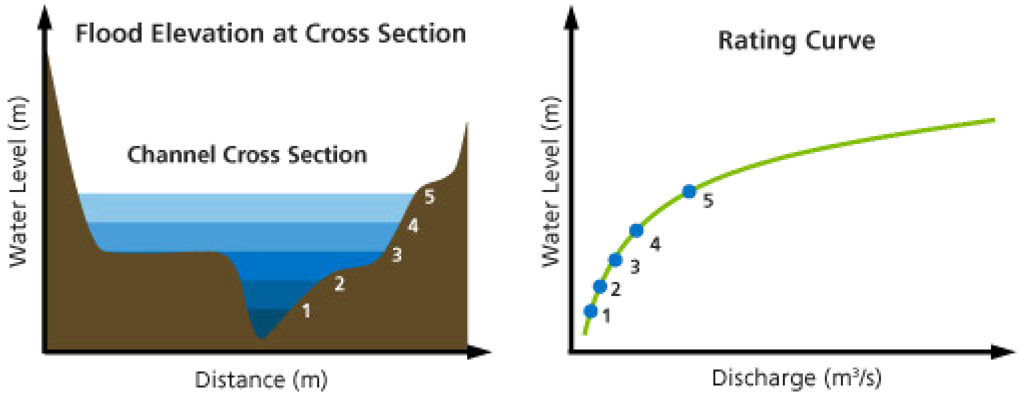


Figure 2.7. Depiction of cross section and rating curve relationship (Cabi 2014).

As discussed previously, measuring large floods is not often achievable, requiring the rating curve to be extrapolated or fitted to indirect discharge estimates (Kuczera & Franks 2019). There are various graphical and statistical methods used in industry to complete this task. A common method is fitting a power function to the available gaugings and/or estimated discharges. A power function is a simplified version of the Manning equation, used for steady uniform flow in open channels. One drawback is that this method cannot model when the flow regime/control changes (see below) or significant changes in the cross-section profile. Beyond this, 1D and 2D hydraulic modelling is also used to better simulate complex streams and floodplains. These detailed methods need to be calibrated to the measured gaugings and require much more time to complete, nevertheless, error and uncertainties related to extrapolation can be greatly reduced (Braca 2008). Braca (2008) recommends a minimum of 12-15 measurements to be used for calibration or fitting a function.

In Palmen and Weeks' (2011) work there was no check completed on the rating curves of the gauges and discharge record was accepted. However, they warned about uncertainty of large floods and stated the difference between maximum gauging and maximum flood varied from 1% to 100%. Despite large uncertainties in larger discharges, which are of most interest in FFA, FFA was assumed to be correct.

Rating Curve Uncertainty

Uncertainty in rating curves is due to a number of factors and is difficult to quantify. Extrapolation can be significant and incrementally grow the further the extrapolation. ARR 2019 Book 3 Section 2.3.7 (Kuczera & Franks 2019) discusses errors related to rating curves and mentions that errors can be considerable. Minimal literature is available that adequately addresses these errors, quantifies them or provides solutions to cater for them in flood frequency analysis (Kuczera & Franks 2019). More detailed methods such as 2D hydraulic modelling can reduce this uncertainty. Furthermore, the stage-discharge relationship is not unique due to unsteady flow conditions (discussed below). However, the rating curve is usually simplified to a single result, falsely suggesting steady flow (flow doesn't change with time). In order to understand the shape of a rating curve one needs to understand the physical phenomena involved (Braca 2008).

Vegetation can affect stream flow significantly and can be cyclical. For large floods the vegetation and debris are knocked down and washed through, allowing less resistance to flow (faster with lower water level), decreasing discharge up to 25% near bank full (Maynard 2014). Then over a time as vegetation grows back, retardance also increases until another large flood. Moreover, after a drought the channel will flow more freely than in wet periods with lots of overgrowth. Maynard (2014) comments that vegetation density is an important consideration in Eastern Queensland streams and is often underestimated in determining resistance by common methods. This all adds to uncertainty in the rating curve.

Flow Controls

The stage-discharge relationship is governed by three flow conditions; section control, channel control, and the transition zone. Section control occurs when there is a restrictive cross section or features such as rock ledge, sand bar, or weir (Braca 2008). The height of water is determined by this point and effects upstream water level. Moreover, at the control location there is local changes in velocity and water surface level. Therefore, the measurement location is often upstream of the control, where water level and slope are stable with changing discharge. Low flows are primarily governed by section control and may change between different section controls as water level increases.

Channel control, especially identifying it, is more complex than section control. The flow is usually larger and, therefore, effected by more features of the channel. Channel size, shape, curvature, slope and roughness, or combinations of these, define the water level upstream (Braca 2008). The effective distance of a channel control is difficult to determine. For example, flat channels have much longer effect than steep channels, which will also have a higher likelihood of other influences within that reach, further complicating flow. For the purposes of developing a robust rating curve, the entire reach would require investigation. The transition zone is where both section and channel control are active.

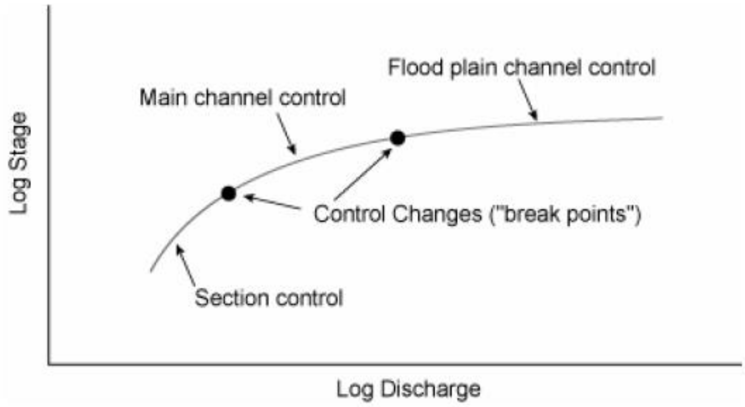


Figure 2.8. Simplified rating curve with different flow controls. (Braca 2008)

When developing a rating curve, it is highly desirable to have multiple ratings for each change of flow control. A change of control will usually present as change in slope or shape of the rating curve (see Figure 2.8). An additional complication is when the physical control changes. Sand beds can be highly unstable and even change after each flood. Stream morphology and erosion will affect the stage-discharge relationship; hence, rating curves are revised regularly to ensure accurate results.

Unsteady flow

Unsteady flow causes a loop rating or hysteresis effect. Unsteady is when the flow changes with time, which all flood events do. As the flood wave travels through the catchment, the water surface will have a leading slope (when flow is increasing) and trailing slope (when flow is decreasing). This will cause different discharges for the same water level. However, under steady flow conditions the water surface will equal channel bed slope and have a single height for discharge. Braca (2008) states that steady state flow occurs only theoretically in prismatic channels. Rapid changes in flow will cause a greater loop effect. The effect is more pronounced for flat streams (Braca 2008 & Maynard 2014).

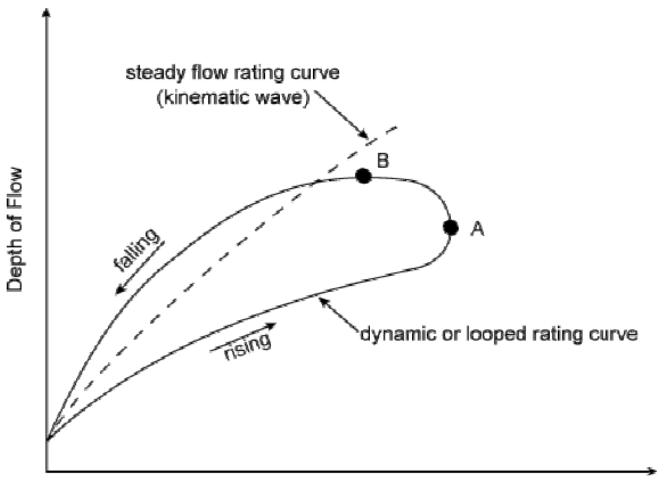


Figure 2.9. Rating curve with unsteady flow (Braca 2008).

2.4 Flood Frequency Analysis

Flood events are spoken of in similar terms as rainfall events. Annual Exceedance Probability (AEP) is also used to describe the probability of a flood exceeding a particular discharge in any one year. Discussion of AEP can be found under Section 2.2.1 'Probability and Recurrence'. The frequency of these floods is estimated through flood frequency analysis (FFA). This process involves fitting a probability distribution to time series of recorded floods and can be rather complex.

2.4.1 Extreme Value Series

The data points used to fit distributions take the form of a time series, of usually maximum values. Frequency analysis may be done for various conditions depending on what is being analysed and what is desired. For drainage design and peak flow estimation, typically, yearly maximum values are desired to predict the probability of that flood or higher occurring each year. For some purposes, analysis monthly or seasonally is required. Figure 2.10 shows how the flood frequency differs seasonally.

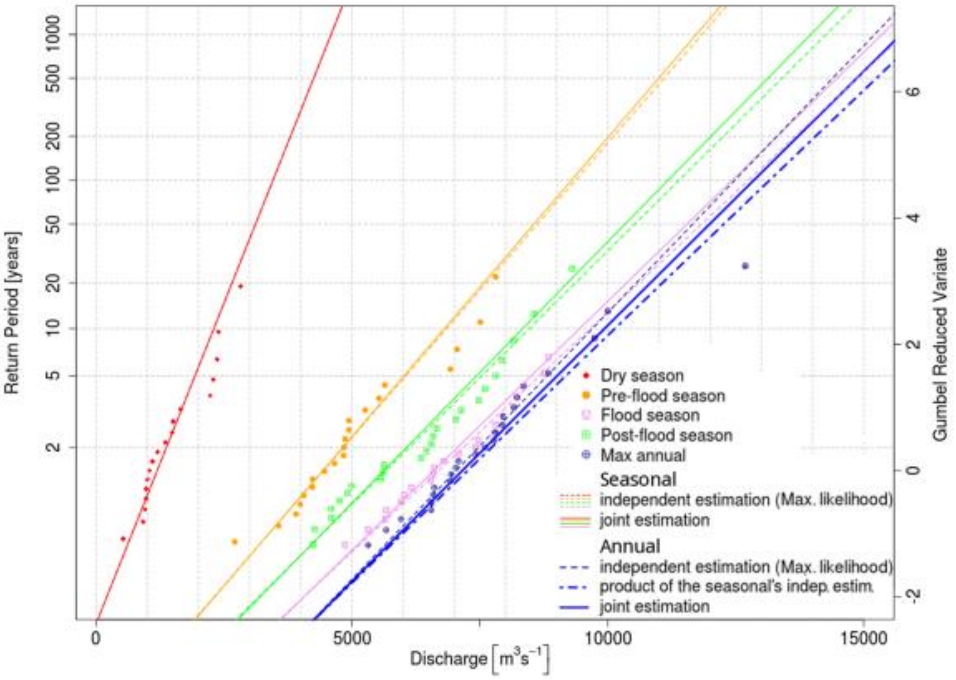


Figure 2.10. Seasonal flood frequency analysis (Baratti et al. 2012).

Peak over threshold (POT) series, also known as partial series, takes a different approach than allotting a time period and finding the maximum. A discharge is selected, the threshold, and all independent floods greater than this discharge is included. The threshold is usually selected based on the number of floods. Research has been conducted to establish the best number of floods for fitting, most of which vary on fitting method, distribution and location. ARR 2019 Book 3 Chapter 2 (Kuczera & Franks 2019) suggests the number of data points equal the number of years of record. POT series requires independence testing of events; methods vary. Two events are not independent if one affects the other.

Annual Maximum (AM) series is the most commonly used and simplest to produce from gauged data. Utilising water year (driest month to driest month) instead of calendar year (Jan-Dec), one can safely assume independence of events. The annual maximum series may incorporate very low flows for dry years, and even though there may be multiple large events within a wet year only the maximum of these is selected. Annual maximum series has been selected for this project and will refer to series of flow unless otherwise stated (for example; AM series of stage).

2.4.2 Missing data

Gauge stations fail for a number of reasons. Hence, there is missing or poor-quality data in the record. This is a difficult problem to deal with when constructing the annual maximum series, with no absolute solution.

There are generally three main options available when determining the annual maximum flood. One, completing regression analysis of other record years (Rahman et al. 2009 p.5-6 & Rahman et al. 2015 p.26), or a similar adjacent catchment (Ball et al. 2019b) and fill in the missing period. This option is only appropriate if good correlation is achieved (Ball et al. 2019b), and has not been considered for this project due to its complexity. Another is to adopt the largest flood available or next largest flood in the case of uncertainty in the measurement. ARR 2019 Book 2 Section 2.3.11 suggests that, if good regression cannot be obtained, that year should be excluded from the annual maximum series unless it is evident the annual maximum flood occurred outside the missing period of record. Baratti et al. (2012) removed 20 out of 45 years of record which had missing or poor data during all or part of the wet season in the study when determining annual maximum series.

It is difficult to judge what the true effect missing data has on the FFA, and will vary for each gauge site. ARR 2019 Book 3 Section 2.2.3 states (Kuczera & Franks 2019);

> "Suspect floods and the years in which they occurred may be omitted in analysis of annual series, but this reduces the sample size and may introduce bias if the suspect floods are all major events."

When removing years from the AM series, the plotting positions (see next section) are affected. If the missing period has the largest flood for the year, which are of most interest, then the frequency curve will underestimate large events. However, if these years are included and you adopt the next smallest (which may vary from zero to large), the frequency of small and medium floods will be increased and will not accurately demonstrate the true frequency of large floods. Consideration of outliers keep plotting positions more or less the same. More details about the assessment of the stream record will be discussed in Section 3.3.

2.4.3 Frequency Plots and Distributions

Probability Distribution Functions

Time series data informs the frequency curve distribution. The most basic frequency curve is one drawn by hand as a by-eye best fit. This method is difficult to use as any event has to be read off the graph, there are inaccuracies in manual methods, it is not reproducible precisely, and regionalisation is not possible. Alternatively, and almost always used, is to find a probability distribution curve to best approximate the data by statistical methods.

There are a number of options for probability models. The best distribution will depend upon many conditions. Typically, only a couple of models are suggested that have been found to perform well in a country or region. United States has adopted Log-Pearson III (LP3), while comparatively the UK now uses Generalised Logistic; pre-1999 the UK used Generalised Extreme Value (GEV) (Kidson & Richards 2005). Restricting the models used generates consistent application and provides a standard for legal and insurance matters. LP3 was used in Palmen and Weeks (2011) and is recommended by ARR 1987, making it very popular even after ARR's revision. Rahman et al. (2013 as cited in Langat et al. 2019) found Log Pearson 3, GEV and Generalised Pareto to be most appropriate throughout Australia. ARR 2019 Book 3 Chapter 2 (Kuczera & Franks 2019) presents these three models, preferring GEV as the best, with their simplified versions Log-Normal, Gumbel and Exponential, respectively. It also describes the possibility of mixture models and non-homogeneous models. The best of LP3 and GEV will be used for quantile estimation as part of this project and are expressed below. The probability notation $P\left(Q \leq q \mid \theta\right)$ describes the probability of a flood, q, is greater than or equal to Q, conditional on theta; theta being input information or parameters.

Distributions have parameters to change the curves' shape and location. Langat et al. (2019) mentions eight different fitting methods to find parameters for the model. The most common is by calculating statistics of the data and translating to the parameters of the model. Each distribution type has different relationships to these statistics.

Statistical moments are comparable to the concept of physical moments of objects. Mean, variance, skew, and less known kurtosis are the first four general statistical moments. Each moment is generally associated to a description. Mean is a measure of the centre of the data, which strongly defines horizontal location. Variance, the square of standard deviation, explains how spread out the bulk of the data is. Skew will inform if most of the data is large or small. Negative skew means the majority of data points are high values. A symmetrical data set has neutral skew and, consequently, mean median and mode are the same value. Kurtosis is more usually, and describes the behaviour or the tails of a distribution.

Log-Pearson III (LP3)

LP3 is a three-parameter probability distribution model for use in extreme value analysis. There is no explicit expression for cumulative distribution. Equation 2.9 uses statistical moments to estimate quantiles,

$$\log(q) = M + K_Y(g)S \tag{2.9}$$

$$K_Y(g) = \begin{cases} \dfrac{2}{g}\left[\left\{\dfrac{g}{6}\left(Z_Y - \dfrac{g}{6}\right)+1\right\}^3 - 1\right] & \text{if } |g| > 0 \\ 0 \ \text{ if } g = 0 \end{cases} \tag{2.10}$$

where, M, S and g are mean, standard deviation, and skewness of the logarithm of the AM series (refer equations 2.11-2.13). Equation 2.10 is a good approximation for frequency factor $K_Y(g)$, where Z_Y is the standard normal cumulative distribution for 1 in Y AEP. Alternatively, tables can be found in Pilgrim & Doran (1987).

Standard Statistical Moments

The following three equations express the standard statistical moments, mean, standard deviation, and skewness, respectively, for a sample. Note, for fitting LP3 distribution logarithm must be applied to the data before calculation.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.11}$$

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{2.12}$$

$$g = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)(N-2)s^3} \tag{2.13}$$

Generalized Extreme Value (GEV)

Similar to LP3, GEV is also a three-parameter model. The parameter τ describes location, α, scale, and κ, shape of the distribution curve. When $\kappa = 0$ the model simplifies to Gumbel two parameter distribution (not presented).

$$P(Q \le q \mid \theta) = e^{\left\{-\left[1-\frac{\kappa(q-\tau)}{\alpha}\right]^{\frac{1}{\kappa}}\right\}} \qquad \begin{array}{l} when\ \kappa > 0, q < \tau + \frac{\alpha}{\kappa}; \\ when\ \kappa > 0, q > \tau + \frac{\alpha}{\kappa} \end{array} \tag{2.14}$$

Standard statistical moments of GEV function are as follows;

$$Mean(q) = \tau + \frac{\alpha}{\kappa}\left[1 - \Gamma(1+\kappa)\right] \qquad \text{for } \kappa > -1 \tag{2.15}$$

$$Variance(q) = \frac{\alpha^2}{\kappa^2}\left[\Gamma(1+2\kappa) - \left[\Gamma(1+\kappa)\right]^2\right] \qquad \text{for } \kappa > -\frac{1}{2} \tag{2.16}$$

where $\Gamma(\ )$ is the gamma function

LH-moments

LP3 uses standard moments skewness, mean and standard deviation of the logarithm of the flow data to fit the curve. However, GEV uses a unique type of statistical moment called L-moment and/or LH-moments to fit to data. L-moments are based on order statistics and are less subject to bias than ordinary moments (Langat et al. 2019). LH-moments are a generalised form of L-moments proposed by Wang (1997) to try and reduce the influence of lower discharges that influence the fit. Wang (1997) introduced a shift parameter η which puts more weight on higher flows, which are usually of more interest. Shift of zero is equivalent to standard L-moment, thus in all cases L-moment equations can be obtained by substituting zero for η. Full description of L/LH-moments is in ARR 2019 and Wang (1997 & 1998).

Estimated LH-moments of the sample data, identified by hats, are calculated from equations 2.17 to 2.20. These equations use combinations, ${}^{m}C_k$, expressed in 2.21.

$$\lambda_1^{\eta} = \frac{1}{{}^{n}C_{\eta+1}} \sum_{i=1}^{n} {}^{i-1}C_{\eta}\, q_{(i)} \tag{2.17}$$

$$\hat{\lambda}_2^{\eta} = \frac{1}{2}\frac{1}{{}^{n}C_{\eta+2}} \sum_{i=1}^{n} \left({}^{i-1}C_{\eta+1} - {}^{i-1}C_{\eta}\,{}^{n-i}C_1\right) q_{(i)} \tag{2.18}$$

$$\hat{\lambda}_3^{\eta} = \frac{1}{3}\frac{1}{{}^{n}C_{\eta+3}} \sum_{i=1}^{n} \left({}^{i-1}C_{\eta+2} - 2\,{}^{i-1}C_{\eta+1}\,{}^{n-i}C_1 + {}^{i-1}C_{\eta}\,{}^{n-i}C_2\right) q_{(i)} \tag{2.19}$$

$$\hat{\lambda}_4^{\eta} = \frac{1}{4}\frac{1}{{}^{n}C_{\eta+4}} \sum_{i=1}^{n} \left({}^{i-1}C_{\eta+3} - 3\,{}^{i-1}C_{\eta+2}\,{}^{n-i}C_1 + 3\,{}^{i-1}C_{\eta+1}\,{}^{n-i}C_2 - {}^{i-1}C_{\eta}\,{}^{n-i}C_3\right) q_{(i)} \tag{2.20}$$

$$\text{where } {}^{m}C_k = \begin{cases} \dfrac{m!}{k!(m-k)!} & \text{if } k \le m \\ 0 & \text{if } k > m \end{cases} \tag{2.21}$$

For practical purposes, the third and fourth moments are standardised into LH-skewness and LH-kurtosis, respectively. Standardisation, or normalisation is when a moment is divided by a form of the second moment, done in order to eliminate location and scale influence on higher order moments. Note this tau in equations 2.22 and 2.23 is different from the location parameter for GEV curve.

$$\text{LH-skewness, } \tau_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta} \tag{2.22}$$

$$\text{LH-kurtosis, } \tau_4^\eta = \frac{\lambda_4^\eta}{\lambda_2^\eta} \tag{2.23}$$

Under the assumption that LH-skewness estimated from the sample data is true, Wang (1997) presented approximations for the shape parameter, kappa. Coefficients to equation 2.24 based on the degree of shift are provided in Table 2.4. Once kappa is obtained, alpha can be calculated by solving 2.26. Further, location parameter, tau, can be determined from equation 2.25.

$$\kappa = a_0 + a_1\left[\tau_3^\eta\right] + a_2\left[\tau_3^\eta\right]^2 + a_3\left[\tau_3^\eta\right]^3 \tag{2.24}$$

Table 2.4. Coefficients to equation 2.24.

η	a_0	a_1	a_2	a_3
0	0.2849	-1.8213	0.8140	-0.2835
1	0.4823	-2.1494	0.7269	-0.2103
2	0.5914	-2.3351	0.6442	-0.1616
3	0.6618	-2.4548	0.5733	-0.1273
4	0.7113	-2.5383	0.5142	-0.1027

$$\lambda_1^\eta = \tau + \frac{\alpha}{\kappa}\left[1 - \Gamma(1+\kappa)(\eta+1)^{-\kappa}\right] \tag{2.25}$$

$$\lambda_2^\eta = \frac{(\eta+2)\alpha\Gamma(1+\kappa)}{2!\kappa}\left[-(\eta+2)^{-\kappa} + (\eta+1)^{-\kappa}\right] \tag{2.26}$$

$$\lambda_3^\eta = \frac{(\eta+3)\alpha\Gamma(1+\kappa)}{3!\kappa}\left[-(\eta+4)(\eta+3)^{-\kappa} + 2(\eta+3)(\eta+2)^{-\kappa} - (\eta+2)(\eta+1)^{-\kappa}\right] \tag{2.27}$$

$$\begin{aligned}\lambda_4^\eta = \frac{(\eta+4)\alpha\Gamma(1+\kappa)}{4!\kappa}\Big[&-(\eta+6)(\eta+5)(\eta+4)^{-\kappa} + 3(\eta+5)(\eta+4)(\eta+3)^{-\kappa} \\ &-3(\eta+4)(\eta+3)(\eta+2)^{-\kappa} + (\eta+3)(\eta+2)(\eta+1)^{-\kappa}\Big]\end{aligned} \tag{2.28}$$

The first three moments are used to estimate parameters for GEV function, leaving the fourth for a goodness-of-fit test to determine if GEV is a reasonable model for the data and which shift to use. Wang (1998) developed this test presented below;

$$z = \frac{\hat{\tau}_4^{\eta} - \tau_4^{\eta}}{\sigma\left(\hat{\tau}_4^{\eta} \mid \hat{\tau}_3^{\eta} = \tau_3^{\eta}\right)} \tag{2.29}$$

where $\hat{\tau}_4^{\eta}$ is the estimated LH-kurtosis based on equations 2.18, 2.20 and 2.23, and τ_4^{η} is the LH-kurtosis based on the fitted GEV parameters. Testing considers the z statistic is approximately normal distributed with mean 0 and variance 1. The standard deviation component (denominator) can be approximated as follows. Wang (1998) reported a fit of R^2 >0.98.

$$\sigma^2\left(\hat{\tau}_4^{\eta} \mid \hat{\tau}_3^{\eta} = \tau_3^{\eta}\right) = \frac{b}{n} + \frac{c}{n^2}, \tag{2.30}$$

where $b = b_0 + b_1\left[\tau_3^{\eta}\right] + b_2\left[\tau_3^{\eta}\right]^2 + b_3\left[\tau_3^{\eta}\right]^3 + b_4\left[\tau_3^{\eta}\right]^4,$ (2.31)

and $c = c_0 + c_1\left[\tau_3^{\eta}\right] + c_2\left[\tau_3^{\eta}\right]^2 + c_3\left[\tau_3^{\eta}\right]^3 + c_4\left[\tau_3^{\eta}\right]^4$ (2.32)

Table 2.5. Coefficients to equation 2.31 & 2.32.

η	b_0	b_1	b_2	b_3	b_4	c_0	c_1	c_2	c_3	c_4
0	0.0745	0.0555	0.0067	-0.3090	0.2240	1.0100	-0.0282	-2.9336	4.0801	-1.0874
1	0.0579	-0.0328	0.1524	-0.4102	0.2672	1.3403	-0.8291	-3.8777	9.5371	-5.7866
2	0.0488	-0.0527	0.1620	-0.3856	0.2566	1.8800	-2.2233	-2.5825	10.435	-7.3887
3	0.0380	-0.0309	0.0254	-0.1233	0.0878	2.6784	-4.8418	3.5255	2.3736	-3.2076
4	0.0241	0.0024	-0.0813	0.0733	-0.0210	3.7793	-8.3485	11.517	-7.9095	1.9459

Plotting of Frequency Curves

The time series of flood peaks is assigned an empirical probability in order to fit a distribution. These plotting positions are given by a general equation 2.33. This formula assumes that α is average over all data points and that the underlying distribution is approximately symmetrical (Pilgrim & Doran 1987).

$$PP(i) = \frac{i - \alpha}{N + 1 - 2\alpha} \tag{2.33}$$

where i is the rank of gauged flood (rank 1 is largest flood), N is the total number of floods in the time series, α is a constant and $PP(i)$ is plotting position in AEP.

The AEP-quantile (discharge for FFA) relationship is not proportional and so the average of one doesn't match the average of the other, as an example. There has been contention over which characteristic should be unbiased. The constant α is 0.0 to obtain unbiased AEP estimates, however, ARR 2019 Book 3 Chapter 2 (Kuczera & Franks 2019) suggests the use of Cunnane's (1978 as cited in Kuczera & Franks 2019) plotting position when α is 0.4 (equation 2.34). 0.4 best suits all typical distribution types and produces almost unbiased quantile estimates.

$$PP(i) = \frac{i - 0.4}{N + 0.2} \tag{2.34}$$

Frequency curves are plotted on graphs with unique axis scales. This is to improve the graph visually and help make judgements on the goodness-of-fit. Log-normal plots have a logarithmic scale on the y-axis to describe discharge, and the x-axis describes the probability and, based on Normal variate, the number of standard deviations from the mean. Log-normal plots have been used for a number of years as they are suited to LP3 distributions. Gumbel plots have been introduced to cater for Gumbel and GEV curves. They generally consist of a linear y-axis scale and use Gumbel variate for the x-axis.

Low Outliers

Outliers are data points that may be incorrect due to errors in data, don't generally fit data trend, or overly influence statistical properties of the data set. Flood series typically have outliers with low or no flow, especially in droughts, which can affect fitting of distribution curves. Therefore, tests are used to identify those floods which adversely influence the curve, especially those effecting the high tail, which is usually of most interest. Traditionally, Grubbs-Beck test is used to test for outliers. Cohn et al. (2013) has improved this method, multiple Grubbs-Beck test (MGBT), which has been adopted in the US and now by ARR 2019. Deficiencies in the original test included the assumption of zero skew in the data and that it only tested for single outliers at a single significance level. Conceptually, MGBT transforms the time series to standardised log-normal, and tests at a specified significance level, if the k smallest flows are likely to resemble the rest of the data. If the calculated p-value for the kth flow is below the significance level then these are considered outliers. ARR 2019 Book 3 Section 2.6.3.9 (Kuczera & Franks 2019) details testing procedure further; an outward sweep from the median at 0.5% significance, and an inward sweep at 10% significance; with the greater being the adopted censored flows. MGBT and this procedure take into consideration masking and swamping effects of the data, and variable mean and spread of the time series.

High Outliers

High outliers are much more difficult to assess than low outliers and is considerably subjective. These are the floods of most interest, so treatment should be carefully chosen. Judgment should be based on statistical evidence and a qualitative assessment of the data. ARR 2019 does not discuss high outliers, however, ARR

1987 provides guidance in Chapter 10 Section 10.7.3 (a) (Pilgrim & Doran 1987, pp. 215-217) and a decision matrix to treat the suspected high outliers (provided below, Table 2.6). Omission of a large flood is regard as an extreme action.

There are four main causes of high outliers and all are treated differently, as they will have differing significance/importance, accuracy, and influence over the FFA;

1. Errors in data,
2. Changes in catchment conditions,
3. Extremely rare event, and
4. Unusual type of phenomenon causing flood.

Examples of 'Changes in catchment conditions' would be sudden change of vegetation due to bush fire, urbanisation effects, storage changes from reservoirs. Although it is suggested to search for any historical information to help identify cause of the flood and quality of data, it is often very difficult to obtain.

The statistical component of the assessment is based on LP3 theory and Grubbs Beck test. Equation 2.35 calculates a discharge threshold, above which is considered an outlier. If the flood is approximately equal to the threshold, statistical evidence is considered weak.

$$X_H = m + \beta K_N s \tag{2.35}$$

where X_H is the threshold for outliers in log units, K_N is given by Table 10.6 ARR 1987 Chapter 10, and β is an adjustment factor based on skew and sample size and given by Table 10.7 ARR 1987 Chapter 10. The tables have not been included due to size.

Table 2.6. Guide for Treatment of Outliers (Pilgrim & Doran 1987, p. 217; Table 10.9).

Cause	Statistical Evidence	Prior Belief		
		Strong	**Moderate**	**Weak**
1. Error in Data	**Strong**	Omit	Omit	Omit or accept
	Moderate	Omit	Modify magnitude or accept	Modify magnitude or accept
	Weak	Modify magnitude or accept	Accept	Accept
2. Change in catchment conditions	**Strong**	Omit	Omit or accept	Omit or accept
	Moderate	Omit or accept	Omit or accept	Accept
	Weak	Accept	Accept	Accept
3. Rare occurrence with very low AEP	**Strong**	Modify probability	Modify probability or omit	Omit
	Moderate	Modify probability	Modify probability or accept	Modify probability or accept
	Weak	Accept or modify probability	Accept	Accept

4. Unusual type of phenomenon	**Strong**	Compound distribution	Compound distribution	Compound distribution or omit
	Moderate	Compound distribution or accept	Compound distribution or accept	Accept or omit
	Weak	Accept	Accept	Accept
Note: In all cases where statistical evidence is nil, accept data.				

TUFLOW Flike

TUFLOW Flike is a flood frequency analysis program developed by Prof. Kuczera of the University of Newcastle Australia (TUFLOW 2015). It is recommended and exemplified in Book 3 Chapter 2 of ARR 2019 (Kuczera & Franks 2019). The program uses advanced statistics and tests to process time series data. The superior multiple Grubbs-Beck test is included and can censor data effectively for historical events, low and high outliers. All recommended probability models and techniques are built in.

Regional information can be included as prior distribution information for Bayesian statistics. This data can be sourced from RFFE tool as a downloaded file which gathers the FFA of those sites that are used in RFFE. This prior information influences the results, following that, if more information is known, a better estimate can be made.

Flike tries to find ‘most probable’ parameters for the distribution by one of two methods. Method 1 involves a ‘hill-climbing’ method; unconstrained quasi-Newton search. This method cannot guarantee a global optimum. Figure 2.11 is an example fitting GEV distribution, which shows contours of good fit for a various shape and scale factors. Notice that if starting with parameters values in the bottom right corner of the graph, climbing the contours will stop at Optimum 1, not continuing onto the global optimum. Alternatively, a more computationally heavy method is applied which uses shuffled complex evolution algorithms to calculate the global optimum.

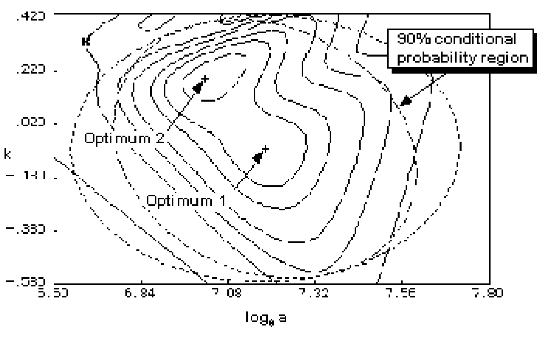


Figure 2.11. Optimisation contours for fitting GEV parameters.

The quantiles estimated, based on the sample data, may not produce the true quantile based on parameters. ARR 2019 describes linear asymmetric and quadratic loss function to cater for this. Flike has adopted the latter. Importance sampling is then used to solve the final model.

2.4.4 Assessment of Frequency Curves

There are many goodness-of-fit tests and methods to select the most likely/representative model for a set of data. Generally, goodness-of-fit tests examine whether a sample came from a theoretical distribution. Some widely used include: Anderson-Darling (AD) test, Kolmogorov-Smirnov (KS) test, and Cramer-Von Mises (CVM) test. For the purposes of this project, the theory behind the KS test will be used for comparing distributions, but not the test in standard form. In practice AD and CVM tests cannot be applied to censored data sets, hence they will not be used. QQ-plots, PP-plots, and frequency curve plotted on log-normal scale will also be used.

Kolmogorov-Smirnov (KS) test statistic specifies the maximum vertical distance between an empirical (sample) and theoretical cumulative distribution function. Therefore, it can be easily seen that a smaller distance will imply a better fit to the data. The tested distribution must be continuous; as we are only dealing with LP3 and GEV, we satisfy this requirement. The standard test statistic is defined below;

$$D = \max_{1 \le i \le N} \left(F\left(x_i\right) - \frac{i-1}{N}, \frac{i}{N} - F\left(x_i\right) \right) \tag{2.36}$$

where D is the Kolmogorov-Smirnov test statistic, N is the total number of data points, and $F\left(x_i\right)$ is the theoretical CDF at each data point.

Two modifications will be made to the KS test for practicality to this application. Firstly, it is typical for the empirical distribution in the KS test to be based on i/N, however, the data plotting positions $PP\left(x_i\right)$ will be used instead (equation 2.37). If the empirical distributions differ between fitting a model and the test, it may cause incorrect or biased results. Secondly, the test will be limited to 50% AEP and above. LH-moments fitting, due to the shift towards large events, cause a very large error for small events. For the current project only events above 50% AEP were of interest. Preliminary trials clearly showed this was necessary and effective.

$$D = \max_{1 \le i \le N} \left(F\left(x_i\right) - PP\left(x_{i-1}\right), PP\left(x_i\right) - F\left(x_i\right) \right) \tag{2.37}$$

Probability-probability plots (PP plot) are a graphic tool to assess how closely two distributions match. For each quantity, such as discharge of an event, the probability of one distribution is plotted against the probability of the other distribution. In the case of frequency analysis, an empirical distribution (i.e. Cunnane's plotting position as cited in Kuczera & Franks 2019) is compared against the fitted theoretical distribution (Langat et al. 2019). If the data follows a 1:1 line, they are exactly expressed by the fitted distribution. Conversely, a quantile-quantile plot (QQ plot) plots discharge against the estimated discharge of the fitted distribution of the same probability. These plots are a good visual tool to assess many distributions and identify areas which may not be well represented. QQ plots tend to exaggerate the tails as small probability changes result in large discharge differences.

2.5 Regression

2.5.1 Model Structure

Regression is a common method of developing an empirical model for a data set based on input parameters. In the case of predicting discharge there are two types of regression; quantile and parameter regression. In parameter regression you find the values of the parameters for a frequency distribution, then calculate the discharge for a given AEP. Quantile regression technique (QRT) uses a set of quantiles for a number of sites to determine an equation for discharge which depends on explanatory variables, such as catchment area, rainfall intensity, and so on. QRT was utilised by Palmen and Weeks (2011) and will be continued in this work. The significance of different explanatory variables was completed by Palmen and Weeks (2011), thus will not be investigated further and is assumed correct.

The prevailing model structure for hydrology equations is to have a number of variables to the power of constants, as in the general equation 2.38. Under quantile regression a different set of coefficients ($a, b, c \ldots$) are required for each quantile of interest.

$$Y = A^a B^b C^c \ldots \tag{2.38}$$

Equation 2.39 shows the specific form that will be used. Catchment area A and rainfall intensity I was found to contribute the most to discharge Q (Palmen and Weeks 2011). A base 10 for the constant is inconsequence as any number will produce the same results.

$$Q = 10^{\alpha} A^{\beta} I^{\gamma} \tag{2.39}$$

Unfortunately, regression of this form of equation is difficult, therefore, it requires transformation into a usable form then regression will be applied to determine the coefficients. Linear equations can easily be obtained by applying logarithm and following standard logarithm rules, as seen below (equation 2.40).

$$\begin{aligned}
\log Q &= \log\left(10^{\alpha} A^{\beta} I^{\gamma}\right) \\
\log Q &= \log\left(10^{\alpha}\right) + \log\left(A^{\beta}\right) + \log\left(I^{\gamma}\right) \\
\log Q &= \alpha \log 10 + \beta \log A + \gamma \log I
\end{aligned} \tag{2.40}$$

In this linear form regression techniques can be easily applied. An addition error term is required for regression to measure the difference between the observation and predicted value. Matrix form, presented below, is a convenient format and utilised by least-squares regression.

$$
\begin{aligned}
&\log Q_i = \alpha \log 10 + \beta \log A_i + \gamma \log I_i + \varepsilon_i \\
&\begin{bmatrix} \log Q_1 \\ \log Q_2 \\ \vdots \\ \log Q_n \end{bmatrix} = \begin{bmatrix} \log 10 & \log A_1 & \log I_1 \\ \log 10 & \log A_2 & \log I_2 \\ \vdots & \vdots & \vdots \\ \log 10 & \log A_n & \log I_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\
&Y = XC + \eta
\end{aligned}
\tag{2.41}
$$

where for each gauge ($i = 1, 2 \ldots n$): Q_i is the estimated AEP discharge for a site; A_i is the catchment area; I_i is the intensity at the AEP and selected duration; ε_i is the error from the observed discharge to the discharge estimated by the regression model; and α , β , and γ are coefficients of the model.

An important mathematical requirement of regression is for the parameters to be independent of each other. The change of one variable does not cause an effect in another. Area and rainfall intensity are independent of each other because no matter how heavy the rain the catchment will not change area. Some other assumptions are that errors from the fitted model to the data are random, uncorrelated with the independent variables, or other observations.

2.5.2 Least Squares Regression

The principle aim of least squares regression is to find the model that minimises the difference between observations and predictions by the model. It is important that all explanatory variables are independent and uncorrelated. Ordinary least squares (OLS) considers all points to have the same influence. However, variations of OLS can have the potential to give weightings to points which will influence the final model. Weightings could have various purposes such as data points with high uncertainty can be weighted lower. Weighted least-squares (WLS) allows for such weightings, as well as Generalised least-squares (GLS).

There are many different varieties of least squares regression. Most differ in the assumption of scedasticity, that is the distribution of the error between the model and data points. If the error residuals have mean of zero and constant variance it is homoscedastic, otherwise if the variance changes it is referred to as heteroskedastic. This characteristic can be visually assessed by residual plots. Ordinary Least Squares assumes homoscedasticity and is the best linear unbiased estimator (BLUE). GLS and others require transformation into a form which OLS works to obtain BLUE regression. Note that best refers to minimum variance of errors.

Ordinary least squares is easily calculated in matrix form (refer equation 2.42). $\hat{C}_{OLS}$ will be a matrix of estimated coefficients for the model.

$$
\begin{aligned}
&Y = XC + \eta \\
&\hat{C}_{OLS} = \left(X'X \right)^{-1} X'Y
\end{aligned}
\tag{2.42}
$$

2.5.3 Bias Correction Factor

A Bias Correction Factor is needed when using transformed equations for regression. Often the regression is completed and coefficients simply substituted into the original equation, forgetting the error term used in regression. The error term has assumed mean of zero as the logarithm in our case, but this does not transform back to a mean of zero in the final equation (Newman 1993). In some applications of log transforms reviewed by Newman (1993), there was found up to 57% bias in predictions. Therefore, the correct formula is shown below (equation 2.43), with the error term absorbed into the constant of the final equation.

$$Q = 10^{\alpha} A^{\beta} I^{\gamma} \times 10^{\varepsilon} = 10^{\alpha+\varepsilon} A^{\beta} I^{\gamma} = \delta A^{\beta} I^{\gamma} \tag{2.43}$$

Newman (1993) suggests equations 2.44 and 2.45 for this error term, assuming residuals are normally distributed. Alternatively, when residuals are not normally distributed a different correction factor is used.

$$10^{\varepsilon} = 10^{\frac{MSE}{2}} \tag{2.44}$$

where MSE is the mean square of the error from the regression,

$$MSE = \frac{\sum_{i=1}^{N} r_i^2}{N-2} \tag{2.45}$$

and r is the regression residual from the ith data point and N is the total number of data points.

2.5.4 Model Evaluation

Residual plots are an effective tool to assess the regression model. The difference between the model's predicted value and observed value, called the residual, is plotted against one or more explanatory variables. If there are any patterns, such as increasing residual with a variable, there may be systemic or model error. Consistent randomness in the residual plot suggests a good fit. If the residuals show an increasing variance and they are becoming more spread out, this may suggest WLS or GLS is required. Residual plots also assist in identifying outliers.

Coefficient of Determination or R^2 measures the amount of variability in the response that is explained by the model (NIST 2013b & Palmen and Weeks 2011). It has a scale from negative infinity to one (negative results are rare), where one is a perfect model with all observations equal to the model. Near zero indicates the model is not much better than the simplest possible model, an average of sample estimates (Drakos 2018). It is calculated by equation 2.46 below,

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2}{\sum_{i=1}^{N} \left(y_i - \overline{y} \right)^2} \tag{2.46}$$

with $\overline{y}$ is mean of the data, y_i is each observation, and $\hat{y}_i$ is the model prediction at for every observation. NIST (2013b) warns that R^2 should not be the only test for model validation as it doesn't guarantee a good fitting model. It is a biased estimator tending to be too high (Frost n.d.).

Adjusted R^2 is a modified version of standard R^2, and is corrected for bias and is usually considered more realistic of the model performance. It also has the ability to compensate for additional variables. However, in this project the number of parameters is consistent in all cases. The adjustment is given in equation 2.47 below.

$$R_{adj}^2 = 1 - \left[\frac{\left(1 - R^2\right)\left(N - 1\right)}{N - k - 1} \right] \tag{2.47}$$

where N is the size of the sample, and k is the number of independent explanatory variables.

The following statistics are common for describing the amount of error of the model. Mean Absolute Error (MAE) simply takes an average of all the errors, irrespective of over- or under-estimation, and is in real units of the dependent variable. Root Mean Squared Error (RMSE) is used to apply more weight to very bad predictions, when errors are large. Mean Absolute Percentage Error (MAPE) and Root Mean Squared Percentage Error (RMSPE) are similar, respectively, but uses the error as a percentage of the observation instead, so are expressed as percentages.

$$MAE = \frac{\sum_{i=1}^{N} \left| y_i - \hat{y}_i \right|}{N} \tag{2.48}$$

$$MAPE = \frac{\sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{N} \tag{2.49}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2}{N}} \tag{2.50}$$

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2}{N}} \tag{2.51}$$

Another simple method to evaluating the accuracy of the model is to count the number of observations within a given range of the model. A greater proportion of data points within this range will imply a more accurate model. The total range of error and distribution of errors should also be reviewed. Extreme discrepancies may indicate the regression was incorrect, outliers or the model is not appropriate.

3 Methodology

3.1 Overview

This chapter will layout the process followed in this project. First the data was gathered; that is stream flow record, catchment characteristics, and design rainfalls. The data was assessed and prepared for use as the Annual Maximum Series. Next, the frequency of floods was analysed and design quantiles calculated. These results were then associated with a design rainfall and catchment area to produce the final model, which was tested and validated. For the sake of brevity, due to the large number of gauges, full details of every gauge are not presented; appendices provide more details.

During this project Australian Rainfall and Runoff 2016 advanced draft was finalised. The final document was reviewed to check any significant changes regarding components of this project and nothing influential was found.

Access to a high-performance computer was provided by DTMR (by remote access) in order to complete the 2D hydraulic modelling. This resulted in using three computer systems, where not all the required software was available. Information management was challenging and multiple backups taken throughout. In addition, data size was substantial, especially dealing with LiDAR – approximately 1.5 terabytes. The software used in this project is sensitive to naming conventions and folder structure, especially MATLAB and TUFLOW. If the files are moved or do not match names, the programs cannot find them. However, for the sake of work flow, specifically when using MATLAB, the folder structure was changed from initial tasks as the project progressed.

3.2 Data Acquisition and Site Selection

3.2.1 Stream Flow Data

All stream flow data was sourced through the Water Monitoring Information Portal (WMIP) (DNRME 2019). Metadata for each station was downloaded to assess for suitability to the project. Daily and monthly stream flow records were then acquired for the selected gauges, along with any cross section, rating curves, and site photos available. Almost all data was downloaded by March/April 2019. It has been observed that since then the website has been modified slightly and some gauges have had various updates.

MATLAB (MathWorks 2019) was used to extract the desired data from WMIP files and format it to be more usable. With over 1200 gauges, data extraction was significantly faster than assessing each manually. Although downloaded files were reasonably consistent in format, a few required manual completion. The rating table, gaugings and metadata were separated into different files, which could then be read into any scripts or functions used later in the project.

3.2.2 Acquire Rainfall IFD tables

1987 Intensity-Frequency-Duration Tables

The Bureau of Meteorology provided an online service for requesting 1987 IFD tables. However, decommissioning is scheduled for June 2020. Using the non-Flash version of this tool, each site selected was requested individually. All tables were specified at the gauge location and not the catchment centroid. The input screen for an example location is depicted below (Figure 3.1). An IFD graph and table, as well as polynomial coefficients for IFD curves were provided by the online tool.

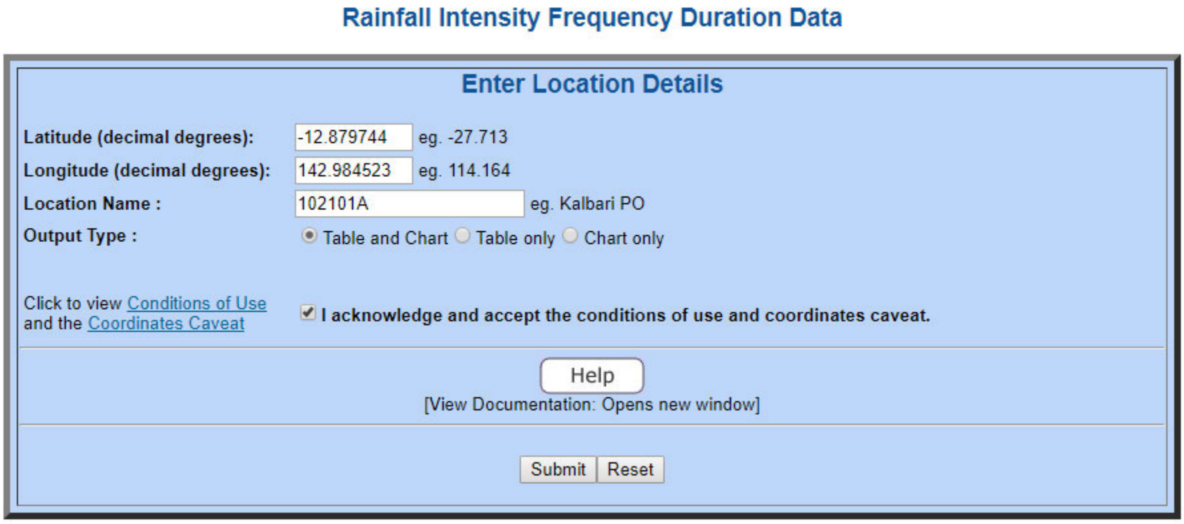


Figure 3.1 Page for 1987 IFD enquires (non-flash version).

2016 Intensity-Frequency-Duration Tables

The new 2016 IFD online tool has much more functionality. Figure 3.2 shows the input page. Multiple points request is limited to 50 points. A template for decimal degrees can be downloaded and populated manually to ensure the tool can read the .csv file. All sites were specified at the gauge location not catchment centroid. Two requests were required for all the gauges being used. The results were then downloaded as a compressed .zip file. Depths, polynomial coefficients, and intensities for rare (>1% AEP), very frequent (in Exceedances per Year), and standard frequencies were provided as results. Durations range from 1 min to 7 days (10,080 mins). The extracted files were .csv format, which can be easily read by MATLAB.

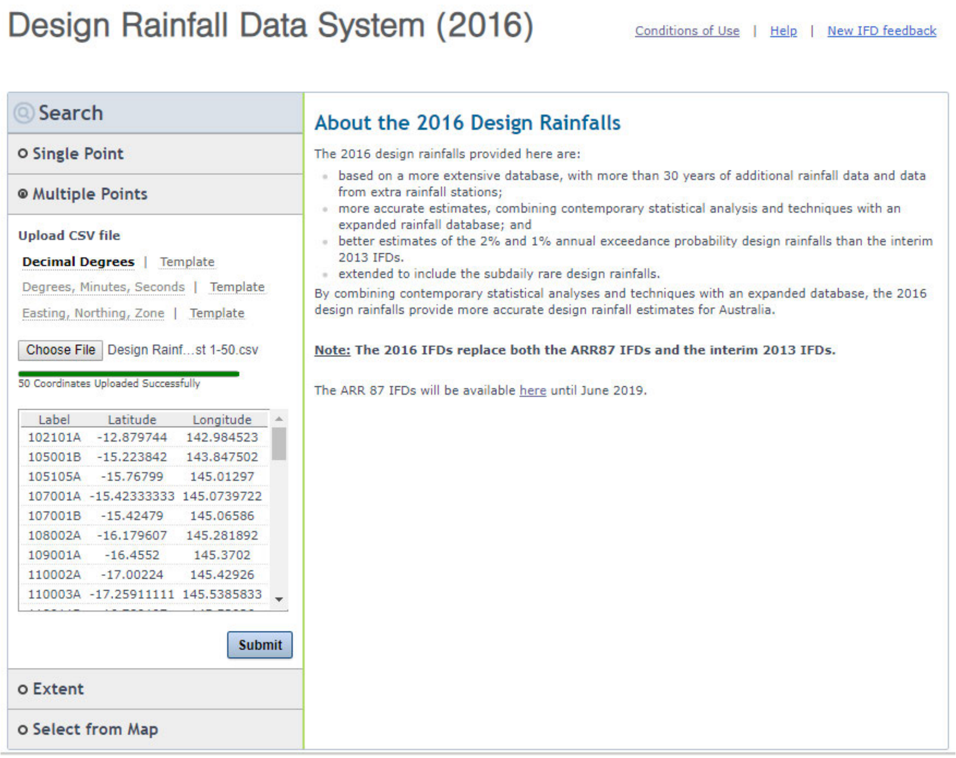


Figure 3.2. 2016 IFD online tool.

3.2.3 Initial selection

The first criteria used to filter gauges was catchment size. Catchments were limited to 1,000km^2, which follows Palmen and Weeks' (2011) paper. ARR 1987 (vol. 1, ch. 5, p. 95) defines small and medium catchments with an upper limit of 25km^2 and 1,000km^2, respectively. In addition, RFFE results are considered lower accuracy when catchments are greater than 1,000km^2. Larger catchments behave differently with different governing hydrology, and may not have the same relationships as small to medium catchments

The length of stream record was also a major consideration when selecting usable sites. Short records may compromise FFA and reduce accuracy and confidence of flow quantile estimates (Kucsera & Franks 2019). Palmen & Weeks (2011) selected all sites with ten or more years of record, and also tested twenty year minimum but found no significant difference. This project has adopted a minimum of fifty years, with some additional inland gauges. This approach was taken for a variety of reasons. The longer time series reduces the influence of errors and outliers, and uncertainty in flood frequency analysis. It also allows for the likelihood of missing data; however, if when interrogating the data there is excessive missing data the gauge will be removed from analysis. More data points help with fitting the frequency distribution curve and making judgements throughout the process; that is more points to view and identify any issues and goodness-of-fit. A longer gauge record encourages confidence in the results.

Some gauges have been rebuilt or moved slightly. These have the potential for their record to be joined together. Only gauges at the same location have been considered to be joined; no adjacent gauges, which is allowed by ARR, due to the complexity and time required. A couple of sites were selected from regional inland areas that have less than fifty years of record, to improve distribution over Queensland.

3.2.4 Sites Removed

Objective was to have about fifty gauge stations for regression. All 97 gauges filtered by initial selection criteria were considered and reviewed. The primary criterion is to have accurate discharge which follows from an accurate rating curve. The influences and uncertainties regarding and importance of precise rating curves is discussed in Section 2.3.3. Several aspects were reviewed when deciding what action to take concerning rating curves provided with gauge data. Those deemed acceptable were adopted as accurate, and discharges in the stream record adopted. Further investigation was taken for suspect sites. These were either found to be acceptable, required revision of the rating curve by 2D hydraulic modelling, or the gauge was excluded from this project.

Preliminary investigation of gauges included visual assessment of the rating curve and gaugings, review of the provided rated cross section, and maps of contours, streams and vegetation. Gauges without a rating curve and cross section provided were immediately removed, because accuracy cannot be reviewed. All coastal gauges were considered far enough inland and high enough elevation to assume no backwater effects from tides. Any gauges downstream or upstream of large reservoirs were also removed, assuming the reservoir causes a variable and unknown effect on the discharge. For example, 146002A which is downstream of Advancetown Lake water supply for the Gold Coast will catch more water after a drought then in wetter periods.

Cross section was an important tool in deselecting gauges. Any gauges that did not show the rated cross section to the height of the maximum record was checked with available LiDAR. Sites with the maximum stage within uniform channel shape and well-defined banks were generally accepted. Flood plains are near impossible to extrapolate the rating curve based on cross section and could interact with other streams on the flood plain, requiring 2D modelling to be adopted for all cases. While gauge 137101A and 422211A did overflow the main channel, gaugings were measured up to 99% of the maximum record. Gaugings were between 3.6-99% of the maximum record stage for sites, with the mean being 51.9%. A wide channel, namely western streams, will cause shallow flow and may have multiple and/or changing flow paths. Only a couple of these were encountered and subsequently excluded. When LiDAR was not available but was required for 2D modelling or reviewing the cross section, the gauge was removed.

If gaugings are close to maximum record stage there is less error and extrapolation. Gauges with large range of extrapolation, especially with changes in cross section profile, have risk of considerable error. These gauges were considered for revision. Variability of gaugings, and gaugings over time (may show change in catchment over time) was also considered in the assessment. Variability in small gaugings is of less consequence.

The process of 2D hydraulic modelling further reduced the number of acceptable gauges. Any sites that had a break-out, where water exits the channel upstream and bypasses the measurement station, were excluded as the channel does not convey the entire discharge of the catchment. This was obvious for some sites by inspection of LiDAR without any hydraulic modelling. A few gauges were excluded because another catchment fed a stream that would interact by backwater effects or across a flood plain. Some consideration was given to the catchment size and potential time of concentration.

Gauge 919201A had a kink in the rating curve and upon further investigation was found to have two main channels that could interact. In addition, the site had no LiDAR available to complete 2D modelling and downstream of the gauge was a bridge at the height of large flood which could cause backwater effects. Therefore, this gauge was removed.

Applying these principles left 51 sites for regression. Appendix E summarises the selection of gauges.

3.2.5 Final Selection

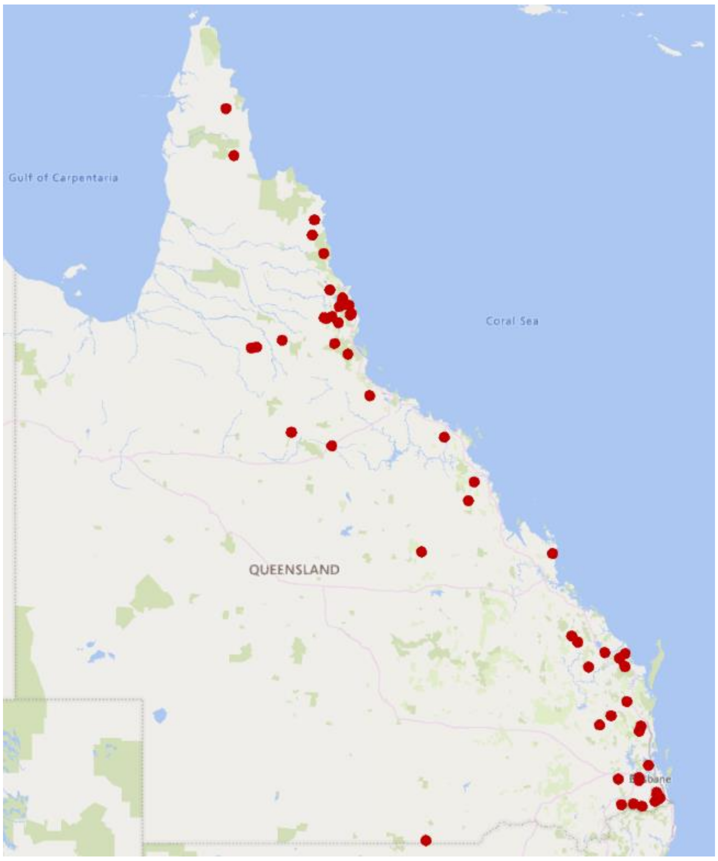


Figure 3.3. Gauges selected for regression.

The above map shows the 51 gauges selected for this project. They are mostly concentrated along the east coast of Queensland. Gauges are very sparse a long way inland, and those available often have very large catchments, or the location is on a flood plain or in very wide rivers. Many did not meet the age criteria as well.

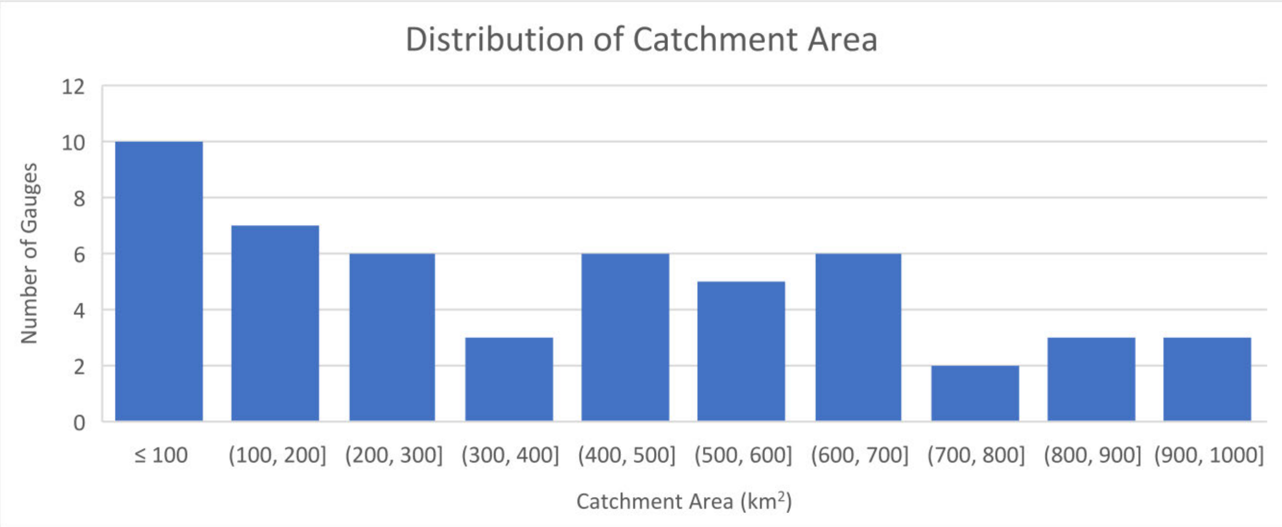


Figure 3.4. Catchment area distribution

The selected gauges have a reasonable distribution of catchment areas ranging from 15.7km^2 to 925km^2, with a mean of 398km^2. There are a few more smaller catchments and so may cause the final model to fit smaller catchments better than larger ones. Rainfall intensity is not so evenly distributed. There is an obvious skew towards smaller intensities. A reason for this may be due to most of the gauges being of more similar region, near the east coast, not spread across regional inland. As an example (Figure 3.5), most gauges have between 4 and 6 mm/hr rainfall for a 72hour duration 2% AEP event.

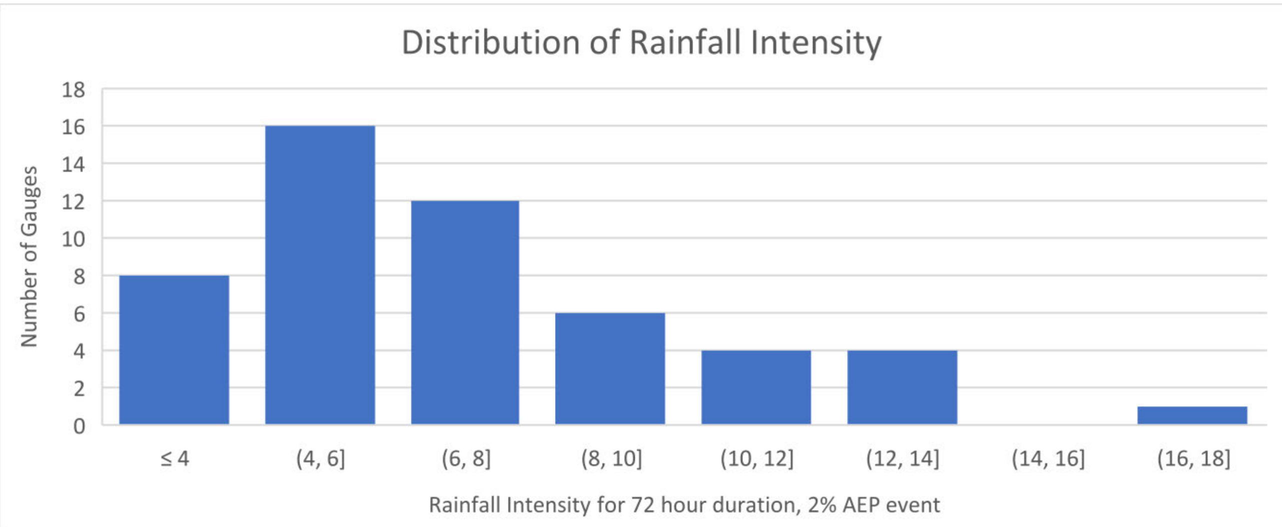


Figure 3.5. Rainfall intensity distribution

3.3 Annual Maximum Series

A spreadsheet was developed to assist identifying the annual maximum series. Daily data record was arranged into each water year, with flow or stage, rainfall, quality code and indication of suspect data points. Water year was assumed to be July to June based on Queensland having a clear summer wet season. This was also confirmed in conduct. Maximums for months were summarised for each year and a typical year also presented. The annual maximum was then confirmed manually, and comments made. Finally, a column graph of the annual maximum series and box and whisker plot were presented to identify error or extreme outliers.

Discharge annual maximum series was derived by three methods;

- Method A – AM series of record discharges (historical rating curves);
- Method B – apply current rating curve to AM series of stages;
- Method C – apply revised rating curve to AM series of stages.

Appendix E has a list of the adopted AM series for FFA. However, some AM series were generated for comparison purposes.

3.3.1 Method A

Annual maximum series was generated for all gauges with available discharge data. The discharge data provided through WMIP system is converted from measured stage to flow rate by the historically applicable rating curve for that period. The accuracy of these rating curves cannot be known and are not readily available to the public. While 116008B, 121001A, and 142202A had a small number of years where discharge data was not available, the number of remaining years were all above 44, thus likely to produce accurate results. Sacrificing a couple of missing years for more accurate discharges is beneficial. For example, if Method B was adopted for gauge 142202A, discharges pre-1995 would be under-estimated between 40-80% (Figure 3.6). Conversely, gauges with greater number of missing discharge record needed Method B to be adopted.

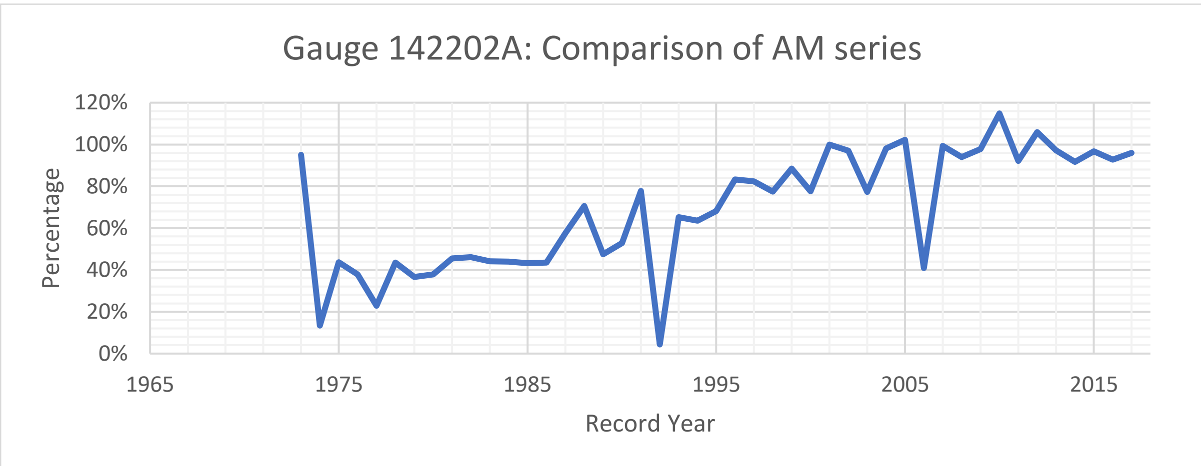


Figure 3.6. Ratio of discharge obtained from Method B over Method A for gauge 142202A. AM series in this graph was not used in FFA.

3.3.2 Method B

For some gauges part or all of the record had missing discharge data where stage was still recorded. This may have been due to missing or inaccurate stage-discharge information. This disparity was often coded 151; 'Data not yet available'. In these cases, a stage instead of discharge AM series was generated and converted using the current DNRME rating curve provided. This method is based on an assumption that the current rating curve represents historical floods accurately. Gauges 130407A and 136006A were the only ones that adopted Method B for FFA. Additional AM series were generated for gauges to compare against Method A and C derived series.

3.3.3 Method C

A number of gauges were identified to revise their rating curves using hydrodynamic modelling. Firstly, to confirm that the revised rating curve could be applied to the entire record period, a check was performed. Method B assumes the current rating curve applies to all years, so by comparing it to the stream discharge record determined by the appropriate historical rating curve (Method A), it can be identified if the stage-discharge relationship has changed significantly over time. An example of consistent stage-discharge relationship for gauge 138002ABC is shown in Figure 3.7. Ratio of small numbers can cause exaggerated percentage, which occurs for low flows because precise conversion can be difficult, hence the large dips in the graph.

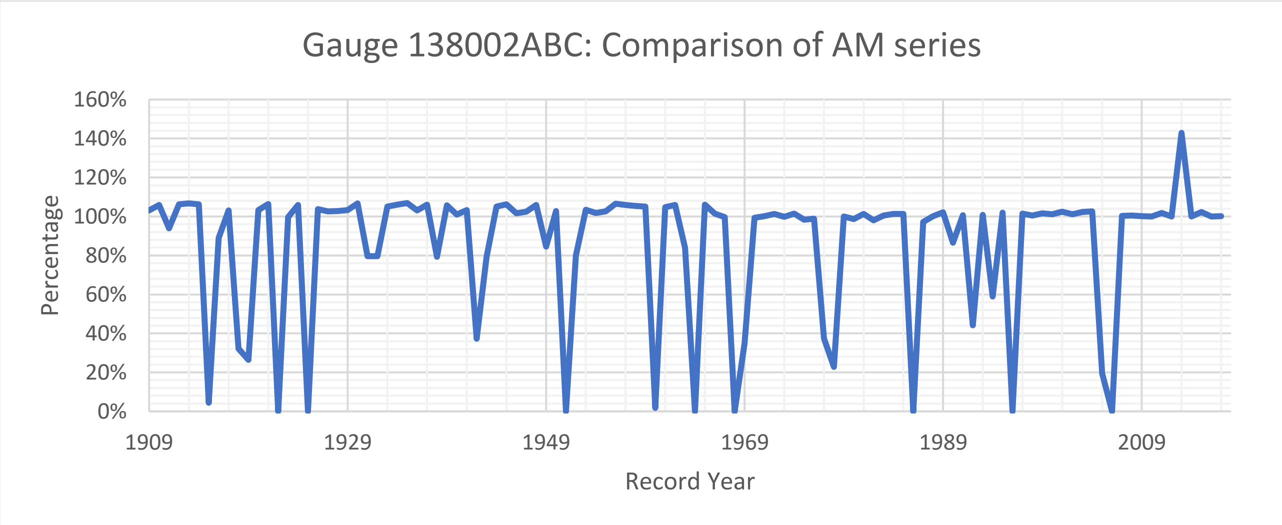


Figure 3.7. Ratio of discharge obtained from Method B over Method A for gauge 138002ABC. AM series in this graph was not used in FFA.

Once the TUFLOW modelling was completed, the new rating curve was applied to the stage AM series. Some modelling showed that DNRME rating curve is appropriate, so Method B was adopted. Section 3.4 will discuss all gauges that were modelled by TUFLOW. Method C was adopted for eleven gauges for FFA. Comment on the historical check are provided in Table 3.1.

Table 3.1. Assessment of historical accuracy for applying revised rating curve.

Gauge	Historically Accurate	Comment
111105A	Fair	Pre-1978 changes a little, maximum of 16%, but most of record is good.
112002A	Unsure	Almost half of discharge record not available for comparison.
129001A	Good	
135004A	Fair	1979-2012 over-estimate discharge by approximately 20%.
136111A	Good	Intermittent years missing discharge data throughout. Never greater than 10% difference.
138002ABC	Good	
145003B	Fair	See Section 3.3.4 and Figure 3.8.
145011A	Poor	Pre-2000 difference range of 45-130%, but is not consistent.
145102B	Unsure	Almost entire discharge record not available.
917104A	Good	
922101B	Good	

3.3.4 Joining Gauge Records

As mentioned previously, gauge records may be joined together. Only those with similar conditions between revisions were joined. Most closed gauges do not have rating curves or other details readily available to review. Only three gauges were joined together: 107001AB, 137001AB, and 138002ABC. Gauge 138002ABC rating curve was revised using TUFLOW as it was found acceptable to apply the revised TUFLOW rating curve to the entire joined record (see Figure 3.7). Gauges 107001AB and 137001AB both followed Method A methodology.

Logan River at Forest Home gauge (145003) was considered to join A and B gauge records and was identified to revise its rating curve which would assume it does not change over time. Figure 3.8 compares AM series as done under Method C to check it can be applied historically. There is large amount of discharge record missing between 1954 and 1973. It is easily seen that the earlier version, gauge A, has a different relationship between stage and discharge; the same stage converts to approximately 20% of the true discharge. Therefore, only B gauge was used in the project. Gauge B between 1973 and 2005 appears to underestimate discharge by approximately 25% if the current rating curve is to be applied, which was similar to some other gauges. The full record of gauge B was adopted and stage converted using the revised TUFLOW stage-discharge curve.

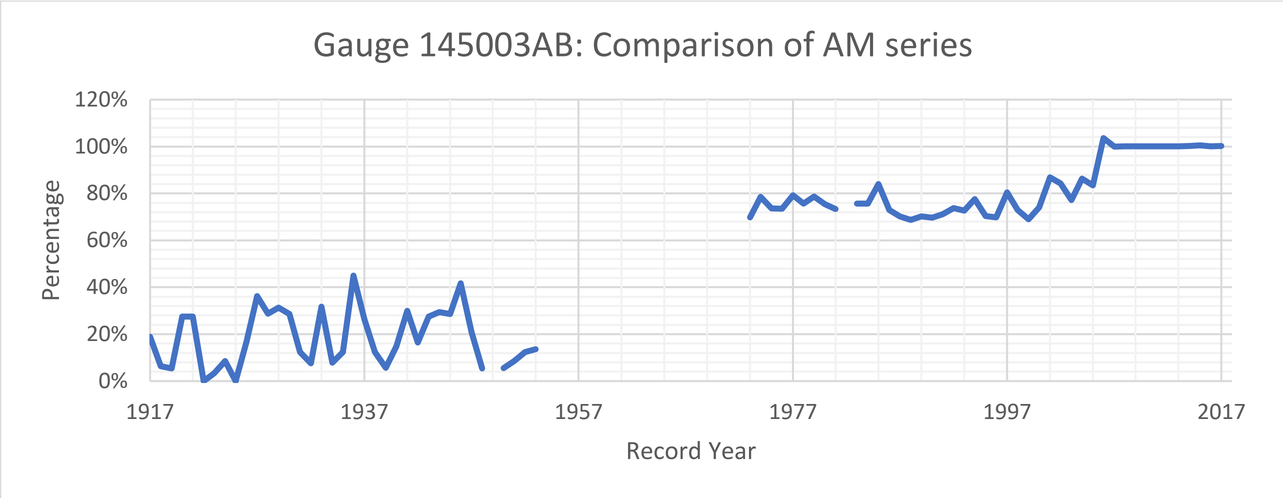


Figure 3.8. Ratio of discharge obtained from Method B over Method A for gauge 145003 revision A and B. AM series in this graph was not used in FFA.

3.3.5 Missing Data

Part of the method prescribed in ARR 2019 Book2, Section 2.3.11 was adopted for dealing with missing or poor-quality data. When an annual maximum is likely within a missing period of data, the whole year was removed from the time series. Regression analysis with nearby catchments to infill missing data is complex and time consuming and has not been used.

A couple of principles were used to judge if a period of missing data likely had an annual peak flow event. Firstly, if the missing period is in the dry season it is unlikely to have the annual peak flood. Rainfall data was consulted, when available, to identify if there was a significant rain event that may cause a large flood within the missing data period. As the flood wave travels past the gauge, flow/stage increases to a peak and slowly reduces over time. The length of time will depend on the catchment size and slope and can be identified by reviewing other peaks in the record. If there is evidence of a peak within a missing period, by increasing flow preceding or decreasing flow following a missing period, the maximum flow/stage is uncertain, and the year is removed from the analysis. All judgements are made relatively, to the current water year and monthly averages and maximums. The quality of data was also checked. Table 3.2 is a short list of quality codes used by WMIP.

Table 3.2. Water Monitoring Information Poral data quality codes.

Code	Suspect	Comment	Code	Suspect	Comment
1		Good (actual)	**69**	Suspect	CITEC - Derived Discharge
9		CITEC - Normal Reading	**79**	Suspect	CITEC - Backwater Record
10		Good (actual)	**130**	Suspect	Not coded value
15	Suspect	Water level below threshold (no flow)	**150**	Suspect	Unknown
19	Suspect	CITEC - No Flow Reading	**151**	Suspect	Data not yet available
20		Fair	**160**	Suspect	Suspect
26		BOM data - Good Daily Read Records	**180**	Suspect	Old Gauge height < instrument threshold
30	Suspect	Poor	**200**	Suspect	Water level below threshold
40	Suspect	Unverified data	**250**	Suspect	Missing Data
59	Suspect	CITEC - Derived height	**255**	Suspect	No data exists
60	Suspect	Estimate			

3.4 Revising Rating Curves

Gauges that were identified to have their rating curve revised were modelled using TUFLOW. Details about TUFLOW models are presented in Appendix C for some gauges. A basic triangular hydrograph was pushed through 2D hydraulic model for each gauge site. This simulated the water level and discharge at the gauge location to determine a rating curve. Column three of Table 3.3 lists the gauges were either the rating curve calculated by 2D hydraulic modelling was adopted, the rating curve was close to that provided by DNRME and so DNRME's was adopted, or the gauge was removed from the project for various reasons. The models provided by DTMR are identified in column four.

Table 3.3. Gauges that were modelled using TUFLOW.

Gauge	Name	Rating curve adopted	Model completed by
109001A	Mossman River at Mossman	Station removed	
111105A	Babinda Creek at The Boulders	Tuflow	
112002A	Fisher Creek at Nerada	Tuflow	
116008B	Gowrie Creek at Abergowrie	DNRME	DTMR
126001A	Sandy Creek at Homebush	Station removed	
129001A	Waterpark Creek at Byfield	Tuflow	DTMR
130207A	Sandy Creek at Clermont	DNRME	
130407A	Nebo Creek at Nebo	DNRME	
135004A	Gin Gin Creek at Brushy Creek	Tuflow	DTMR
136111A	Splinter Creek at Dakiel	Tuflow	
136202D	Barambah Creek at Litzows	DNRME	DTMR
136301B	Stuart River at Weens Bridge	DNRME	DTMR
137001AB	Elliott River at Elliott	DNRME	DTMR
138002ABC	Wide Bay Creek at Brooyar	Tuflow	

138111A	Mary River at Moy Pocket	DNRME	DTMR
142001A	Caboolture River at Upper Caboolture	Station removed	
143107A	Bremer River at Walloon	DNRME	
143209AB	Laidley Creek at Mulgowie	Station removed	DTMR
145003B	Logan River at Forest Home	Tuflow	
145011A	Teviot Brook at Croftby	Tuflow	
145102B	Albert River at Bromfleet	Tuflow	DTMR
422319B	Dalrymple Creek at Allora	Station removed	DTMR
917104A	Etheridge River at Roseglen	Tuflow	
922101B	Coen River at Racecourse	Tuflow	

In order to learn TUFLOW, some specific training was provided by DTMR and the online module 1 tutorial was completed, in addition to the TUFLOW manual being consulted throughout. Nebo Creek at Nebo (130407A) gauge was used for training. LiDAR was sourced primarily from DTMR, also the ELVIS (Elevation Information System) website which came generally as 1m or 5m grids. DTMR also provided GIS data and other required software.

The gauge coordinates provided through the WMIP system are not precise, although it is usually easy to identify the gauge hut on aerial imagery. It is important to locate the rated cross section so that stage measurements match. Cross sections through the LiDAR surveyed surface were cut until a good match to that provided by DNRME was found. Sand bed streams made it difficult to find the cross section and some didn't match well, as expected. The LiDAR data was also verified by comparing heights to permanent bench marks. When completing Coen River at Racecourse (922101B), the zero-gauge height stated in WMIP was identified to be approximately six metres low, and DNRME North Region was informed. DTMR identified gauge 116008B was out by two metres. DNRME confirmed these errors and made a point to correct the system.

3.4.1 TUFLOW

TUFLOW has three main products for hydrodynamic modelling; TUFLOW Classic, TUFLOW HPC, and TUFLOW FV. TUFLOW Flike is a flood frequency analysis tool and will be discussed in section 3.5.1. TUFLOW engines essentially solve the free surface shallow water flow equations (BMT group Ltd 2018 & TUFLOW 2017). Classic and HPC engines utilise implicit finite difference and explicit finite volume methods, respectively, to simulate a fixed grid in 2D. This makes it ideal for modelling flooding of rivers and creeks. Alternatively, TUFLOW FV solve a flexible irregular triangular/quadrilateral partitioned mesh in 2D or 3D. HPC refers to Heavily Parallelised Computing where calculations are completed simultaneously across multiple processing cores. HPC has capability to utilise GPU hardware (graphics card) opposed to standard CPU, which further reduces run time. Detailed theory of 2D hydraulic modelling will not be discussed as part of this project.

TUFLOW HPC build 2018-03-AC with GPU module was employed for this project. In this report TUFLOW will refer to HPC version unless otherwise stated. Control files or scripts and GIS files are used as inputs to TUFLOW engine, furthermore third-party software is used for an interface. MapInfo version 15 with Vertical Mapper add-in served as GIS package and Notepad++ for script files. DTMR provided required software and high-performance computer to run hydraulic models. Simulations tended to run within an hour.

3.4.2 TUFLOW Modelling Process

TUFLOW was run using a batch file. The batch file locates and executes the TUFLOW licence, and prompts the master script, *.tcf. In all control files, relative path references are used which stress the importance of a common folder structure and naming convention. The typical folder structure and most of the files can be copied and then modified to suit the new gauge site. The folder structure suggested by TULFOW wiki was adopted. Naming convention generally had the type of GIS file followed by short name of the gauge and version number. For example, 2d_loc_Logan_03.mif is the name of the GIS layer that defines location for Logan gauge 145003 and is the third model iteration. Figure 3.9 is a suggested work-flow diagram for basic 2D modelling and was generally followed. The various control files will be discussed briefly below.

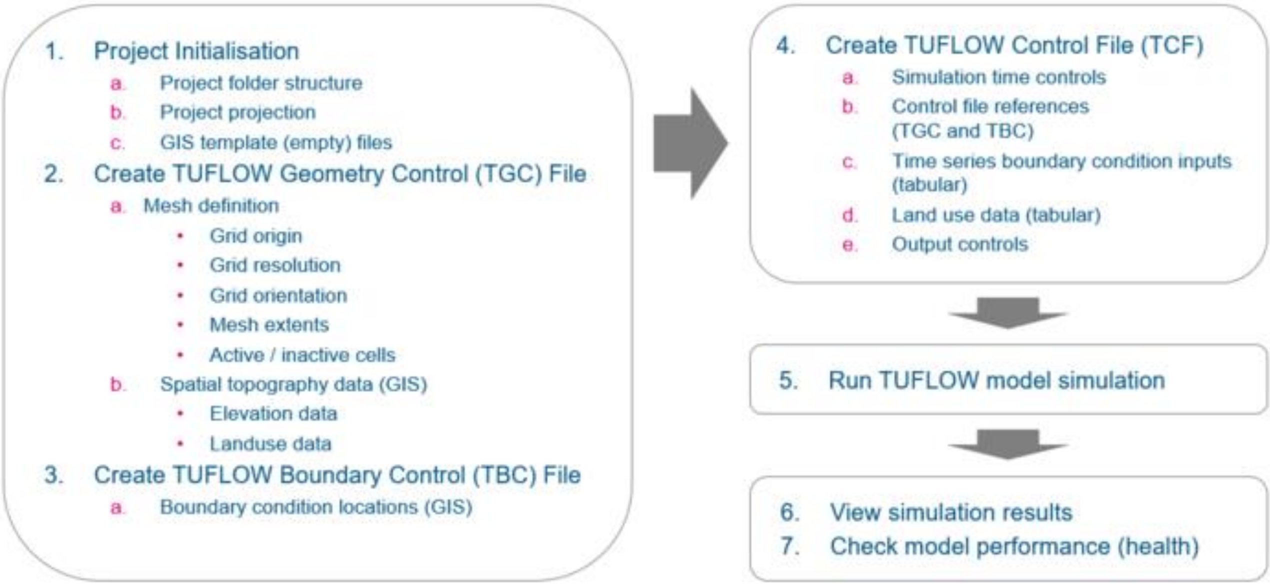


Figure 3.9. Example of TUFLOW modelling workflow diagram (TUFLOW 2018)

*TUFLOW Control File - *.tcf*

TUFLOW control file acts as the master script which calls different control files and sets parameters for TUFLOW to run. The TUFLOW build and solving engines/modules are all set. The coordinate system is also established. MGA94 coordinate system was used for all sites, specifying zone 54, 55, or 56 appropriately. A preliminary run of the file with these definitions creates a subdirectory with all default GIS files needed, that have predefined attributes and layers. Once complete, the modelling proper may begin.

Along with calling other control files, which will be discussed individually, parameters for the simulation are defined but are not limited to: time step, outputs, velocity cut-off depth, negative depth approach, and so on. A time step of half the grid size was used.

The GIS layer for outputs, 2d_po, is what defines where and what type of measurements are taken. Discharge and water surface level is measured as the volume of water crossing a polyline and water height in metres AHD at a point. For constructing the rating curve, discharge was measured across the cross section at the gauge, and heights on both banks and at the deepest point in the cross section. Discharge was also measured upstream and downstream of the gauge for checking. Any other locations that were of interest or where a breakout might occur were also captured. In addition to the specific locations identified by the GIS layer, TUFLOW writes many grids covering the entire simulation including: stream power, water height, discharge, velocity, maximum water height, and so on.

*Geometry Control File - *.tgc*

Geometry control file defines all the geometry and properties of the catchment such as grid definitions, material areas, terrain elevations. Firstly, the location is defined by 2d_loc file, where a line specifies the bottom left corner and orientation of the domain. The length and width of the domain are then defined in metres. Anything outside this domain will be dismissed. Simulation calculations within this domain are prescribed by active cells, designated by polygons in 2d_code file. It is important to locate good upstream and downstream limits. These desirably have a well-defined cross section within a uniform reach and far enough away to not influence hydraulics at the gauge site (i.e. the natural flow controls drive the hydraulics not the model's boundary conditions).

TUFLOW requires a Digital Elevation Model (DEM) in ascii grid format for terrain data. However, LiDAR is usually provided in tiles and often not in the correct format. Vertical Mapper plugin for MapInfo has the capability to merge tiles and export in the proper format. A cell size in metres and default height is set prior to DEM. Cell size was typically 10m grid, however, grid cell size was reduced if necessary to improve computations or pick up detail, especially for deep narrow streams. The default height is used to fill any holes in the LiDAR survey and is arbitrary, but must be above any expected flood level. The terrain data can be further manipulated by using other calls. In some cases, the LiDAR couldn't pick up the bottom of a creek due to standing water. To approximate the channel under water, by lowering grid points, 2d_zsh file is used. A polyline is drawn along the stream with attributes that describe a typical profile of the channel. The depth and width to modify the terrain was approximated based on the cross-section comparison and observing the reach. Any modifications must be called after DEM.

Catchment roughness is defined by drawing polygons around regions of similar roughness properties. Aerial imagery was examined to delineate the different roughness materials. Material regions were defined by vegetation density and land use. A command will overwrite anything called previously, therefore, multiple commands can be made if additional details are needed. This is of particular use for materials. A stream will usually consist of the stream bed and main passage of water and riparian area, adjacent vegetated region, including banks and overflow channel. In some cases where riparian and main stream were difficult to draw, the stream was saved as a separate file and called later to overwrite any cells assigned in the stream area. Materials are indexed and roughness values specified in Materials File called from the Control File. Each polygon is attributed the appropriate material index.

*Materials Control File - *.tmf*

The materials file defines the properties for each material. Infiltration losses were not used. As discussed in the literature review, roughness can be thought of as a resistance to flow, thus increasing water depth to conserve mass. There is the ability to define a varying roughness with depth. This is to account for the stream cleaning in a flood, grass laying flat once flow is strong enough, rough stream bed and/or overhanging trees to name a few. Manning's roughness coefficient, 'n', was used exclusively for this project and will be discussed in the next section.

*Boundary Conditions Control File - *.tbc*

Once the catchment is defined by geometry and materials, inputs into the model need to be defined. The boundary conditions are where water is inserted or outlet from the catchment, or interactions between 1D and 2D, or separate 2D and 2D models. They are specified as polylines with attributes for the type of boundary condition. Upstream used a basic triangle hydrograph to push through the catchment, discharge-time relationship. The maximum recorded flow gave a guide to the flow required. Primarily 6 hours was given until peak flow and another 6 hours to zero flow input, however, this was changed based on the site and catchment. Downstream boundary was defined as stage-discharge relationship which requires approximation of the stream slope and was obtained by cutting cross sections and measuring between them. TUFLOW calculates what is needed at these boundaries based on cross section and flow.

*Event Control File - *.tef*

The event file is more applicable when running multiple events for different AEPs. Completing rating curves only needed one event because stage-discharge is not expected to change based on the size of flood, except unsteady flow effects. The event file specifies where results and check files are saved, when to start and how long to run the simulation.

Simulations

The first model was mainly to check that setup was correct and all files referenced properly, ensure the extents of the model encompasses wetted areas and show flow paths, and identify additional details needed or problems raised. If the extents of the model are significantly bigger than the wetted area, it can be reduced to decrease computational time. Minimal effort was given to the first run and only one default material was assigned to the entire area. Any breakouts were usually identified after the first run, and the gauge removed from the project.

Further runs involved fine-tuning the model by adding appropriate details such as: cutting channels, assigning materials, or increasing input discharge. When a good representative model was developed, calibration and adjustments were made to the material roughness until satisfied (discussed in the next section). DTMR checked all models and helped fine tune the model, specifically reasonableness of roughness values. A simulation log was kept to keep track of model iterations and files, problems, and decision making.

Part of the modelling process includes reviewing outputs and performing checks. The output and check files can be read into MapInfo to review in comparison to your model. One check file of use was the grid used by TUFLOW where each cell listed all the final properties assigned it, such as elevation or material. Every cell can be investigated at each point on the corners, sides and centre. As TUFLOW runs calculations, it runs quality checks for stability and repeats and/or reduces the timestep if necessary. When this happens a lot, it suggests the calculations are 'numerically "on-the-edge"' (TUFLOW 2017). Control Number Factor variable is used to adjust the sensitivity and is called from the *.tcf file. Occasionally, this variable required adjustment for stability.

3.4.3 Calibration and Roughness

Calibration and validation are the main methods to eliminate errors in a good model and adjust parameters to produce accurate results. Often, input information may be uncertain, indefinite or unknown. Lohani (n.d.) explains, "Model calibration in general involves manipulation of a specific model to reproduce the response of the catchment under study within some range of accuracy." Interactive or user-driven methods, as well as automated systems, are used to complete calibration (Anderson 2002b; Lohani n.d.). Typically, trial and error technique is dominant as the manual optimisation method. A thorough understanding of the catchment and the representational models used is required to undertake manual process. Computers use various algorithms and numerical optimisation methods. Although significantly more work can be completed in the same time frame, equalling less money and resources, computers have their own problems. They are sensitive to data quality; restricted in the number of criterion/variables they can assess, thus potentially resulting in a local optimum instead of a global optimum; and algorithms require various assumptions. Moreover, manual process allows judgements to be made and usually retains the physical basis of the parameters.

Calibration in relation to rating curves involves adjusting parameters in order to obtain similar results as measured gaugings. Usually gaugings only make up the lower portion of the rating curve requiring extrapolation to large floods, hence the TUFLOW modelling. Roughness was the primary parameter for calibration. Starting from an expected roughness value, the roughness was refined until the rating curve passed through known gaugings. It is important to note that any parameter being calibrated should remain within a range resembling the real world; something that is smooth cannot have a high roughness in order to make the model fit specific data. Unusual results or large variations may indicate problems, as any variations should be realistic and explainable (Anderson 2002b). Along with aerial imagery and any available site photos, three sources were consulted to help select material roughness; Brisbane City Council Natural Channel Design Guidelines Appendix C (2003), Phillips & Tadayon (2006), and Arcement & Schneider (1989). Manning's roughness is also discussed in the literature review Section 2.2.3.

In some cases, the gauge location was near cropping land. As crops vary seasonally it was difficult to select the most appropriate roughness for the crop area, especially dense sugar cane which has significant difference between full growth and harvested roughness. This may be a reason for two lines of gaugings, however, if may be difficult to know the true effect, or account for it. Gaugings can also vary due to a change in the measuring instrument or location of measurement in the stream reach. The age and quality of gaugings were considered in calibration. Another difficulty is that the 2D modelling will not simulate very small floods precisely due to the grid size and time step adopted. However, these floods are not critical to this project. It was also noted that rising water, the front of the wave, can be very discordant at small flows, not smooth, due to it being faster than receding part of the wave for a fixed time step. Also looked at the point water overflows main channel, because if the channel estimation and extrapolation is reasonable it should be close to DNRME rating curve. Gaugings above overflow is highly desirable and should be matched.

3.4.4 Stage-Discharge Curve

After selecting the most reasonable TUFLOW model, MATLAB was used to find the best fit curve to be the revised rating curve. MATLAB code can be found in Appendix B.3. Curve fitting app was used to generate some preliminary code. Over the region of gaugings, where TUFLOW model was calibrated to, the DNRME rating curve was adopted, up to when TUFLOW started to depart. A smoothed spline was fitted to either the upper, lower, or between the two limbs (rising or falling path from unsteady flow) of the TUFLOW results, depending on the most realistic. A spline was used to more closely match TUFLOW results. Alternatively, multiple power or polynomial functions could be used with short transition between each, however, this requires much more work and little hydraulic justification was found in literature, especially reasoning that TUFLOW should reflect the actual function of the stream. The final rating curves are also presented in Appendix C. Any sites that used DNRME's curve entirely was linearly interpolated to obtain stages, which was considered sufficient as it is defined at 100mm intervals.

3.5 Flood Frequency Analysis

3.5.1 Flike

TUFLOW Flike (version 5.0.251.0) program, as recommended by ARR 2019 Book 3 Chapter 2 (Kuczera & Franks 2019), was used for flood frequency analysis. ARR 2019 Book 3 Chapter 2 provides informative examples on how to use Flike in accordance with ARRs guidelines. The following fitting methods were completed;

- GEV by LH moments for H = 0 and optimized H (note, H = 0 is standard L-moments);
- LP3 by Bayesian method with no prior information;
- LP3 by Bayesian method with no prior information but censored (if outliers found); and
- LP3 by Bayesian method with Gaussian prior distributions and censoring (if outliers found), using RFFE data for that site.

Annual maximum series was saved as *.csv file format to input into Flike. The list of floods is the minimum required for Flike but has the ability to include year labels for the data points. RFFE tool provides prior regional information from 'Nearby' gauges used in the RFFE methodology; downloadable at the bottom of results page. Refer Section 3.8 for obtaining RFFE results. No additional historical flood data was sourced or considered.

The report Flike produces presents the input data, any censored floods, LH- or statistical moments, a number of standard quantiles of the fitted model and some miscellaneous details. Code was developed to quickly read these output files and efficiently assess the best frequency curve (explained in next section). GEV and LP3 models have different report formats, and there were slight differences between gauges even for the same method. The changes were primarily in relation to the standard quantiles reported and MATLAB code was adjusted for that gauge. The graphs automatically produced by Flike differ depending on the probability model. The plot scale is Gumbel and Log-normal for GEV and LP3, respectively. These graphs were not used for goodness-of-fit assessment, which is explained in the next section 3.5.2.

3.5.2 High Outliers

Four gauges where suspected of having high outliers; these are not identified by MGBT. As ARR 2019 does not describe procedures in assessing high outliers, ARR 1987 Chapter 10 Section 10.7.3 (a) was consulted. Gauges 113004A, 143107A, and 116014A showed no statistical evidence for outliers and so no action was taken. A gauge, 116013A, 5.6km away which had similar characteristics to 116014A was also compared. A similar large event occurred at the same times in its record and so supported that it was not a data error. Gauge 915011A, however, did show strong evidence of a high outlier. The outlier occurred February 2002 and was the second biggest flood on record. The quality code was 59 – 'CTEC – Derived Height', so quality was also suspect. Figures 3.10 and 3.11 are frequency curves for with and without the suspected outlier. The flood was omitted.

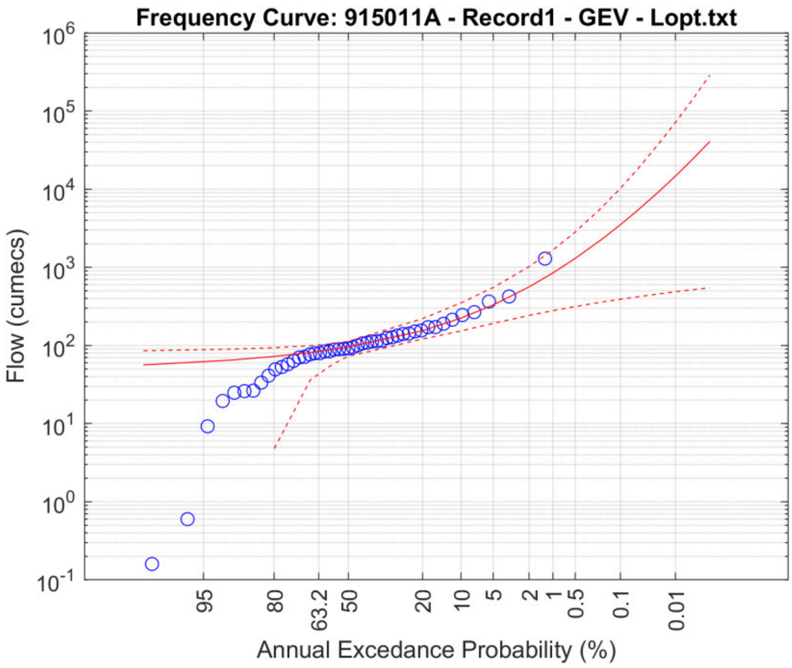


Figure 3.10. Gauge 915011A frequency curve with suspected high outlier.

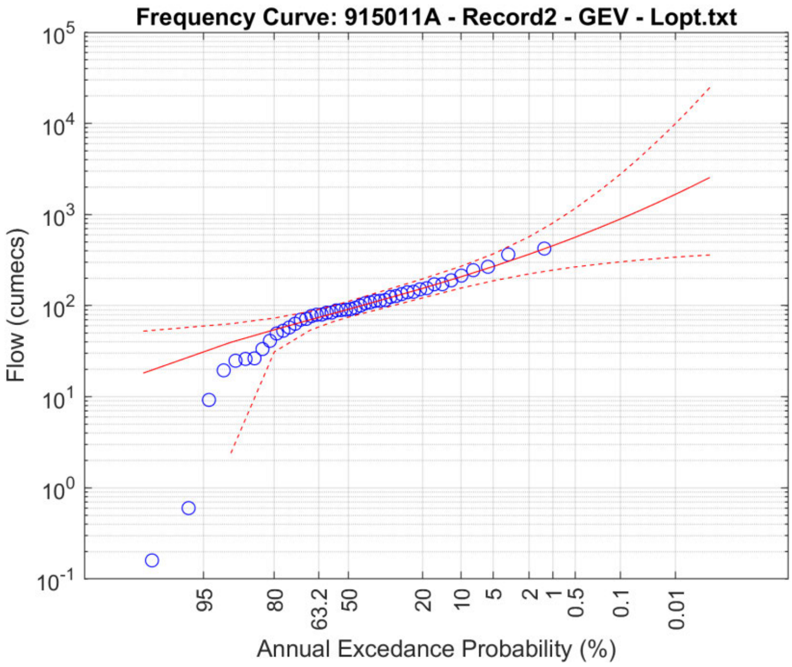


Figure 3.11. Gauge 915011A frequency curve without suspected high outlier.

3.5.3 Distribution Selection

The selection of the best distribution followed several principles. Primarily, all the observations should be within confidence limits. Generally, the curve should not project upwards beyond the data especially on log scale. The KS test has been measured between empirical CDF based on plotting position and proposed frequency distribution curve. The smaller the distance, the better the fit; each model was ranked accordingly with one being the best and plotted as column graphs to compare the selected distributions. KS test was particularly used when frequency curves were close. MATLAB was utilised to quickly compare Flike results. ReadFlike.m function (Appendix B.5) reads the Flike file and creates Log-normal plots, QQ plots and performs the KS test.

ReadFlike.m has been written as a script where the user changes the inputs manually. The script requires station numbers to locate the Flike files (in standard folder structure) where on first run will list all available. The user then selects the files wanted for comparison by index numbers are then selected in 'Selection' variable. The outputs, such as writing results to file or plotting graphs, are controlled by turning on or off variables with 1 or 0, respectively. The script checks the destination folder for any files to reduce the risk of overwriting any work. Variables were summarised into structure variable type for later use and also saved to Excel format when MATLAB was not available.

The script proper essentially loops through all the Flike files selected and reads them, produces graphs, and calculates KS test. Flike files are formatted differently for LP3 and GEV distributions; local functions (not as a separate file) were written to read each: ReadLP3 and ReadGEV. Frequency curve plots were plotted at log-normal scale and labelled with standard AEPs. This was easily achieved in MATLAB by `semilogy` plot function and `norminv` for the x axis.

To obtain quantiles estimates at each observation for the QQ plot and PP plot, FitFLIKE.m (Appendix B.6) was written. It was found that when using the reported parameters for the fitted distribution to estimate quantiles, they did not match the reported quantiles. Therefore, a curve was fitted to the reported quantiles based on the distribution function. Custom Equation option in Curve Fitting app was used to generate initial code. The equation used by curve fitting was rearranged GEV and LP3 equations. However, LP3 could not be rearranged and usable for MATLAB coding for converting quantiles to probability, so a smoothing spline fit to log of the quantiles was used and had better than R^2 of 0.99. A minimum of 0.01 discharge was applied to avoid issues associated with logarithms of close to zero. GEV and LP3 parameters, probabilities and quantiles were able to be calculated accurately, however, the reported quantiles were always used for all other calculations. Residual plot of the fitted curves could be consulted if any issues were suspected.

3.5.4 Quality Checks

In order to check there were not any gross errors or mistakes, the design quantiles were compared against RFFE input data. Input data into RFFE tool is available for the nearby catchments used (refer Section 3.8). The quantiles of Method A or B was used for comparison, as Method C used revised rating curve which could automatically be significantly different.

All but three gauges are part of RFFE database; 110002A, 130407A, and 917114A. These were assumed to be correct. Gauges where the 2% AEP flood discharge was greater than 25% difference were reviewed. This method did identify a couple of mistakes, which were recalculated. For most other gauges the difference could be attributed to the difference in record length: RFFE FFA was completed by 2011/2012 (Rahman et al. 2015). However, the discrepancy for some gauges could not be reconciled. More discussion can be found in Section 4.6.1. Table 3.4 summarises the gauges with greater than 25% discrepancy even after review.

Table 3.4. List of gauges greater than 25% discrepancy.

	Record Length	RFFE record length	RFFE Tool		Method A or B		Ratio (2%)
			2%	1%	2%	1%	
136006A	53	46	795.9	1059.3	1358.3	1992.6	1.71
136108A	56	49	430.5	565.5	899.9	1341.0	2.09
136202D	54	91	1221.7	1518.4	1694.	2228.9	1.39
137101A	53	45	1891	3089.8	814.8	939.8	0.43
138110A	59	52	4304.2	5488.6	3172.2	3693.2	0.74
142202A	51	46	1505.7	1896.8	1087.4	1295.2	0.72
143107A	57	50	1090.7	1202.6	1473	1813.7	1.35
143108A	57	50	911.2	1073.7	1475.1	1939.7	1.62
145003B	65	90	1314.1	1675.3	874.7	1200.2	0.67
422211A	51	32	185.51	197.87	334.8	389.8	1.80
915011A	47	40	627.4	821.3	366.3	455.9	0.58

3.6 Regression

The model regression was completed using MATLAB. The calculations and theory are expressed in Section 2.5. Continuing on from many of the previous processes, the MATLAB script ‘RegressionData.m’ assembled all the data required into a complete structure and saved. This variable was then loaded into other scripts to run various regression options. Appendix B.8-11 has MATLAB code used.

All design probability events were treated independently, namely 50%, 20%, 10%, 5%, 2%, and 1% AEPs. The first option (RegressionBaseCase.m) used rainfall intensity 72hour duration, 2% AEP frequency for all probabilities, the same as Palmen & Weeks (2011). This regression was tested for outliers by removing potential outlier gauges to test their influence. Second option (RegressionOption2.m) sought to improve the model by selecting the best duration at each frequency. This was completed by looping through all durations and reviewing test statistics and performance. The third option (RegressionOption3.m) also looked to improve the set of equations by finding a single optimum duration-frequency for all equations and followed a similar method of looping through possible selections. Assessment of the best model is discussed in the Section 4.4.

3.7 Palmen & Weeks (2011) Method

As discussed in the Literature Review section 2.1.3, Palmen & Weeks (2011) developed a method for estimating catchment discharge with minimal inputs. Calculations are limited to 2, 5, 10, 20, 50, and 100-year ARIs. The only two inputs are catchment area and rainfall intensity. The intensity required is 72-hour, 50-year ARI from 1987 IFD tables; as the method was developed prior to ARRs update. Obtaining the gauge 1987 IFD table is outlined in section 3.2.2.

The process has been exemplified with gauge 102101A. The station's 1987 IFD table is provided as Table 3.5, where the 72-hour, 50-year ARI rainfall intensity is underlined. The station has a catchment area of 651 km2.

Table 3.5 1987 IFD table for gauge 102101A.

Duration	1 year	2 years	5 years	10 years	20 years	50 years	100 years
5 mins	164	205	242	263	295	338	371
6 mins	153	192	226	246	276	316	348
10 mins	129	162	190	207	233	267	293
20 mins	102	128	150	163	183	210	230
30 mins	86.1	108	127	138	154	177	194
1 hour	58.7	73.7	86.9	94.7	106	122	134
2 hours	35.2	44.5	53.2	58.5	66.1	76.5	84.6
3 hours	25	31.7	38.4	42.5	48.4	56.4	62.6
6 hours	13.6	17.4	21.6	24.2	27.8	32.8	36.8
12 hours	7.81	10	12.6	14.2	16.5	19.6	22.1
24 hours	4.99	6.41	8.02	9.02	10.4	12.4	13.9
48 hours	3.35	4.28	5.27	5.87	6.74	7.91	8.85
72 hours	2.47	3.15	3.85	4.29	4.91	**5.75**	6.42

With the intensity selected (5.75mm/h), and catchment area, discharge is calculated using the equations presented in section 2.1.3. The results for station 102101A are summarised in Table 3.6, along with equivalent AEP.

Table 3.6 Discharge estimation by Palmen & Weeks (2011) method for gauge 102101A.

ARI (years)	AEP (%)	Flow (m^3/s)
2	39.4	265
5	18.1	639
10	9.5	918
20	4.9	1305
50	2	1803
100	1	2197

3.8 Online Regional Flood Frequency Estimation Tool

RFFE requires catchment outlet and approximate centroid location, and catchment area. These parameters are used to obtain rainfall data by location from BoM, and develop a shape factor. Figure 3.12 shows the inputs for station 102101A. The interactive map allows the user to move outlet and catchment centroid points to desired location or input longitudes and latitudes on the left-hand panel.

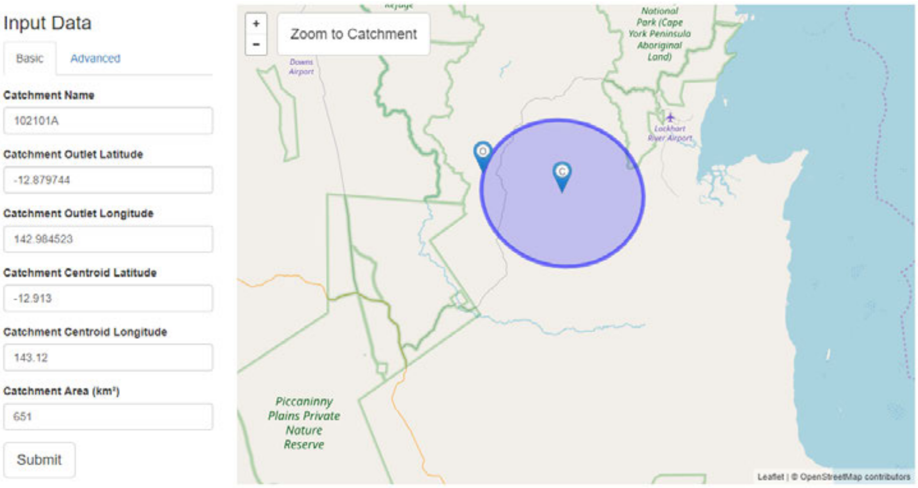


Figure 3.12 Example input to RFFE for gauge 102101A.

The blue ellipse shows an approximated catchment shape based on the control points; outlet and centroid locations. If the ellipse is too oblong it will turn red, as the accuracy of this method becomes questionable.

To locate the approximate centroid without rigorous calculations and defining the exact catchment boundaries, QldGlobe was consulted. In QldGlobe the station was located to review the catchment and estimate a centroid location by eye. Various QldGlobe layers were used to assist identifying the catchment: drainage basins, watercourses, surface water monitoring, elevation contours. In some cases, the ellipse was not able to resemble the catchment shape. In these cases, the centroid was moved until the ellipse turned blue in order to get a result.

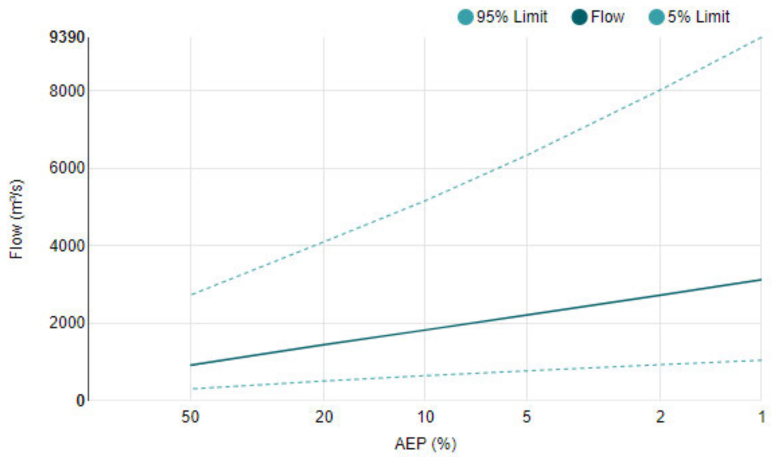


Figure 3.13 Discharge estimation graph by RFFE for gauge 102101A.

Table 3.7. Discharge estimation by RFFE for gauge 102101A.

AEP (%)	Estimated Flow (m^3/s)	Lower Confidence Limit (5%) (m^3/s)	Upper Confidence Limit (95%) (m^3/s)
50	920	309	2730
20	1450	514	4110
10	1830	652	5170
5	2220	776	6340
2	2730	936	8030
1	3130	1050	9390

Figure 3.13 and Table 3.7 shows the RFFE estimate for gauge station 102101A, with 95% and 5% confidence limits. Additional details and statistics are also provided by RFFE online tool (not shown). The results are able to be downloaded as a text file. The results file is also used by Flike for prior information input, see section 3.5.1.

Three additional graphs presented as part of RFFE results show how an input catchment relates to the gauges in the RFFE database used for calculations. If the subject catchment is a long distance from the other sites' characteristics, the results may be less accurate. As demonstrated in Figure 3.14 below, the gauge 102101A shape factor is a little low compared to the other stations used in the RFFE analysis, however, is still reasonable. The other nearby gauges used for calculations (in this case four others) are provided by the RFFE online tool as a download through 'nearby' button. This information has been used for quality checks for FFA.

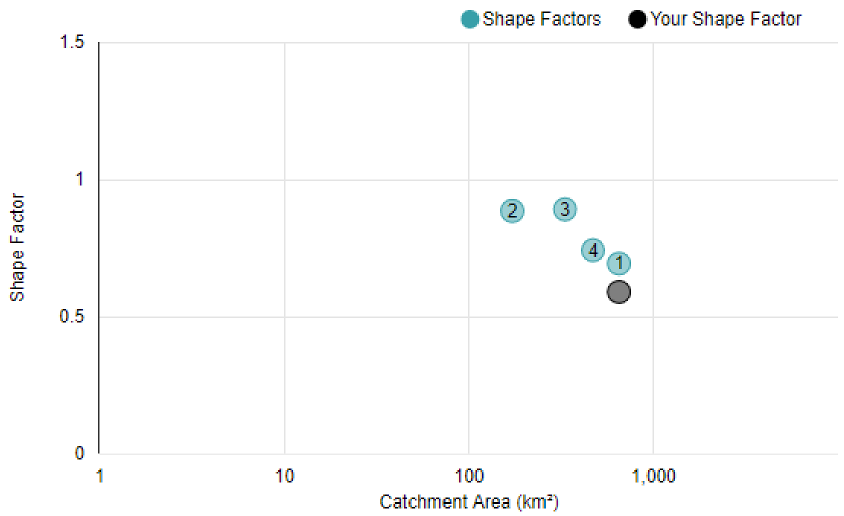


Figure 3.14 Shape Factor vs Catchment Area RFFE results for gauge 102101A. (Your Shape Factor = manual input for gauge 102101A, Other gauges from database are: 1 = 102101; 2 = 922101; 3 = 926002; 4 = 104001)

4 Results and Discussion

4.1 Overview

This chapter will briefly detail some results for the revised rating curves and flood frequency analysis. It will further report optimisation and the final model from regression. The performance of this model will then be examined.

4.2 Revision of Rating Curves

Due to the large amount of details involved in 2D hydraulic modelling, only some gauges have been added in Appendix C and D and rating curves results have only been summarised for all gauges. In the methodology, Table 3.3 presents what rating curve was adopted after completion of TUFLOW modelling of the 24 identified suspect gauges. The gauges reviewed ranged from Cape York to south east corner of Queensland. There were five gauges removed from the project due to bypass flow, interaction with other streams (meaning additional runoff not attributed to the catchment of the gauge), or other issues. In eight cases the 2D hydrodynamic modelling produced very similar results to DNRMEs rating curves, and as a result Method A or B was adopted.

The remaining eleven deviated significantly. The percentage difference from DNRMEs extrapolated curves is plotted in Figure 4.1 for each equivalent design probability. Positive and negative means that DNRME under-estimated and over-estimated, respectively. Additional graphs and data are available from Appendix D. Note that there is still some uncertainty in 2D modelling of rating curve extrapolation due to few gaugings available for calibration and subjective nature of selecting roughness values. However, error regarding extrapolation is greatly reduced by this method of modelling.

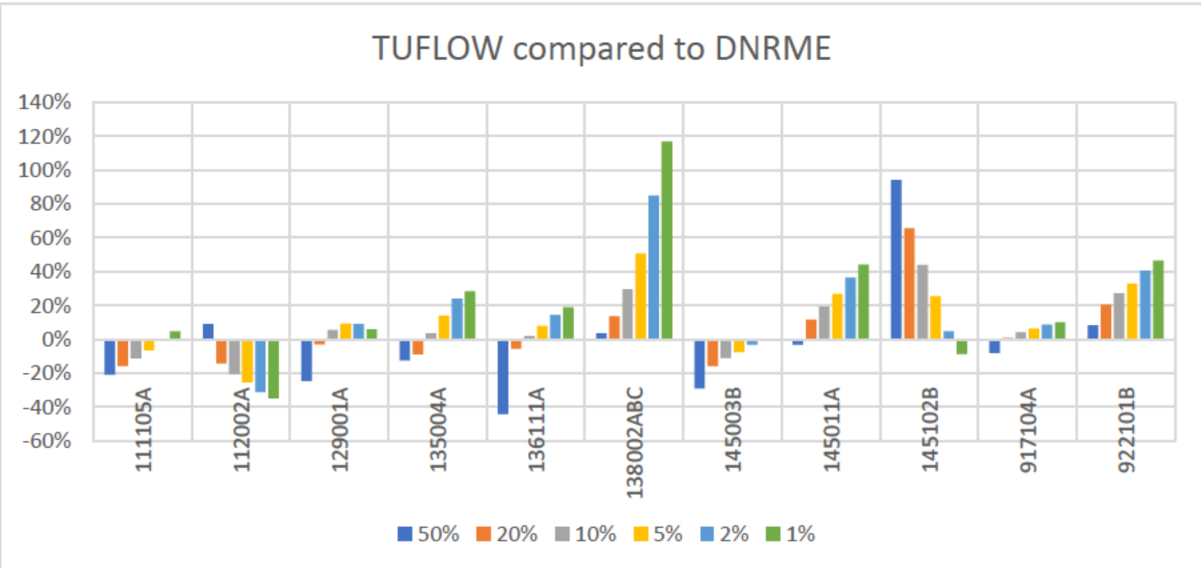


Figure 4.1. Discharge difference between TUFLOW simulation and DNRME extrapolated rating curves.

The largest difference belongs to 138002 gauge with discharge 1.16 times bigger than DNRME's estimate for flood at 1% AEP. See rating curve Figure 4.2.

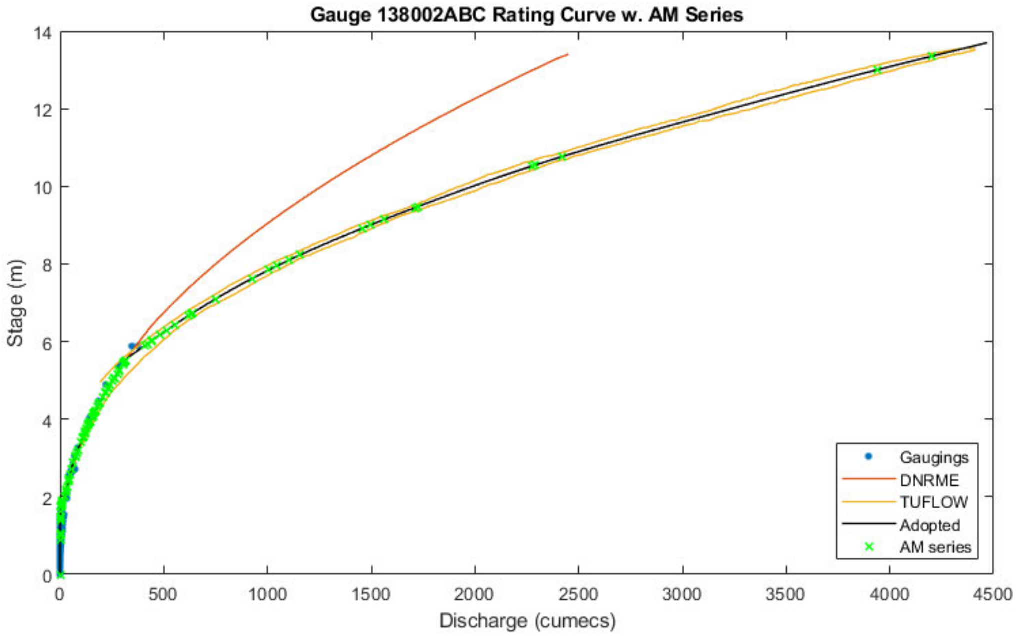


Figure 4.2. Revised rating curve for 138002ABC gauge.

There are a variety rating curves that actually differed at medium floods then matched DNRME more closely for large events. An example of this is gauge 136111A; rating curve shown below. The effect of unsteady flow conditions can also be clearly seen by the rising and falling curve of TUFLOW model (yellow).

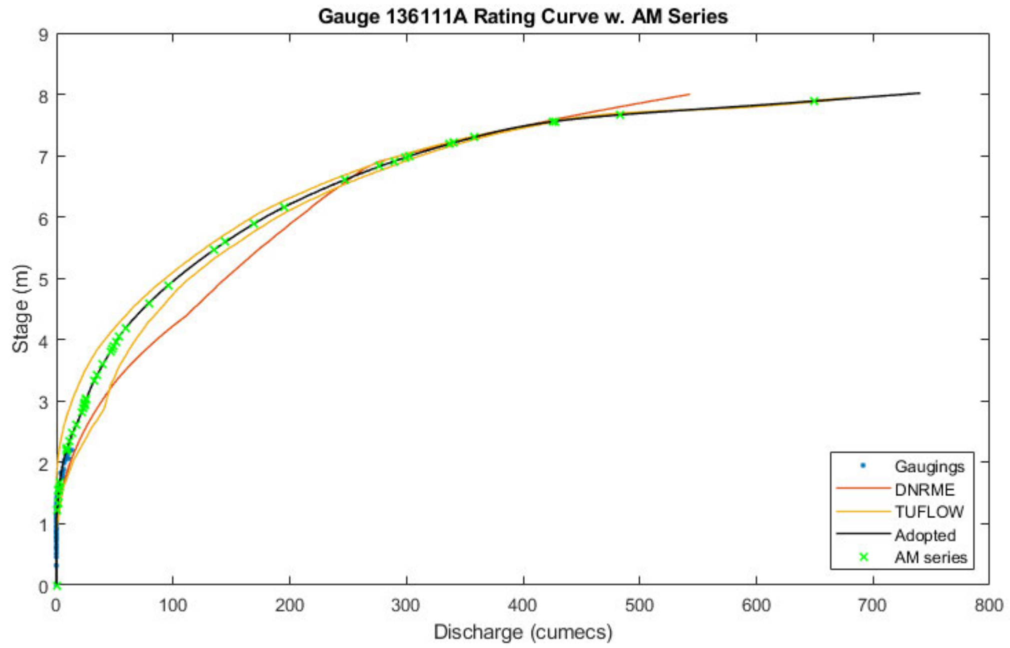


Figure 4.3. Revised rating curve for 136111A gauge.

4.3 Flood Frequency Analysis

After gaining appreciation of many frequency curves, assessment became easier, although the whole process is rather subjective. The most useful aid was log-normal graph. The KS test did give some answers that were not correct when looking at the frequency curve, but it was useful at times judging border line cases. QQ-plots did assist a little as an alternative presentation, but PP-plots were not useful.

Overall, GEV was selected most often, 32 times compared to 19 for LP3. This supports recommendation by ARR 2019. However, in some cases GEV was clearly problematic. Figure 4.4 shows that there was no particular region that suited LP3 or GEV.

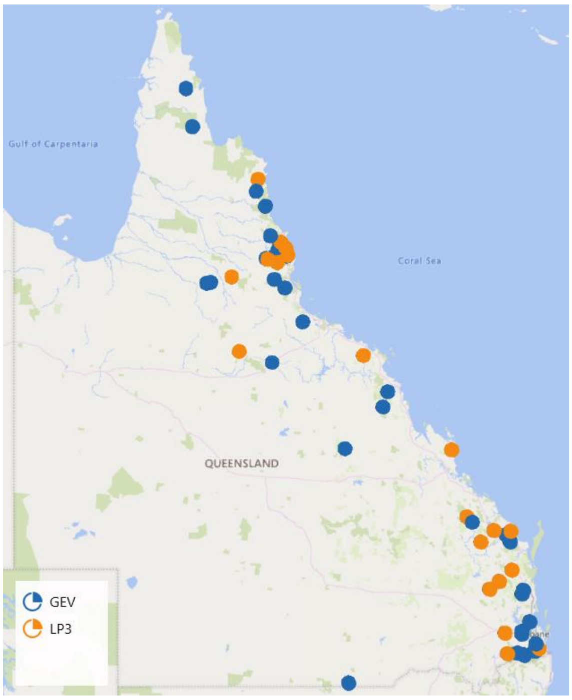


Figure 4.4. Distribution selected for gauges. Blue indicate GEV selected and Orange indicate LP3.

LH-moment fitting was generally superior to L-moments, see Figure 4.4. Of those where GEV was chosen, twenty were optimised to Shift 4, significantly more to any other degree of shift (Figure 4.5). On two occasions, gauge 143212A and 136006A, LH-moment fitting was optimised to a shift too great so that the flood for 50% AEP was significantly under-estimated but had good estimates at rarer events. There would be concern for this optimisation method if extracting any more frequency event discharges. If these are of interest, it may be better to use POT series, another distribution, or complete LH-moment fitting by manually. This problem is why KS test was modified (refer Section 2.4.4).

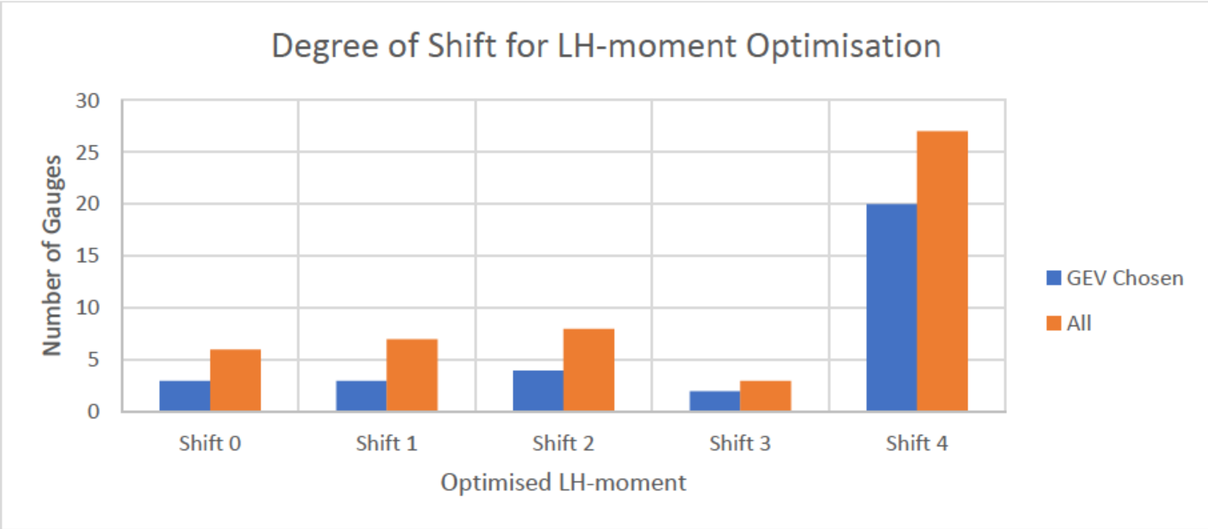


Figure 4.5. Optimisation of LH-moment fitting.

In almost all cases RFFE prior information did not produce a good fit. Twenty-five gauges had a positive regional skew for prior information, where all gauge data was negatively skewed. The regional skew was all less then plus or minus 0.15. However, the data skew varied from -0.02 to -1.63.

Multiple Grubbs Beck Test (MGBT) was only applicable to LP3 distribution to remove potentially influential low flows. The effect of this tended to straighten out the frequency curve, consequently reducing the amount of skew. In five out of the 19 where LP3 was selected, this was not desirable and produced a worse fit, not matching the rare events well.

In the graph below (Figure 4.6) the results of MGBT are plotted for all gauges despite the best fitted distribution. The number of years censored by MGBT is plotted against the total length of AM series. For additional information the final selected distribution has been coloured; blue indicates those where GEV was superior, and orange where LP3 was chosen. For many gauges a large number of floods were removed: 12 gauges censored more than 25% of the AM series with 5 of them removing approximately half – the limit prescribed by ARR 2019. Only ten gauges had results of zero floods to be censored. The significance of these findings and effectiveness of MGBT was not investigated further.

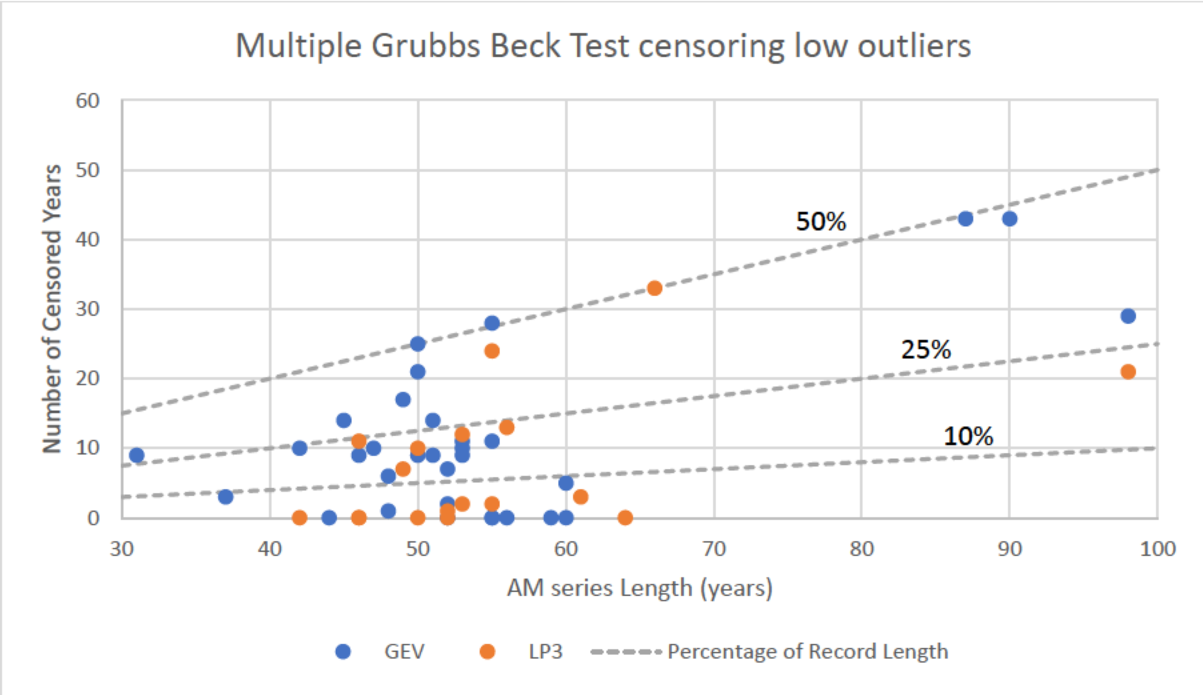


Figure 4.6. Assessment of Multiple Grubbs Beck Test censoring low outliers.

4.4 Regression and Optimisation

The final empirical model was obtained by performing ordinary least squares regression. Fifty-one gauges were utilised and each probability event, 50%, 20%, 10%, 5%, 2%, and 1% AEPs, treated independently. The final equations are in the form:

$$Q = \delta A^{\beta} I^{\gamma} \qquad \text{(2.43 repeated)}$$

where A is the catchment area, I is rainfall intensity with duration and frequency chosen, and Q is the resultant estimated peak discharge for a probability event, with regression process finding coefficients; δ, β, and γ.

Completing regression by systematically removing potential outliers found that removing them did not significantly improve the performance of the regression. Therefore, all catchments were considered typical with no outliers. If there was a consistent gauge outlier then this catchment may be atypical and excluded. Gauge 422211A, which is a long way inland compared to other gauges, was also tested for influence and found to have minimal influence on the regression. Residual plots showed that the data had no substantial trend or change in variance, so was deemed homoscedastic, rendering Weighted Least Squares or Generalised Least Squares regression not necessary.

The different options varied only by the rainfall intensity selection. Results of the regression of the three options are presented in the following sections. There were not enough gauges to test if categorising the catchments into small and medium sizes would give better results, nor dividing Queensland into north and south regions. Moreover, Palmen & Weeks (2011) found that breaking into regions gave no improvement to the results.

4.4.1 Option 1

The Option 1 model used 72 hour duration, 2% AEP rainfall intensity for all equations. This is the same frequency-duration as Palmen & Weeks equations. Regression coefficients for equation 2.43 are presented in Table 4.1.

Table 4.1. Coefficients for Option 1 equations

AEP	δ	β	γ	Rainfall Intensity Selection (Duration, Frequency)
50%	0.305	0.638	1.520	72 hours, 2% AEP
20%	1.913	0.600	1.168	72 hours, 2% AEP
10%	3.989	0.586	1.025	72 hours, 2% AEP
5%	6.930	0.575	0.917	72 hours, 2% AEP
2%	12.391	0.563	0.806	72 hours, 2% AEP
1%	17.959	0.556	0.736	72 hours, 2% AEP

4.4.2 Option 2

Option 2 involves finding the optimum duration for each frequency event. Figure 4.7 shows an example of 10% AEP regression model results. Minimising Mean Absolute Percentage Error (MAPE) and Root Mean Squared Percentage Error (RMSPE), and maximising R^2 and adjusted R^2, gives an optimum duration of approximately 36hours. All probabilities produced similar shaped graphs except 50% AEP, which did not have a turning point but flattened out at the longer durations.

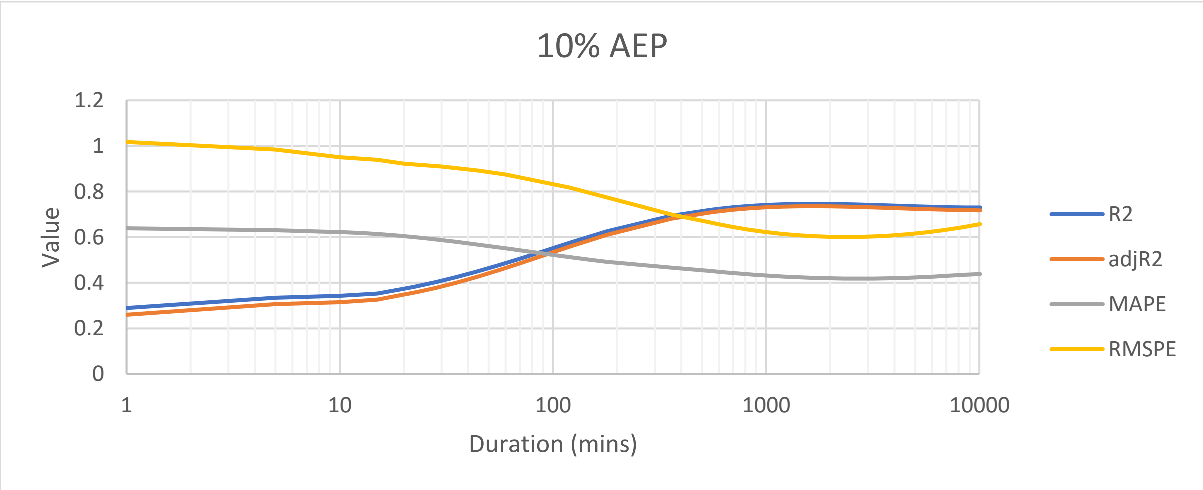


Figure 4.7. Regression statistics for Option 2, 10% AEP; statistical tests.

Results for the number of gauges within 10%, 20%, and 50% of the true quantiles, as well as an even number over- and under-estimated, had much less smooth shape across the durations, see Figure 4.8.

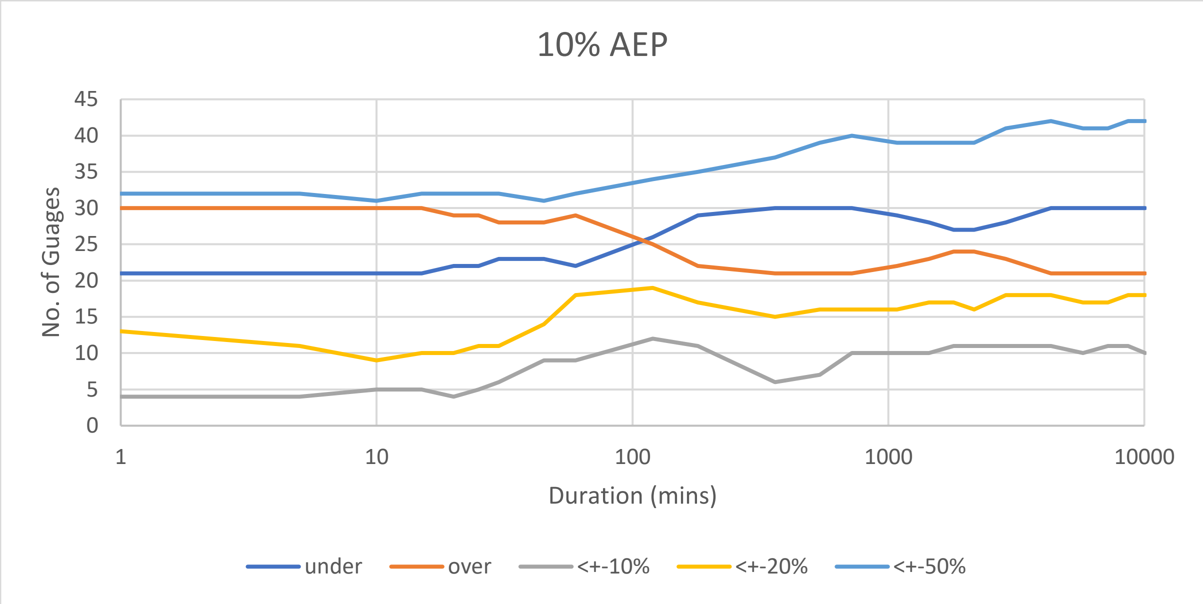


Figure 4.8. Regression statistics for Option 2, 10% AEP; gauges within percentage error.

Overall, the optimum duration was around 48 hours or 36 hours. It was considered desirable to have a consistent duration for all probabilities for simplicity of use, and 36 hours was selected. Coefficients of the regression and rainfall selection are presented in Table 4.2 below. As a basic correlation model that is not founded on time of concentration or similar phenomenon, does not necessarily optimise to different durations depending on the magnitude of the flood.

Table 4.2. Coefficients for Option 2 equations

AEP	δ	β	γ	Rainfall Intensity Selection (Duration, Frequency)
50%	0.381	0.676	1.771	36 hours, 50% AEP
20%	1.619	0.618	1.307	36 hours, 20% AEP
10%	2.784	0.599	1.145	36 hours, 10% AEP
5%	4.247	0.585	1.025	36 hours, 5% AEP
2%	6.954	0.567	0.898	36 hours, 2% AEP
1%	9.668	0.556	0.818	36 hours, 1% AEP

4.4.3 Option 3

Another option for trying to optimise the regression model was optimise the best single rainfall intensity for all probabilities. Instead of testing all possibilities available in standard IFD tables produced by BoM, 12-96 hours durations were selected for 10% and 2% AEP frequencies.

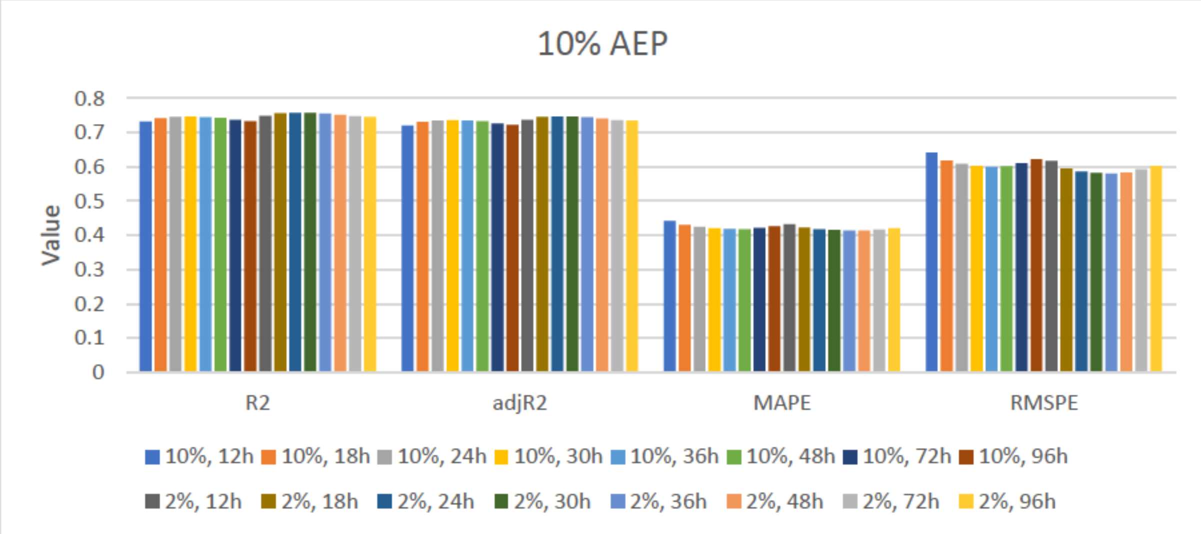


Figure 4.9(a). Regression statistics for Option 3, 10% AEP.

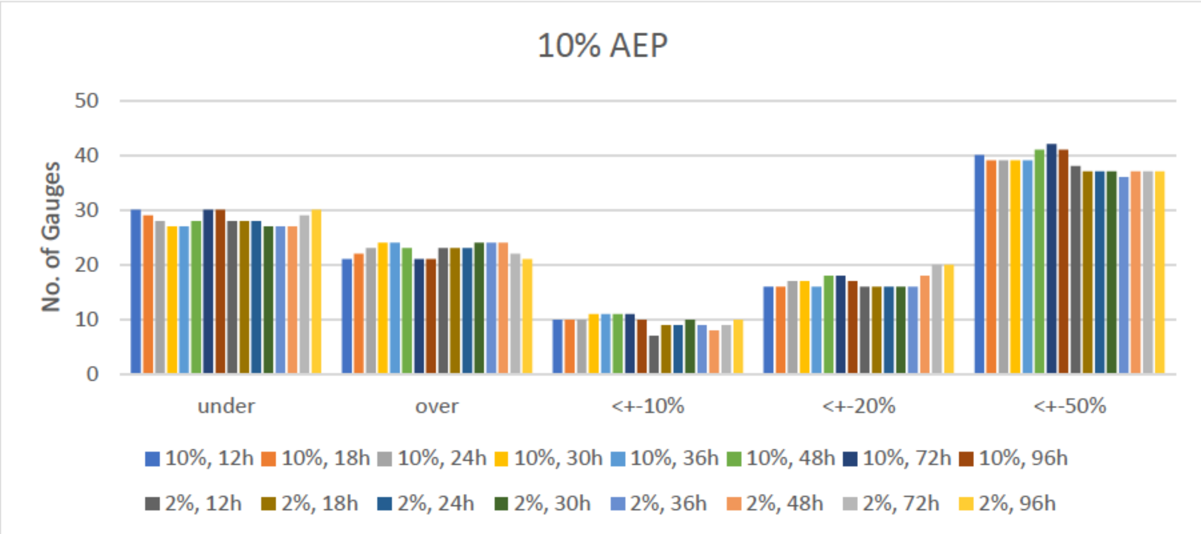


Figure 4.9(b). Regression statistics for Option 3, 10% AEP.

All probabilities had similar shape to 10% AEP results presented above. Consistent with Option 2, an optimum duration of 36 hours is clear. The 10% AEP frequency was slightly better. Thus 36 hours duration, 10% AEP rainfall intensity was selected, and coefficients from regression can be found in Table 4.3 below.

Table 4.3. Coefficients for Option 3 equations

AEP	δ	β	γ	Rainfall Intensity Selection (Duration, Frequency)
50%	0.172	0.640	1.597	36 hours, 10% AEP
20%	1.215	0.603	1.233	36 hours, 10% AEP
10%	2.662	0.588	1.084	36 hours, 10% AEP
5%	4.799	0.577	0.972	36 hours, 10% AEP
2%	8.912	0.566	0.857	36 hours, 10% AEP
1%	13.226	0.558	0.785	36 hours, 10% AEP

4.4.4 Final Model

The three models' performance against true quantiles from FFA were compared by several methods and tests. Box and whisker plots of the percentage error are presented below for each design probability (Figure 4.10). These were used to get an idea of the distribution of error. Quantile-Quantile plots against FFA, not included in the main body of this report, showed that the large over-estimations belonged to small floods: not frequent events but catchments, by combination of catchment characteristics and rainfall, that produce smaller than average floods. The most accurate probability for each model is 20% AEP event, and the largest variance is 50% AEP event. All estimates are within +350% and -75% of the true discharge.

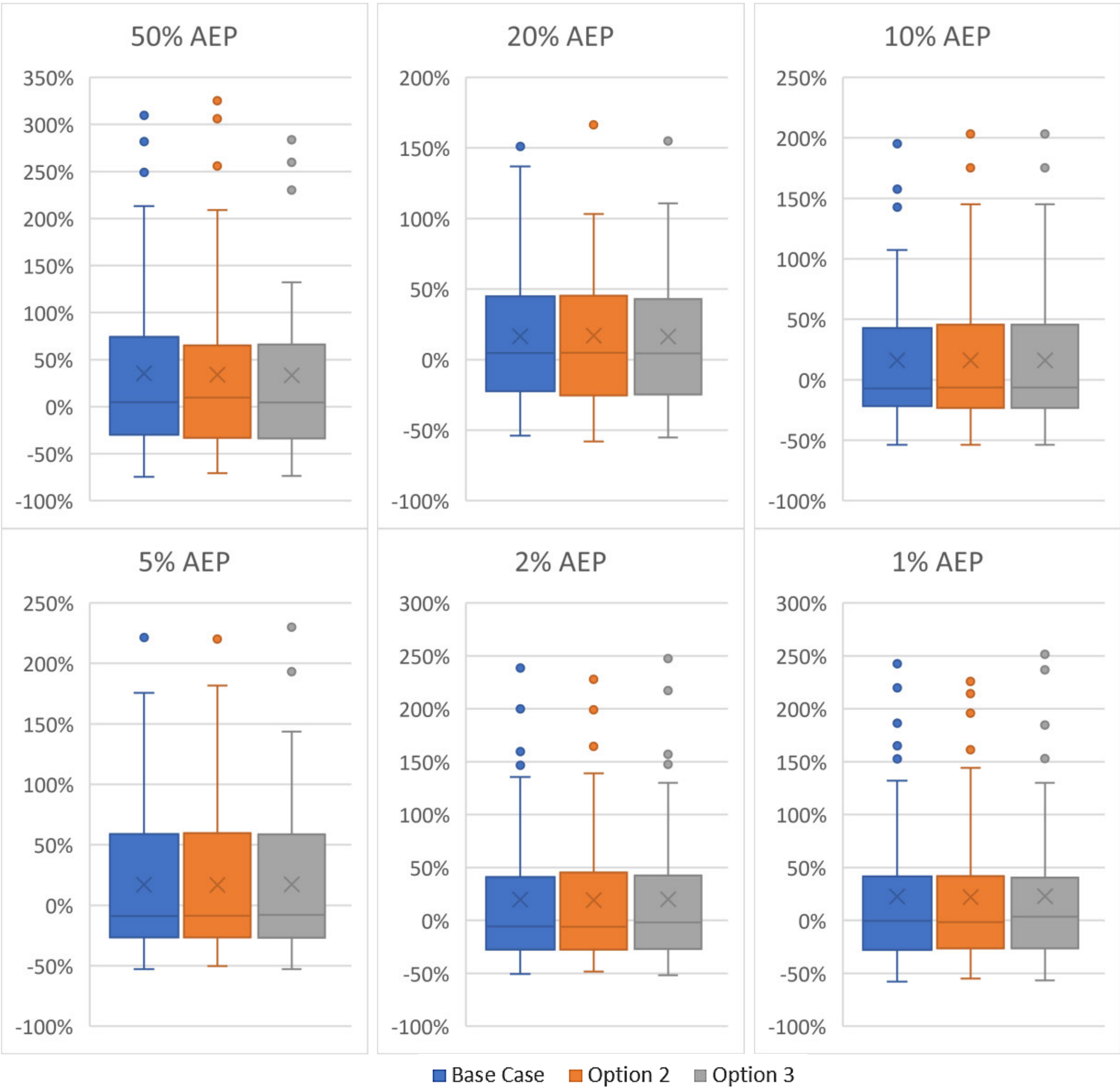


Figure 4.10. Box and whisker plots of percentage error from FFA quantiles.

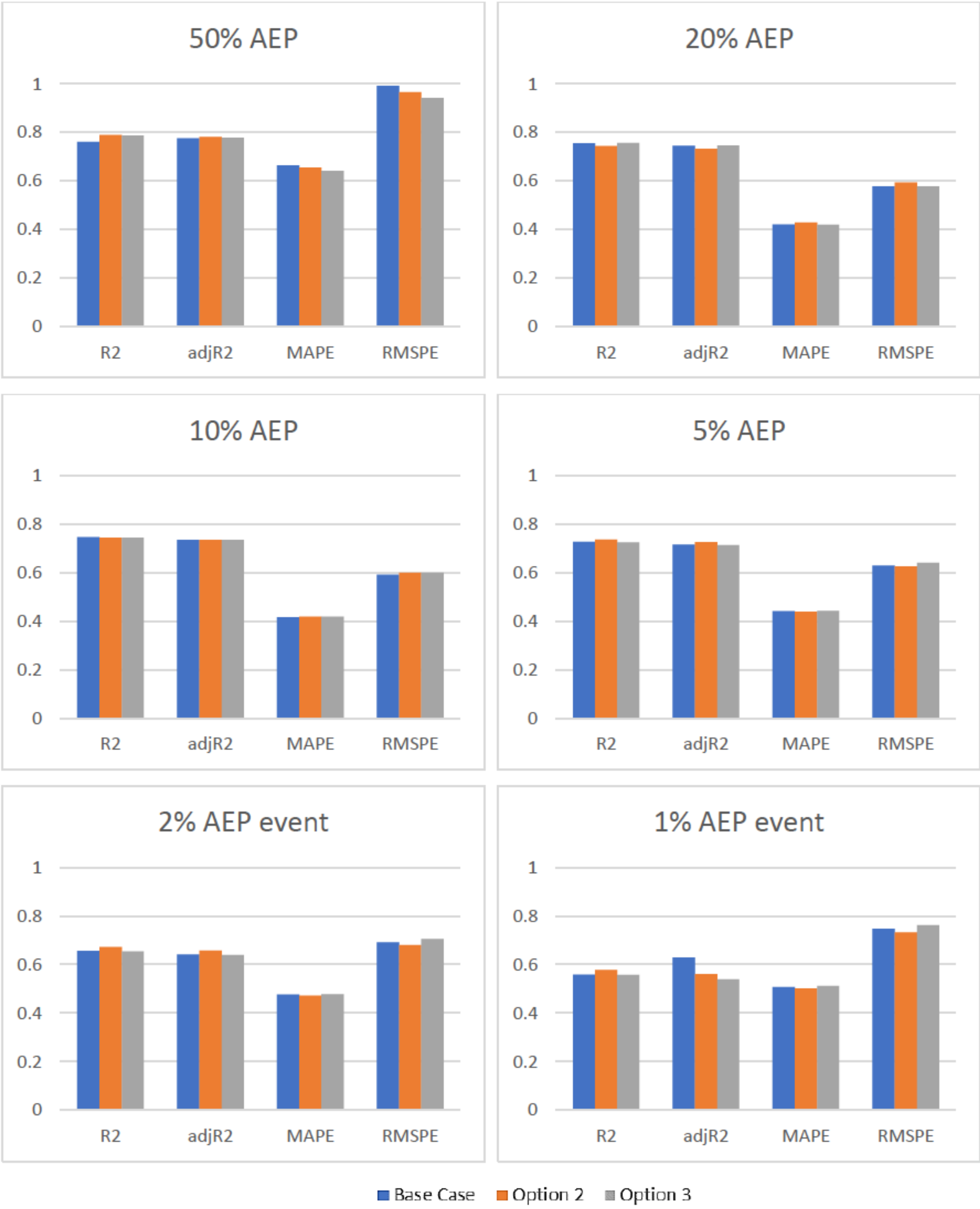


Figure 4.11. Compare statistics for options.

In consideration of the total range, median, mean, interquartile range were all taken into account when reviewing the models. Median and mean was consistent for all models. Further tests of R^2, adjusted R^2, MAPE, and RMSPE were completed, presented in Figure 4.11. The three options where indistinguishable in performance without a clear superior model. For ease of use, selecting only one intensity for all equations, and minimal improvement seen by trying to optimise the rainfall intensity selection, the Option 1 was adopted as the final model. The final two parameter equations are as follows;

$$Q_{50\%} = 0.305 A^{0.638} I^{1.520} \tag{4.1}$$

$$Q_{20\%} = 1.913 A^{0.600} I^{1.168} \tag{4.2}$$

$$Q_{10\%} = 3.989 A^{0.586} I^{1.025} \tag{4.3}$$

$$Q_{5\%} = 6.930 A^{0.575} I^{0.917} \tag{4.4}$$

$$Q_{2\%} = 12.391 A^{0.563} I^{0.806} \tag{4.5}$$

$$Q_{1\%} = 17.959 A^{0.556} I^{0.736} \tag{4.6}$$

where A is the catchment area, I is the design 72hours, 2% AEP rainfall intensity (for all equations), and Q is the resultant estimated discharge for given AEP event. Interesting to note the area coefficient does not change much across the design probability equations. This is expected because the area does not change at different probabilities where rainfall intensity does.

4.4.5 General Testing and Statistics

The final model was examined by common statistical tests and analysis of error between the predictions and the true discharge quantile calculated from FFA. There seems to be a general bias for small floods to be over-estimated and larger floods to be slightly under-estimated. Figure 4.12 shows 1% AEP, however, this appears to be the case for all probabilities. One possible improvement to reduce this is to add more variables to get a better representation of catchment variability and conditions. Another may be to use a calculated duration for rainfall intensity based on catchment characteristics. This is different to the optimisation method used in this project in that each catchment will have a unique rainfall selection not one predefined for each probability or the entire set of equations. However, this would add complexity and labour.

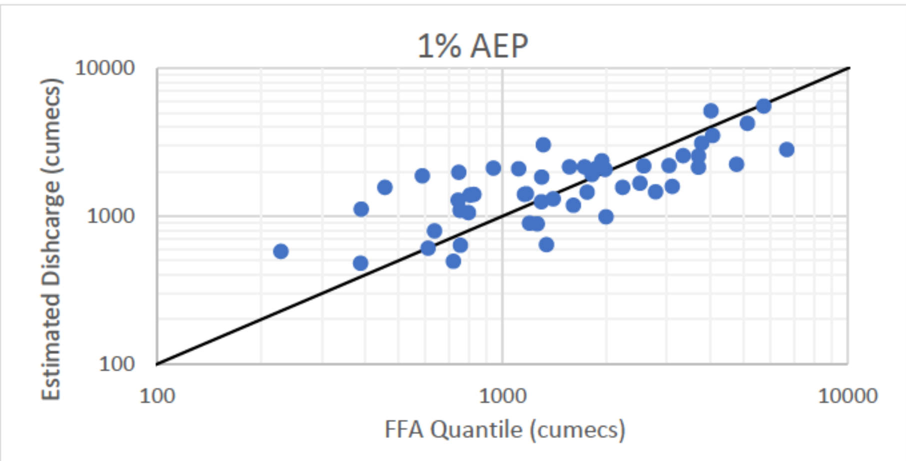


Figure 4.12. Comparison of FFA and final model, 1% AEP.

Although there is bias towards small and large floods, the number of gauges over- and under-estimated were quite even. Approximately 75% of gauges where estimated within 50% of the true quantile discharge. The final model predicted on average within 41-67% of the true value. R^2 indicated there was a reasonable fit but reduced to fair for less frequent events. Box and whisker plot of the percentage error can be found under Section 4.5.3 Figure 4.15.

Table 4.4. Statistics for Final model

AEP	R^2	Adj R^2	MAPE	RMSPE	Under	Over	< ±10%	< ±20%	< ±50%
50%	0.762	0.752	66.5%	99.3%	24	27	8	11	27
20%	0.756	0.745	42.1%	57.8%	23	28	8	20	39
10%	0.747	0.737	41.6%	59.2%	29	22	9	20	38
5%	0.728	0.717	44.2%	63.0%	29	22	8	15	37
2%	0.656	0.642	47.7%	69.2%	26	25	6	14	38
1%	0.559	0.540	50.7%	74.8%	26	25	6	13	36

4.4.6 Comparison to Palmen & Weeks (2011)

It may appear that Palmen & Weeks' original model performed better than the final model; RMSPE and adjusted R^2 being worse than Palmen & Weeks. This may suggest a greater variance in floods or a poorer model fit. Firstly, ARI is not equivalent to AEP so mostly cannot compare effectively. However, 2% and 1% AEPs are very close to 50year and 100year ARIs.

Compared to Palmen & Weeks there are many differences in the data and methodology and consequently the output. There have been a least 8-9 years additional stream records, new techniques and distributions suggested by updated version of ARR 2019. Palmen & Weeks (2011) used 289 gauges across Queensland, significantly more than 51 gauges used in this project. This would have utilised a wider range of catchments and more data points to fit the model better. Finally, IFD design rainfalls have been updated.

Comparison to 1987 IFDs

Between 1987 and 2016 design rainfall intensities there is a slight increase of +0.75% on average across all sites, which could be attributed to sampling error or difference in calculation techniques. Figure 4.13 is a graph of the 1987 and 2016 rainfall intensity data for the sites, labelled with increase or decrease in mm/hr. Updated design IFDs, 72hour duration, 2% AEP rainfall intensities across the sites varies from -21.5% to +29.5% compared to ARR 1987 for 72hour, 1 in 50year ARI event. Using the maximum difference of +29.5% (which may not be the largest difference across all of Queensland) would mean if you use the new IFDs with Palmen & Weeks equations it would produce a 26.1% larger discharge for 1% AEP event and 50.7% larger for 50% AEP event. This is clear that Palmen & Weeks (2011) equations are no longer usable with 1987 IFD rainfall information being decommissioned.

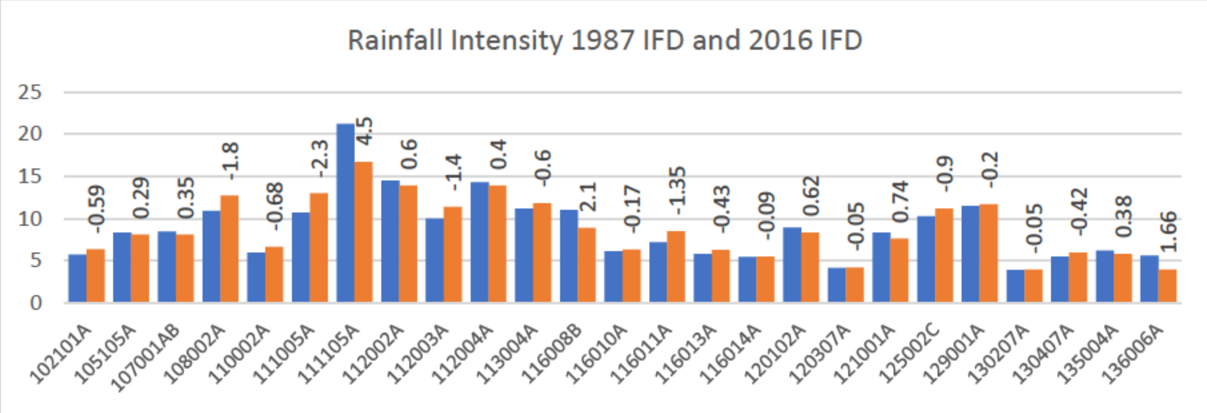


Figure 4.13(a). Comparison of 1987 and 2016 rainfall intensities.

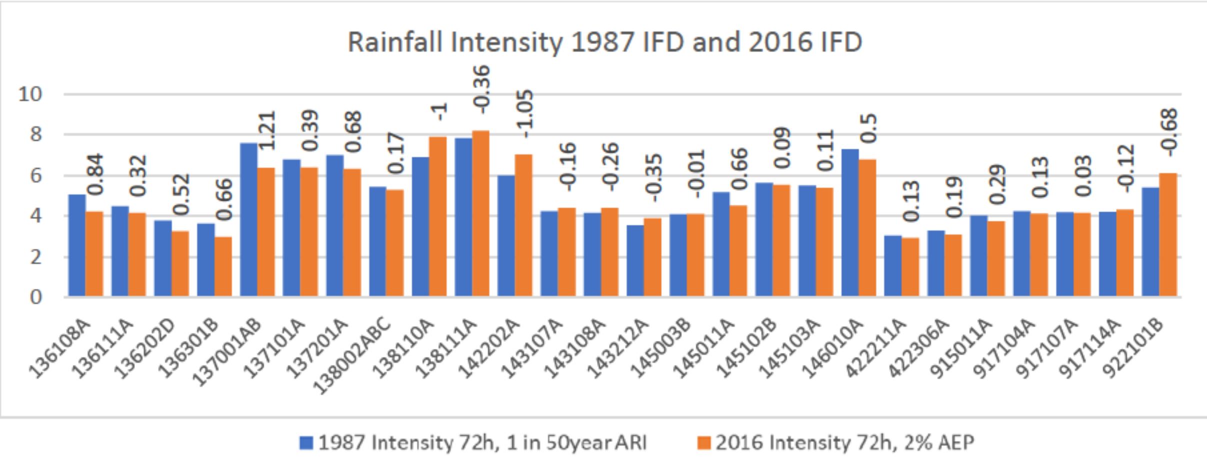


Figure 4.13(b). Comparison of 1987 and 2016 rainfall intensities.

4.5 Regional Flood Frequency Estimation Model Review

Regional Flood Frequency Estimation tool was used throughout this project; for regional information for FFA, quality checks, and comparison of the final study model. All except one gauge were in Region 1 calculation area: 422211A was located in Fringe 1 region. Three gauges were not available as part of RFFE database; 110002A, 130407A, and 917114A. As mentioned previously in Section 4.3, regional statistics for use as prior information in FFA did not improve distribution fitting.

4.5.1 Quality Checks

When completing the quality checks on FFA quantiles, refer Section 3.5.4, some anomalies in RFFE input data were identified. Most gauges simply need the RFFE database to be updated to account for additional 7-8 years of data. Several large floods in this period would have changed the frequency curve shape and produced different quantile discharges. Gauge 136108A for example, had three of the top four largest floods within the last eight years, and the largest previously may have been considered an outlier.

Those that differed dramatically could not be explained by the shorter record length. The tool uses a region of influence and interpolation principles to derive flood estimates. If the catchment size is the same as and location is within 50m of a gauge used for the input data, the tool should produce very similar results to that input data, yet this is not the case. Furthermore, the outputs are more closely in accordance with calculated quantiles in this project.

The maximum flood on record for gauge 137101A is 998.5m^3/s and the 1% AEP discharge is 939.8m^3/s. However, the RFFE 1% AEP is reported as 3089.8m^3/s, 3.1 times larger than the largest flood. Despite this substantial disparity, the output of RFFE gave a more reasonable answer of 822m^3/s. It appears RFFE combined records for gauge 145003 revision A and B, which may not be appropriate due to significant changes in stage-discharge relationship, refer Section 3.3.4 and Figure 3.8. Although it is unknown if there were any corrections made during RFFE calculations. Gauge 915011A matched RFFE input data closely with the entire gauge record. However, after excluding one flood based on the high outlier test, the revised 2% AEP quantile was dropped to 58% of the RFFE quantile, suggesting RFFE used all floods.

4.5.2 Shape Factor

Manually positioning the centroid of the catchment was subjectively challenging. Figure 4.14 below compares shape factor obtained by manual input to the true shape factor for the catchment from RFFE database. The shape factor on average was under-estimated. RFFE has limitations on the input shape factors to be within 0.5 and 1.1. However, six gauges required shape factors below 0.5 and five above 1.1. These have been shown in red. In order to get a result from RFFE the shape factor was reduced or increased. Apart from those not in the database, 422211A did not have a shape factor due to it being in Fringe 1 calculation region which does not require it.

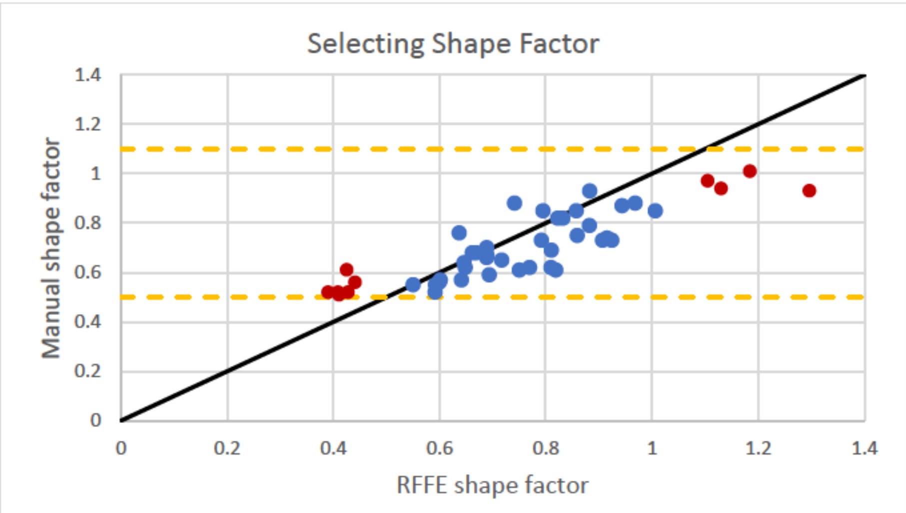


Figure 4.14. Shape factor for RFFE tool.

4.5.3 Study Model Compared to Regional Flood Frequency Estimation Model

Statistics and box and whisker plots describing errors from FFA discharges are presented below in Table 4.5 and Figure 4.15. The study model appeared to be slightly better for 2% and 1% AEP events but slightly worse for 20%, 10%, and 5% AEP events, according to MAPE, RMSPSE, and box and whisker plots. However, the study model estimated within 10%, 20%, and 50% of the true quantiles more than RFFE. RFFE was not as balanced for the number of gauges that were over- and under-estimated compared to the study model. The study model had more gauges closer to the true value but RFFE had smaller average error overall. A possible reason is that RFFE can represent a wider range of catchment conditions due to it being a more complex model.

Neither model could estimate 50% AEP well. Oddly, RFFE has more gauges within the given percentiles but has larger average error, significantly more over-estimations, and excessive errors up to 680%. It also appears that the interquartile range is smaller than study model. RFFE estimated a lot well but some are excessively wrong, tending to over-estimate. Overall, the final model had similar performance to RFFE, except for 50% AEP.

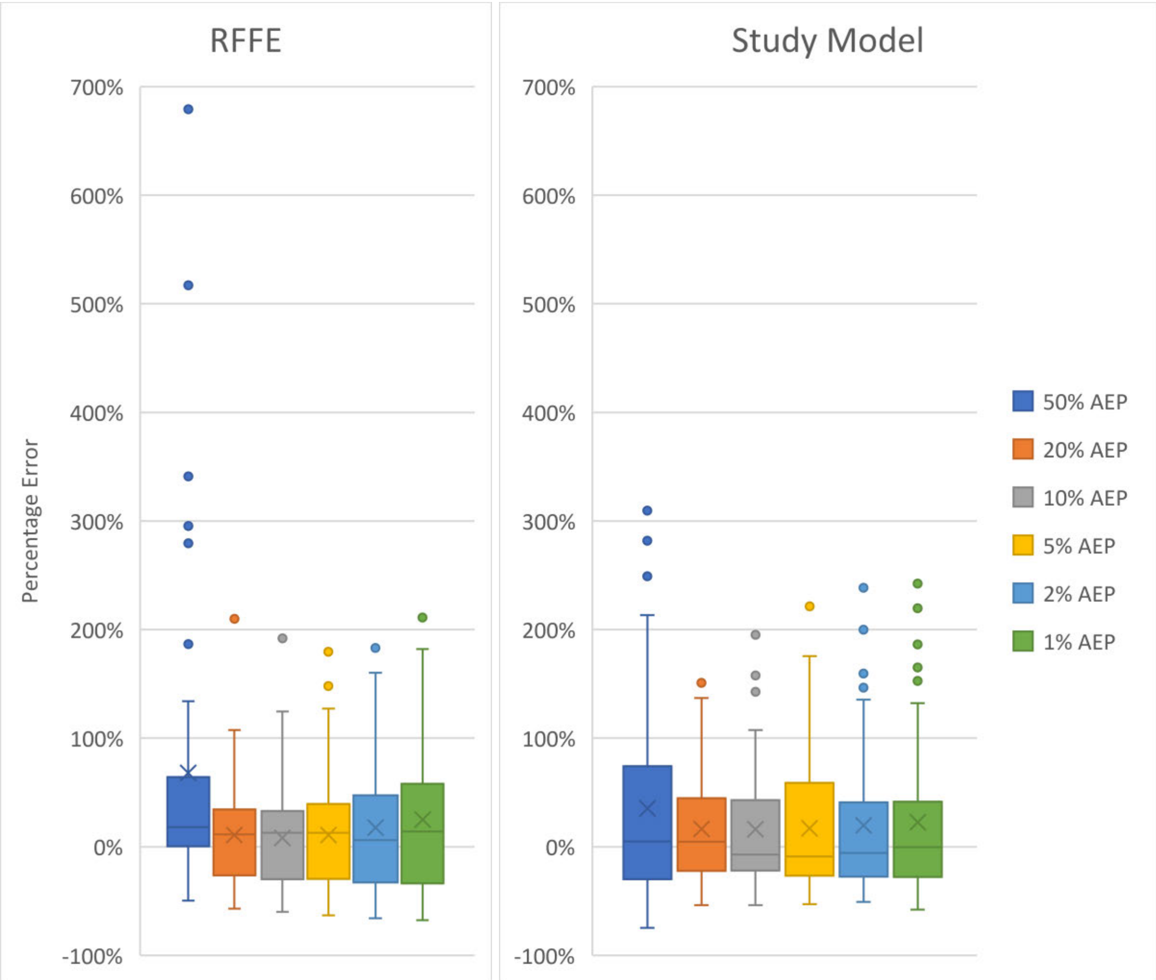


Figure 4.15. Box and whisker plots of percentage error from FFA quantiles comparing RFFE and Study model.

Table 4.5. Statistics for RFFE model.

AEP	MAPE	RMSPE	Under	Over	< ±10%	< ±20%	< ±50%
50%	76.0%	150.2%	11	40	11	23	32
20%	38.4%	51.0%	21	30	7	16	38
10%	38.7%	49.9%	22	29	3	14	38
5%	42.2%	54.7%	24	27	2	14	35
2%	48.4%	64.4%	23	28	5	12	34
1%	54.8%	75.9%	23	28	5	12	34

4.6 Case Studies

A further evaluation of the study model was completed by reviewing eleven randomly selected gauges (see Table 4.7). All models were plotted on frequency graphs and quantiles compared to determine the best model compared to FFA. Gauge 138110A is presented below in Figure 4.16 and Table 4.6, and all gauges are attached in Appendix F.

Table 4.6. Model estimates for gauge 138110A

AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	369.7	79.8	667.6	370	363.7		**39.4%**	282.6
20%	1299	899.9	1740.1	622	873.5		**18.1%**	656.9
10%	1898	1381	2457.3	821	1241		**9.52%**	929.7
5%	2460.7	1786.7	3184	1040	1617		**4.88%**	1301
2%	3172.2	2186.6	4320	1350	2138		**1.98%**	1773
1%	3693.2	2394.2	5365.5	1610	2556		**1%**	2142

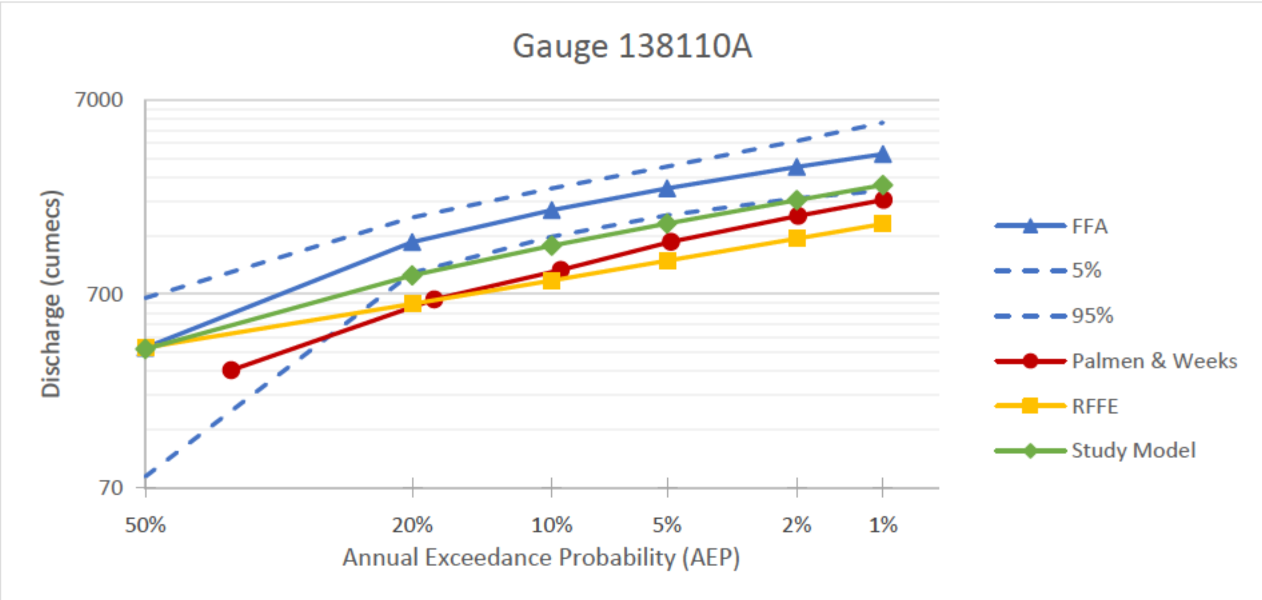


Figure 4.16. Frequency plot comparing models for gauge 138110A.

The study model was the most accurate for this gauge station, however, for many probabilities the estimated are outside 90% confidence interval of the flood frequency analysis. Both RFFE and the study model was accurate for 50% AEP event. RFFE had the shape of a straight line on log-normal graph, which was typical for all eleven gauges. The study model typically had more of a curve shape in comparison. Palmen & Weeks was always below the study model, for the eleven stations reviewed, and had even greater curve shape, except gauge 136202D which matched the study model closely.

A summary of the eleven gauges are presented in Table 4.7 where the superior model is indicated. Gauges 111005A and 146010A were difficult to identify the best model due to RFFE and the study model were similar distance from FFA. The study model was regard superior six times out of the eleven gauges reviewed.

Table 4.7. Superior model for reviewed gauges

Gauge ID	Study Model	RFFE Model
102101A		✓
107001AB	✓	
111005A	✓	
116013A		✓
120307A		✓
136108A	✓	
136202D	✓	
138110A	✓	
143107A		✓
143108A		✓
146010A	✓	

5 Conclusion and Recommendations

This project was undertaken to update Palmen & Weeks (2011) method using current guidelines, primarily Australian Rainfall and Runoff 2019, and data, and has resulted in a usable set of six two parameter equations to estimate peak discharge for a catchment at different design probabilities, see equations 4.1-4.6.The project was successful in this regard and also showed additional incites.

An attempt was made to improve the model by selecting different duration-frequencies for rainfall intensity. Results showed not improvement. The performance of the new model is similar to RFFE model, which is recommended by ARR 2019. Predictions where within 41-67% on average across all 51 gauges used for regression. The model tended to slightly over-estimate small floods and under-estimate large floods for all frequency events. Neither, RFFE and final model performed well for 50% AEP frequency.

This model is suited for preliminary design of culverts, causeways, and small bridges and model verification. The model is an alternative to RFFE and was found to have similar performance. The level of accuracy and confidence required for final design may not be achievable. It will also not perform well for complex and atypical catchments with unusual catchment shapes, terrain or storage. Catchments greater than 1000km² or very small catchments Urbanised catchment with lots of impervious surfaces or constructed storage or drainage are not typical of those used in this project.

Additional Findings and Further Work

It was found that Generalised Extreme Value probability distribution for flood frequency analysis is typical for Queensland, supporting ARR 2019 Book 2 Chapter 2 suggestion. However, Log-Pearson III was necessary at times. Multiple Grubbs Beck Test (MGBT) triggered some concern when up to half the annual maximum series was removed for 10% of the gauges used. It is unknown if the test truly identified 50% of observations as outliers or if it would have identified more if the test was not limited to less than the median observation. This may require further investigation. In addition, removing this many observations may produce a frequency curve that does not reflect reality. Regional parameter information provided by RFFE was not useful for flood frequency analysis. In almost all cases it produces a poor fit.

It was found the RFFE database requires updating. The model itself may also require a review. RFFE is based on region of influence and interpolation for nearby gauges. So, if the catchment area, and location is the same as one used by the tool (zero regionalisation), it is expected to output the same quantile discharges as the input. For some gauges the input data and output were significantly different. Furthermore, the output was closer to FFA and the study model estimates. The mechanics of RFFE was not looked at as part of this study to identify the cause of this disparity.

The work completed, database of flood quantiles and revised rating curves may also be submitted to DNRME or RFFE to update their systems, or indicate things to review. The database may also be used and extended as a foundation to further improve the model. During the project some erroneous data for two gauges was discovered and DNRME was consulted and made changes to online portal system.

Further work will likely involve further testing of the model for its limitations and performance. Reducing the record length criteria from fifty years will increase the gauges used for regression. This may give a better regression result and sensitivity of the regression method. Reducing the record length to forty year will increase the number of gauges to 172, before any checks. It is suggested that investigation into a different equations structure or addition of more parameters (that are hydrologically significant) may provide a more flexible model to cater for the influence of more catchment characteristics. However, additional parameters will increase the labour involved in calculations. Moreover, the type and complexity of those parameters must be also considered. Catchment shape or stream length or slope are two possible parameters to include.

6 References

26 - Prior and posterior predictive distribution – an introduction 2014, online video, Ox educ, viewed 14 May 2019, <https://www.youtube.com/watch?v=R9NQY2Hyl14>.

Arcement, GJ Jr., & Schneider, VR 1989, 'Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains', *United States Geological Survey Water-Supply Paper 2339*, US Department of The Interior, Denver, CO, United States.

Ball, J, Babister, M, Retallick, M, Ling, F & Thyer, M 2019a, 'Book 1: Chapter 2: Fundamental Issues', *Australian Rainfall & Runoff: A guide to flood estimation,* Commonwealth of Australia (Geoscience Australia).

Ball, J, Jordan, P, Seed, A, Nathan, R, Leonard, M & Weinmann, E 2019b, 'Book 2: Chapter 2: Rainfall Models, *Australian Rainfall & Runoff: A guide to flood estimation*, Commonwealth of Australia (Geoscience Australia).

Baratti, E, Montanari, A, Castellarin, A, Salinas, JL, Viglione, A, & Bezzi, A 2012, 'Estimating the flood frequency distribution at seasonal and annual time scales', *Hydrology and Earth System Sciences*, vol. 16, pp. 4651-4660.

Bates, B, McLuckie, D, Westra, S, Johnson, F, Green, J, Mummery, J, & Abbs, D 2019, 'Book 1: Chapter 6: Climate Change Considerations', *Australian Rainfall & Runoff: A guide to flood estimation*, Commonwealth of Australia (Geoscience Australia).

BMT Group Ltd 2018, *TUFLOW*, DMT group Ltd, viewed 4 September 2019, <https://www.tuflow.com/Default.aspx>.

Braca, G 2008, 'Stage-discharge relationships in open channels: Practices and problems', *FORALPS Technical Report 11*, University of Trento, Department of Civil and Environmental Engineering, Trento, Italy.

Bradshaw, CJA 2012, 'Little left to lose: deforestation and forest degradation in Australia since European colonization', *Journal of Plant Ecology*, vol. 5, no. 1, pp. 109-120.

Brisbane City Council (BCC) 2003, 'Appendix C – Manning's Roughness', *Natural Channel Design Guidelines*, issue B, Brisbane City Council, Brisbane, QLD, viewed 27 May 2019, <https://www.brisbane.qld.gov.au/planning-and-building/planning-guidelines-and-tools/planning-guidelines/technical-documents>.

Bureau of Meteorology 2019a, *BOM Webinar: Using the new design rainfalls for Australia*, online video, viewed 9 May 2019, <https://www.youtube.com/watch?v=OLdeAE1kROk&list=PLbKuJrA7Vp7mnYycssks_17ykBiK-h6ip>.

Bureau of Meteorology 2019b, 'Brisbane: Comparison between 2016 IFDs and ARR87 IFDs', Australian Government, viewed 6 May 2019, <http://www.bom.gov.au/water/designRainfalls/document/Brisbane_2016IFDs.pdf>.

Bureau of Meteorology 2019c, 'Frequently Asked Questions', Australian Government, viewed 9 May 2019, <http://www.bom.gov.au/water/designRainfalls/index.shtml#faq>.

Bureau of Meteorology 2019d, 'Glossary', *Intensity -Frequency-Duration (AR&R87),* Australian Government, viewed 29 May 2019, <http://www.bom.gov.au/water/designRainfalls/ifd-arr87/glossary.shtml>.

Bureau of Meteorology 2019e, 'How were the 2016 design rainfalls estimated', Australian Government, viewed 6 May 2019, <http://www.bom.gov.au/water/designRainfalls/index.shtml#estimation>.

Cabi, N 2014, *Rating curve produced from hydraulic model results,* digital image of cross section and rating curve, AIR Worldwide, viewed 20 September 2019, <https://www.air-worldwide.com/Publications/AIR-Currents/2014/Where-Does-the-Water-Go-/>.

Cohn, TA, England, JF, Berenbrock, CE, Mason, RR, Stedinger, JR, & Lamontagne, JR 2013, 'A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series', *Water Resources Research*, vol. 49, pp. 5047-5058.

Cunnane, C 1985, 'Unbiased plotting positions – a review', *Journal of Hydrology*, vol. 37, pp. 205-222.

Department of Natural Resources, Mines and Energy (DNRME) 2019, *Water Monitoring Information Portal*, Queensland Government, accessed April 2019, <https://water-monitoring.information.qld.gov.au/>.

Design Rainfall Data System (2016) n.d., computer software, Bureau of Meteorology, Australian Government, accessed 28 April 2019, <http://www.bom.gov.au/water/designRainfalls/revised-ifd/>.

Devi, GK, Ganasri, BP, Dwarakish, GS 2015, 'A review on hydrological models', *Aquatic Procedia*, vol. 4, pp. 1001-1007.

Don Ho 2019, *Notepad++*, version 7.7.1, computer software, Don Ho.

Drakos, G 2018, 'How to select the Right Evaluation Metric for Machine Learning Models: Part 1 Regression Metrics', Towards Data Science, viewed 10 October 2019, <https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-1-regrression-metrics-3606e25beae0>.

Frost, J n.d., 'Five Reasons Why Your R-squared can be Too High', Statistics By Jim, viewed 10 October 2019, <https://statisticsbyjim.com/regression/r-squared-too-high/>.

Gaffney, A, & Babister, M 2019, 'Queensland Regional Flood Frequency using the Quantile Regression Technique (ARR2016)', power point, WMAwater.

Geoscience Australia 2019a, 'Australian Rainfall and Runoff', Commonwealth of Australia (Geoscience Australia), viewed 11 September 2019, <http://arr.ga.gov.au/>.

Geoscience Australia 2019b, 'Revision Projects', Commonwealth of Australia (Geoscience Australia), viewed 11 September 2019, <http://arr.ga.gov.au/arr-guideline/revision-projects>.

Geoscience Australia 2019c, 'RFFE Regions', map, pdf document, Commonwealth of Australia (Geoscience Australia) viewed 11 September 2019, <http://www.arr-software.org/pdfs/RFFE_Regions.pdf>.

Green, J, Johnson, F, Beesley, C, & The, C 2019, 'Book 2: Chapter 3: Design Rainfall', *Australian Rainfall & Runoff: A guide to flood estimation*, Commonwealth of Australia (Geoscience Australia).

Hatton, T 2011, *Australian drainage divisions*, digital image of drainage divisions, Australia State of the Environment, Commonwealth of Australia, viewed 7 September 2019, <https://soe.environment.gov.au/theme/inland-water/topic/australias-water-resources-and-use>.

Jamal, H 2017, 'History and origin of hydrology', AboutCivil, viewed 17 October 2018, <https://www.aboutcivil.org/origin-history-of-hydrology.html>.

Kiem, AS, Franks, SW, & Kuczera, G 2004, 'Multi-decadal variability of flood risk', *Geophysical Research Letters*, vol. 30, no. 2, p.1035.

Kidson, R & Richards, KS 2005, 'Flood frequency analysis: assumptions and alternatives', *Progress in Physical Geography*, vol. 29, no. 3, pp. 392-410.

Kuczera, G & Franks, S 2019, 'Book 3: Chapter 2: At-Site Flood Frequency Analysis', *Australian Rainfall & Runoff: A guide to flood estimation*, Commonwealth of Australia (Geoscience Australia).

Langat, PK, Kumar, L, & Koech, R 2019, 'Identification of the Most Suitable Probability Distribution Models for Maximum, Minimum, and Mean Streamflow', *Water*, vol. 11, no. 734.

Lohani, A n.d., 'Rainfall-runoff analysis and modelling', National Institute of Hydrology, Roorkee, India, viewed 10 October 2018, <http://nihroorkee.gov.in/NHP/traning_Modules/AKL_RAINFALL-RUNOFF%20MODELLING-LECTURE%20(1).pdf>.

Maynard, R 2014, 'Burnett River Floods and Gauging & Rating Implications', paper presented at 17th Australian Hydrographers Association Conference, Sydney, NSW, 28-31 October.

MathWorks 2019, *MATLAB*, version R2018b student, computer software, MathWorks, Natick, Massachusetts, USA.

Millington, N, Das, S, & Simonovic, SP 2011, 'The Comparison of GEV, Log-Pearson Type 3 and Gumbel Distributions in the Upper Thames River Watershed under Global Climate Models', *Water Resources Research Report*, report no. 77, Department of Civil and Environmental Engineering & The University of Western Ontario, London, Ontario, Canada.

National Institute of Standards and Technology (NIST) 2013a, '1.3.5.14. Anderson-Darling Test', *NIST/SEMATECH e-Handbook of Statistical Methods*, US Department of Commerce, viewed 3 May 2019, <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm>.

National Institute of Standards and Technology (NIST) 2013b, '1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test', *NIST/SEMATECH e-Handbook of Statistical Methods*, US Department of Commerce, viewed 3 May 2019, <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm>.

National Institute of Standards and Technology (NIST) 2013b, ‘4.4.4 How can I tell if a model fits my data?’, *NIST/SEMATECH e-Handbook of Statistical Methods*, US Department of Commerce, viewed 21 August 2019, <https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd44.htm>.

Newman, MC 1993, ‘Regression analysis of log-transformed data: Statistical bias and its correction’, *Environmental Toxicology and Chemistry,* vol. 12, pp. 1129-1133.

Palmen, LB & Weeks, WD 2011, ‘Regional flood frequency for Queensland using the quantile regression technique’, *Australian Journal of Water Resources,* vol. 15, no. 1, pp. 47-57.

Philips, JV, & Tadayon, S 2006, ‘Selection of Manning’s Roughness Coefficient for Natural and Constructed Vegetated and NonVegetated Channels, and Vegetation Maintenance Plan Guidelines for Vegetated Channels in Central Arizona’, *U.S. Geological Survey Scientific Investigations Report 2006-5108,* Reston, Virginia.

Pilgrim, DH & Doran, DG 1987, 'Chapter 10 - Flood Frequency Analysis', Australian Rainfall and Runoff: A Guide to Flood Estimation, vol. 1, ch. 10, revised edn. 1987, pp. 195-236, The Institution of Engineers, Australia, Barton, ACT.

Pitney Bowes 2016, *MapInfo Professional,* version 15.0, computer software, Pitney Bowes, Macquarie Park, NSW.

QGIS Brighton 2017, *QGIS,* version 2.6.1, computer software, QGIS Brighton.

Queensland Government 2019, *Queensland Globe,* computer software, State of Queensland 2019, accessed 14 October 2019,
< https://qldglobe.information.qld.gov.au/>.

Rahman, A & Haddad, K 2019a, ‘About the Technique’, Engineers Australia, viewed 15 September 2019, <https://rffe.arr-software.org/about.html>.

Rahman, A & Haddad, K 2019b, ‘Limits of Applicability’, Engineers Australia, viewed 11 September 2019, <https://rffe.arr-software.org/limits.html>.

Rahman, A & Haddad, K 2019c, ‘Method Changelog, Engineers Australia, viewed 15 September 2019, <https://rffe.arr-software.org/changelog.html>.

Rahman, A, Haddad, K, Kuczera, G & Weinmann, E 2009, 'Project 5 Regional Flood Methods: Stage 1 Report', *Australian Rainfall and Runoff Revision Projects*, The University of Western Sydney, Penrith South, NSW.

Rahman, A, Haddad, K, Kuczera, G & Weinmann, E 2019, 'Book3: Chapter 3: Regional Flood Methods', *Australian Rainfall & Runoff: A guide to flood estimation*, Commonwealth of Australia (Geoscience Australia).

Rahman, A, Haddad, K, Haque, M, Kuczera, G & Weinmann, E 2015, 'Project 5 Regional Flood Methods: Stage 3 Report', *Australian Rainfall and Runoff Revision Projects*, The University of Western Sydney, Penrith South, NSW.

Rahman, AS, Rahman, A, Zaman, MA, Haddad, K, Ahsan, A & Imteaz, M 2013, 'A study on selection of probability distributions for at-site flood frequency analysis in Australia', *Natural Hazards*, no. 69, pp. 1803-1813.

Rainfall Intensity Frequency Duration Data Online System n.d., computer software, Bureau of Meteorology, Australian Government, accessed 25 May 2019, <http://www.bom.gov.au/cgi-bin/hydro/has/CDIRSWebBasic>.

Rui, X-f, Liu, N-n, Li, Q-l, & Liang, X 2013, 'Present and future of hydrology', *Water Science and Engineering*, vol. 6, no. 3, pp. 241-249.

Teng, J, Jakeman, AJ, Vaze, J, Croke, BFW, Dutta, D & Kim, S 2017, 'Flood inundation modelling: A review of methods, recent advances and uncertainty analysis', Environmental Modelling & Software, vol. 90, pp. 201-216.

TUFLOW 2015, *Flike TUFLOW*, BMT Group, viewed 18 May 2019, <https://flike.tuflow.com/>.

TUFLOW 2017, *TUFLOW User Manual: Build 2017-09-AC*, BMT Group Ltd.

TUFLOW 2018, 'Tutorial Module01', *TUFLOW Wiki*, BMT Group Ltd, viewed 11 September 2019, <https://wiki.tuflow.com/index.php?title=Tutorial_Module01>.

Using the new design rainfalls for Australia 2019, webinar, Bureau of Meteorology, Australian Government, viewed 16 April 2019,

<https://www.youtube.com/watch?v=OLdeAE1kROk&list=PLbKuJrA7Vp7mnYycssks_17ykBiK-h6ip>.

Wang, QJ 1997, ‘LH moments for statistical analysis of extreme events’, *Water resources research*, vol. 33, no. 12, pp. 2841-2848.

Wang, QJ 1998, ‘Approximate goodness-of-fit tests of fitted generalized extreme value distributions using LH moments’, *Water resources research*, vol. 34, no. 12, pp. 3497-3502.

WBM Pty Ltd 2011, *TUFLOW Flike*, version 5.0.251.0, computer software, BMT Group Ltd, Teddington, UK.

WBM Pty Ltd 2018, *TUFLOW HPC*, build 2018-03-AC with GPU module, computer software, BMT Group Ltd, Teddington, UK.

Westra, S, Evans, J, Mehrotra, R, Sharma, A 2013, ‘A conditional disaggregation algorithm for generating fine time-scale rainfall data in a warmer climate’, *Journal of Hydrology*, no. 479, pp. 86-99.

Zhang, L, Dawes, WR & Walker, GR 1999, ‘Predicting the effect of vegetation changes on catchment average water balance’, technical report 99/12, Cooperative Research Centre for Catchment Hydrology, Canberra, ACT, 1999.

Appendices

A.1 Project Specification

ENG4111/4112 Research Project
Project Specification

For: Samuel Walker

Title: Estimation of peak flow in Queensland using Quantile Regression Technique and ARR 2016.

Major: Civil Engineering

Supervisors: Rezaul Chowdhury

Enrolment: ENG4111 – EXT S1, 2019
ENG4112 – EXT S2, 2019

Project Aim: To update and potentially improve overall performance of the QRT method in estimating peak flow in Queensland presented by Palmen and Weeks (2011), utilising the latest ARR 2016 rainfall intensities and procedures.

Programme: Version 1, 20th March 2019

1. Acquire and prepare stream gauge data for stations throughout Queensland.

2. Identify stations that may have poor quality rating curves. Conduct 2D hydraulic modelling to produce improved rating curves with assistance from TMR. TMR will provide a number of rating curves.

3. Complete Flood Frequency Analysis for each site. Conduct checks for outliers. Fit Generalised Extreme Value using L/LH-moments and Log-Person 3 distributions to data. Compare and select best distribution and estimate quantiles for various AEPs.

4. Conduct regression with quantile estimate flow, catchment area and rainfall intensity (72hours duration, 2% AEP). Investigate if other rainfall intensity duration/frequencies selection produce more accurate results.

5. Assess accuracy, sensitivity and limitations of final equations. Randomly select sites and compare final equations to real data, Palmen and Weeks (2011) findings and other common methods.

If time and resources permit:

6. Investigate whether further improvement is found by breaking catchment size into small and medium sized.

7. Investigate whether further improvement is found by adding more, relevant parameters to regression equations.

Appendix B – MATLAB code

Table of Contents

B.1 DataExtraction

Name: DataExtraction.m

Type: Script

Details:

Takes raw data from files downloaded from WMIP system and extracts what is needed and saves text files and MATLAB variables. Based on old folder structure.

```matlab
% Records downloaded files for each site into table for review of what is missing.
%
% Created: Sam Walker
% Date: 27/03/2019
% Version: 1.0

clear
format compact

%% DIRECTORIES
p = pwd;

% Open Sites
popen = [p(1:3), '1. Data Collection\WMIP raw data\Open']
% Closed Sites
pclosed = [p(1:3), '1. Data Collection\WMIP raw data\Closed']

% Save location
psave = [p(1:3), '1. Data Collection\WMIP Extracted']


% Delete contents of destination
pd = dir(psave);
for pdk = 3:length(pd)
    if pd(pdk).isdir == 1
        rmdir([pd(pdk).folder,'\',pd(pdk).name],'s')
    else
        delete([pd(pdk).folder,'\',pd(pdk).name])
    end
end

%% BASINS
b = [dir(popen);dir(pclosed)];
s = 0;
for k = 1:length(b)
    if any(strcmp(b(k).name,{'.','..'}))
    elseif b(k).isdir == 1
        s = s+1;
        Basin{s,1} = b(k).name(1:end-6);
        Basin{s,2} = [''' ', b(k).name(end-2:end)];
        Basin{s,3} = [b(k).folder,'\',b(k).name];

        % temporary for use in script
        if strcmp(b(k).folder,pclosed)
            Basin{s,4} = 'closed';
        else
            Basin{s,4} = 'open';
        end
    end
end
[~,BB] = unique(Basin(:,1));
OUTbasin = Basin(BB,1:2);
OUTbasin = [OUTbasin, num2cell(zeros(size(OUTbasin,1),2))];

%% GAUGES
```

```
ss = 0;
for kk = 1:length(Basin)
    pp = Basin(kk,3);
    g = dir(pp{1});
    cl = 0;
    op = 0;
    for ii = 1:length(g)
        if any(strcmp(g(ii).name,{'.','..'}))
        elseif g(ii).isdir == 1
            ss = ss+1;
            Gauge{ss,1} = [char(Basin(kk,3)), '\',g(ii).name];
            Gauge{ss,2} = g(ii).name;
            Gauge{ss,3} = char(Basin(kk,1));
            Gauge{ss,4} = char(Basin(kk,2));
            Gauge{ss,5} = char(Basin(kk,4));

            % Counter for no. sites per gauge
            idx = find(strcmp(OUTbasin(:,1),Basin(kk,1)));
            if strcmp(Basin(kk,4),'open')
                OUTbasin{idx,3} = op + 1;
                op = op+1;
            else
                OUTbasin{idx,4} = cl + 1;
                cl = cl+1;
            end
        end
    end

end
m = size(Gauge(:,1),1);

% Check for duplicate Gauges
munique = size(unique(Gauge(:,2)),1);
[G,GG,GGG] = unique(Gauge(:,2));
d = hist(GGG,unique(GGG));
Gdup = G(d>1);

% Sort
Gauge = Gauge(GG,:);

%% OUTPUT BASINS
% Print to screen
n = length(unique(Basin(:,1)));
fprintf('Number of Basins: %d\n',n)
if m > munique
    fprintf('Number of Stations: %d (all)   %d (unique)\nDuplicate Gauges:\n',m,munique)
    disp(Gdup)
    error('Please delete duplicate gauges.')
end
fprintf('Number of Stations: %d\n',m)

%% SITES
AllMeta = table;
AllSummary = table;
```

```matlab
for kkk = 1:m % m % change start end numbers when testing through instead of processing everything all the time
    fprintf('No.: %d\t\t processing: %s\n',kkk,char(Gauge(kkk,2)))

    %% Make folder
    mkdir([psave,'\',char(Gauge(kkk,2))])

    %% Files
    gg = dir(char(Gauge(kkk,1)));

    % Num of folders
    Gauge{kkk,6} = sum([gg.isdir])-2;

    % Total num of files
    Gauge{kkk,7} = size(gg,1)-sum([gg.isdir]);

    gg = struct2cell(gg)';
    % Site Details
    if any(strcmp(gg(:,1),{'Site Details Summary.txt'})),   Gauge{kkk,8} = 1;
    else,   Gauge{kkk,8} = 0;
    end

    % Rating Table
    if any(strcmp(gg(:,1),{'Rating Table.txt'})),   Gauge{kkk,9} = 1;
    else,   Gauge{kkk,9} = 0;
    end

    % Cross Section
    if any(strcmp(gg(:,1),{'Cross Section.png'})),  Gauge{kkk,10} = 1;
    else,   Gauge{kkk,10} = 0;
    end

    % Rating Curve
    if any(strcmp(gg(:,1),{'Rating Curve.png'})),   Gauge{kkk,11} = 1;
    else,   Gauge{kkk,11} = 0;
    end

    % Rating Curve with Data
    if any(strcmp(gg(:,1),{'Rating Curve with Data.png'})), Gauge{kkk,12} = 1;
    else,   Gauge{kkk,12} = 0;
    end

    %% Extract Rating Table
    if cell2mat(Gauge(kkk,9)) == 1    % test if Rating Table.txt exists
        RatingTable = ExtRT(char(Gauge(kkk,1)));

%         % Write to files
        save([psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Rating Table.mat'],'RatingTable')
        writetable(RatingTable,[psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Rating Table.txt'])
    end

    %% Extract from Site Details Summary
    if cell2mat(Gauge(kkk,8)) == 1
```

```matlab
        [Meta,Gaugings,SDS] = ExtSDS(char(Gauge(kkk,1)),char(Gauge(kkk,2)),...
            char(Gauge(kkk,3)),char(Gauge(kkk,4)));

        Summary = table;
        Summary.Gauge = Meta.Gauge;
        Summary.Lat = Meta.Lat;
        Summary.Long = Meta.Long;
        Summary.StartDate = Meta.StartDate;
        Summary.EndDate = Meta.EndDate;
        Summary.Years = Meta.Years;
        Summary.MaxGaugeStage = Meta.MaxGaugeStage;
        Summary.MaxRecordStage = Meta.MaxRecordStage;
        Summary.Area = Meta.Area;

        % write to files
        save([psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Summary.mat'],'Summary')
        writetable(Summary,[psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Summary.txt'])

        save([psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Metadata.mat'],'Meta')
        writetable(Meta,[psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Metadata.txt'])

        save([psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Gaugings.mat'],'Gaugings')
        writetable(Gaugings,[psave,'\',char(Gauge(kkk,2)),'\',char(Gauge(kkk,2)),' - Gaugings.txt'])
    end

    %% Add tables together
    AllMeta = [AllMeta;Meta];
    AllSummary = [AllSummary;Summary];
end

%% OUTPUT SITES

save([psave,'\All - Metadata.mat'],'AllMeta')
writetable(AllMeta,[psave,'\All - Metadata.txt'])

save([psave,'\All - Summary.mat'],'AllSummary')
writetable(AllSummary,[psave,'\All - Summary.txt'])

hdngs = {'FilePath','Gauge','Basin','BasinCode','OpenClosed',...,
    'NoOfFolders','NoOfFiles','SiteDetails','RatingTable',...
    'CrossSection','RatingCurve','RatingCurveWithData'};
OutGauge = cell2table(Gauge,'VariableNames',hdngs);
save([psave,'\All - Files Summary.mat'],'OutGauge')
writetable(OutGauge,[psave,'\All - Files Summary.txt'])

hdngs = {'Basin','BasinCode','OpenSites','ClosedSites'};
OutBasin = cell2table(OUTbasin,'VariableNames',hdngs);
save([psave,'\Basin Summary.mat'],'OutBasin')
writetable(OutBasin,[psave,'\Basin Summary.txt'])
```

```
fprintf('Number of Basins: %d\n',n)
fprintf('Number of Stations: %d\n\n',m)
fprintf('Saved file path: %s\n\n',psave)

%% FUNCTION: Etract data from Site Details Summary.txt
function [Meta,Gaugings,SDS] = ExtSDS(path,gauge,basin,basincode)
    clear('Gaugings','Meta','RThist')

    % open file
    fSDS = fopen([path,'\Site Details Summary.txt'],'r');
    SDS = textscan(fSDS,'%s','Delimiter','\n');
    fclose(fSDS);
    SDS = SDS{1,1};

    % Remove empty rows and repeated titles
    SDS = SDS(~cellfun('isempty',SDS'));
    SDS = regexprep(SDS,'Page \d+','Page 1'); % remove page numbers
    SDS = unique(SDS,'stable');

    %% Gaugeings
    idx = find(contains(SDS,'GAUGINGS'))+3;
    if isempty(idx)
        Gaugings = table;
    else
        Gaugings = table;
        temp = [];
        k = 1;
        while contains(char(SDS(idx)), {'100 Level (Metres)'})
            temp2 = strsplit(char(SDS(idx))); % if row is short
            if size(temp2,2)<size(temp,2)
                break
            else
                temp = [temp;strsplit(char(SDS(idx)))];
            end
%             strsplit(char(SDS(idx)))
%             temp = [temp;strsplit(char(SDS(idx)))];

            if length(char(temp(end,8))) == 8
                Date(k) = datetime([char(temp(end,7)), char(temp(end,8))],'InputFormat','dd/MM/yyyyHH:mm:ss');
            else
                Date(k) = datetime([char(temp(end,7)), char(temp(end,8))],'InputFormat','dd/MM/yyyyHH:mm');
            end

            idx = idx+1;
            k = k+1;
        end
        Gaugings.No = str2double(temp(:,9));
        Gaugings.Date = Date';
        Gaugings.Stage = str2double(temp(:,10));
        Gaugings.Flow = str2double(temp(:,11));
        Gaugings.PercentDev = str2double(temp(:,12));
%         fprintf('...\t\t\t Gaugings exists\n')
```

```
    end

    %% Removed due to complexity and irregularity of files
%     %% Rating Table History
%     idx = find(contains(SDS,'RATING TABLES'))+3;
%     if isempty(idx)
%         RThist = table;
%     else
%         RThist = table;
%         temp = [];
%
%         tsttxt = {'140 Discharge (Cumecs)','136 Storage Vol. (ML)',...
%             '137 Storage Area  (Ha)','103 Level Calib.(Metres)'};
%         tst = strsplit(char(SDS(idx)),tsttxt);
%         while contains(strtrim(tst(1)), {'100.00 Level (Metres)','130.00 SW EL AHD'})
%             temp = [temp;strsplit(char(tst(2)))];
%
%             idx = idx+1;
%             tst = strsplit(char(SDS(idx)),tsttxt);
%         end
%         RThist.No = str2double(temp(:,5));
%         RThist.Date = datetime([char(temp(:,2)), char(temp(:,3))],'InputFormat','dd/MM/yyyyHH:mm');
%         fprintf('...\t\t Rating Table history exists\n')
%     end

    %% Metadata
    Meta = table;
    Meta.Gauge = gauge;

    temp = strsplit(char(SDS(contains(SDS,'Site Name:')==1)),':');
    Meta.Name = string(strtrim(temp(2)));

    Meta.Basin = string(basin);
    Meta.BasinCode = basincode;

    temp = strsplit(char(SDS(contains(SDS,'Commence:')==1)),':');
    Meta.StartDate = datetime(char(strtrim(temp(2))),'InputFormat','dd/MM/yyyy');

    temp = strsplit(char(SDS(contains(SDS,'Cease:')==1)),':');
    temp = char(strtrim(temp(2)));
    if isempty(temp)
        Meta.EndDate = datetime(2019,1,1);
    else
        Meta.EndDate = datetime(temp,'InputFormat','dd/MM/yyyy');
    end

    Meta.Years = round(years(Meta.EndDate-Meta.StartDate));

    % Eating Northing Zone
    temp = strsplit(char(SDS(contains(SDS,'Grid Ref:')==1)));
    if isempty(char(temp))
        Meta.Zone = NaN;
        Meta.East = NaN;
```

```
        Meta.North = NaN;
    else
        Meta.Zone = str2double(temp(4));
        Meta.East = str2double(temp(6));
        Meta.North = str2double(temp(8));
    end

    % Latitude
    temp = strsplit(char(SDS(contains(SDS,'Latitude:')==1)));
    if isempty(char(temp))
        Meta.Lat = NaN;
    else
        Meta.Lat = str2double(temp(2));
    end

    % Longitude
    temp = strsplit(char(SDS(contains(SDS,'Longitude:')==1)));
    if isempty(char(temp))
        Meta.Long = NaN;
    else
        Meta.Long = str2double(temp(2));
    end

    % Elevation
    temp = strsplit(char(SDS(contains(SDS,'Elevation:')==1)));
    if isempty(char(temp))
        Meta.Elev = NaN;
    else
        Meta.Elev = str2double(temp(2));
    end

    % Stream Distance from station to mouth in km
    temp = strsplit(char(SDS(contains(SDS,'Stream Distance:')==1)));
    if isempty(char(temp))
        Meta.SDist = NaN;
    else
        Meta.SDist = str2double(temp(3));
    end

    % Control
    temp = strsplit(char(SDS(contains(SDS,'Control:')==1)),':');
    if isempty(char(temp))
        Meta.Control = NaN;
    else
        Meta.Control = string(strtrim(temp(2)));
    end

    % Catchment Area
    temp = strsplit(char(SDS(contains(SDS,'Catchment Area:')==1)));
    if isempty(char(temp))
        Meta.Area = NaN;
    else
        Meta.Area = str2double(temp(3));
    end

```

```
    % Max gauaged flow
    if isempty(Gaugings)
        Meta.MaxGaugeStage = NaN;
        Meta.MaxGaugeFlow = NaN;
        Meta.MaxGaugeDate = NaT;
    else
        Meta.MaxGaugeStage = max(Gaugings.Stage);
        Meta.MaxGaugeFlow = max(Gaugings.Flow);
        idx = find(Gaugings.Stage == max(Gaugings.Stage));
        Meta.MaxGaugeDate = Gaugings.Date(idx(1));
    end

    % Max Record Stage
    idx = find(strcmp(SDS,'PERIOD OF RECORD'));
    if isempty(idx)
        Meta.MaxRecordStage = NaN;
    else
        tst = 1;
        n = 0;
        while tst == 1 && idx < length(SDS)
            if contains(SDS(idx),'Level')
                tst = 0;
                temp = strsplit(char(SDS(idx)));
                Meta.MaxRecordStage = str2double(temp(9));
            else
                Meta.MaxRecordStage = NaN; % can't find row
            end
            idx = idx+1;
            n = n+1;
        end
    end

    % Max Record Flow
    idx = find(strcmp(SDS,'PERIOD OF RECORD'));
    if isempty(idx)
        Meta.MaxRecordFlow = NaN;
    else
        tst = 1;
        n = 0;
        while tst == 1 && idx < length(SDS)
            if contains(SDS(idx),'Discharge')
                tst = 0;
                temp = strsplit(char(SDS(idx)));

                Meta.MaxRecordFlow = str2double(temp(9));
            else
                Meta.MaxRecordFlow = NaN; % can't find row
            end
            idx = idx+1;
            n = n+1;
        end
    end

    % Max Record Date
    idx = find(strcmp(SDS,'PERIOD OF RECORD'));
```

```matlab
    if isempty(idx)
        Meta.MaxRecordDate = NaT;
    else
        tst = 1;
        n = 0;
        while tst == 1 && idx < length(SDS)
            if contains(SDS(idx),'Level')
                tst = 0;
                temp = strsplit(char(SDS(idx)));
                Meta.MaxRecordDate = datetime(char(strtrim(temp(10))),'InputFormat','dd/MM/yyyy');
            else
                Meta.MaxRecordDate = NaT; % can't find row
            end
            idx = idx+1;
            n = n+1;
        end
    end

    % Minimum Peak Discharge
    temp = strsplit(char(SDS(contains(SDS,'Min Peak Discharge:')==1)));
    if isempty(char(temp))
        Meta.MinPeakFlow = NaN;
    else
        Meta.MinPeakFlow = str2double(temp(4));
    end

    % Time between Peaks
    temp = strsplit(char(SDS(contains(SDS,'Time between Peaks:')==1)),':');
    if isempty(char(temp))
        Meta.TimeBtwPeaks = "";
    else
        Meta.TimeBtwPeaks = string(strtrim(temp(2)));
    end

    % Downstream from Dam
    temp = strsplit(char(SDS(contains(SDS,'from Dam:')==1)),':');
    if isempty(char(temp))
        Meta.Dam = "";
    else
        Meta.Dam = string(strtrim(temp(2)));
    end

    % Zero Gauge
    temp = strsplit(char(SDS(contains(SDS,'Zero Gauge:')==1)));
    if isempty(char(temp))
        Meta.ZeroGauge = NaN;
    else
        Meta.ZeroGauge = str2double(temp(3));
    end

    % Comment
    idx = find(contains(SDS,'Comment:'));
    if isempty(idx)
        Meta.Comment = NaN;
```

```
    else
        tst = 1;
        n = 0;
        tsttxt = {'STATION DESCRIPTION','PERIOD OF RECORD'};
        while tst == 1 && n <10
            if contains(SDS(idx),tsttxt)
                tst = 0;
            elseif n == 9
                Meta.Comment = 'too many comments';
            elseif n == 0
                temp = strsplit(char(SDS(idx)),':');
                temp = char(strtrim(temp(2)));
            else
                temp = [temp,' ', char(strtrim(char(SDS(idx))))];
            end
            idx = idx+1;
            n = n+1;
        end
        temp = string(temp);
        Meta.Comment = temp;
    end

end

%% FUNCTION: Extract Rating Table
function RTout = ExtRT(path)
    % open file
    fRT = fopen([path,'\Rating Table.txt'],'r');
    RT = textscan(fRT,'%s','Delimiter','\n');
    fclose(fRT);
    RT = RT{1,1};

    % Remove empty rows and repeated titles
    RT = RT(~cellfun('isempty',RT'));
    RT = regexprep(RT,'Page \d+','Page 1'); % remove page numbers
    RT = unique(RT,'stable');

    % Turn into matrix
    inds = find(contains(RT,'G.H.'))+2;  % at the moment overlooking negative stages
    inde = find(contains(RT,'Notes'));

    RTout = table('Size',[(inde-inds-1)*10,3],...
        'VariableTypes',{'double','double','cellstr'},...
        'VariableNames',{'Stage','Flow','Tags'});
    for k = 1:inde-inds-1
        RTtxt = strsplit(char(RT(k + inds)));

        RTchar = regexp(RTtxt,'([a-z_A-Z]+)','match');
        RTchar = [RTchar(2:end)]';

        RTflow = regexp(RTtxt,'([+-]?\d+.\d+)','match');
        RTflow = str2double([RTflow{:}]);
        RTflow = [RTflow'; zeros(11-size(RTflow,2),1)];

```

```
        RTstage = RTflow(1):0.1:RTflow(1)+0.9;

        RTidx = (k-1)*10+1;
        RTout.Stage(RTidx:RTidx + 9) = RTstage';
        RTout.Flow(RTidx:RTidx + 9) = RTflow(2:end);
        RTout.Tags(RTidx:RTidx + size(RTchar,1)-1) = RTchar;
    end
    % remove zeros from end
    RTmaxidx = find(RTout.Flow == max(RTout.Flow));
    RTout(RTmaxidx+1:end,:) = [];
end
```

B.2 SelectedDataExtraction

Name: SelectedDataExtraction.m

Type: Script

Details:

Looks at index file for selected gauges and extracts data, copies files, and graphs to new location. Summaries all data.

```
% Reads and writes daily and monthly data to better format. Copies other
% files of each gauge. Only does for list of selected sites.
%
% Getting ready for data preparation
%
% Created: Sam Walker
% Date: 28/03/2019
% Version: 1.0

tic
clear
format compact

%% DIRECTORIES   (fill out)
p = pwd;

% List of gauges selection
plist = [p(1:3),'1. Data Collection\Gauges Selection\Gauges Selection - v1.1.xlsx'];

% List of all files
pAllFiles = [p(1:3),'1. Data Collection\WMIP Extracted\All - Files Summary.mat'];

% Copy files from data extraction process
pcopy = [p(1:3), '1. Data Collection\WMIP Extracted'];

% Desitination folder
%%% Always save to new folder then manually drag out and pick merge or
%%% overwrite through windows explorer!
psave = [p(1:3),'2. Data Preparation\Selection v1.1'];

%% Delete contents of destination
pd = dir(psave);
for pdk = 3:length(pd)
    if pd(pdk).isdir == 1
        rmdir([pd(pdk).folder,'\',pd(pdk).name],'s')
    else
        delete([pd(pdk).folder,'\',pd(pdk).name])
    end
end


%% DATA EXTRACTION & WRITING
Gauges = load(pAllFiles);
Gauges = Gauges.OutGauge;
List = readtable(plist,'Sheet','Selection');
for k = 1:size(List,1)
    site = char(List{k,1});
    m = find(strcmp(Gauges.Gauge, site));
    fprintf('No.: %d\t Processing:\t %s\n',k,site)

    %% Monthly Data
    % Read file
    praw = char(Gauges.FilePath(m));
    praw = [p(1:3),praw(4:end)];
```

```
pday = [praw,'\cf',site,' monthly\',site,'.csv'];
MD = readtable(pday,'Delimiter',',','ReadVariableNames',false);
MonthData = table;

% Date
MonthData.Date = datetime(MD{5:end-2,1},'InputFormat','ss:mm:hh dd/MM/yyyy');

% Stage
a = find(contains(MD{3,:},'Level'));
b = find(contains(MD{4,:},'Max'));
if ~isempty(a) || ~isempty(b)
    col = b(ismember(b,a));
    MonthData.Stage = str2double(MD{5:end-2,col});
    MonthData.StageQuality = str2double(MD{5:end-2,col+1});
end

% Flow
a = find(contains(MD{3,:},'Discharge'));
b = find(contains(MD{4,:},'Max'));
if ~isempty(a) || ~isempty(b)
    col = b(ismember(b,a));
    MonthData.Flow = str2double(MD{5:end-2,col});
    MonthData.FlowQuality = str2double(MD{5:end-2,col+1});
end

% Rain
a = find(contains(MD{3,:},'Rainfall'));
b = find(contains(MD{4,:},'Total'));
if ~isempty(a) || ~isempty(b)
    col = b(ismember(b,a));
    MonthData.Rain = str2double(MD{5:end-2,col});
    MonthData.RainQuality = str2double(MD{5:end-2,col+1});
end

%% Daily Data
% Read file
pday = [char(Gauges.FilePath(m)),'\cf',site,' daily\',site,'.csv'];
DD = readtable(pday,'Delimiter',',','ReadVariableNames',false);
DayData = table;

% Date
DayData.Date = datetime(DD{5:end-2,1},'InputFormat','ss:mm:hh dd/MM/yyyy');

% Stage
a = find(contains(DD{3,:},'Level'));
b = find(contains(DD{4,:},'Max'));
if ~isempty(a) || ~isempty(b)
    col = b(ismember(b,a));
    DayData.Stage = str2double(DD{5:end-2,col});
    DayData.StageQuality = str2double(DD{5:end-2,col+1});
end

% Flow
a = find(contains(DD{3,:},'Discharge'));
b = find(contains(DD{4,:},'Max'));
```

```matlab
    if ~isempty(a) || ~isempty(b)
        col = b(ismember(b,a));
        DayData.Flow = str2double(DD{5:end-2,col});
        DayData.FlowQuality = str2double(DD{5:end-2,col+1});
    end

    % Rain
    a = find(contains(DD{3,:},'Rainfall'));
    b = find(contains(DD{4,:},'Total'));
    if ~isempty(a) || ~isempty(b)
        col = b(ismember(b,a));
        DayData.Rain = str2double(DD{5:end-2,col});
        DayData.RainQuality = str2double(DD{5:end-2,col+1});
    end

    %% WRITE TO FILE
    mkdir([psave,'\',site])

    % Write MonthData
    save([psave,'\',site,'\',site,' - Monthly Data.mat'],'MonthData')
    writetable(MonthData,[psave,'\',site,'\',site,' - Monthly Data.txt'])

    % Write DayData
    save([psave,'\',site,'\',site,' - Daily Data.mat'],'DayData')
    writetable(DayData,[psave,'\',site,'\',site,' - Daily Data.txt'])

    % Copy files
    copyfile([pcopy,'\',site,'\*'],[psave,'\',site])
    if Gauges.CrossSection(m) == 1
        copyfile([praw,'\Cross Section.png'],[psave,'\',site])
    end
    if Gauges.RatingCurve(m) == 1
        copyfile([praw,'\Rating Curve.png'],[psave,'\',site])
    end
    if Gauges.RatingCurveWithData(m) == 1
        copyfile([praw,'\Rating Curve with Data.png'],[psave,'\',site])
    end

    clear('DD','DayData','MD','MonthData') % to make sure nothing hangs over loops
end

toc
```

B.3 RatingCurveFitting_138002ABC

Name: RatingCurveFitting_138002ABC.m

Type: Script

Details:

Fits a spline to TUFLOW model results and DNRME (only for small floods). Plots rating curve. Input data is manual copied into the script. Uses ReduceDensity.m to remove some points in the TUFLOW output so spline fits better.

Shown is an example of gauge 138002ABC. Saved into each gauge folder.

```matlab
clear
clc
close all

%% INFORMAION
% Gauge: 138002ABC - Wide Bay Creek at Brooyar
% Tuflow model completed by Sam Walker.

%% DATA 1-09-2019
% Data was copied into variables then copied here.

% [Stage, Flow]
DNRME = [0,0;0.100000000000000,0;...];

% [Flow, Level mAHD]
TUFLOW = [0,43.9618000000000;0,43.9618000000000;...];

ZeroGauge = 42.373;

% INPUTS
H_change_d = 5.5;     % change point from DNRME
H_change_t = 5.6;     % change point to Tuflow

Limbs = 3;            % 1= upper limb ; 2= lower limb ; 3= both limbs
SplineFit = 0.8;    % modify how tight spline is fit to data
GraphTitle = 'Gauge 138002ABC';

% Stage AM Series
Stage = [7.09000000000000;5.18000000000000;2.59000000000000;...];

% Gaugings
Gaugings_A = {'05-May-1909 00:00:00',0.300000000000000,0.0520000000000000,9999.99000000000;'30-Oct-1909 00:00:00',0.330000000000000,0.0570000000000000,9999.99000000000;...];
Gaugings_B = {'27-May-1962 12:30:00',0.570000000000000,0.0790000000000000,1.82000000000000;'19-Jul-1962 10:00:00',0.580000000000000,0.102000000000000,10.5200000000000;...];
Gaugings_C = {'07-Aug-1967 16:20:00',0.920000000000000,0.620000000000000,1.38000000000000;'19-Oct-1967 13:00:00',0.760000000000000,0.0670000000000000,-1.39000000000000;...];

H_g_A = [Gaugings_A{:,2}]';
Q_g_A = [Gaugings_A{:,3}]';
```

```
H_g_B = [Gaugings_B{:,2}]';
Q_g_B = [Gaugings_B{:,3}]';
H_g_C = [Gaugings_C{:,2}]';
Q_g_C = [Gaugings_C{:,3}]';

%% ASSIGN VARIABLES
H_d = DNRME(:,1);
Q_d = DNRME(:,2);

H_t = TUFLOW(:,2) - ZeroGauge;
Q_t = TUFLOW(:,1);
H_t_all = H_t;
Q_t_all = Q_t;

figure(1)
plot(Q_d,H_d,Q_t_all,H_t_all)
legend('DNRME','TUFLOW','Location','southeast')
xlabel('Discharge (cumecs)')
ylabel('Stage (m)')
title([GraphTitle,' Compare DNRME and TUFLOW model'])

%% Adopt DNRME for low flows and TUFLOW or large flows.
idx_maxQ_t = find(Q_t == max(Q_t));
if Limbs == 1
    Q_t = Q_t(idx_maxQ_t:end);
    H_t = H_t(idx_maxQ_t:end);
elseif Limbs == 2
    Q_t = Q_t(1:idx_maxQ_t);
    H_t = H_t(1:idx_maxQ_t);
end

H_large = H_t(H_t >= H_change_t);
Q_large = Q_t(H_t >= H_change_t);

%% FITTING DNRME
[xData, yData] = prepareCurveData( H_d, Q_d);
ft = 'linearinterp';
[fitDNRME, gof_DNRME] = fit( xData, yData, ft, 'Normalize', 'on'
);

%% FITTING TUFLOW
[xData, yData] = prepareCurveData( H_large, Q_large );

% Set up fittype and options.
```

```
ft = fittype( 'smoothingspline' );
opts = fitoptions( 'Method', 'SmoothingSpline' );
opts.Normalize = 'on';

% Adjust to smooth or roughen spline fit.
opts.SmoothingParam = SplineFit;

% Fit model to data.
[fitTUFLOW, gof_TUFLOW] = fit( xData, yData, ft, opts );

% Plot fit with data.
figure(10);
subplot( 2, 1, 1 );
h = plot( fitTUFLOW, xData, yData );
legend( h, 'TUFLOW model', 'Fitted Spline', 'Location', 'SouthEast' );
title('Fitting to TUFLOW model')
% Label axes
xlabel Stage
ylabel Flow
grid on

% Plot residuals.
subplot( 2, 1, 2 );
h = plot( fitTUFLOW, xData, yData, 'residuals' );
legend( h, 'TUFLOW model - residuals', 'Zero Line', 'Location', 'NorthEast' );
title('Fitting to TUFLOW model - Residuals')
% Label axes
xlabel Stage
ylabel('Flow residuals')
grid on

%% FITTING MID
% To get smooth join between DNRME and TUFLOW curves use 4 points to fit a
% cubic. To make smooth find point at 5% of gap to use for fit.

gap = (H_change_t - H_change_d).*0.05; % make smooth by 5% of gap

H_mid = [H_change_d, H_change_d+gap, H_change_t-gap, H_change_t];
Q_mid = [fitDNRME([H_change_d, H_change_d+gap]);fitTUFLOW([H_change_t-gap, H_change_t])]';

```

```matlab
[xData, yData] = prepareCurveData( H_mid, Q_mid );

% Set up fittype and options.
ft = fittype( 'poly3' );

% Fit model to data.
[fitMID, gof_MID] = fit( xData, yData, ft );

%% FINALISE
H = 0:0.01:max(H_large)*1.01;
Q = [fitDNRME(H(H <= H_change_d));...
    fitMID(H(H > H_change_d & H < H_change_t));...
    fitTUFLOW(H(H >= H_change_t))];

figure(2)
plot(Q_d,H_d,Q_t_all,H_t_all,Q,H,'k')
legend('DNRME','TUFLOW','Adopted','Location','southeast')
xlabel('Discharge (cumecs)')
ylabel('Stage (m)')
title([GraphTitle,' Rating Curve'])

%% CONVERT DATA
Flow(Stage <= H_change_d) = fitDNRME(Stage(Stage <= H_change_d));
Flow(Stage > H_change_d & Stage < H_change_t) = fitMID(Stage(Stage > H_change_d & Stage < H_change_t));
Flow(Stage >= H_change_t) = fitTUFLOW(Stage(Stage >= H_change_t));
Flow = Flow';

Flow_DNRME = fitDNRME(Stage);

figure(4)
plot(Q_g_A,H_g_A,'.',Q_g_B,H_g_B,'.',Q_g_C,H_g_C,'.','MarkerSize',12), hold on
plot(Q_d,H_d,Q_t_all,H_t_all,Q,H,'k',Flow,Stage,'xg')
legend('Gaugings A','Gaugings B', 'Gaugings C','DNRME','TUFLOW','Adopted','AM series','Location','southeast')
xlabel('Discharge (cumecs)')
ylabel('Stage (m)')
title([GraphTitle,' Rating Curve w. AM Series'])


%% SEE 'Flow' VARIABLE FOR AM SERIES.
%% SEE 'Flow_DNRME' VARIABLE FOR AM SERIES BASED ON RATING CURVE PROVIDED.
```

B.4 ReduceDensity

Name: ReduceDensity.m

Type: Function

Details:

Used to remove additional points within a range.

```
function [I,C] = ReduceDensity(X,method,d)
% Filters out points in X based.
%   X = vector to be filtered
%   method = specifies which method to filter
%       "dim" will keep the closest point to regular spacing at d.
%           d vector is measured from min(X)to max(X)
%       "count" will keep one in every d values in X

if method == "dim"
    % Spacing dim
    % Point closest to dim is kept
    % dim vector is measured from min(X)to max(X).
    J = zeros(size(X));
    for v = min(X):d:max(X)
        [~,a] = min(abs(X-v));
        J(a) = 1;
    end
    I = find(J);
elseif method == "count"
    % Spacing count
    % One in evey count points is kept
    I = 1:d:length(X);
else
    I = 0;
    warning('ReduceDensity did not run')
    return
end
C = X(I);

end
```

B.5 ReadFlike.m

Name: ReadFlike.m

Type: Script

Details:

Function is described in report main body.

Reads FLIKE output files and converts into useful format. Also performs test and creates plots to review the best fit distribution.

Uses FitFlike.m to fit distributions to get additional quantiles and probabilities and xlcolnum2.m for saving to excel.

```matlab
% Reads Flike files and produces Kolmogorov-Smirnov test, QQ plot, PP plot
% for a site.
%
%
% Created: Samuel J Walker
% Date: 14/08/2019
% Version: 2.2

clear
clc
close all
format compact
format long

%% INPUTS
WriteToFile = 0; % 1=write to file; 0=DO NOT write to file
PrintCurveFit = 0; % Plot curve fitting to the extimated quantiles (in order to estmiate any quanitle/AEP)
PrintFigs = 1; % Plot (and write) PP plots, QQ plots, Frequency Curve, CDF

Station = '915011A';

Selection = [15,16,18];
PrintSelection = [15,16,18];

%% Folders and files
p = pwd;
pread = [p(1:3), '2. Data Prep & FFA\', Station, '\FLIKE'];

psavedir = [p(1:3), '2. Data Prep & FFA\', Station, '\FFA\Record'];
psave = [psavedir,'\',Station, ' - '];

%% Delete contents of destination
if WriteToFile == 1
    pd = dir(psavedir);
    if size(pd,1) > 2
        warning('  Save folder has files in it.')

        askCon = input('Do you want to stop? [Y/N]   ','s');
        while ~any(askCon == ["Y", "y", "N", "n"])
            askCon = input('Do you want to stop? [Y/N]   ','s');
        end
        if any(askCon == ["Y", "y"])
            return
        end

        askDel = input('Do you want to delete them? [Y/N]   ','s');
        while ~any(askDel == ["Y", "y", "N", "n"])
            askDel = input('Do you want to delete them? [Y/N]   ','s');
        end
        if any(askDel == ["Y", "y"])
            for pdk = 3:length(pd)
                if pd(pdk).isdir == 1
                    rmdir([pd(pdk).folder,'\',pd(pdk).name],'s')
                else
```

```
                    delete([pd(pdk).folder,'\',pd(pdk).name])
                end
            end
        end
    end
end

%% READ FILE
pfiles = dir(pread);
m = 1;
Files = {};
for k = 1:size(pfiles,1)
    pfile = pfiles(k).name;
    if contains(pfile,'.txt')
        Files(m,:) = {m, pfile};
        m = m+1;
    end
end
disp('Files available:')
disp(Files)
disp('  ')
disp('Select files to process: ...')

if isempty(Selection)
    return
end

%% CALCULATIONS
disp(Selection)
mm = 1;
for kk = Selection
    pfile = char(Files(kk,2))
    % open file
    fread = fopen([pread,'\',pfile],'r');
    flike = textscan(fread,'%s','Delimiter','\n');
    fclose(fread);
    flike = flike{1,1};

    %% READ FILE
    if contains(pfile,'GEV')
        % GEV Extraction
        [Data, FitDist, shift, StartPoints] = ReadGEV(flike);
        Type = {'GEV Q-P', 'GEV P-Q'};
        NoOutliers = 0;
    else
        % LP3 Extraction
        [Data, FitDist, NoOutliers, Cen, StartPoints] = ReadLP3(flike);
        Type = {'LP3 Q-P', 'LP3 P-Q'};
        shift = 0;
    end

    %% FIT THROUGH EXPECT QUANTILES
    [fitQP, gofQP] = FitFLIKE(FitDist.AEP,FitDist.ExpFlow,char(Type(1)), StartPoints, PrintCurveFit);
    % NOTE: if LP3 the input to fitQP has to be log(Q)
```

```
    [fitPQ, gofPQ] = FitFLIKE(FitDist.AEP,FitDist.ExpFlow,char(Type(2)),StartPoints, PrintCurveFit);

    %% TESTING
    % Fitted estimates
    n = size(Data,1);
    Data.EstQ = fitPQ(Data.PP);
    % NOTE: if LP3 the input to fitQP has to be log(Q)
    if contains(pfile,'GEV')
        Data.EstAEP = fitQP(Data.Flow);
    else
        Data.Flow(Data.Flow < 0.01) = 0.01; % so not to get -Inf when log
        Data.EstAEP = fitQP(log(Data.Flow));
    end

    %% Kolmogorov-Smirnov test
    KSrank = fliplr(Data.Rank');
    KSdata = fliplr(Data.Flow');
    KSpp = fliplr(Data.PP');
    KSest = fliplr(Data.EstAEP');

    KSp = KSpp-KSest;
    KSn = KSest(2:end)-KSpp(1:end-1);
    KS = max([KSp,KSn]);

    % KS test from 50% to 1%
    KSrank2 = KSrank(KSpp >= 0.5);
    KSdata2 = KSdata(KSpp >= 0.5);
    KSpp2 = KSpp(KSpp >= 0.5);
    KSest2 = KSest(KSpp >= 0.5);

    KSp2 = KSpp2-KSest2;
    KSn2 = KSest2(2:end)-KSpp2(1:end-1);
    KS2y = max([KSp2,KSn2]);

    %% PLOTS
    if PrintFigs == 1
        % Plot against data, with confidence limits
        figh(1) = figure('Name',['Frequency Curve: ', pfile]);
        semilogy(norminv(Data.PP,0,1),Data.Flow,'ob'), hold on
        semilogy(norminv(FitDist.AEP,0,1),FitDist.ExpFlow,'-r')
        semilogy(norminv(FitDist.AEP,0,1),FitDist.CL5,'--r')
        semilogy(norminv(FitDist.AEP,0,1),FitDist.CL95,'--r')
        grid on
        title(['Frequency Curve: ', pfile])
        ylabel('Flow (cumecs)')

        % AEP %
        Xtick = [95,80,63.2,50,20,10,5,2,1,0.5,0.1,0.01];
        xticks(norminv(1-Xtick./100))
        xticklabels(Xtick)
        xtickangle(90)
        xlabel('Annual Excedance Probability (%)')

```

```
%     % AEP 1 in
%     Xtick = [1,1.44,2,5,10,20,50,100,200,500];
%     xticks(norminv(1-1./Xtick))
%     xtickangle(90)
%     xticklabels(Xtick)
%     xlabel('Annual Excedance Probability (1 in X)')

    hold off

    % PP Plot
    figh(2) = figure('Name',['PP Plot: ', pfile]);
    plot([0,1],[0,1],'-k',Data.PP,Data.EstAEP,'o')
    title(['PP Plot: ', pfile])
    xlabel('Data Probability')
    ylabel('Fitted Distribution Probability')

    % QQ Plot
    maxQQ = max([Data.Flow;Data.EstQ],[],'all');
    figh(3) = figure('Name',['QQ Plot: ', pfile]);
    plot([0,maxQQ],[0,maxQQ],'-k',Data.Flow,Data.EstQ,'o')
    title(['QQ Plot: ', pfile])
    xlabel('Data Quantile')
    ylabel('Fitted Distribution Quantile')

    % CDF comparison
    EDF(KSrank.*2-1) = KSpp;
    EDF(KSrank(1:end-1).*2) = KSpp(1:end-1);
    q(KSrank.*2-1) = KSdata;
    q(KSrank(1:end-1).*2) = KSdata(2:end);

    figh(4) = figure('Name',['Empicial and Model CDF: ', pfile]);
    plot(q,EDF), hold on
    plot(KSdata,KSest)
    legend('Empirical','Model','Location','southeast')
    axis([0, inf, 0, 1])
    title(['Empicial and Model CDF: ', pfile])
    xlabel('Flow (cumecs)')
    ylabel('Cummulative Probability')
    hold off
end

%% GROUP INTO STRUCTURE
FFAresults(mm).Index = mm;
FFAresults(mm).FileNo = kk;
FFAresults(mm).Name = pfile;
FFAresults(mm).Data = Data;
FFAresults(mm).Outliers = NoOutliers;
FFAresults(mm).Shift = shift;
FFAresults(mm).FitDist = FitDist;

if contains(pfile,'GEV')
    % GEV parameters
    para = coeffvalues(fitPQ);
    FFAresults(mm).GEVa = para(1);
    FFAresults(mm).GEVk = para(2);
```

```
        FFAresults(mm).GEVt = para(3);

        % LP3 paramters
        FFAresults(mm).LP3m = 0;
        FFAresults(mm).LP3s = 0;
        FFAresults(mm).LP3g = 0;
    else
        % GEV parameters
        FFAresults(mm).GEVa = 0;
        FFAresults(mm).GEVk = 0;
        FFAresults(mm).GEVt = 0;

        % LP3 paramters
        para = coeffvalues(fitPQ);
        FFAresults(mm).LP3m = para(2);
        FFAresults(mm).LP3s = para(3);
        FFAresults(mm).LP3g = para(1);
    end
    FFAresults(mm).KStest = KS;
    FFAresults(mm).KStest2y = KS2y;

    %% WRITE TO FILE
    if WriteToFile == 1
        %% All Data
        colnum = (mm-1)*8+1;
        colstr = xlcolnum2alpha(colnum);

        xlswrite([psave, 'Flike Comparison.xlsx'],{pfile},'Data',[colstr, '2'])
        writetable(Data,[psave, 'Flike Comparison.xlsx'],'Sheet','Data','Range',[colstr, '3'])
        %% All FitDist
        colnum = (mm-1)*6+1;
        colstr = xlcolnum2alpha(colnum);

        xlswrite([psave, 'Flike Comparison.xlsx'],{pfile},'Quantiles',[colstr, '2'])
        writetable(FitDist,[psave, 'Flike Comparison.xlsx'],'Sheet','Quantiles','Range',[colstr, '3'])
        %% Write Plots
        % Fitting curve residual ...

        savefig(figh,[psave, pfile(1:end-4),' - graphs.fig']);

        print(figh(1),[psave, pfile(1:end-4),' - Frequency Curve.png'],'-dpng','-r300');
        print(figh(2),[psave, pfile(1:end-4),' - PP plot.png'],'-dpng','-r300');
        print(figh(3),[psave, pfile(1:end-4),' - QQ plot.png'],'-dpng','-r300');
        print(figh(4),[psave, pfile(1:end-4),' - CDF.png'],'-dpng','-r300');
    end
    %% Reset loop
    % close all
    mm = mm+1;
    clear pfile fread flike Data FitDist shift StartPoints ...
        Type NoOutliers Cen fitQP gofQP fitPQ gofPQ n Rank ...
        AD KS CVM maxQQ colnum colstr figh EDF q;
```

```
end
%% Comparison of all
FFAresults = struct2table(FFAresults);

[~,a] = sort([FFAresults.KStest],'ascend');
temp(a) = 1:length(a);
FFAresults.KSrank = temp';

[~,b] = sort([FFAresults.KStest2y],'ascend');
temp(b) = 1:length(b);
FFAresults.KSrank2y = temp';

% Test comparison
figh_bar(1) = figure('Name','Kolmogorov-Smirnov Test');
hbar = bar([FFAresults.KStest,FFAresults.KStest2y]);
title('Kolmogorov-Smirnov Test for each model')
legend('All data', 'Greater-equal to 50% event','Location','northeast')
set(gca, 'Xticklabel', [])
text(1:length(Selection),zeros(1,length(Selection)),num2cell(Selection,1),'HorizontalAlignment','center','VerticalAlignment','top')


figh_bar(2) = figure('Name','Goodness-Of-Fit Rank (Kolmogorov-Smirnov Test)');
bar([FFAresults.KSrank,FFAresults.KSrank2y])
title('KoGoodness-Of-Fit Rank (Kolmogorov-Smirnov Test)')
legend('All data', 'Greater-equal to 50% event','Location','northeast')
set(gca, 'Xticklabel', [])
text(1:length(Selection),zeros(1,length(Selection)),num2cell(Selection,1),'HorizontalAlignment','center','VerticalAlignment','top')

% To write or Not to write
if WriteToFile == 1
    writetable(FFAresults(:,[1,2,4,5,7:end]),[psave, 'Flike Comparison.xlsx'],'Sheet','Summary','Range','A2')
    save([psave, 'Flike Comparison.mat'],'FFAresults')

    % bar graph
    savefig(figh_bar(1),[psave,'Kolmogorov-Smirnov Test.fig']);
    print(figh_bar(1),[psave,'Kolmogorov-Smirnov Test.png'],'-dpng','-r300');

    % bar graph rank
    savefig(figh_bar(2),[psave,'Kolmogorov-Smirnov Test Ranking.fig']);
    print(figh_bar(2),[psave,'Kolmogorov-Smirnov Test Ranking.png'],'-dpng','-r300');
end

%% Comparison Graphs
if isempty(PrintSelection)
    return
end

figh_PP = figure('Name','PP Plot Comparison');

figh_QQ = figure('Name','QQ Plot: 3 best + Worst');

```

```
figh_FC = figure('Name','Frequency Curve Comparison');
semilogy(norminv(FFAresults.Data{1,1}.PP,0,1),FFAresults.Data{1,1}.Flow,'o')

idx = find(any(PrintSelection == FFAresults.FileNo,2));
maxQQ = 0;
for kkk = idx'
    % Frequency Plot
    figure(figh_FC)
    hold on
    semilogy(norminv(FFAresults.Data{kkk,1}.PP,0,1),FFAresults.Data{kkk,1}.EstQ,'-')

    % PP plot
    figure(figh_PP)
    hold on
    plot(FFAresults.Data{kkk,1}.PP,FFAresults.Data{kkk,1}.EstAEP,'-o')

    % QQ plot
    maxQQ = max([FFAresults.Data{kkk,1}.Flow;FFAresults.Data{kkk,1}.EstQ;...
        maxQQ],[],'all');

    figure(figh_QQ)
    hold on
    plot(FFAresults.Data{kkk,1}.Flow,FFAresults.Data{kkk,1}.EstQ,'-o')

    % Create legend names
    lgnd{kkk} = ['[',mat2str(FFAresults.FileNo(kkk)), '] ', FFAresults.Name{kkk}];
end

figure(figh_FC)
hold on
set(gca, 'YScale', 'log');
grid on
title('Frequency Curves Comparison')
ylabel('Flow (cumecs)')
legend(['Data',lgnd], 'Location','southeast')
Xtick = [95,80,63.2,50,20,10,5,2,1,0.5,0.1,0.01];
xticks(norminv(1-Xtick./100))
xticklabels(Xtick)
xtickangle(90)
xlabel('Annual Excedance Probability (%)')
axis([norminv(0.005), norminv(1-0.005), 1, max(FFAresults.Data{1,1}.Flow)*1.1])
hold off

figure(figh_PP)
hold on
plot([0,1],[0,1],'-k')
grid on
title('PP Plot Comparison')
xlabel('Data Probability')
ylabel('Fitted Distribution Probability')
legend(lgnd, 'Location','southeast')
axis([0,1,0,1])
hold off

```

```
figure(figh_QQ)
hold on
plot([1,maxQQ],[1,maxQQ],'-k')
%set(gca, 'XScale', 'log', 'YScale', 'log')
grid on
title('QQ Plot Comparison')
xlabel('Data Quantile')
ylabel('Fitted Distribution Quantile')
legend(lgnd, 'Location','southeast')
axis([0,maxQQ,0,maxQQ])
hold off

% To write or Not to write
if WriteToFile == 1
    savefig(figh_FC,[psave,'Frequency Curves Comparison.fig']);
    print(figh_FC,[psave,'Frequency Curves Comparison.png'],'-dpng','-r300');

    savefig(figh_PP,[psave,'PP Plot Comparison.fig']);
    print(figh_PP,[psave,'PP Plot Comparison.png'],'-dpng','-r300');

    savefig(figh_QQ,[psave,'QQ Plot Comparison.fig']);
    print(figh_QQ,[psave,'QQ Plot Comparison.png'],'-dpng','-r300');
end


%% GEV Extraction
function [Data, FitDist, shift, StartPoints] = ReadGEV(flike)

    % LH moment shift
    ishift = find(contains(flike,{'Optimized L moment shift','Nominated L momemt shift'}));
    shift = strsplit(char(flike(ishift)),'= ');
    shift = str2double(shift(2));

    % Data
    isdata = find(contains(flike,'Rank'))+2;
    iedata = find(contains(flike,'Value'))-2;
    Data = str2double(strsplit(char(flike(isdata)),' '));
    for kkk = 2:iedata-isdata+1
        Data = [Data; str2double(strsplit(char(flike(isdata+kkk-1)),' '))];
    end
    Data = [max(Data(:,1))-Data(:,1)+1,Data(:,1:3), 1-1./Data(:,3)];
    Data = array2table(flipud(Data),'VariableNames',{'Observation','Rank','Flow','PPyrs','PP'});

    % Expected Quantiles
    isquant = find(contains(flike,'AEP 1 in Y'))+2;
    iequant = length(flike);
    FitDist = str2double(strsplit(char(flike(isquant)),' '));
    for kkk = 2:iequant-isquant+1
        FitDist = [FitDist; str2double(strsplit(char(flike(isquant+kkk-1)),' '))];
    end
    FitDist = [FitDist(:,1), 1-1./FitDist(:,1), FitDist(:,2:end-1)];
    FitDist = array2table(FitDist,'VariableNames',{'AEPyrs','AEP','ExpFlow','CL5','CL95'});
```

```

    % Parameters
    iPara = find(contains(flike,'Parameter'))+2;
    alpha = str2double(strsplit(char(flike(iPara+1))));
    alpha = alpha(3);
    kappa = str2double(strsplit(char(flike(iPara+2))));
    kappa = kappa(3);
    tau = str2double(strsplit(char(flike(iPara))));
    tau = tau(3);
    StartPoints = [alpha,kappa,tau];
end

%% LP3 Extraction
function [Data, FitDist, NoOutliers, Cen, StartPoints] = ReadLP3(flike)
    % Data
    isdata = find(contains(flike,'Gauged Annual Maximum Discharge Data'))+4;
    Data = str2double(strsplit(char(flike(isdata)),' '));
    kkk = isdata+1;
    while ~isempty(char(flike(kkk)))
        Data = [Data; str2double(strsplit(char(flike(kkk)),' '))];
        kkk = kkk + 1;
    end
    Data = [Data(:,1),max(Data(:,1))-Data(:,1)+1, Data(:,2), Data(:,5), Data(:,4)];
    Data = array2table(Data,'VariableNames',{'Observation','Rank','Flow','PPyrs','PP'});

    % Censored
    iscen = kkk+4;
    iecen = find(contains(flike,'Censored Data'))-3;
    if isempty(iecen)
        Cen = nan;
        NoOutliers = 0;
    else
        Cen = str2double(strsplit(char(flike(iscen)),' '));
        for kkk = 2:iecen-iscen+1
            Cen = [Cen; str2double(strsplit(char(flike(iscen+kkk-1)),' '))];
        end
        Cen = Cen(:,1:2);
        Cen = array2table(Cen,'VariableNames',{'Observation','Flow'});
        NoOutliers = size(Cen,1);
    end

    % Expected Quantiles
    isquant = find(contains(flike,'AEP 1 in Y'))+3+1;
    iequant = isquant + 17;
    FitDist = str2double(strsplit(char(flike(isquant)),' '));
    for kkk = 2:iequant-isquant+1
        FitDist = [FitDist; str2double(strsplit(char(flike(isquant+kkk-1)),' '))];
    end
    FitDist = [FitDist(:,1), 1-1./FitDist(:,1), FitDist(:,2:4)];
    FitDist = array2table(FitDist,'VariableNames',{'AEPyrs','AEP','ExpFlow','CL5','CL95'});

    % Parameters
```

```
    iPara = find(contains(flike,'Summary of Posterior Moments'))+3;
    M = str2double(strsplit(char(flike(iPara))));
    M = M(5);
    S = str2double(strsplit(char(flike(iPara+1))));
    S = exp(S(7));
    G = str2double(strsplit(char(flike(iPara+2))));
    G = G(5);
    StartPoints = [G,M,S];
end
```

B.6 FitFlike

Name: FitFlike.m

Type: Function

Details:

Used by ReadFlike.m to fit distribution to quantiles reported by Flike. Needed to get additional quantiles and probabilities for QQ plots, PP plots and true parameter values.

```matlab
function [fitresult, gof] = FitFLIKE(P, Q, Type, StartPoints, Plot)
%CREATEFITS(P,Q)
%  Create fits.
%
%  Data for 'GEV P-Q' fit:
%      X Input : P
%      Y Output: Q
%  Data for 'LP3 P-Q' fit:
%      X Input : P
%      Y Output: Q
%  Data for 'GEV Q-P' fit:
%      X Input : Q
%      Y Output: P
%  Data for 'LP3 Q-P' fit:
%      X Input : Q
%      Y Output: P
%       A Smoothing Spline was used as LP3 does not have a cdf expression.
%  Output:
%      fitresult : a cell-array of fit objects representing the fits.
%      gof : structure array with goodness-of fit info.
%
%  See also FIT, CFIT, SFIT.

%  Original Auto-generated by MATLAB Curve Fitting Toolbox on 04-May-2019 11:43:52

%% Initialization.

% Initialize arrays to store fits and goodness-of-fit.
fitresult = cell( 1, 1 );
gof = struct( 'sse', cell( 1, 1 ), ...
    'rsquare', [], 'dfe', [], 'adjrsquare', [], 'rmse', [] );
[pData, qData] = prepareCurveData( P, Q );

if Type == 'GEV P-Q'
    %% Fit: 'GEV P-Q'.

    % Set up fittype and options.
    ft = fittype( 't+a/k*(1-(-log(x))^k)', 'independent', 'x', 'dependent', 'y' );
    opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
```

```
    opts.Display = 'Off';
    if ~isempty(StartPoints)
        opts.StartPoint = StartPoints;
    end

    % Fit model to data.
    [fitresult, gof] = fit( pData, qData, ft, opts );

    if Plot == 1
        % Create a figure for the plots.
        figure( 'Name', 'GEV P-Q' );

        % Plot fit with data.
        subplot( 2, 1, 1 );
        h = plot( fitresult, pData, qData );
        legend( h, 'Q vs. P', 'GEV P-Q', 'Location', 'NorthEast' );
        % Label axes
        xlabel P
        ylabel Q
        grid on

        % Plot residuals.
        subplot( 2, 1, 2 );
        h = plot( fitresult, pData, qData, 'residuals' );
        legend( h, 'GEV P-Q - residuals', 'Zero Line', 'Location', 'NorthEast' );
        % Label axes
        xlabel P
        ylabel Q
        grid on
        hold off
    end

elseif Type == 'LP3 P-Q'
    %% Fit: 'LP3 P-Q'.

    % Set up fittype and options.
    ft = fittype( 'exp(m+(2/g*((g/6*(norminv(x)-g/6)+1)^3-1))*s)', 'independent', 'x', 'dependent', 'y' );
    opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
    opts.Display = 'Off';
    if ~isempty(StartPoints)
        opts.StartPoint = StartPoints;
```

```
    end

    % Fit model to data.
    [fitresult, gof] = fit( pData, qData, ft, opts );

    if Plot == 1
        % Create a figure for the plots.
        figure( 'Name', 'LP3 P-Q' );

        % Plot fit with data.
        subplot( 2, 1, 1 );
        h = plot( fitresult, pData, qData );
        legend( h, 'Q vs. P', 'LP3 P-Q', 'Location', 'NorthEast' );
        % Label axes
        xlabel P
        ylabel Q
        grid on

        % Plot residuals.
        subplot( 2, 1, 2 );
        h = plot( fitresult, pData, qData, 'residuals' );
        legend( h, 'LP3 P-Q - residuals', 'Zero Line', 'Location', 'NorthEast' );
        % Label axes
        xlabel P
        ylabel Q
        grid on
        hold off
    end

elseif Type == 'GEV Q-P'
    %% Fit: 'GEV Q-P'.

    % Set up fittype and options.
    ft = fittype( 'exp(-(1-k*(x-t)/a)^(1/k))', 'independent', 'x', 'dependent', 'y' );
    opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
    opts.Display = 'Off';
    if ~isempty(StartPoints)
        opts.StartPoint = StartPoints;
    end

    % Fit model to data.
```

```
    [fitresult, gof] = fit( qData, pData, ft, opts );

    if Plot == 1
        % Create a figure for the plots.
        figure( 'Name', 'GEV Q-P' );

        % Plot fit with data.
        subplot( 2, 1, 1 );
        h = plot( fitresult, qData, pData );
        legend( h, 'P vs. Q', 'GEV Q-P', 'Location', 'NorthEast' );
        % Label axes
        xlabel Q
        ylabel P
        grid on

        % Plot residuals.
        subplot( 2, 1, 2 );
        h = plot( fitresult, qData, pData, 'residuals' );
        legend( h, 'GEV Q-P - residuals', 'Zero Line', 'Location', 'NorthEast' );
        % Label axes
        xlabel Q
        ylabel P
        grid on
        hold off
    end

elseif Type == 'LP3 Q-P'
    %% Fit: 'LP3 Q-P'.
    [~, lqData] = prepareCurveData( log(P), log(Q) );

    % Set up fittype and options.
    ft = fittype( 'smoothingspline' );

    % Fit model to data.
    [fitresult, gof] = fit( lqData, pData, ft );

    if Plot == 1
        % Create a figure for the plots.
        figure( 'Name', 'LP3 Q-P' );

        % Plot fit with data.
        subplot( 2, 1, 1 );
```

```
        h = plot( fitresult, lqData, pData );
        legend( h, 'P vs. log(Q)', 'LP3 Q-P', 'Location', 'NorthEast' );
        % Label axes
        xlabel log(Q)
        ylabel P
        grid on

        % Plot residuals.
        subplot( 2, 1, 2 );
        h = plot( fitresult, lqData, pData, 'residuals' );
        legend( h, 'LP3 Q-P - residuals', 'Zero Line', 'Location', 'NorthEast' );
        % Label axes
        xlabel log(Q)
        ylabel P
        grid on
    end

end

```

B.7 xlcolnum2alpha

Name: xlcolnum2alpha.m

Type: Function

Details:

Used by ReadFlike.m. When saving to Excel, has columns in alpha characters, therefore, need to convert a column number to alpha code.

```
function colalpha = xlcolnum2alpha(colnum)


if colnum <= 26
    % One letter
    colalpha = char(colnum + 64);
elseif colnum > 26*27
    % three letters
    a1 = mod(colnum-1,26)+1;
    a2 = mod(floor((colnum-a1)/26)-1,26)+1;
    a3 = mod(floor((colnum-a1-a2*26)/26^2)-1,26^2)+1;
    colalpha = char([a3,a2,a1] +64);
else
    % two letters
    a1 = mod(colnum-1,26)+1;
    a2 = mod(floor((colnum-a1)/26)-1,26)+1;
    colalpha = char([a2,a1] +64);
end
```

B.8 RegressionData

Name: RegressionData.m

Type: Script

Details:

Assembles all the data needed for regression into one structure.

```matlab
% Assemble Data for Regression
% Version: 1.0

clear
clc
close all
format compact
format short

WriteToFile = 1;

% Folders and files
p = pwd;
pindex = [p(1:3), '3. Regression\Regression - all.xlsx'];

pread = [p(1:3), '2. Data Prep & FFA\'];
psave = [p(1:3), '3. Regression\Data 17-09-2019\'];

[~,~,Index] = xlsread(pindex);

%% ASSEMBLE DATA
m = 1;
for k = 2:size(Index,1)
    if ~isnan(Index{k,4})
        Station = Index{k,1};
        StnData(m).Station = Station;
        StnData(m).Name = Index{k,2};

        %% Assign Gauge Metadata
        load([pread,Station,'\Site Data\',Station(1:7),' - Metadata.mat'])
        StnData(m).Metadata = Meta;

        %% Assign Area
        StnData(m).Area = Meta.Area;

        %% Assign Quantiles
        StnData(m).DistSelection = Index{k,4};
        StnData(m).Updated = Index{k,5};
        %StnData(m).Updated = datetime(Index{k,4},'InputFormat','dd/MM/yyyy');
        if isnan(Index{k,3})
            subdir = [];
        else
            subdir = [Index{k,3}, '\'];
        end

        load([pread,Station,'\FFA\',subdir,Station,' - Flike Comparison.mat'])
        StnData(m).FFAresults = FFAresults;
        StnData(m).OptShift = FFAresults.Shift(2);
        StnData(m).Out1 = max(FFAresults.Outliers);
        StnData(m).Out2 = min(FFAresults.Outliers(FFAresults.Outliers > 0));
        StnData(m).AMlen = size(FFAresults.Data{1,1},1);
        StnData(m).RecordLen = Meta.Years;

        idx = find(strcmp(FFAresults.Name, StnData(m).DistSelection));
        FitDist = FFAresults.FitDist{idx};
```

```
        StnData(m).Quantiles = FitDist;

        StnData(m).Q2y = FitDist.ExpFlow(FitDist.AEPyrs == 2);
        StnData(m).Q5y = FitDist.ExpFlow(FitDist.AEPyrs == 5);
        StnData(m).Q10y = FitDist.ExpFlow(FitDist.AEPyrs == 10);
        StnData(m).Q20y = FitDist.ExpFlow(FitDist.AEPyrs == 20);
        StnData(m).Q50y = FitDist.ExpFlow(FitDist.AEPyrs == 50);
        StnData(m).Q100y = FitDist.ExpFlow(FitDist.AEPyrs == 100);

        %% Assign 2016 IFD table
        [~,~,rawRain] = xlsread([pread,Station,'\Rainfall\intensities_',Station(1:7),'_ifds.csv']);

        idx = find(strcmp(rawRain,'Duration'));
        IFD = [0, [rawRain{idx,3:end}]];
        IFD = [IFD; cell2mat(rawRain(idx+1:end,2:end))];
        StnData(m).IFDtable = IFD;

        % 50% AEP
        StnData(m).I1m2y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.5);
        StnData(m).I5m2y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.5);
        StnData(m).I10m2y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.5);
        StnData(m).I15m2y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.5);
        StnData(m).I20m2y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.5);
        StnData(m).I25m2y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.5);
        StnData(m).I30m2y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.5);
        StnData(m).I45m2y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.5);
        StnData(m).I60m2y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.5);
        StnData(m).I2h2y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.5);
        StnData(m).I3h2y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.5);
        StnData(m).I6h2y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.5);
        StnData(m).I9h2y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.5);
        StnData(m).I12h2y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.5);
        StnData(m).I18h2y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.5);
        StnData(m).I24h2y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.5);
        StnData(m).I30h2y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.5);
        StnData(m).I36h2y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.5);
        StnData(m).I48h2y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.5);
        StnData(m).I72h2y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.5);
        StnData(m).I96h2y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.5);
        StnData(m).I120h2y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.5);
        StnData(m).I144h2y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.5);
        StnData(m).I168h2y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.5);

        % 20% AEP
        StnData(m).I1m5y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.2);
        StnData(m).I5m5y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.2);
        StnData(m).I10m5y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.2);
        StnData(m).I15m5y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.2);
        StnData(m).I20m5y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.2);
        StnData(m).I25m5y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.2);
        StnData(m).I30m5y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.2);
        StnData(m).I45m5y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.2);
        StnData(m).I60m5y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.2);
        StnData(m).I2h5y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.2);
```

```
        StnData(m).I3h5y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.2);
        StnData(m).I6h5y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.2);
        StnData(m).I9h5y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.2);
        StnData(m).I12h5y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.2);
        StnData(m).I18h5y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.2);
        StnData(m).I24h5y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.2);
        StnData(m).I30h5y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.2);
        StnData(m).I36h5y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.2);
        StnData(m).I48h5y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.2);
        StnData(m).I72h5y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.2);
        StnData(m).I96h5y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.2);
        StnData(m).I120h5y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.2);
        StnData(m).I144h5y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.2);
        StnData(m).I168h5y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.2);

        % 10% AEP
        StnData(m).I1m10y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.1);
        StnData(m).I5m10y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.1);
        StnData(m).I10m10y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.1);
        StnData(m).I15m10y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.1);
        StnData(m).I20m10y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.1);
        StnData(m).I25m10y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.1);
        StnData(m).I30m10y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.1);
        StnData(m).I45m10y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.1);
        StnData(m).I60m10y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.1);
        StnData(m).I2h10y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.1);
        StnData(m).I3h10y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.1);
        StnData(m).I6h10y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.1);
        StnData(m).I9h10y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.1);
        StnData(m).I12h10y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.1);
        StnData(m).I18h10y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.1);
        StnData(m).I24h10y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.1);
        StnData(m).I30h10y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.1);
        StnData(m).I36h10y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.1);
        StnData(m).I48h10y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.1);
        StnData(m).I72h10y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.1);
        StnData(m).I96h10y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.1);
        StnData(m).I120h10y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.1);
        StnData(m).I144h10y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.1);
        StnData(m).I168h10y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.1);

        % 5% AEP
        StnData(m).I1m20y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.05);
        StnData(m).I5m20y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.05);
        StnData(m).I10m20y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.05);
        StnData(m).I15m20y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.05);
        StnData(m).I20m20y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.05);
        StnData(m).I25m20y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.05);
        StnData(m).I30m20y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.05);
        StnData(m).I45m20y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.05);
        StnData(m).I60m20y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.05);
        StnData(m).I2h20y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.05);
        StnData(m).I3h20y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.05);
        StnData(m).I6h20y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.05);
        StnData(m).I9h20y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.05);
```

```
        StnData(m).I12h20y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.05);
        StnData(m).I18h20y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.05);
        StnData(m).I24h20y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.05);
        StnData(m).I30h20y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.05);
        StnData(m).I36h20y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.05);
        StnData(m).I48h20y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.05);
        StnData(m).I72h20y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.05);
        StnData(m).I96h20y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.05);
        StnData(m).I120h20y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.05);
        StnData(m).I144h20y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.05);
        StnData(m).I168h20y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.05);

        % 2% AEP
        StnData(m).I1m50y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.02);
        StnData(m).I5m50y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.02);
        StnData(m).I10m50y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.02);
        StnData(m).I15m50y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.02);
        StnData(m).I20m50y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.02);
        StnData(m).I25m50y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.02);
        StnData(m).I30m50y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.02);
        StnData(m).I45m50y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.02);
        StnData(m).I60m50y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.02);
        StnData(m).I2h50y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.02);
        StnData(m).I3h50y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.02);
        StnData(m).I6h50y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.02);
        StnData(m).I9h50y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.02);
        StnData(m).I12h50y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.02);
        StnData(m).I18h50y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.02);
        StnData(m).I24h50y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.02);
        StnData(m).I30h50y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.02);
        StnData(m).I36h50y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.02);
        StnData(m).I48h50y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.02);
        StnData(m).I72h50y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.02);
        StnData(m).I96h50y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.02);
        StnData(m).I120h50y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.02);
        StnData(m).I144h50y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.02);
        StnData(m).I168h50y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.02);

        % 1% AEP
        StnData(m).I1m100y = IFD(IFD(:,1) == 1,IFD(1,:) == 0.01);
        StnData(m).I5m100y = IFD(IFD(:,1) == 5,IFD(1,:) == 0.01);
        StnData(m).I10m100y = IFD(IFD(:,1) == 10,IFD(1,:) == 0.01);
        StnData(m).I15m100y = IFD(IFD(:,1) == 15,IFD(1,:) == 0.01);
        StnData(m).I20m100y = IFD(IFD(:,1) == 20,IFD(1,:) == 0.01);
        StnData(m).I25m100y = IFD(IFD(:,1) == 25,IFD(1,:) == 0.01);
        StnData(m).I30m100y = IFD(IFD(:,1) == 30,IFD(1,:) == 0.01);
        StnData(m).I45m100y = IFD(IFD(:,1) == 45,IFD(1,:) == 0.01);
        StnData(m).I60m100y = IFD(IFD(:,1) == 60,IFD(1,:) == 0.01);
        StnData(m).I2h100y = IFD(IFD(:,1) == 2*60,IFD(1,:) == 0.01);
        StnData(m).I3h100y = IFD(IFD(:,1) == 3*60,IFD(1,:) == 0.01);
        StnData(m).I6h100y = IFD(IFD(:,1) == 6*60,IFD(1,:) == 0.01);
        StnData(m).I9h100y = IFD(IFD(:,1) == 9*60,IFD(1,:) == 0.01);
        StnData(m).I12h100y = IFD(IFD(:,1) == 12*60,IFD(1,:) == 0.01);
        StnData(m).I18h100y = IFD(IFD(:,1) == 18*60,IFD(1,:) == 0.01);
        StnData(m).I24h100y = IFD(IFD(:,1) == 24*60,IFD(1,:) == 0.01);
```

```
        StnData(m).I30h100y = IFD(IFD(:,1) == 30*60,IFD(1,:) == 0.01);
        StnData(m).I36h100y = IFD(IFD(:,1) == 36*60,IFD(1,:) == 0.01);
        StnData(m).I48h100y = IFD(IFD(:,1) == 48*60,IFD(1,:) == 0.01);
        StnData(m).I72h100y = IFD(IFD(:,1) == 72*60,IFD(1,:) == 0.01);
        StnData(m).I96h100y = IFD(IFD(:,1) == 96*60,IFD(1,:) == 0.01);
        StnData(m).I120h100y = IFD(IFD(:,1) == 120*60,IFD(1,:) == 0.01);
        StnData(m).I144h100y = IFD(IFD(:,1) == 144*60,IFD(1,:) == 0.01);
        StnData(m).I168h100y = IFD(IFD(:,1) == 168*60,IFD(1,:) == 0.01);

        %% Next Loop
        clear Meta Station FFAresults;
        m = m+1;
    end
end

%% Writing to file
if WriteToFile == 1
    pd = dir(psave);
    % Delete contents
    for pdk = 3:length(pd)
        if pd(pdk).isdir == 1
            rmdir([pd(pdk).folder,'\',pd(pdk).name],'s')
        else
            delete([pd(pdk).folder,'\',pd(pdk).name])
        end
    end

    % Save
    save([psave, 'All Gauge Data.mat'],'StnData')
end
```

B.9 RegressionBaseCase

Name: RegressionBaseCase.m

Type: Script

Details:

Performs regression for Option 1. Also has ability to test outliers by entering gauge ID into Exclude variables for each probability.

```
% Regression Test 1
% Version: 1.1

clear
clc
close all
format compact
format short

% Folders and files
p = pwd;
load([p(1:3), '3. Regression\Data 17-09-2019\All Gauge Data.mat']);
StnData = struct2table(StnData);
psave = [p(1:3), '3. Regression\Run 17-09-2019\'];

n = size(StnData,1);

A = [StnData.Area];
I = [StnData.I72h50y];
%% =====  50% AEP   =====
Q2y = [StnData.Q2y];

% Gauge 102101A identified to be removed 21/09/2019.
Exclude = []; % gauge numbers
m(1) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

[beta(:,1),estQ2y(idx),R2(:,1),adjR2(:,1),MAE(:,1),MAPE(:,1),RMSE(:,1),RMSPE(:,1),pm20(:,1),pm50(:,1)] = OLS(Q2y(idx),A(idx),I(idx),'50%');
res2y = estQ2y' - Q2y;

%% =====  20% AEP   =====
Q5y = [StnData.Q5y];

% no improvement by omitting any gauges 21/09/2019.
Exclude = []; % gauge numbers
m(2) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

[beta(:,2),estQ5y(idx),R2(:,2),adjR2(:,2),MAE(:,2),MAPE(:,2),RMSE(:,2),RMSPE(:,2),pm20(:,2),pm50(:,2)] = OLS(Q5y(idx),A(idx),I(idx),'20%');
res5y = estQ5y' - Q5y;

%% =====  10% AEP   =====
Q10y = [StnData.Q10y];

% no improvement by omitting any gauges 21/09/2019.
```

```
Exclude = []; % gauge numbers
m(3) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

[beta(:,3),estQ10y(idx),R2(:,3),adjR2(:,3),MAE(:,3),MAPE(:,3),RMSE(:,3),RMSPE(:,3),pm20(:,3),pm50(:,3)] = OLS(Q10y(idx),A(idx),I(idx),'10%');
res10y = estQ10y' - Q10y;

%% =====  5% AEP   =====
Q20y = [StnData.Q20y];

% no improvement by omitting any gauges 21/09/2019.
Exclude = []; % gauge numbers
m(4) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

[beta(:,4),estQ20y(idx),R2(:,4),adjR2(:,4),MAE(:,4),MAPE(:,4),RMSE(:,4),RMSPE(:,4),pm20(:,4),pm50(:,4)] = OLS(Q20y(idx),A(idx),I(idx),'5%');
res20y = estQ20y' - Q20y;

%% =====  2% AEP   =====
Q50y = [StnData.Q50y];

% Some improvement by omitting 121001A but not much, 21/09/2019.
Exclude = []; % gauge numbers
m(5) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];
[beta(:,5),estQ50y(idx),R2(:,5),adjR2(:,5),MAE(:,5),MAPE(:,5),RMSE(:,5),RMSPE(:,5),pm20(:,5),pm50(:,5)] = OLS(Q50y(idx),A(idx),I(idx),'2%');
res50y = estQ50y' - Q50y;

%% =====  1% AEP   =====
Q100y = [StnData.Q100y];

Exclude = []; % gauge numbers
m(6) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

[beta(:,6),estQ100y(idx),R2(:,6),adjR2(:,6),MAE(:,6),MAPE(:,6),RMSE(:,6),RMSPE(:,6),
```

```
6),pm20(:,6),pm50(:,6)] = OLS(Q100y(idx),A(idx),I(idx),'1%');
res100y = estQ100y' - Q100y;

%% Outputs
disp('% AEP:')
disp([50,20,10,5,2,1])
disp('beta: ')
disp(beta)
disp('R^2: ')
disp(R2)
disp('Adjusted R^2: ')
disp(adjR2)
disp('Root Mean Square Error: ')
disp(RMSE)
disp('Root Mean Square Percentage Error: ')
disp(RMSPE)
disp('Mean Absolute Error: ')
disp(MAE)
disp('Mean Absolute Percentage Error: ')
disp(MAPE)
disp('No. +-20%: ')
disp(pm20)
disp('% +-20%: ')
Percent_pm20 = pm20./m.*100;
disp(Percent_pm20)
disp('No. +-50%: ')
disp(pm50)

%% OLS base case
function [betaBCF,estQ,R2,adjR2,MAE,MAPE,RMSE,RMSPE,pm20,pm50] = OLS(Q,A,I,AEP)
    n = size(Q,1)

    base = ones(n,1).*10;
    lbase = log10(base);
    lQ = log10(Q);
    lA = log10(A);
    lI = log10(I);
    X = [lbase,lA,lI];
    beta = (X'*X)^(-1)*X'*lQ;

    % Bias Correction Factor
    res = lQ - log10(10.^beta(1) .*A.^beta(2) .*I.^beta(3));
    MSE = sum(res.^2)./(n-2);
    betaBCF = [beta(1) + MSE./2; beta([2,3])];

    estQ = 10.^betaBCF(1) .*A.^betaBCF(2) .*I.^betaBCF(3);

    R2 = 1 - sum((Q - estQ).^2) / sum((Q - mean(Q)).^2);
    adjR2 = 1-(1-R2).*(n-1)./(n-2-1);
    MAE = sum(abs(Q - estQ))./n;
    MAPE = sum(abs(Q - estQ)./Q)./n;
    RMSE = sqrt(sum((Q - estQ).^2)./n);
    RMSPE = sqrt(sum(((Q - estQ)./Q).^2)./n);
    pm20 = sum(abs(Q - estQ) <= estQ.*0.2);
    pm50 = sum(abs(Q - estQ) <= estQ.*0.5);
```

```

%     vecA = (0:1000/20).*20;
%     vecI = 0:20;
%     matA = repmat(vecA,length(vecI),1);
%     matI = repmat(vecI',1,length(vecA));
%     matC = matA.*0 + 10;
%     matQ = matC.^beta(1,k) .*matA.^beta(2,k) .*matI.^beta(3,k);
%     matQ_all(:,:,k) = matQ;
%

    % Plot Curve
%     f1 = figure('Name',[AEP
%     plot3(A,I,Q,'ok'), hold on
%     surf(matA,matI,matQ)
%     colormap(f1,'jet')
%     grid on
%     xlabel('Area')
%     ylabel('Intensity')
%     zlabel('Discharge')
%     hold off
    norm95 = norminv(0.95,0,std(estQ-Q));
    norm5 = norminv(0.05,0,std(estQ-Q));

    % Residuals
    figure
    subplot(2,1,1)
    plot(A,estQ-Q,'ok',[0,1000],[0,0],'r-',[0,1000],[norm95,norm95],'b--',[0,1000],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Area'])
    subplot(2,1,2)
    plot(I,estQ-Q,'ok',[0,max(I)*1.1],[0,0],'r-',[0,max(I)*1.1],[norm95,norm95],'b--',[0,max(I)*1.1],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Intensity'])
end
```

B.10 RegressionOption2

Name: RegressionOption2.m

Type: Script

Details:

Performs regression for option 2. Loops through all durations for each frequency AEP.

```matlab
% Regression Option 2
% Version: 1.1

clear
clc
close all
format compact
format short

% Folders and files
p = pwd;
load([p(1:3), '3. Regression\Data 17-09-2019\All Gauge Data.mat']);
StnData = struct2table(StnData);
psave = [p(1:3), '3. Regression\Run 17-09-2019\'];

n = size(StnData,1);

A = [StnData.Area];
Iname = [];
%% =====  50% AEP   =====
Q2y = [StnData.Q2y];

Exclude = []; % gauge numbers
m(1) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 21:44
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,1} = StnData.Properties.VariableNames{kk};

    [beta2y(:,f),estQ2y,R2(1,f),adjR2(1,f),MAE(1,f),MAPE(1,f),RMSE(1,f),RMSPE(1,f),under(1,f),over(1,f),pm10(1,f),pm20(1,f),pm50(1,f)] = OLS(Q2y(idx),A(idx),I(idx),'50%');
    res2y = estQ2y' - Q2y(idx);
end

%% =====  20% AEP   =====
Q5y = [StnData.Q5y];

Exclude = []; % gauge numbers
m(2) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 45:68
```

```matlab
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,2} = StnData.Properties.VariableNames{kk};

    [beta5y(:,f),estQ5y,R2(2,f),adjR2(2,f),MAE(2,f),MAPE(2,f),RMSE(2,f),RMSPE(2,f),under(2,f),over(2,f),pm10(2,f),pm20(2,f),pm50(2,f)] = OLS(Q5y(idx),A(idx),I(idx),'20%');
    res5y = estQ5y' - Q5y(idx);
end

%% =====  10% AEP   =====
Q10y = [StnData.Q10y];

Exclude = []; % gauge numbers
m(3) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 69:92
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,3} = StnData.Properties.VariableNames{kk};

    [beta10y(:,f),estQ10y,R2(3,f),adjR2(3,f),MAE(3,f),MAPE(3,f),RMSE(3,f),RMSPE(3,f),under(3,f),over(3,f),pm10(3,f),pm20(3,f),pm50(3,f)] = OLS(Q10y(idx),A(idx),I(idx),'10%');
    res10y = estQ10y' - Q10y(idx);
end

%% =====  5% AEP   =====
Q20y = [StnData.Q20y];

Exclude = []; % gauge numbers
m(4) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 93:116
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,4} = StnData.Properties.VariableNames{kk};

    [beta20y(:,f),estQ20y,R2(4,f),adjR2(4,f),MAE(4,f),MAPE(4,f),RMSE(4,f),RMSPE(4,f),under(4,f),over(4,f),pm10(4,f),pm20(4,f),pm50(4,f)] = OLS(Q20y(idx),A(idx),I(idx),'5%');
    res20y = estQ20y' - Q20y(idx);
end
```

```matlab

%% =====  2% AEP   =====
Q50y = [StnData.Q50y];

Exclude = []; % gauge numbers
m(5) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 117:140
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,5} = StnData.Properties.VariableNames{kk};

    [beta50y(:,f),estQ50y,R2(5,f),adjR2(5,f),MAE(5,f),MAPE(5,f),RMSE(5,f),RMSPE(5,f),under(5,f),over(5,f),pm10(5,f),pm20(5,f),pm50(5,f)] = OLS(Q50y(idx),A(idx),I(idx),'2%');
    res50y = estQ50y' - Q50y(idx);
end

%% =====  1% AEP   =====
Q100y = [StnData.Q100y];

Exclude = []; % gauge numbers
m(6) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = 141:164
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,6} = StnData.Properties.VariableNames{kk};

    [beta100y(:,f),estQ100y,R2(6,f),adjR2(6,f),MAE(6,f),MAPE(6,f),RMSE(6,f),RMSPE(6,f),under(6,f),over(6,f),pm10(6,f),pm20(6,f),pm50(6,f)] = OLS(Q100y(idx),A(idx),I(idx),'1%');
    res100y = estQ100y' - Q100y(idx);
end


%% OLS
function [betaBCF,estQ,R2,adjR2,MAE,MAPE,RMSE,RMSPE,under,over,pm10,pm20,pm50] = OLS(Q,A,I,AEP)
    n = size(Q,1);

    base = ones(n,1).*10;
    lbase = log10(base);
```

```
    lQ = log10(Q);
    lA = log10(A);
    lI = log10(I);
    X = [lbase,lA,lI];
    beta = (X'*X)^(-1)*X'*lQ;

    % Bias Correction Factor
    res = lQ - log10(10.^beta(1) .*A.^beta(2) .*I.^beta(3));
    MSE = sum(res.^2)./(n-2);
    betaBCF = [beta(1) + MSE./2; beta([2,3])];

    estQ = 10.^betaBCF(1) .*A.^betaBCF(2) .*I.^betaBCF(3);

    R2 = 1 - sum((Q - estQ).^2) / sum((Q - mean(Q)).^2);
    adjR2 = 1-(1-R2).*(n-1)./(n-2-1);
    MAE = sum(abs(Q - estQ))./n;
    MAPE = sum(abs(Q - estQ)./Q)./n;
    RMSE = sqrt(sum((Q - estQ).^2)./n);
    RMSPE = sqrt(sum(((Q - estQ)./Q).^2)./n);
    under = sum(estQ - Q < 0);
    over = sum(estQ - Q > 0);
    pm10 = sum(abs(Q - estQ) <= estQ.*0.1);
    pm20 = sum(abs(Q - estQ) <= estQ.*0.2);
    pm50 = sum(abs(Q - estQ) <= estQ.*0.5);

%     vecA = (0:1000/20).*20;
%     vecI = 0:20;
%     matA = repmat(vecA,length(vecI),1);
%     matI = repmat(vecI',1,length(vecA));
%     matC = matA.*0 + 10;
%     matQ = matC.^beta(1,k) .*matA.^beta(2,k) .*matI.^beta(3,k);
%     matQ_all(:,:,k) = matQ;
%

    % Plot Curve
%     f1 = figure('Name',[AEP
%     plot3(A,I,Q,'ok'), hold on
%     surf(matA,matI,matQ)
%     colormap(f1,'jet')
%     grid on
%     xlabel('Area')
%     ylabel('Intensity')
%     zlabel('Discharge')
%     hold off
    norm95 = norminv(0.95,0,std(estQ-Q));
    norm5 = norminv(0.05,0,std(estQ-Q));

    % Residuals
    figure
    subplot(2,1,1)
    plot(A,estQ-Q,'ok',[0,1000],[0,0],'r-',[0,1000],[norm95,norm95],'b--',[0,1000],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Area'])
    subplot(2,1,2)
    plot(I,estQ-Q,'ok',[0,max(I)*1.1],[0,0],'r-',[0,max(I)*1.1],[norm95,
```

```
norm95],'b--',[0,max(I)*1.1],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Intensity'])
    hold off
end
```

B.11 RegressionOption3

Name: RegressionOption3.m

Type: Script

Details:

Performs regression for Option3. Loops through given intensity selections.

```
% Regression Option 3
% Version: 1.1

clear
clc
close all
format compact
format short

% Folders and files
p = pwd;
load([p(1:3), '3. Regression\Data 17-09-2019\All Gauge Data.mat']);
StnData = struct2table(StnData);
psave = [p(1:3), '3. Regression\Run 17-09-2019\'];

n = size(StnData,1);

A = [StnData.Area];
Iname = [];
% Itensity_columns = [82:89, 130:137];
%% =====  50% AEP   =====
Q2y = [StnData.Q2y];

Exclude = []; % gauge numbers
m(1) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,1} = StnData.Properties.VariableNames{kk};

    [beta2y(:,f),estQ2y,R2(1,f),adjR2(1,f),MAE(1,f),MAPE(1,f),RMSE(1,f),RMSPE(1,f),under(1,f),over(1,f),pm10(1,f),pm20(1,f),pm50(1,f)] = OLS(Q2y(idx),10,A(idx),I(idx),'50%');
    res2y = estQ2y' - Q2y(idx);
end

%% =====  20% AEP   =====
Q5y = [StnData.Q5y];

Exclude = []; % gauge numbers
m(2) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
```

```
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,2} = StnData.Properties.VariableNames{kk};

    [beta5y(:,f),estQ5y,R2(2,f),adjR2(2,f),MAE(2,f),MAPE(2,f),RMSE(2,f),RMSPE(2,f),under(2,f),over(2,f),pm10(2,f),pm20(2,f),pm50(2,f)] = OLS(Q5y(idx),10,A(idx),I(idx),'20%');
    res5y = estQ5y' - Q5y(idx);
end

%% =====  10% AEP   =====
Q10y = [StnData.Q10y];

Exclude = []; % gauge numbers
m(3) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,3} = StnData.Properties.VariableNames{kk};

    [beta10y(:,f),estQ10y,R2(3,f),adjR2(3,f),MAE(3,f),MAPE(3,f),RMSE(3,f),RMSPE(3,f),under(3,f),over(3,f),pm10(3,f),pm20(3,f),pm50(3,f)] = OLS(Q10y(idx),10,A(idx),I(idx),'10%');
    res10y = estQ10y' - Q10y(idx);
end

%% =====  5% AEP   =====
Q20y = [StnData.Q20y];

Exclude = []; % gauge numbers
m(4) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,4} = StnData.Properties.VariableNames{kk};

    [beta20y(:,f),estQ20y,R2(4,f),adjR2(4,f),MAE(4,f),MAPE(4,f),RMSE(4,f),RMSPE(4,f),under(4,f),over(4,f),pm10(4,f),pm20(4,f),pm50(4,f)] = OLS(Q20y(idx),10,A(idx),I(idx),'5%');
    res20y = estQ20y' - Q20y(idx);
```

```matlab
end

%% =====  2% AEP   =====
Q50y = [StnData.Q50y];

Exclude = []; % gauge numbers
m(5) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,5} = StnData.Properties.VariableNames{kk};

    [beta50y(:,f),estQ50y,R2(5,f),adjR2(5,f),MAE(5,f),MAPE(5,f),RMSE(5,f),RMSPE(5,f),under(5,f),over(5,f),pm10(5,f),pm20(5,f),pm50(5,f)] = OLS(Q50y(idx),10,A(idx),I(idx),'2%');
    res50y = estQ50y' - Q50y(idx);
end

%% =====  1% AEP   =====
Q100y = [StnData.Q100y];

Exclude = []; % gauge numbers
m(6) = n - length(Exclude);
idx = 1:n;
for k = 1:length(Exclude)
    idx(strcmp(StnData.Station, Exclude(k))) = -1;
end
idx(idx == -1) = [];

f = 0;
for kk = Itensity_columns
    f = f + 1;
    I = [StnData{:,kk}];
    Iname{f,6} = StnData.Properties.VariableNames{kk};

    [beta100y(:,f),estQ100y,R2(6,f),adjR2(6,f),MAE(6,f),MAPE(6,f),RMSE(6,f),RMSPE(6,f),under(6,f),over(6,f),pm10(6,f),pm20(6,f),pm50(6,f)] = OLS(Q100y(idx),10,A(idx),I(idx),'1%');
    res100y = estQ100y' - Q100y(idx);
end


%% OLS
function [betaBCF,estQ,R2,adjR2,MAE,MAPE,RMSE,RMSPE,under,over,pm10,pm20,pm50] = OLS(Q,b,A,I,AEP)
    n = size(Q,1)

    base = ones(n,1).*b;
```

```
    lbase = log10(base);
    lQ = log10(Q);
    lA = log10(A);
    lI = log10(I);
    X = [lbase,lA,lI];
    beta = (X'*X)^(-1)*X'*lQ;

    % Bias Correction Factor
    res = lQ - log10(10.^beta(1) .*A.^beta(2) .*I.^beta(3));
    MSE = sum(res.^2)./(n-2);
    betaBCF = [beta(1) + MSE./2; beta([2,3])];

    estQ = 10.^betaBCF(1) .*A.^betaBCF(2) .*I.^betaBCF(3);

    R2 = 1 - sum((Q - estQ).^2) / sum((Q - mean(Q)).^2);
    adjR2 = 1-(1-R2).*(n-1)./(n-2-1);
    MAE = sum(abs(Q - estQ))./n;
    MAPE = sum(abs(Q - estQ)./Q)./n;
    RMSE = sqrt(sum((Q - estQ).^2)./n);
    RMSPE = sqrt(sum(((Q - estQ)./Q).^2)./n);
    under = sum(estQ - Q < 0);
    over = sum(estQ - Q > 0);
    pm10 = sum(abs(Q - estQ) <= estQ.*0.1);
    pm20 = sum(abs(Q - estQ) <= estQ.*0.2);
    pm50 = sum(abs(Q - estQ) <= estQ.*0.5);

%     vecA = (0:1000/20).*20;
%     vecI = 0:20;
%     matA = repmat(vecA,length(vecI),1);
%     matI = repmat(vecI',1,length(vecA));
%     matC = matA.*0 + 10;
%     matQ = matC.^beta(1,k) .*matA.^beta(2,k) .*matI.^beta(3,k);
%     matQ_all(:,:,k) = matQ;
%

    % Plot Curve
%     f1 = figure('Name',[AEP
%     plot3(A,I,Q,'ok'), hold on
%     surf(matA,matI,matQ)
%     colormap(f1,'jet')
%     grid on
%     xlabel('Area')
%     ylabel('Intensity')
%     zlabel('Discharge')
%     hold off
    norm95 = norminv(0.95,0,std(estQ-Q));
    norm5 = norminv(0.05,0,std(estQ-Q));

    % Residuals
    figure
    subplot(2,1,1)
    plot(A,estQ-Q,'ok',[0,1000],[0,0],'r-',[0,1000],[norm95,norm95],'b--',[0,1000],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Area'])
    subplot(2,1,2)
```

```
    plot(I,estQ-Q,'ok',[0,max(I)*1.1],[0,0],'r-',[0,max(I)*1.1],[norm95,norm95],'b--',[0,max(I)*1.1],[norm5,norm5],'b--')
    title([AEP,' AEP: OLS - Residual by Intensity'])
    hold off
end
```

Appendix C – TUFLOW Models

Table of Contents

Only some models have been included as examples to reduce volume of appendices.

Summary of all TUFLOW models

Gauge	Name	Rating curve adopted	Modelled by
109001A	Mossman River at Mossman	Station removed	
111105A	Babinda Creek at The Boulders	Tuflow	
112002A	Fisher Creek at Nerada	Tuflow	
116008B	Gowrie Creek at Abergowrie	DNRME	TMR
126001A	Sandy Creek at Homebush	Station removed	
129001A	Waterpark Creek at Byfield	Tuflow	TMR
130207A	Sandy Creek at Clermont	DNRME	
130407A	Nebo Creek at Nebo	DNRME	
135004A	Gin Gin Creek at Brushy Creek	Tuflow	TMR
136111A	Splinter Creek at Dakiel	Tuflow	
136202D	Barambah Creek at Litzows	DNRME	TMR
136301B	Stuart River at Weens Bridge	DNRME	TMR
137001AB	Elliott River at Elliott	DNRME	TMR
138002ABC	Wide Bay Creek at Brooyar	Tuflow	
138111A	Mary River at Moy Pocket	DNRME	TMR
142001A	Caboolture River at Upper Caboolture	Station removed	
143107A	Bremer River at Walloon	DNRME	
143209AB	Laidley Creek at Mulgowie	Station removed	TMR
145003B	Logan R at Forest Home	Tuflow	
145011A	Teviot Brook at Croftby	Tuflow	
145102B	Albert River at Bromfleet	Tuflow	TMR
422319B	Dalrymple Creek at Allora	Station removed	TMR
917104A	Etheridge River at Roseglen	Tuflow	
922101B	Coen River at Racecourse	Tuflow	

C.1 Gauge 111105A

Name:	Babinda Creek at The Boulders
Longitude:	-17.34716
Latitude:	145.8726
Control:	Sand Gravel
Catchment Area:	39 km^2
Maximum Gauging Stage:	3.72m
Maximum Record Stage:	6.898m
Zero Gauge Height:	15.702 mAHD

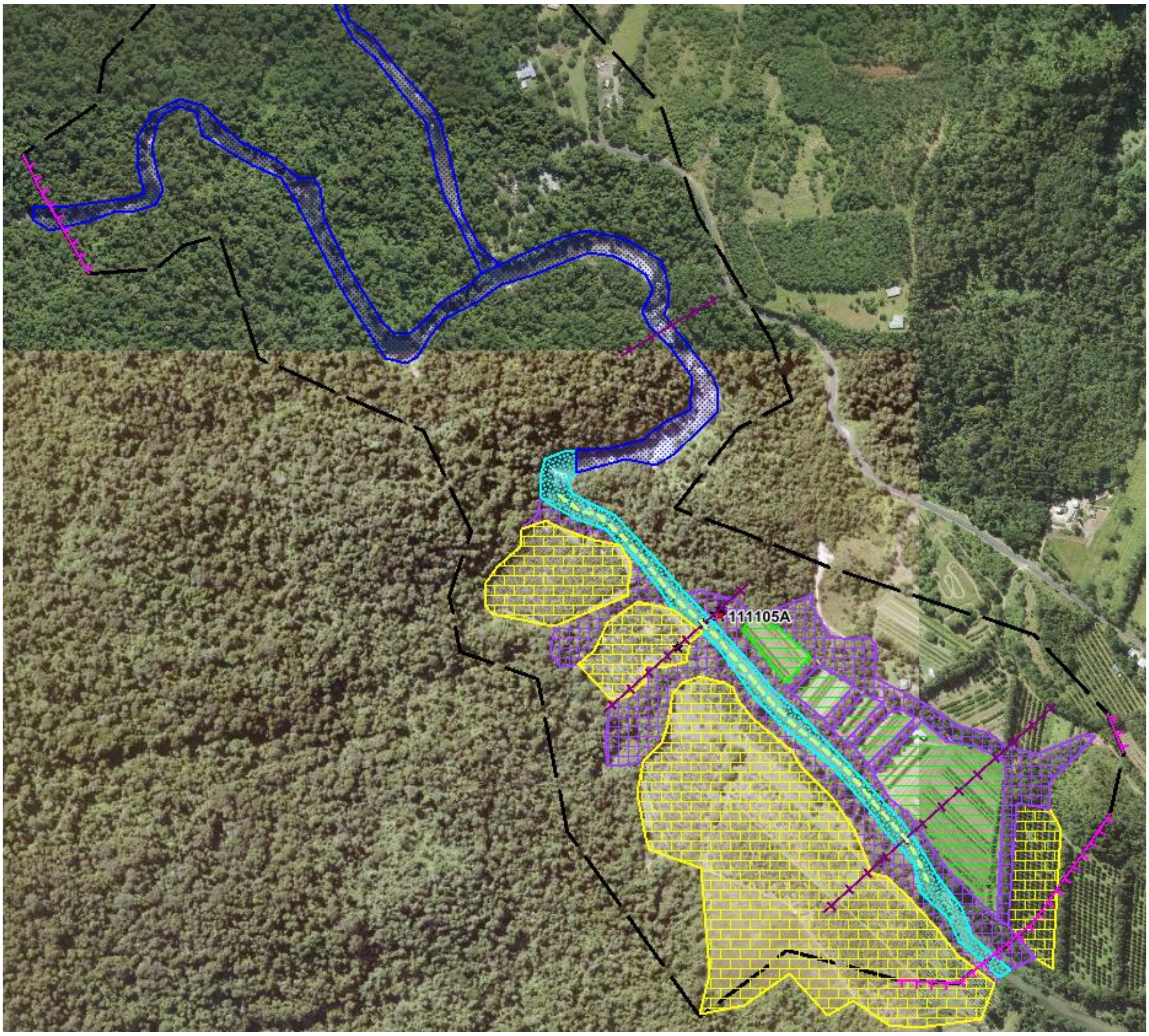
Comments

Completed by:	Samuel Walker
Action:	Adopt TUFLOW rating curve

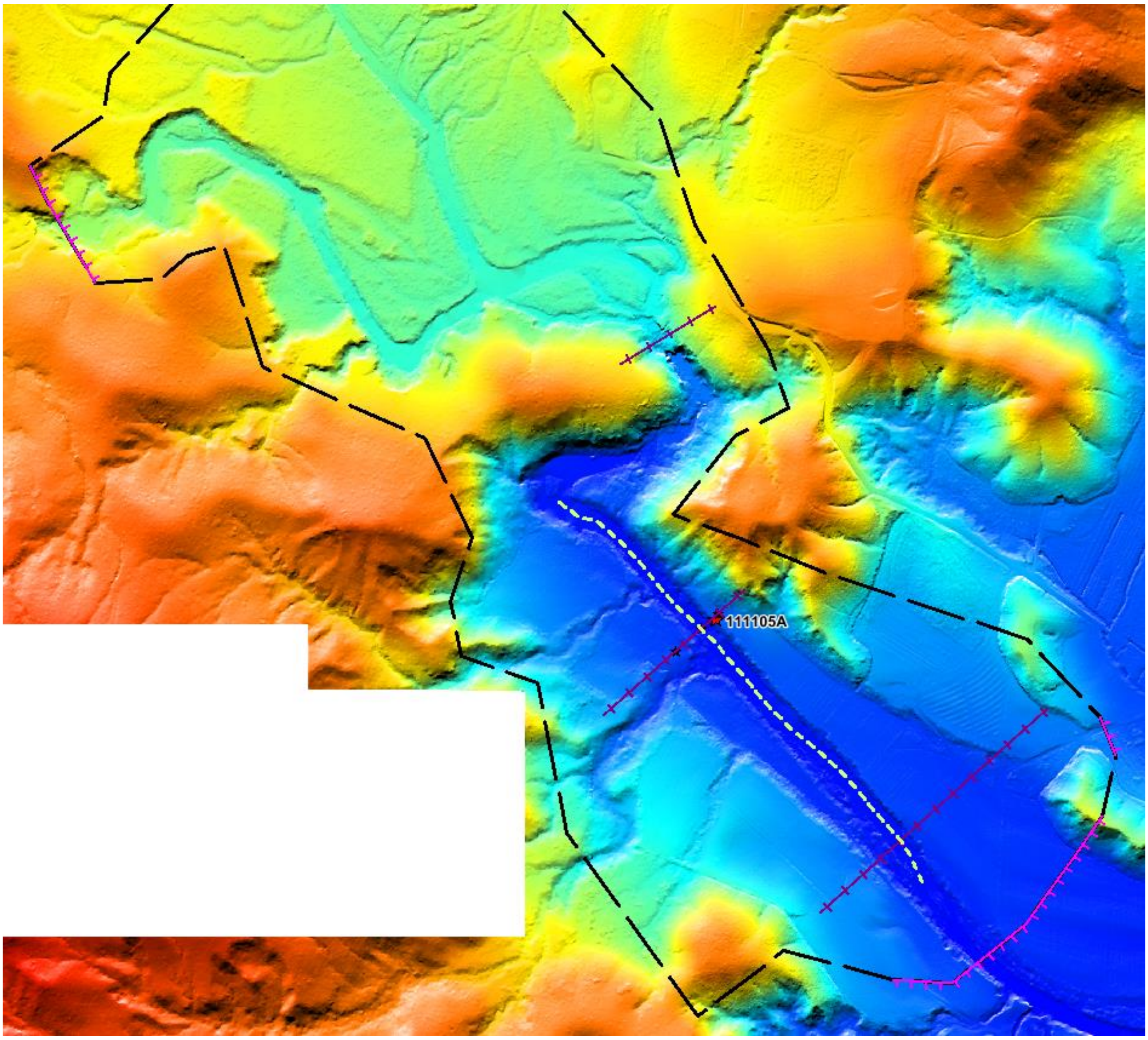
- Quite sudden change from rough creek bed (Blue) to smooth (Cyan).
- Cut channel along light green dashed line, -0.5m deep, 25m width.
- Upstream in the hills there are some waterfalls prior to the start of the model.
- Entered flow into the dominate upstream branch.
- Upper stream has large gravel/boulders. There is a sudden drop in roughness when the slope changes and sand accumulate.

Materials

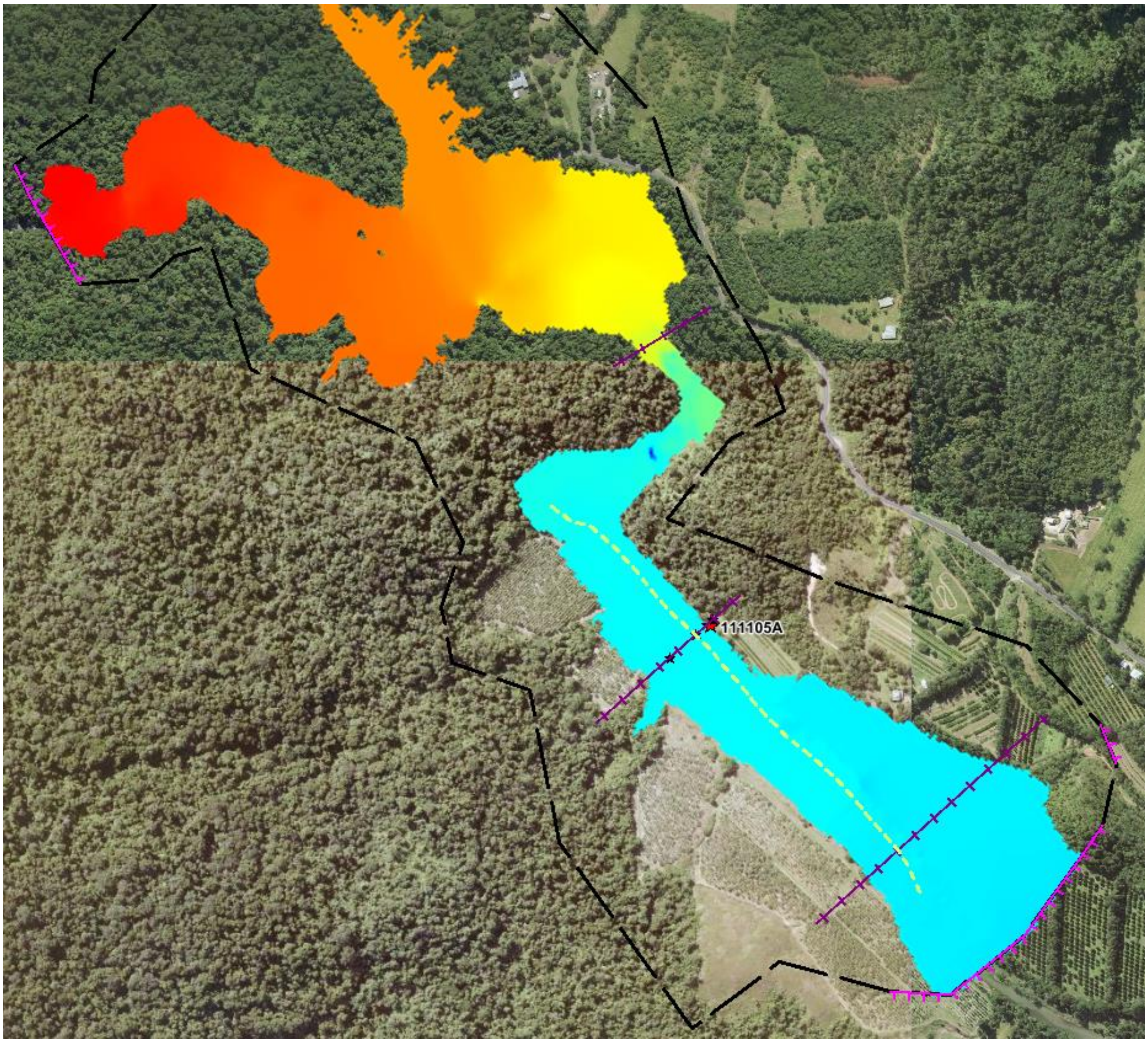
Index	Roughness	Colour	Description
1	0.15	Default	Very high dense rainforest
2	0.06	Blue	Rocky rough creek
3	0.035	Cyan	Smooth creek
4	0.04	Yellow	Orchards
5	0.04	Green	Crops
6	0.06	Violet	Medium-high dense trees, banks, between crops



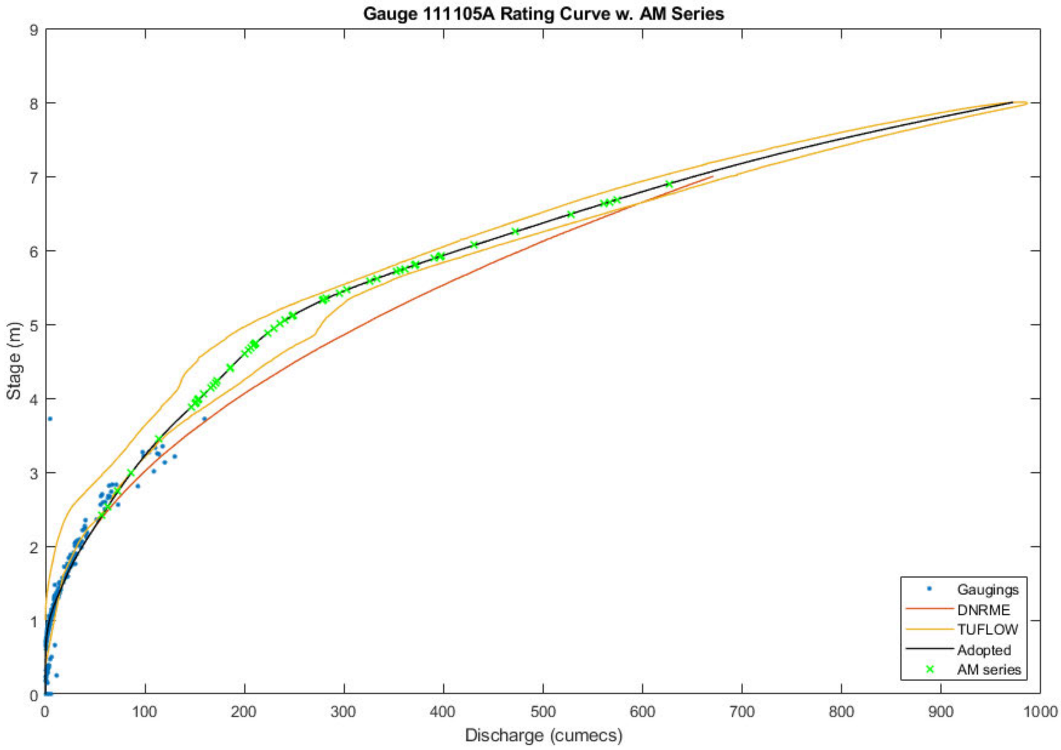
Areal Imagery with Material Polygons



Terrain with relief shading



Maximum water surface height



Rating Curve graph

C.2 Gauge 126001A

Name:	Sandy Creek at Homebush
Longitude:	-21.2832888
Latitude:	149.0225055
Control:	Control weir
Catchment Area:	326 km^2
Maximum Gauging Stage:	8.8m
Maximum Record Stage:	14.766m
Zero Gauge Height:	11.952m AHD

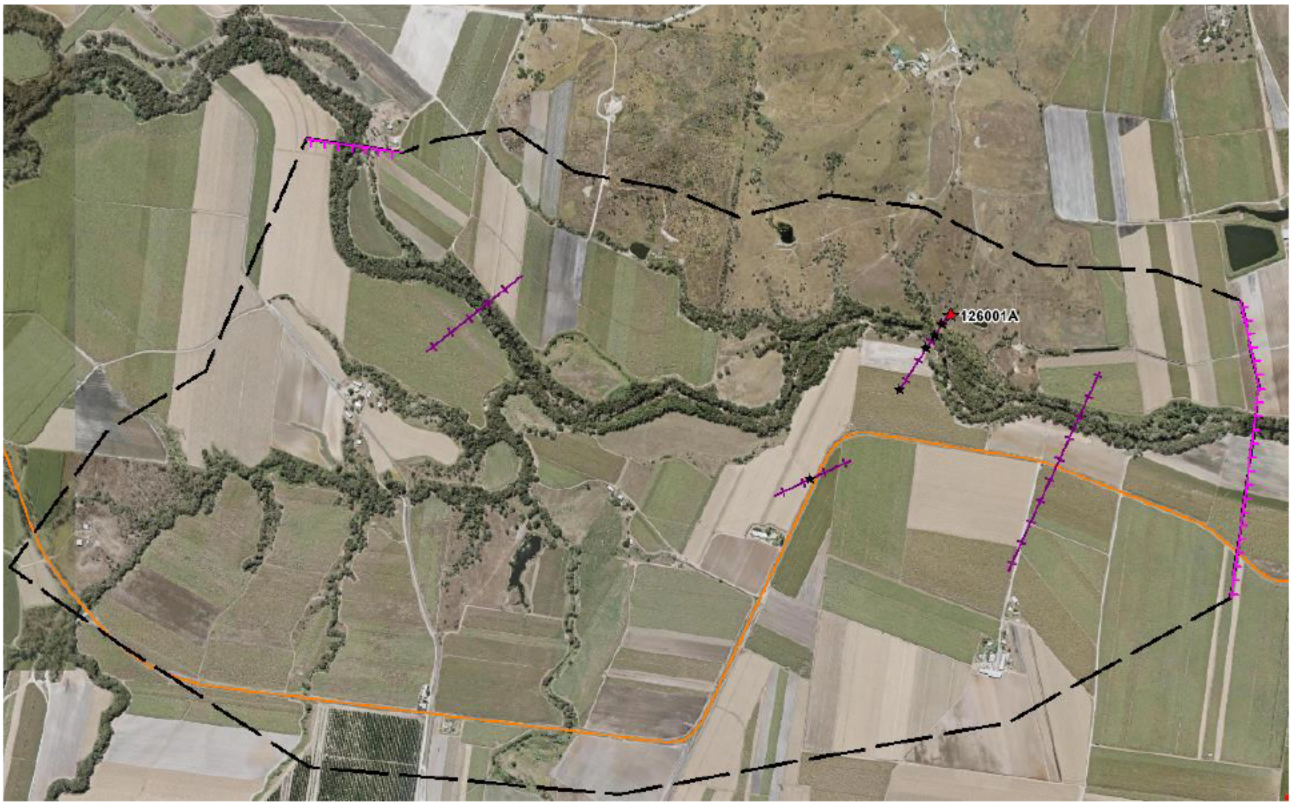
Comments

Completed by:	Samuel Walker
Action:	Gauge was removed from project

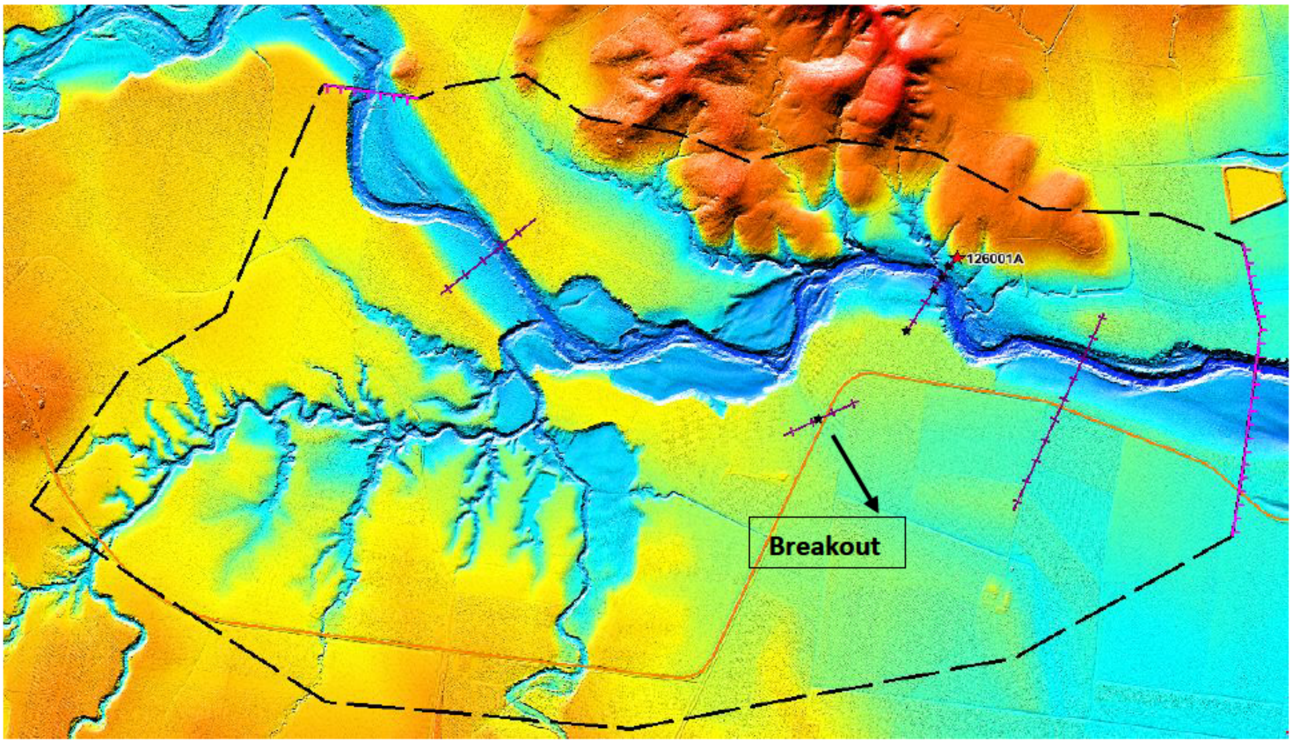
- There are outbreaks and spreading over flood plain with influences from farms and their drainage. Hence, only one run was completed, and no materials applied.

Materials

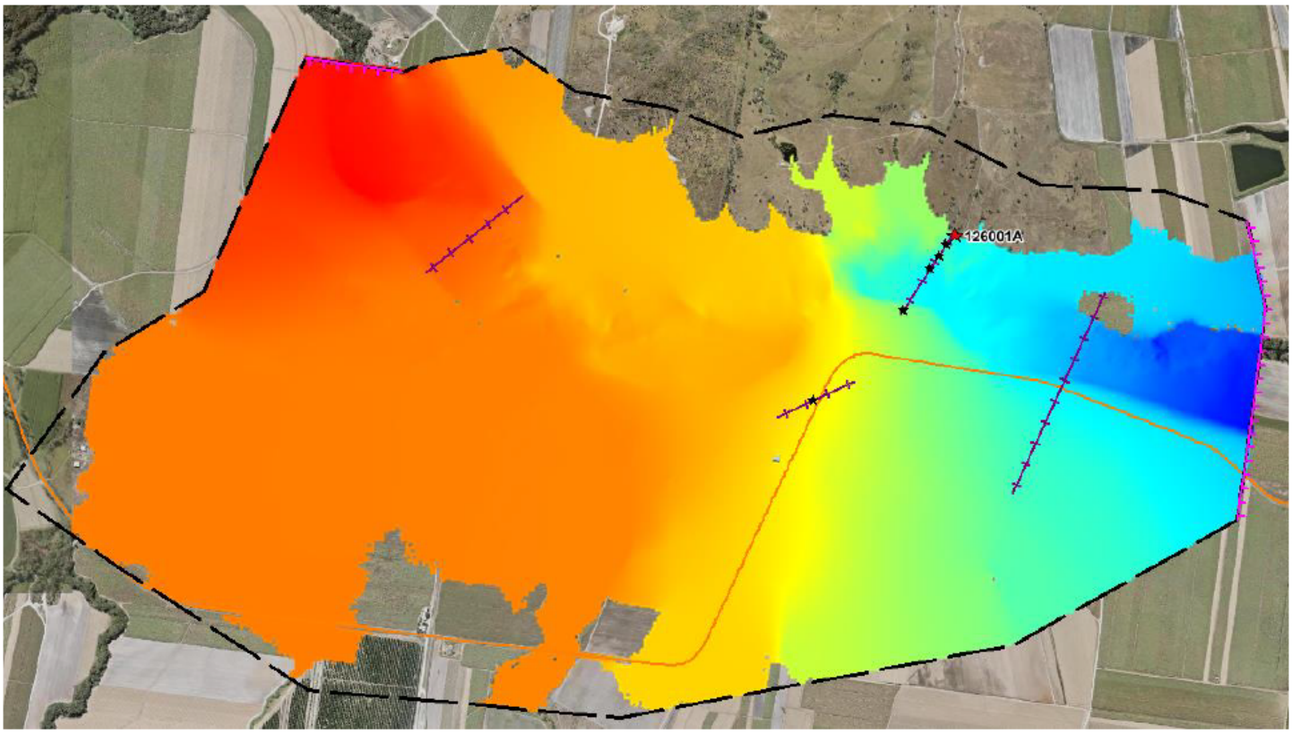
Index	Roughness	Colour	Description
1	0.05	Default	first run to identify model extents and issues



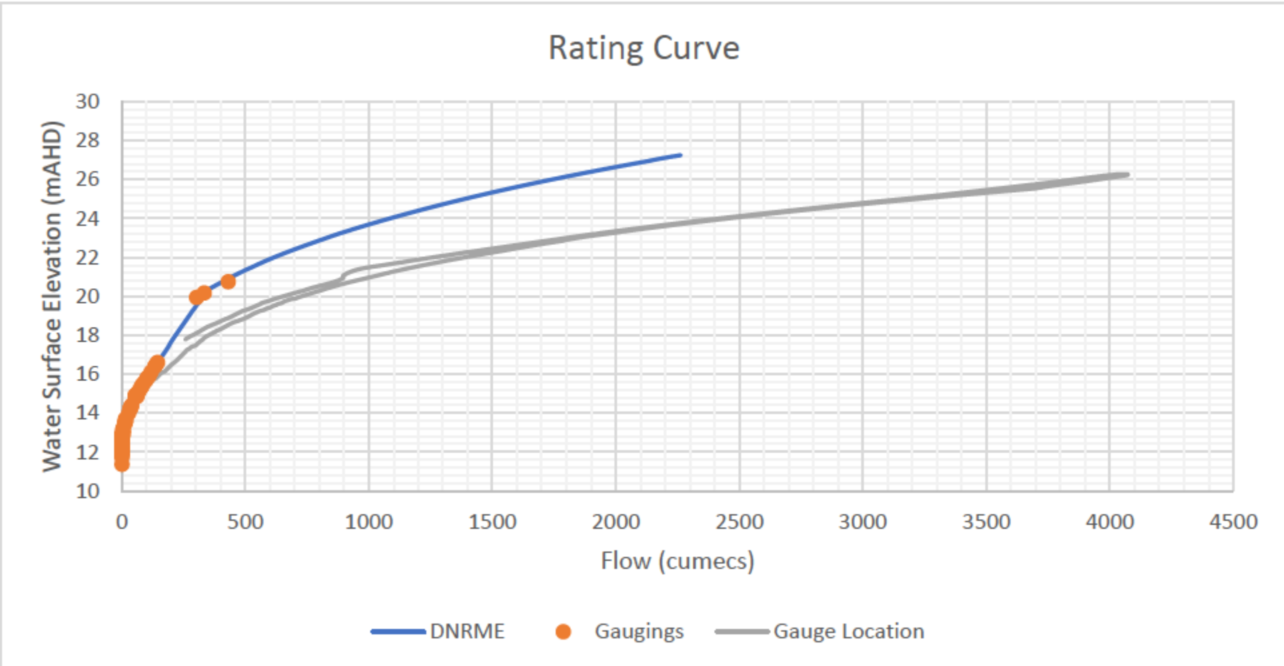
Areal Imagery



Terrain with relief shading



Maximum water surface height



Rating Curve graph – Run 01

C.3 Gauge 136111A

Name:	Splinter Creek at Dakiel
Longitude:	-24.74543
Latitude:	151.26025
Control:	Gravel
Catchment Area:	139 km^2
Maximum Gauging Stage:	2.2m
Maximum Record Stage:	7.89m
Zero Gauge Height:	291.651 mAHD

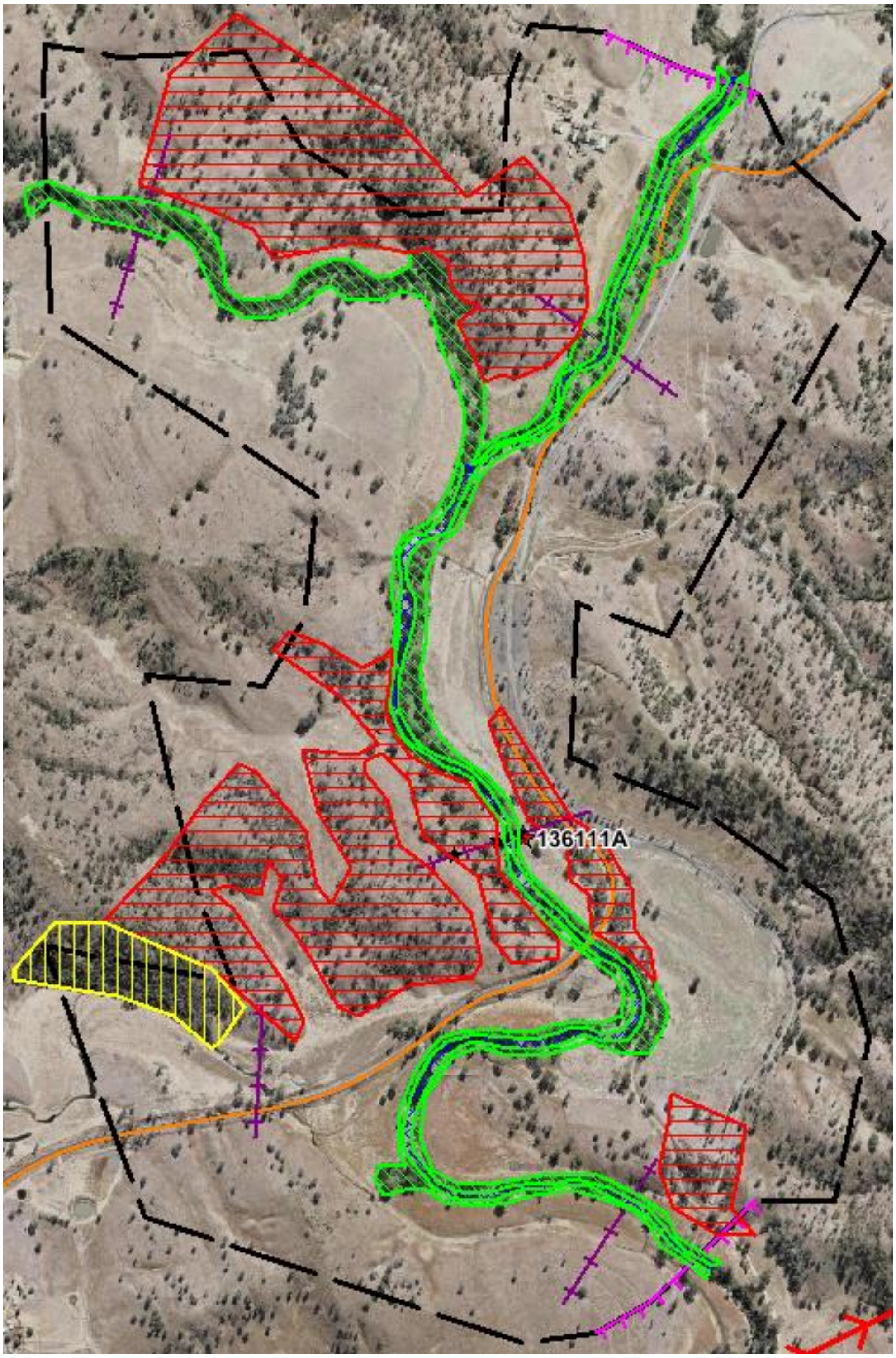
Comments

Completed by:	Samuel Walker
Action:	Adopt TUFLOW rating curve

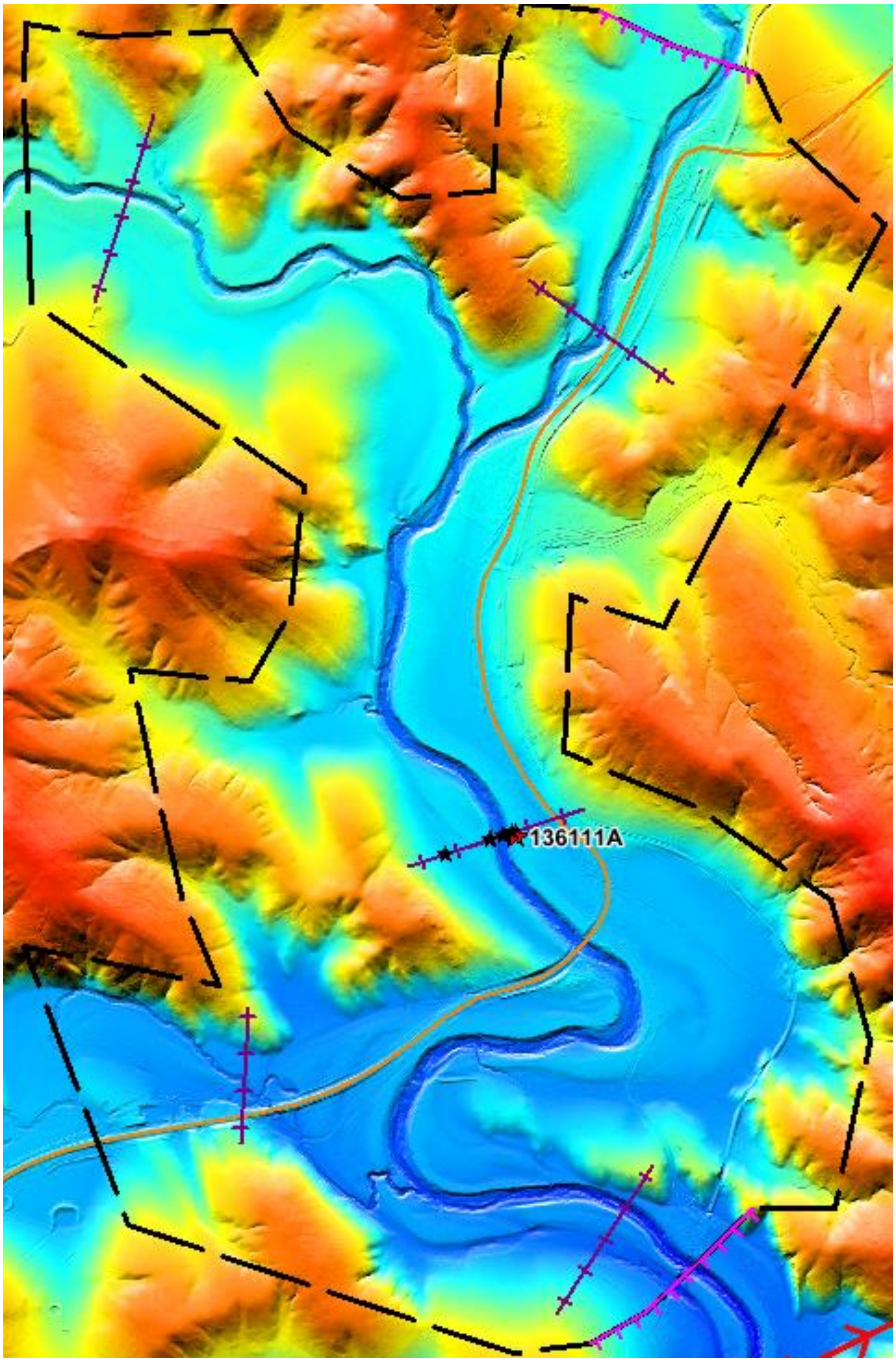
- There was a very small breakout near the gauge however, this was insignificant and only for a short time: the entire cross section would be rated, and quantity of water bypassing was negligible compared to the main channel.
- Another catchment of approximately 25km^2 joins 1.27km downstream of gauge. It was considered to not influence the rating curve significantly due to the catchment would have a different peak time and the distance downstream.

Materials

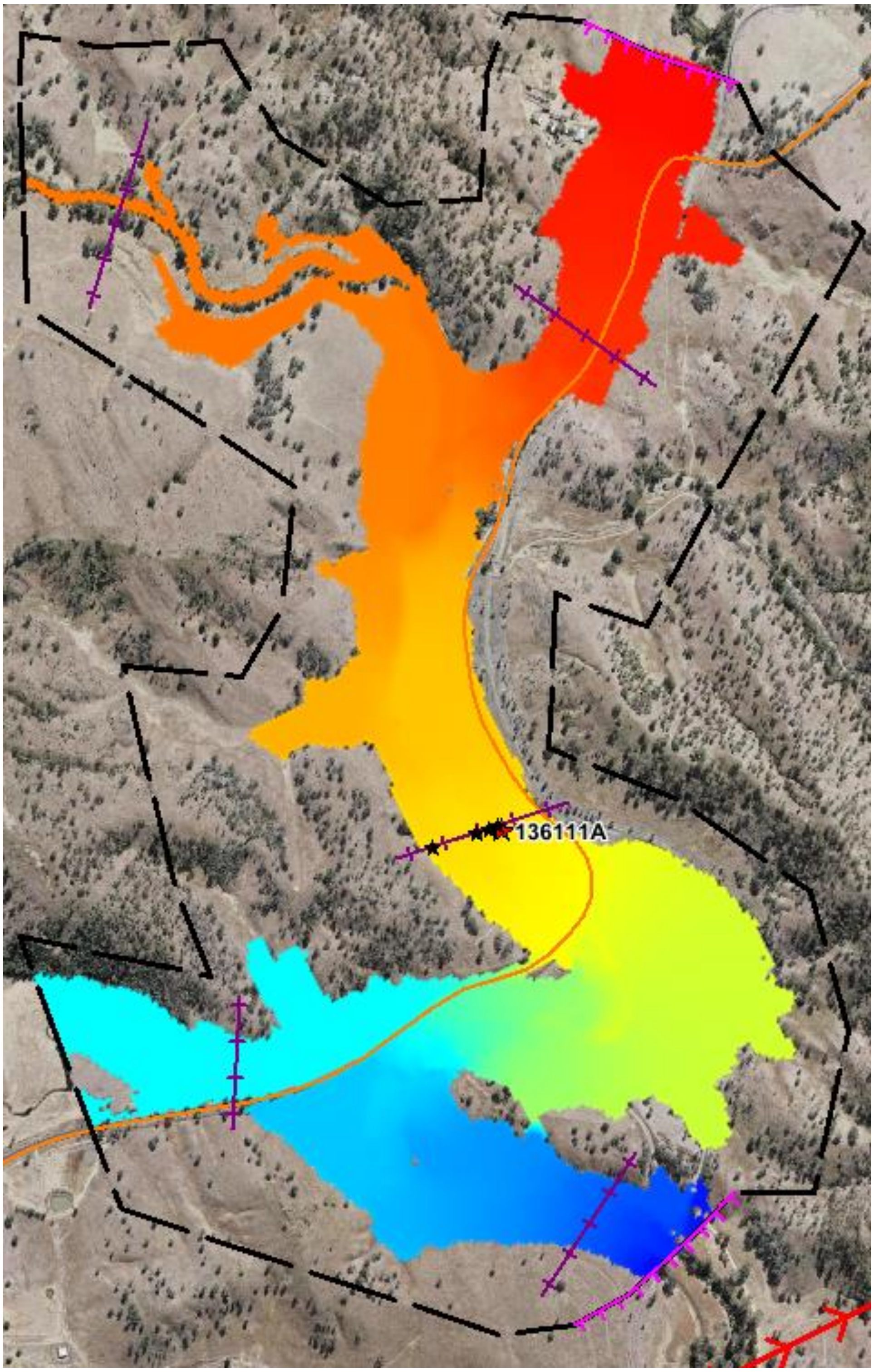
Index	Roughness	Colour	Description
1	0.04	Default	Short grass and sparse trees
2	0.06	Blue	Creek and water
3	0.08	Green	Riparian and creek banks
4	0.065	Red	Low density trees
5	0.08	Yellow	Medium density trees



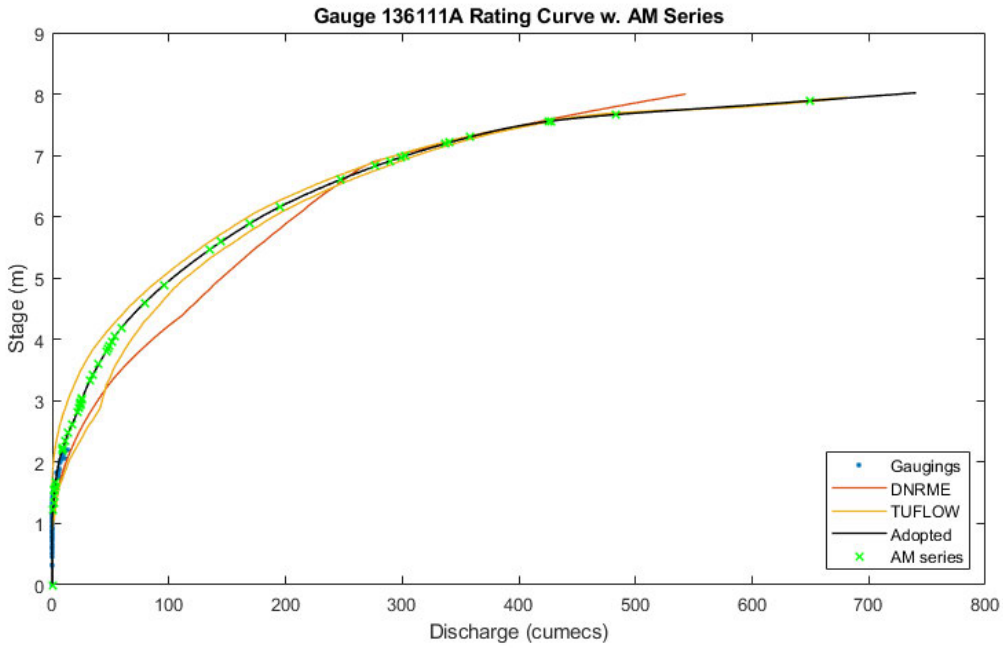
Areal Imagery with Material Polygons



Terrain with relief shading



Maximum water surface height



Rating Curve graph

C.4 Gauge 138002ABC

Name:	Wide Bay Creek at Brooyar
Longitude:	-26.005383
Latitude:	152.411502
Control:	Sand Gravel
Catchment Area:	655 km^2
Maximum Gauging Stage:	5.88m
Maximum Record Stage:	13.35m
Zero Gauge Height:	42.373 mAHD

Comments

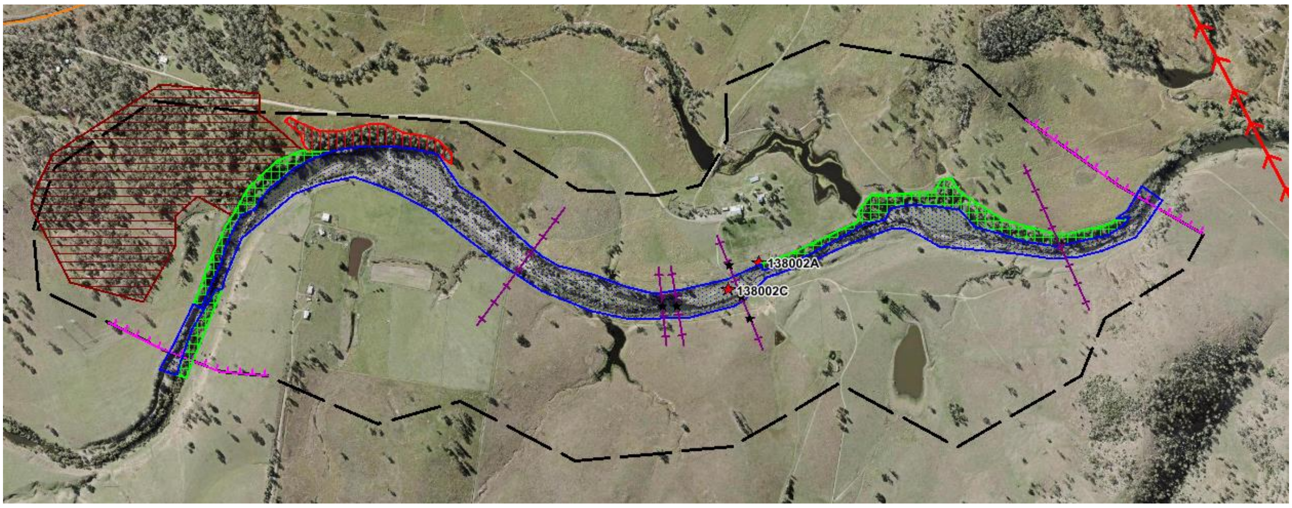
Completed by:	Samuel Walker
Action:	Adopt TUFLOW rating curve

- Had some trouble finding the rated cross section. DNRME confirmed it was cross section 7, see image below.
- Stream control is the road through the creek which governs water height.
- The stream bed was rather variable with pools, vegetation and sandy areas.

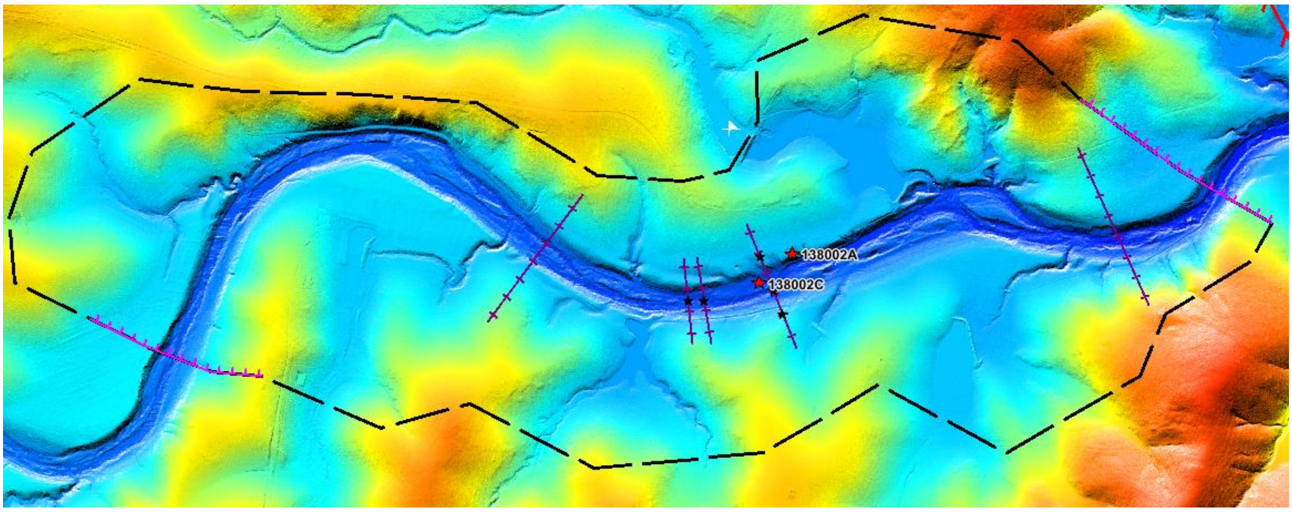
Materials

Index	Roughness	Colour	Description
1	0.03	Default	Pasture and short grass
2	0.05	Blue	Creek and water
3	0.07	Maroon	Medium dense tree
4	0.12	Red	High dense trees
5	0.10	Green	Rough creek / Riparian

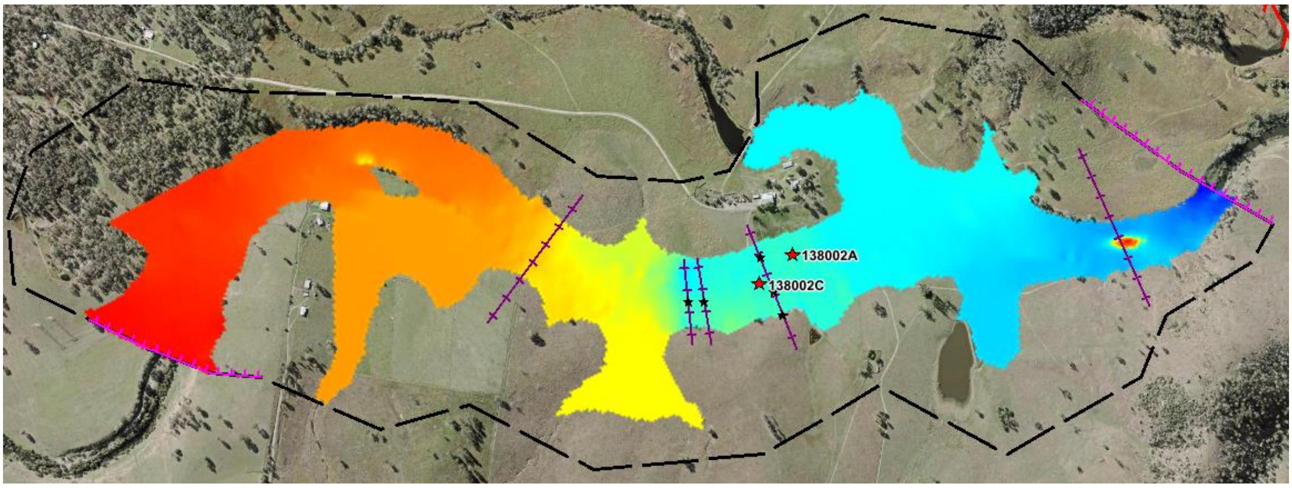
Finding rated cross section.



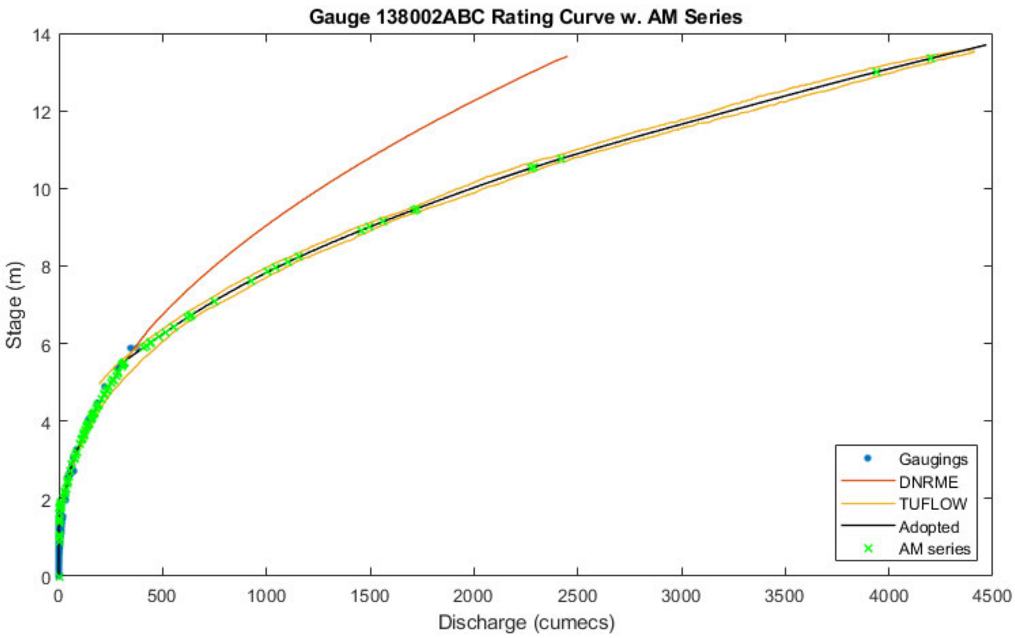
Areal Imagery with Material Polygons



Terrain with relief shading



Maximum water surface height



Rating Curve graph

C.5 Gauge 142001A

Name:	Caboolture River at Upper Caboolture
Longitude:	-27.0978
Latitude:	152.8906
Control:	Sand Gravel
Catchment Area:	94 km^2
Maximum Gauging Stage:	7.08m
Maximum Record Stage:	12.914m
Zero Gauge Height:	6.716 mAHD

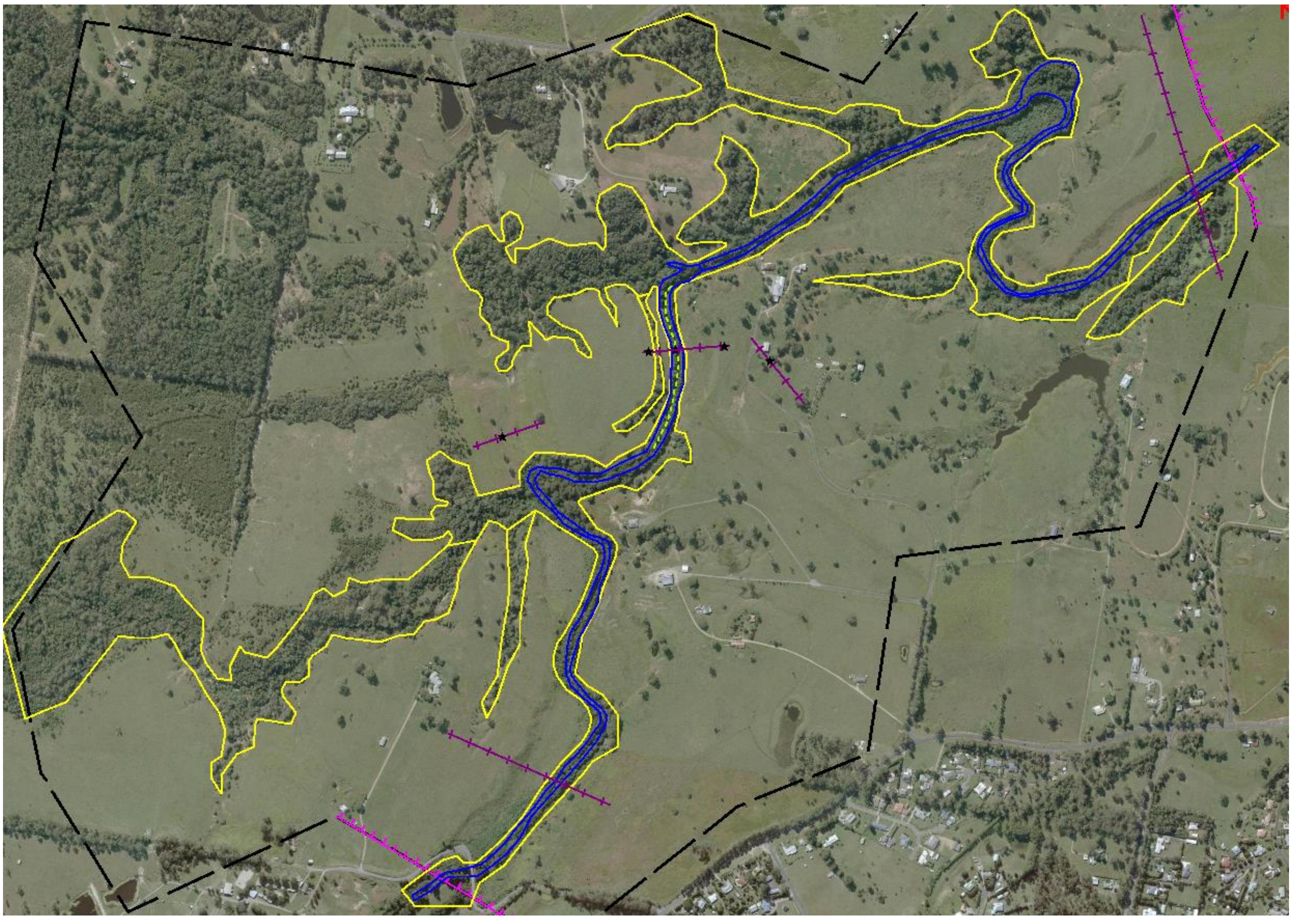
Comments

Completed by:	Samuel Walker
Action:	Gauge removed from project

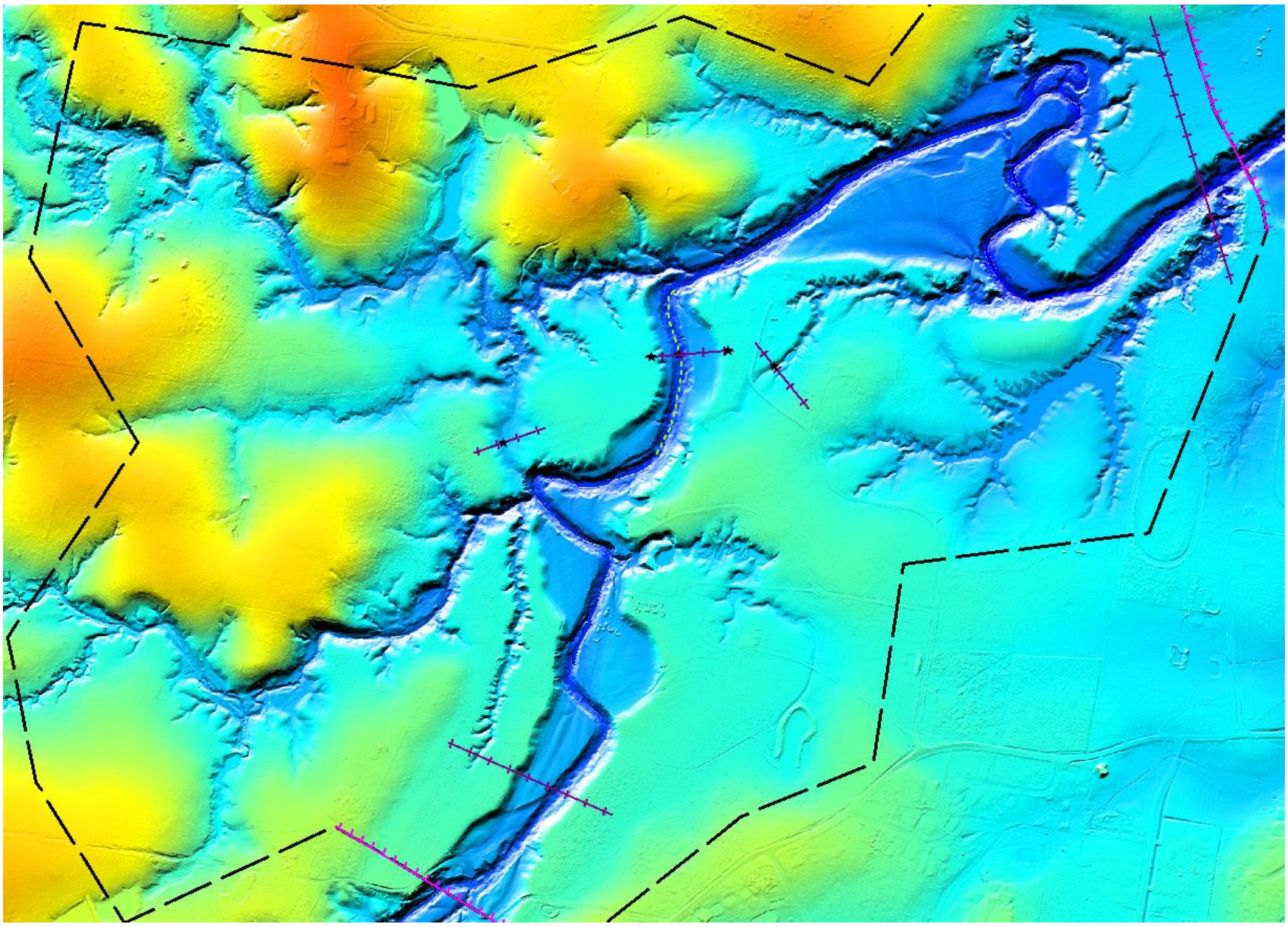
- A stream joins 160m downstream of gauge location. Basic hydrology using Rational method was used to calculate an approximate discharge to check its effect on the gauge. Catchment was 484ha, had a time of concentration of 215min, and runoff coefficient of 0.7. This gave an approximate discharge of 45.1m^3/s for a 1% AEP. The side catchment will have a different peak time then the gauge catchment. Even with the proximity this stream it is will unlikely to have significant backwater effects at the gauge.
- Bypass flow occurs
- Appears roughness needs to increase which will cause greater depth, contributing to bypass flow. Once bypass flow was identified as being a problem the model was not developed further.

Materials

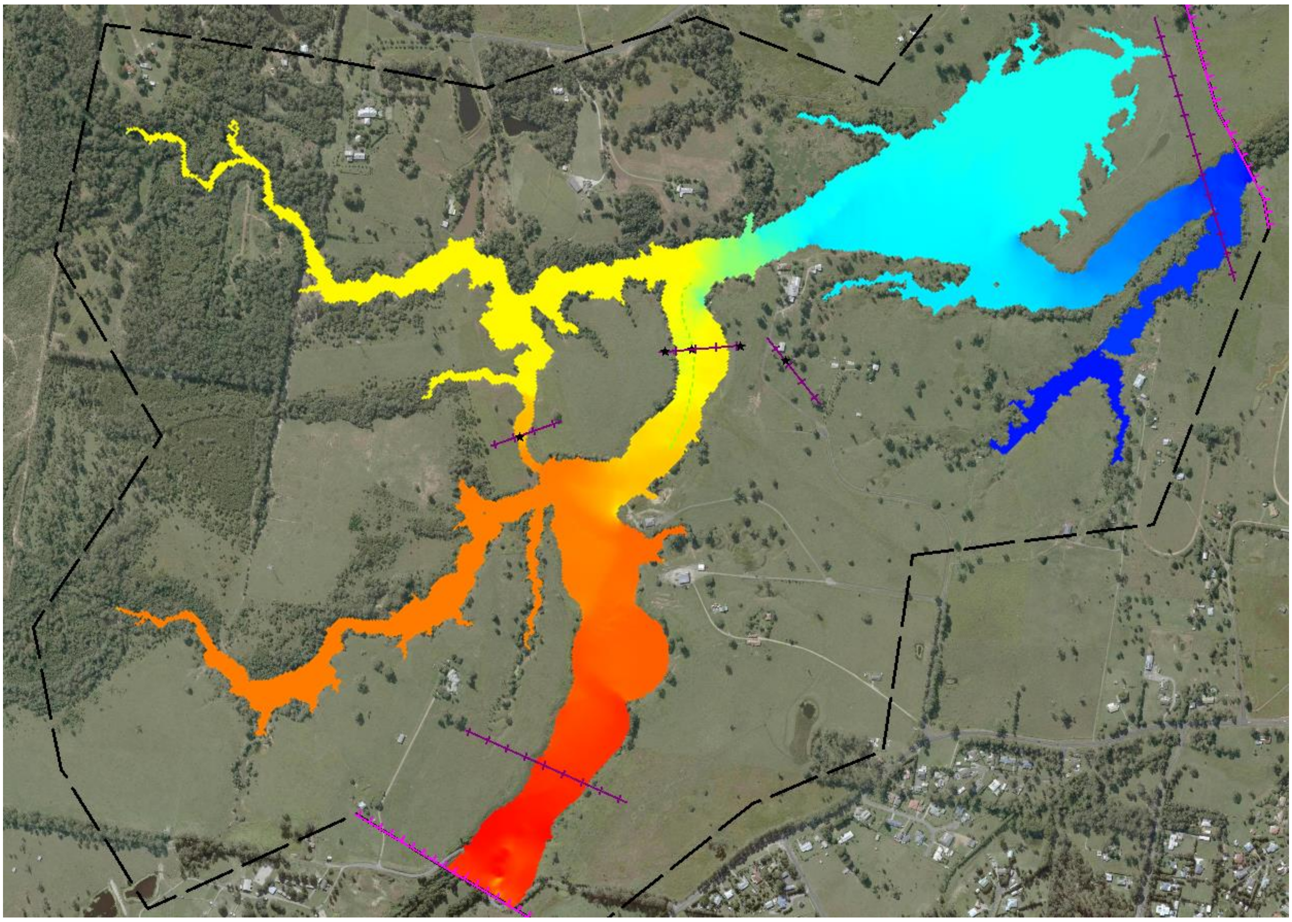
Index	Roughness	Colour	Description
1	0.03	Default	Pasture and short grass
2	0.04	Blue	Creek and water
3	0.10	Yellow	Riparian and creek banks, and high-density trees



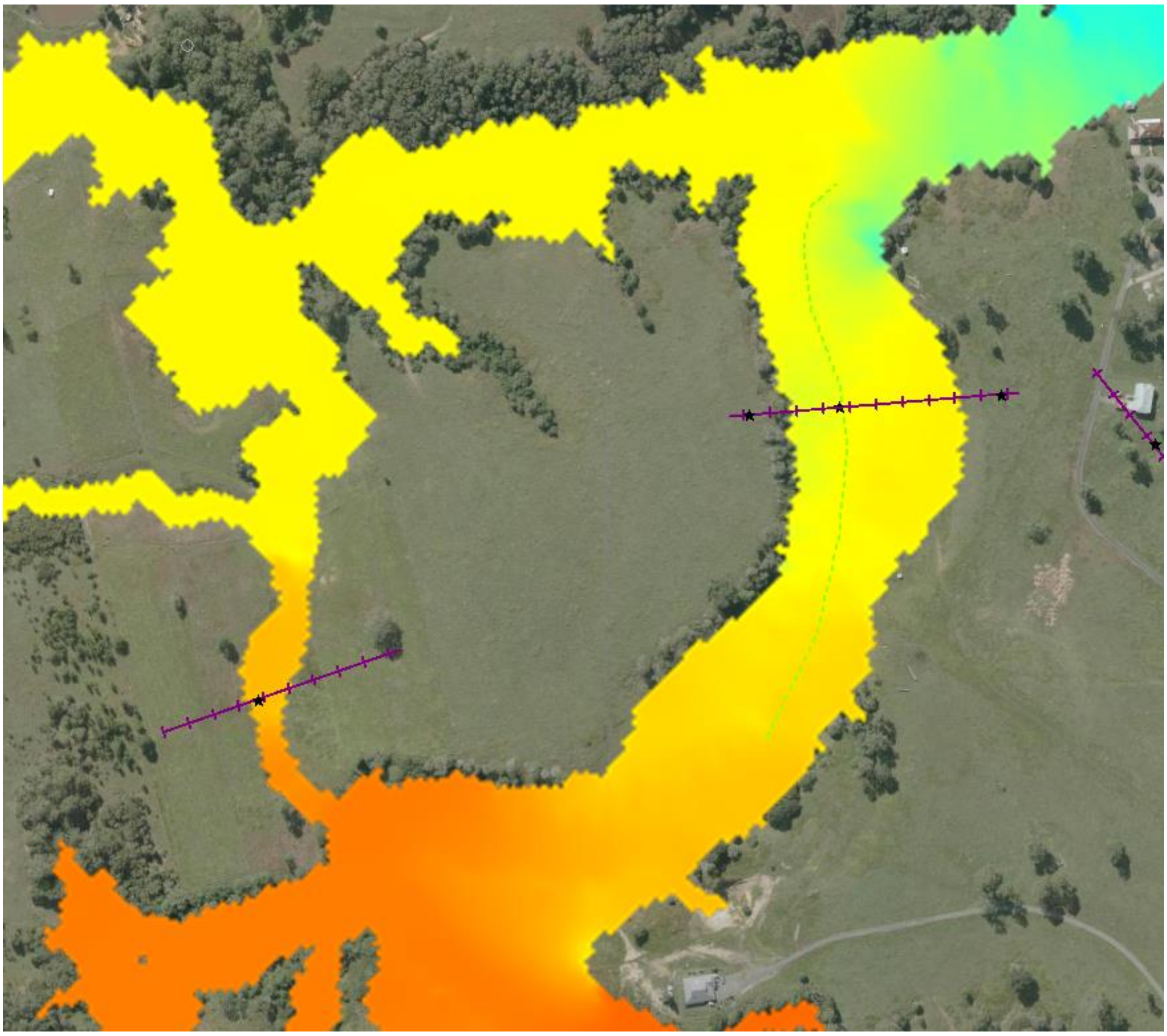
Areal Imagery with Material Polygons



Terrain with relief shading



Maximum water surface height



Maximum water surface height – zoom into bypass

C.6 Gauge 143107A

Name:	Bremer River at Walloon
Longitude:	-27.601912
Latitude:	152.693968
Control:	Control weir
Catchment Area:	622 km^2
Maximum Gauging Stage:	8.96m
Maximum Record Stage:	11.284m
Zero Gauge Height:	16.462m AHD

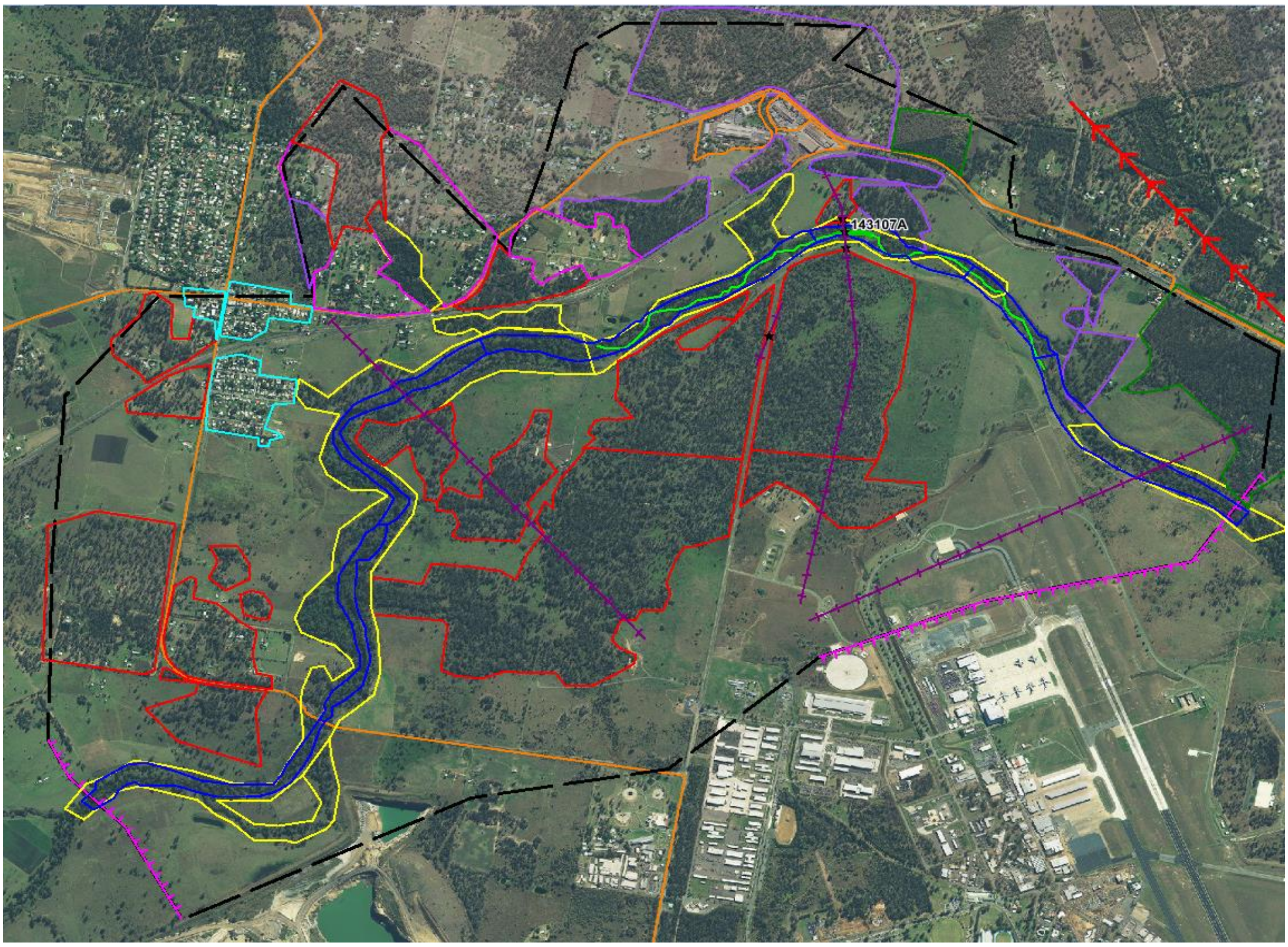
Comments

Completed by:	Samuel Walker
Action:	Adopt DNRME rating curve

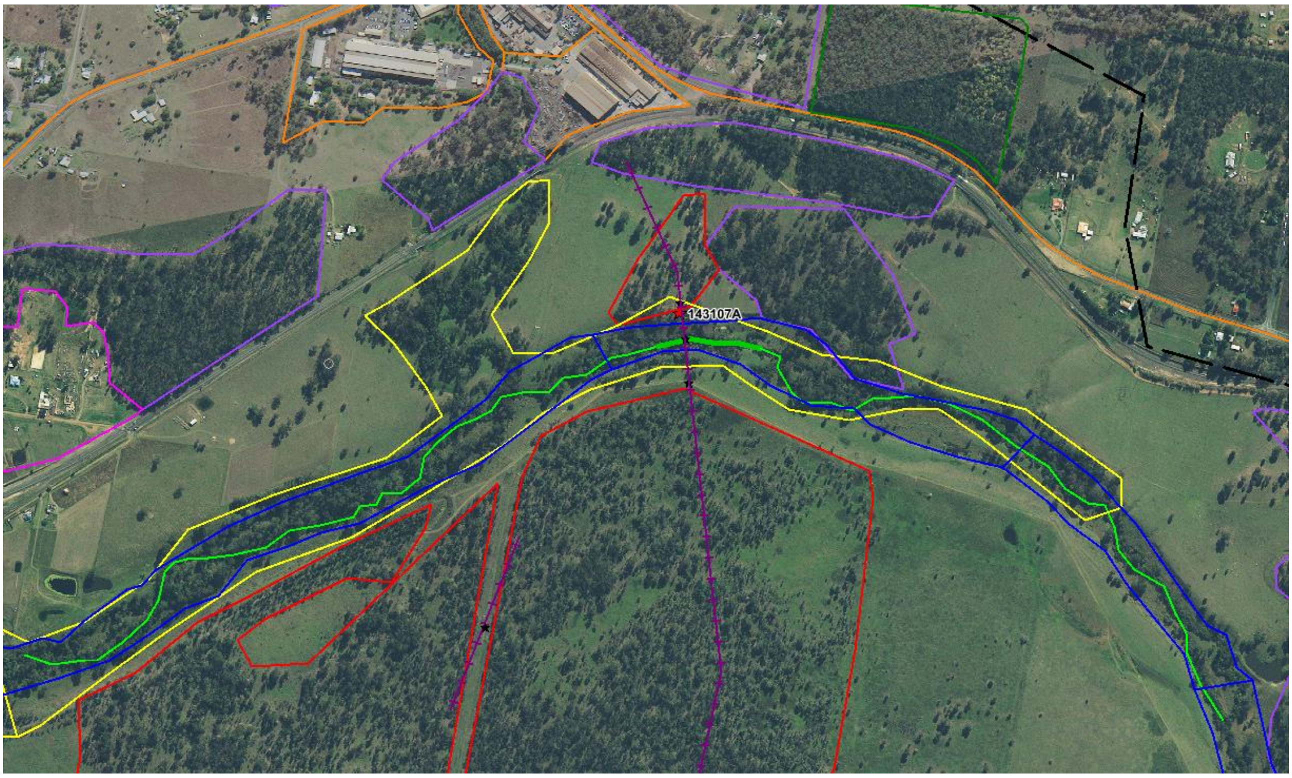
- Complex model with many different roughness areas.
- Stream was lowered (see light green lines on maps);
    - Preceding gauge: -1.1m deep, 5m wide
    - At gauge (bold line): -1.1m deep, 15m wide
    - Following gauge: -1.1m deep, 5m wide
- Rating curve match DNRME closely, so it was adopted.

Materials

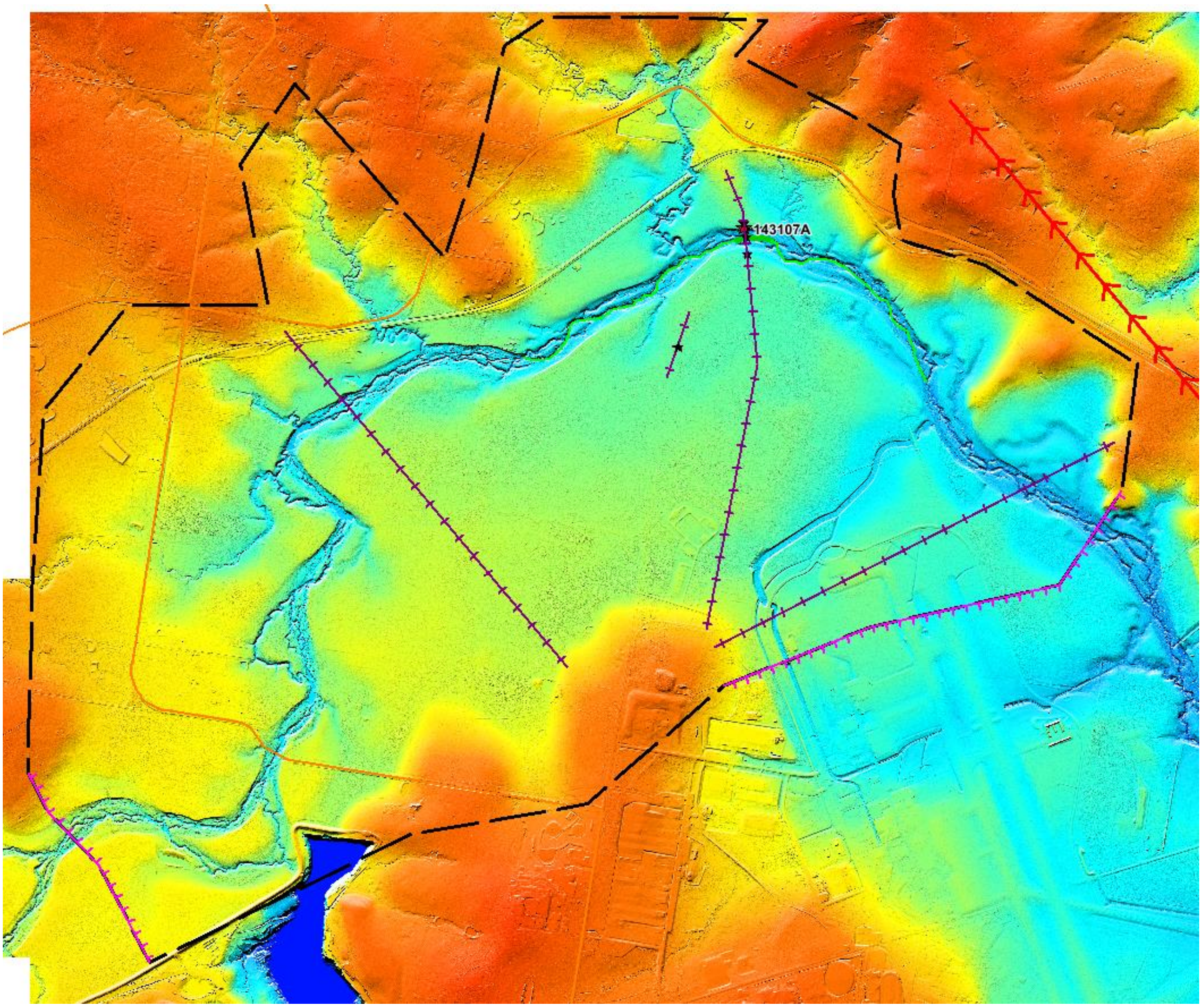
Index	Roughness	Colour	Description
1	0.035	Default	Cleared grassed areas
2	0.10	Blue	Creek and water
3	0.08	Yellow	Riparian and creek banks
4	0.07	Red	Low density trees
5	0.05	Magenta	Rural residential
6	0.07	Violet	Medium density trees
7	0.05	Orange	Shops and paved areas
8	0.10	Dark Green	High density trees
9	0.08	Cyan	Residential



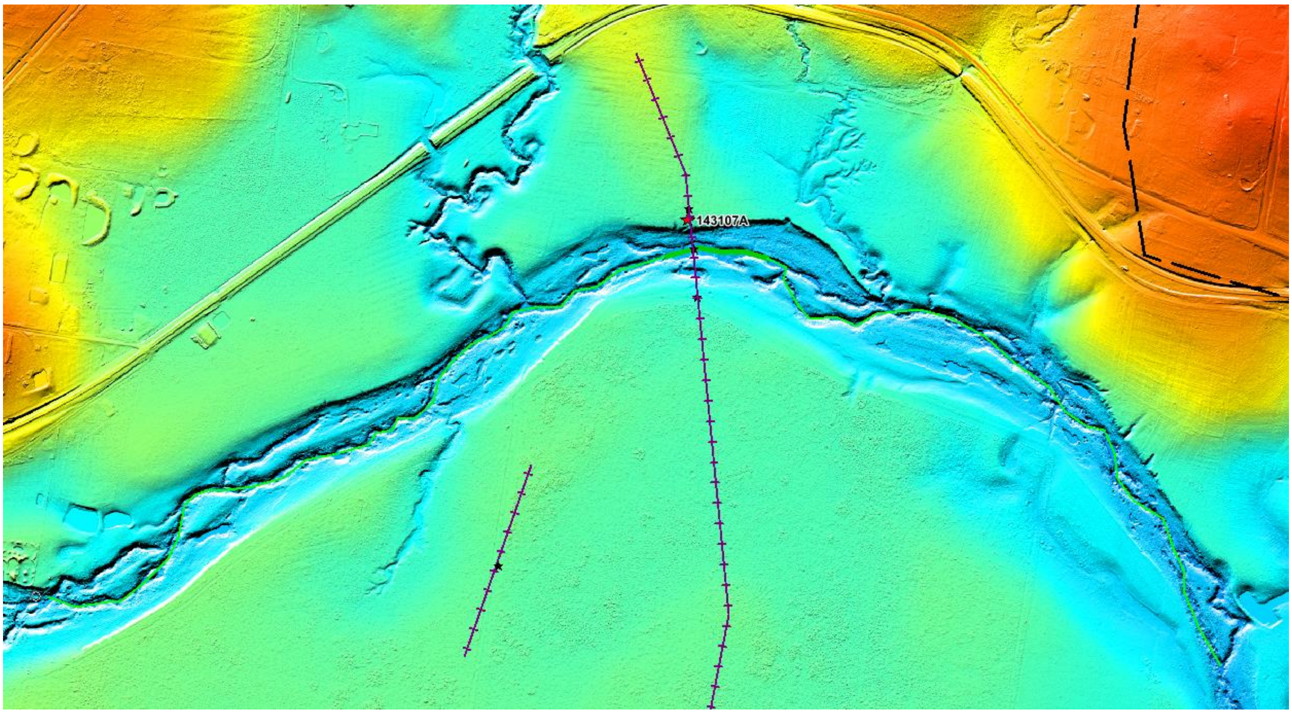
Areal Imagery with Material Polygons



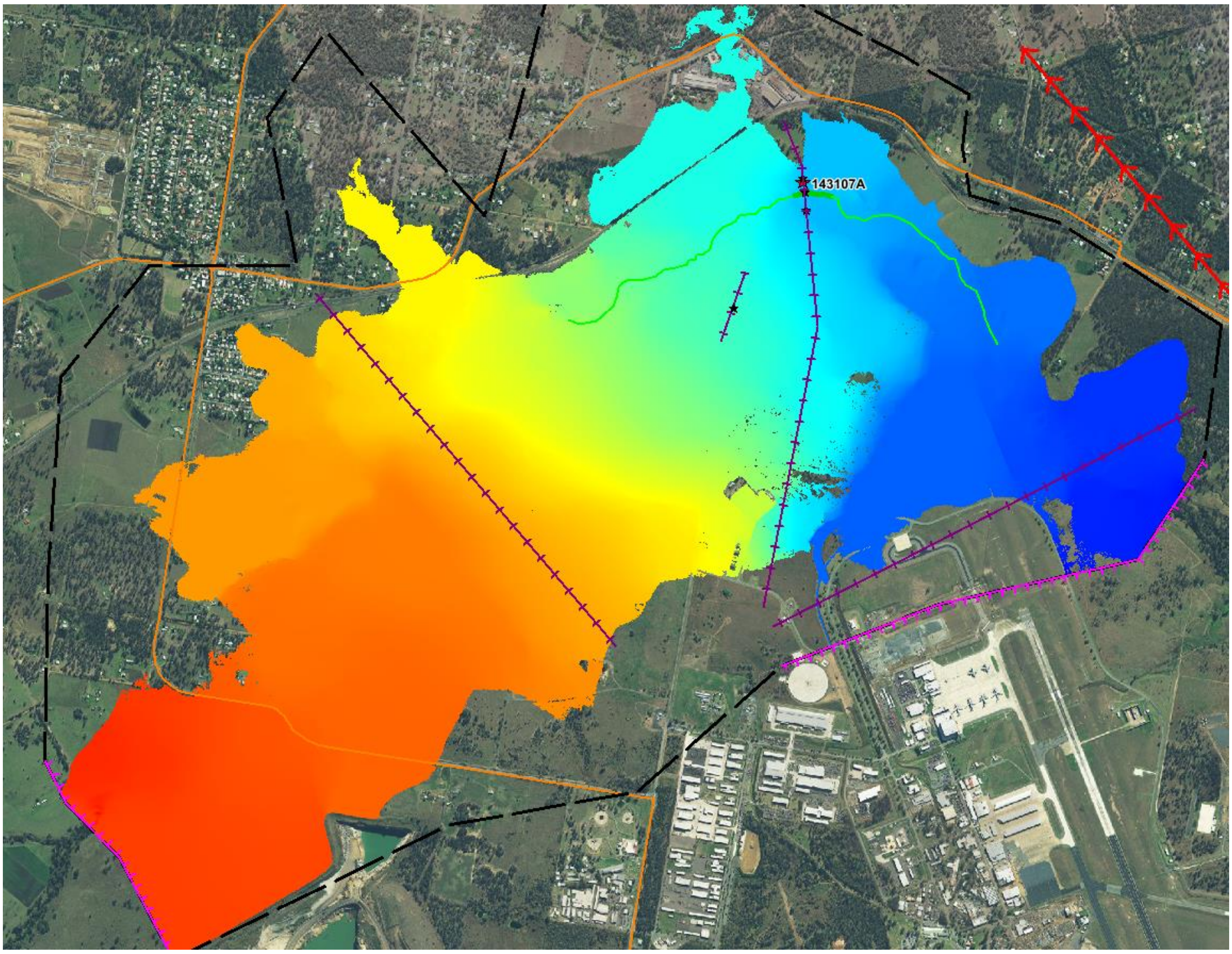
Areal Imagery with Material Polygons – Zoomed in at gauge



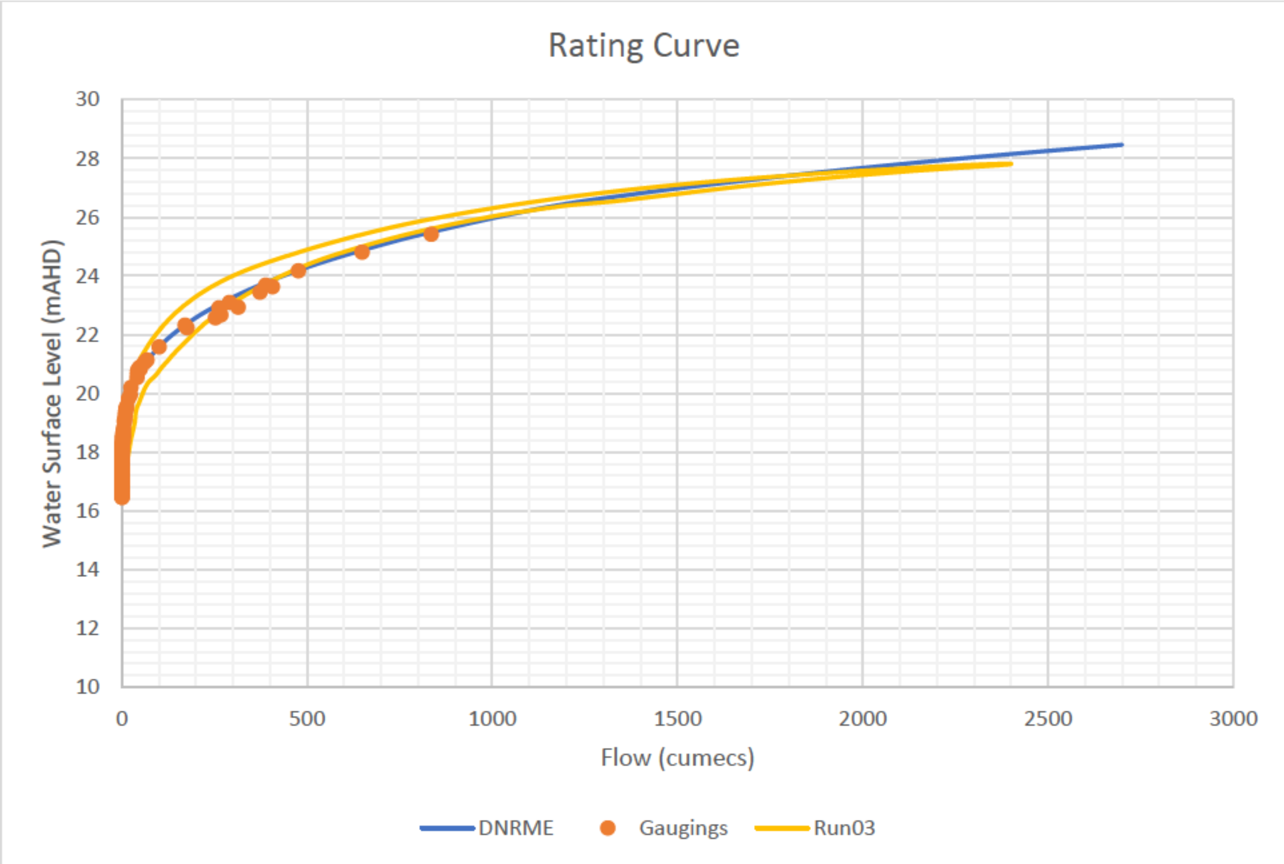
Terrain with relief shading



Terrain with relief shading – Zoomed in at gauge



Maximum water surface height



Rating Curve graph

C.7 Gauge 145011A

Name:	Teviot Brook at Croftby
Longitude:	-28.148048
Latitude:	152.57005
Control:	Control weir
Catchment Area:	83 km^2
Maximum Gauging Stage:	3.575m
Maximum Record Stage:	7.27m
Zero Gauge Height:	161.394m AHD

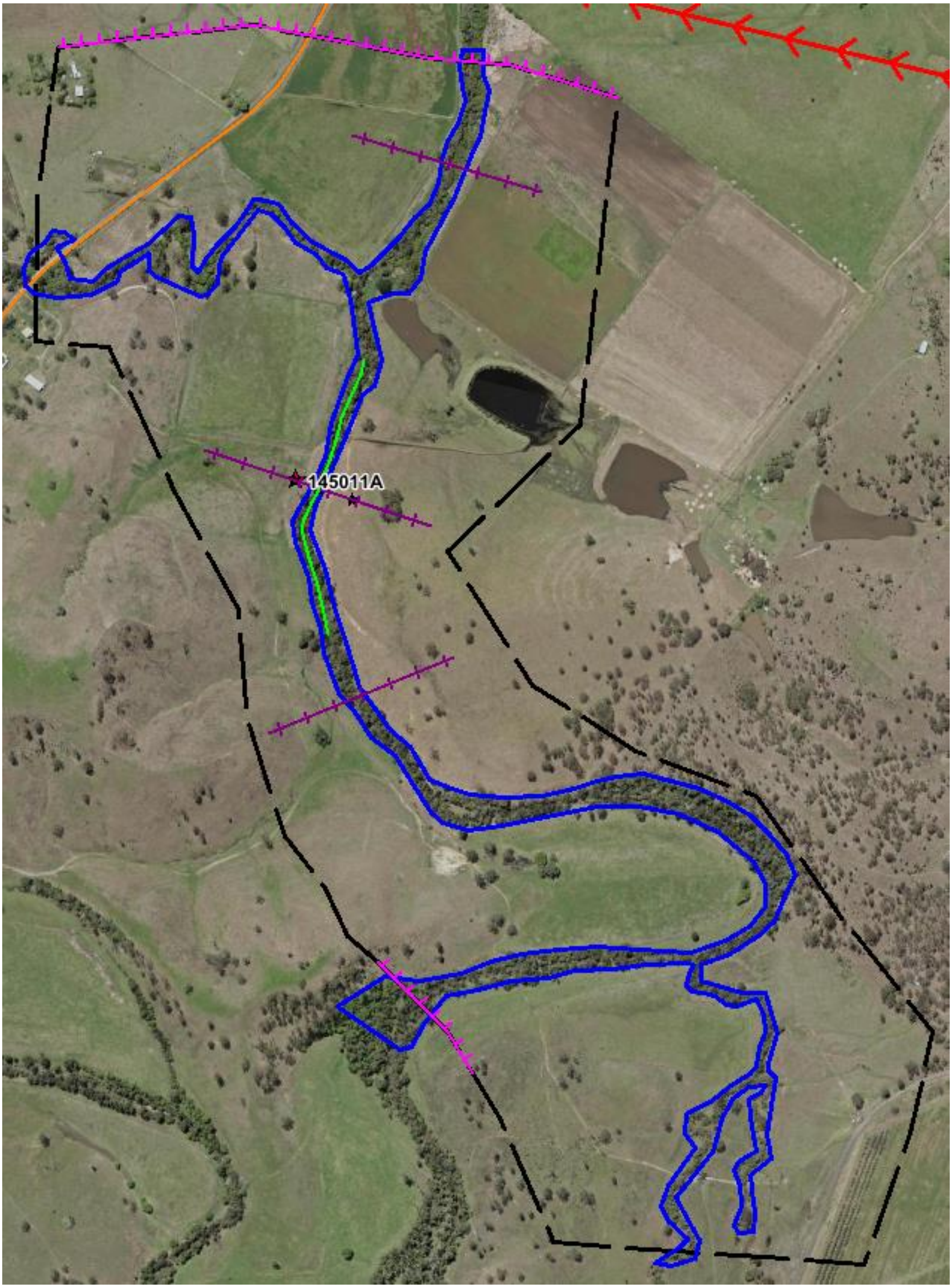
Comments

Completed by:	Samuel Walker
Action:	Adopt TUFLOW rating curve

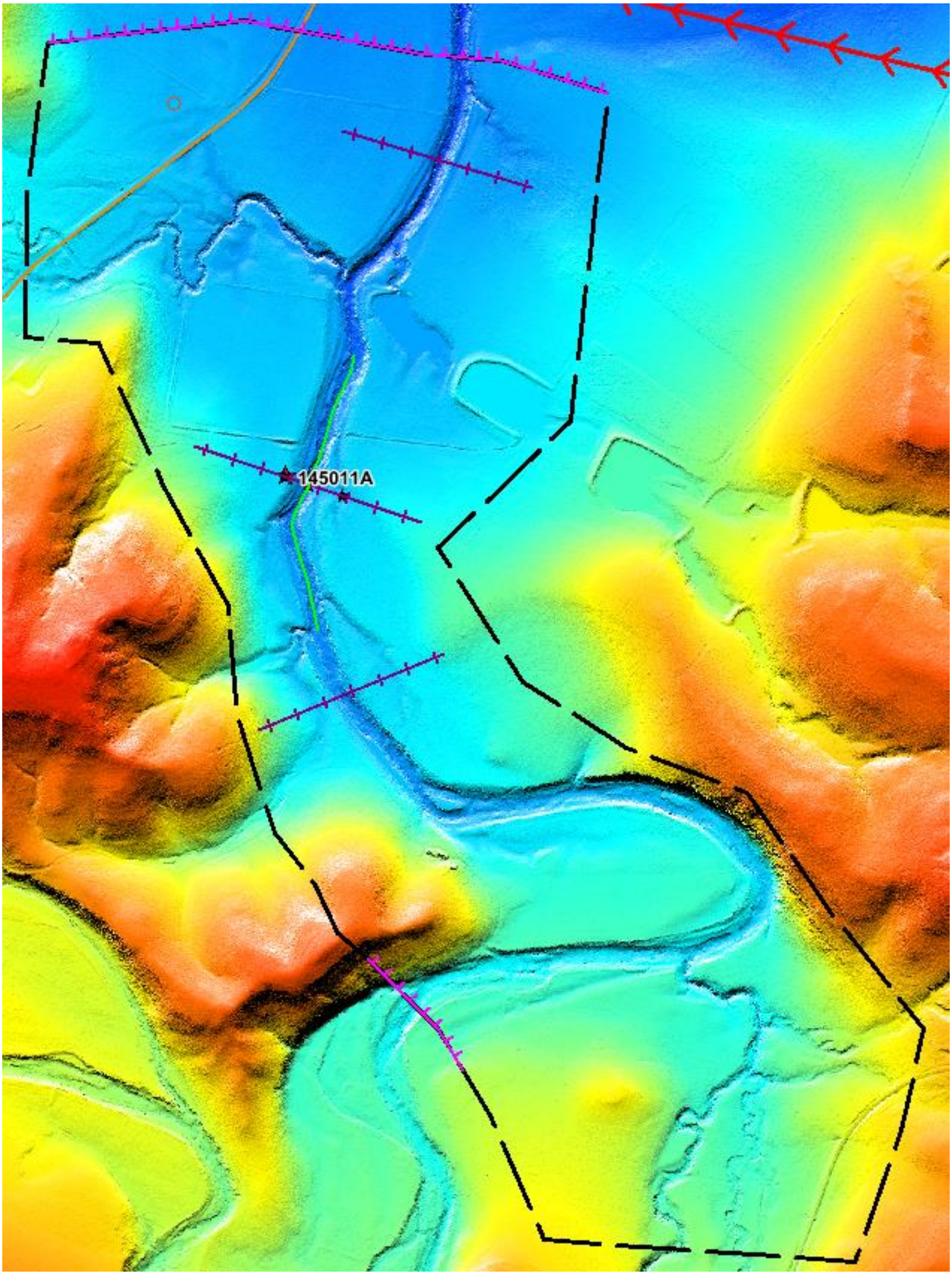
- Stream was rather narrow and site photos showed vegetated therefore creek had only one roughness assigned instead of the bed and riparian areas separated.
- Stream was lowered (see light green lines on maps); -0.5m deep, 5m wide.

Materials

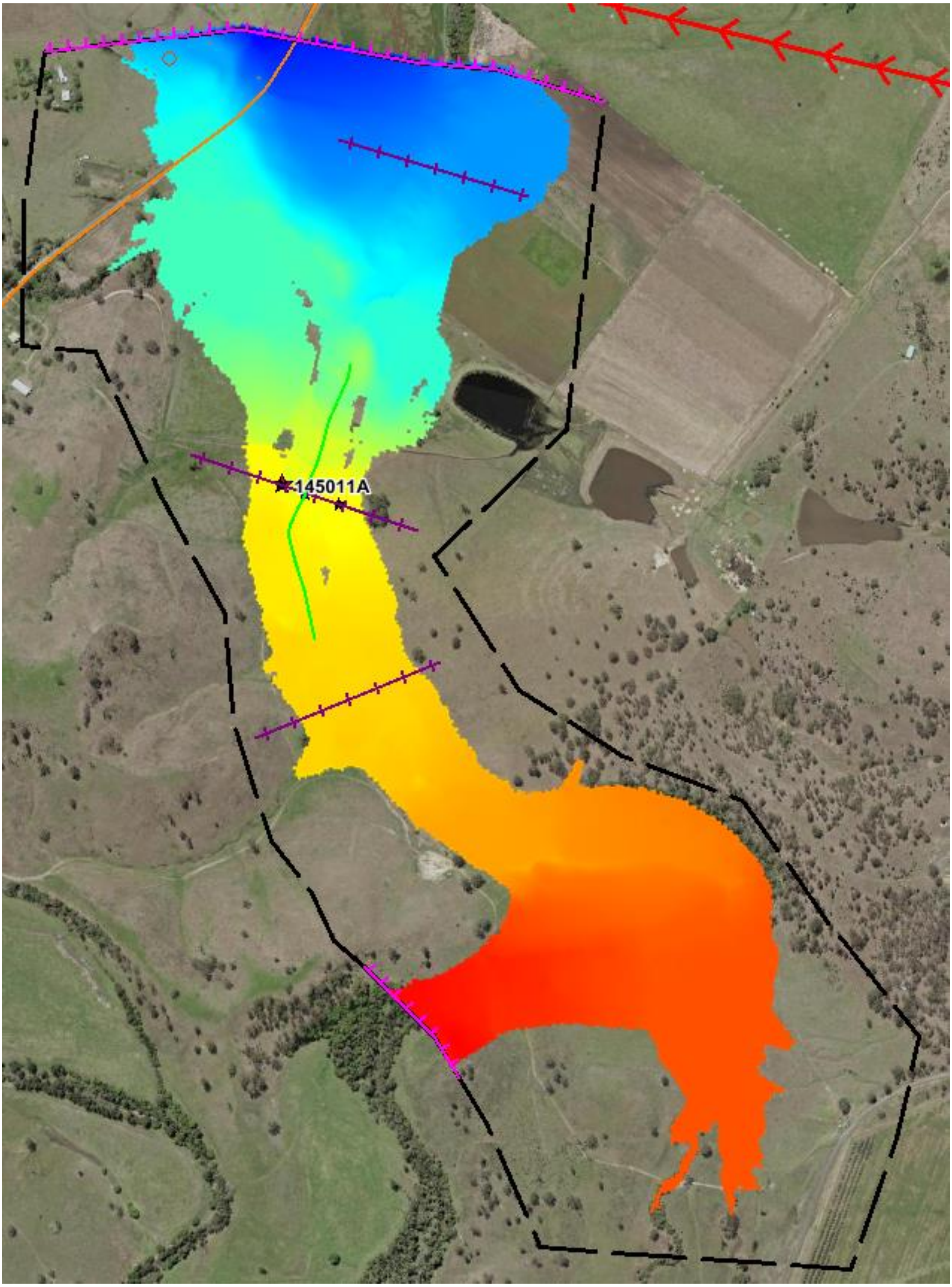
Index	Roughness	Colour	Description
1	0.04	Default	Pasture, and few trees
2	0.06	Blue	Creek and riparian



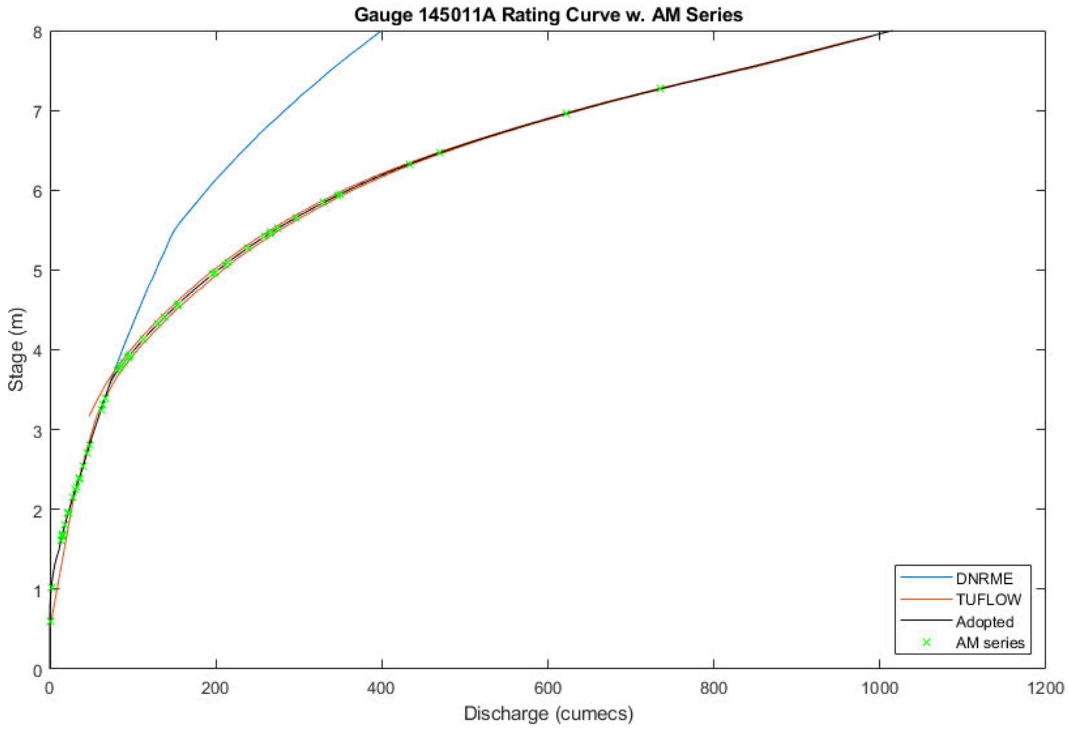
Areal Imagery with Material Polygons



Terrain with relief shading



Maximum water surface height



Rating Curve graph

C.8 Gauge 922101B

Name:	Coen River at Racecourse
Longitude:	-13.955092
Latitude:	143.174689
Control:	Sand and Rock Outcrops
Catchment Area:	172 km^2
Maximum Gauging Stage:	2.71m
Maximum Record Stage:	10.416m
Zero Gauge Height:	168.676m AHD (Adopted 174.6m AHD)

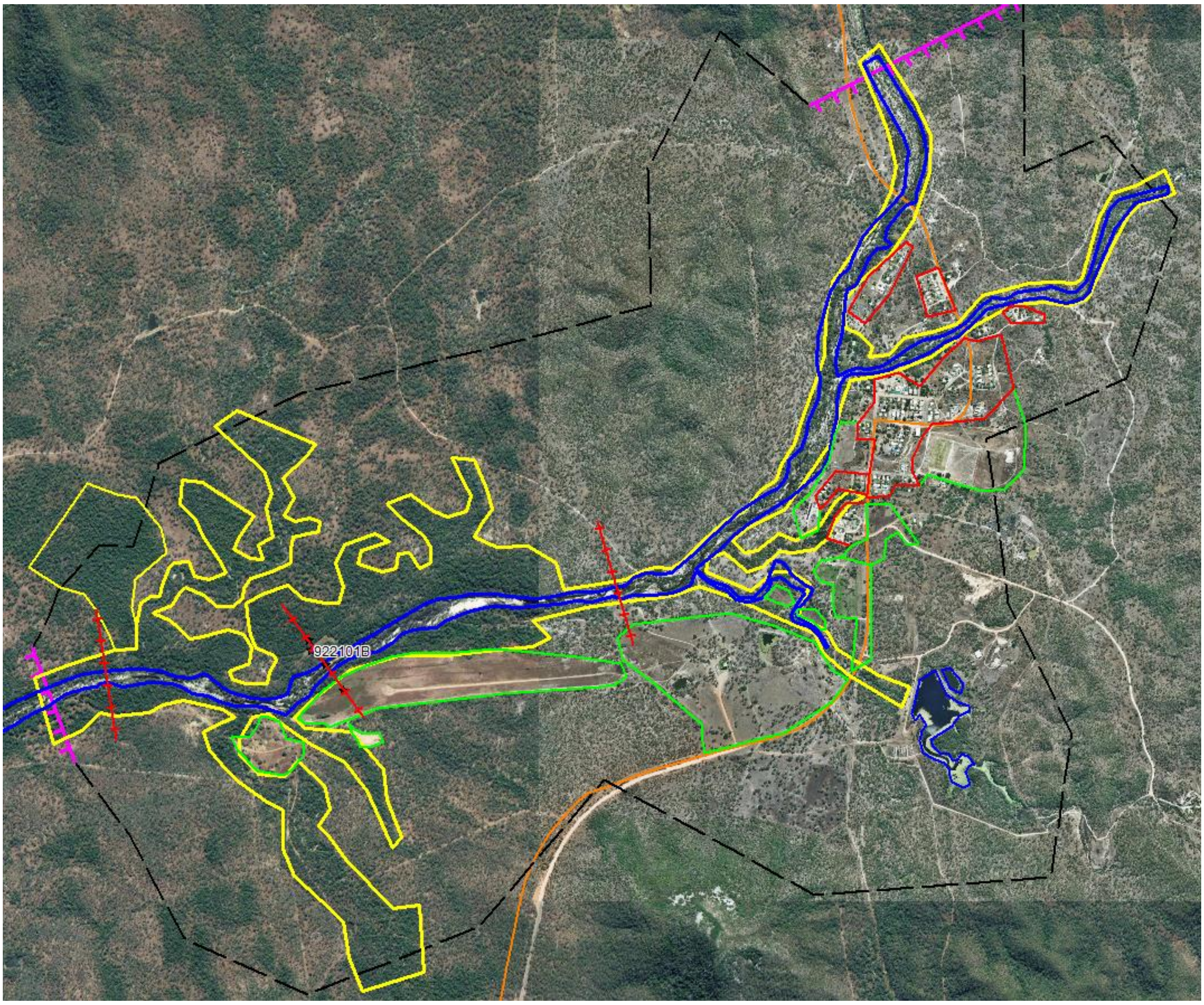
Comments

Completed by:	Samuel Walker
Action:	Adopt TUFLOW rating curve

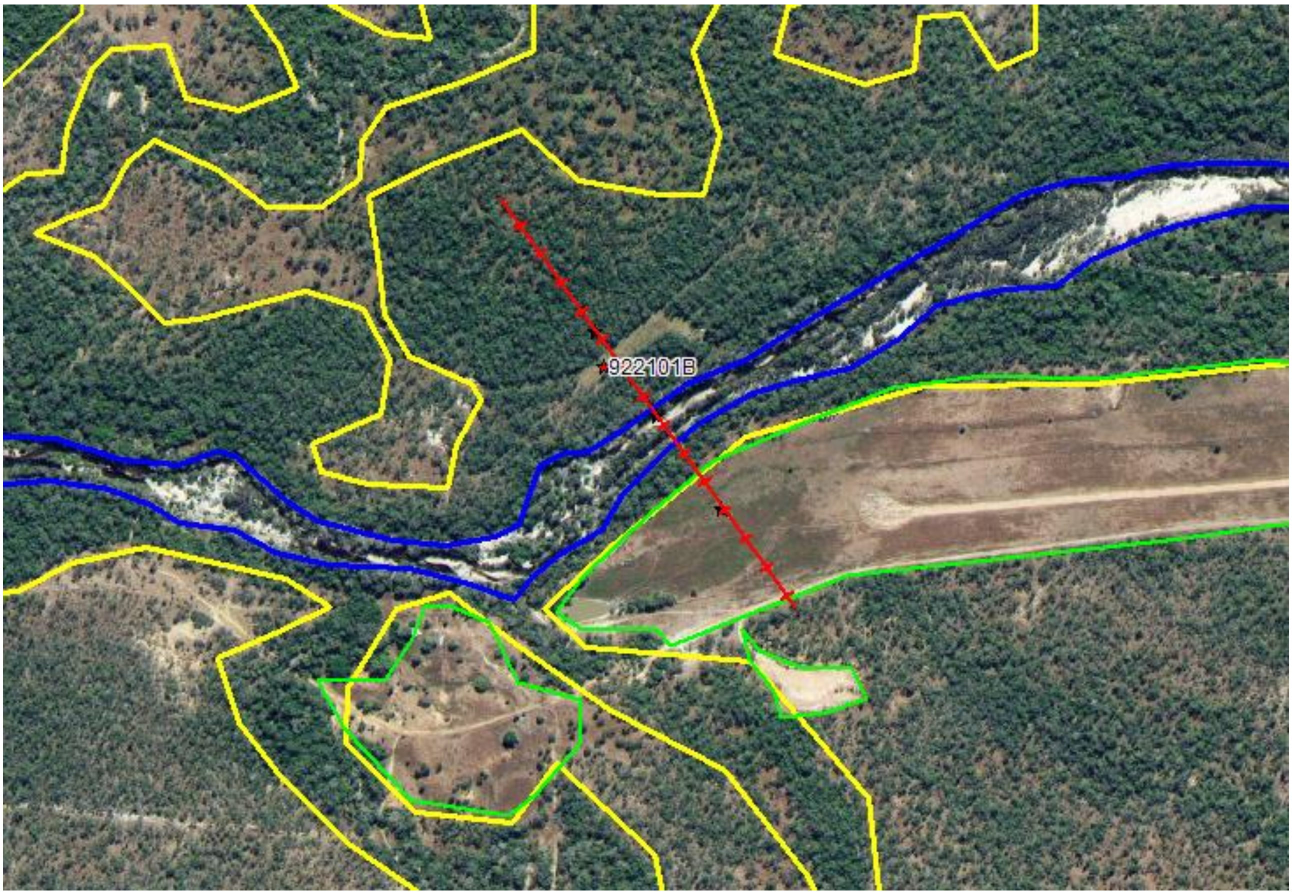
- Lidar was checked against survey marks and found that Zero Gauge Height reported on WMIP was wrong. This was confirmed by DNRME.

Materials

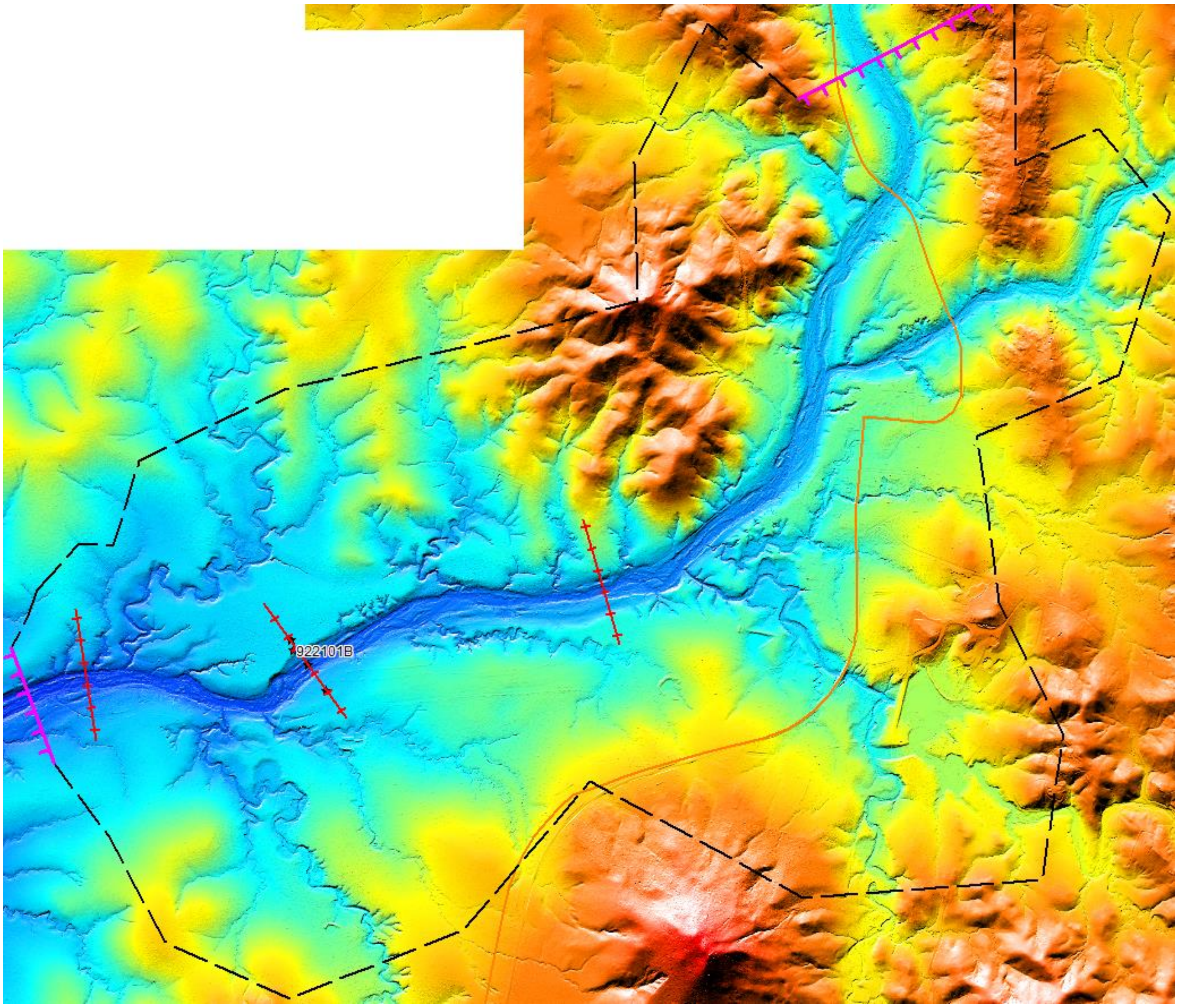
Index	Roughness	Colour	Description
1	0.06	Default	Medium density forested
2	0.08	Blue	Creek and water
3	0.03	Green	Cleared, ovals, minimal to no trees
4	0.15	Yellow	Riparian areas, and high-density vegetation
5	0.05	Red	Residential areas



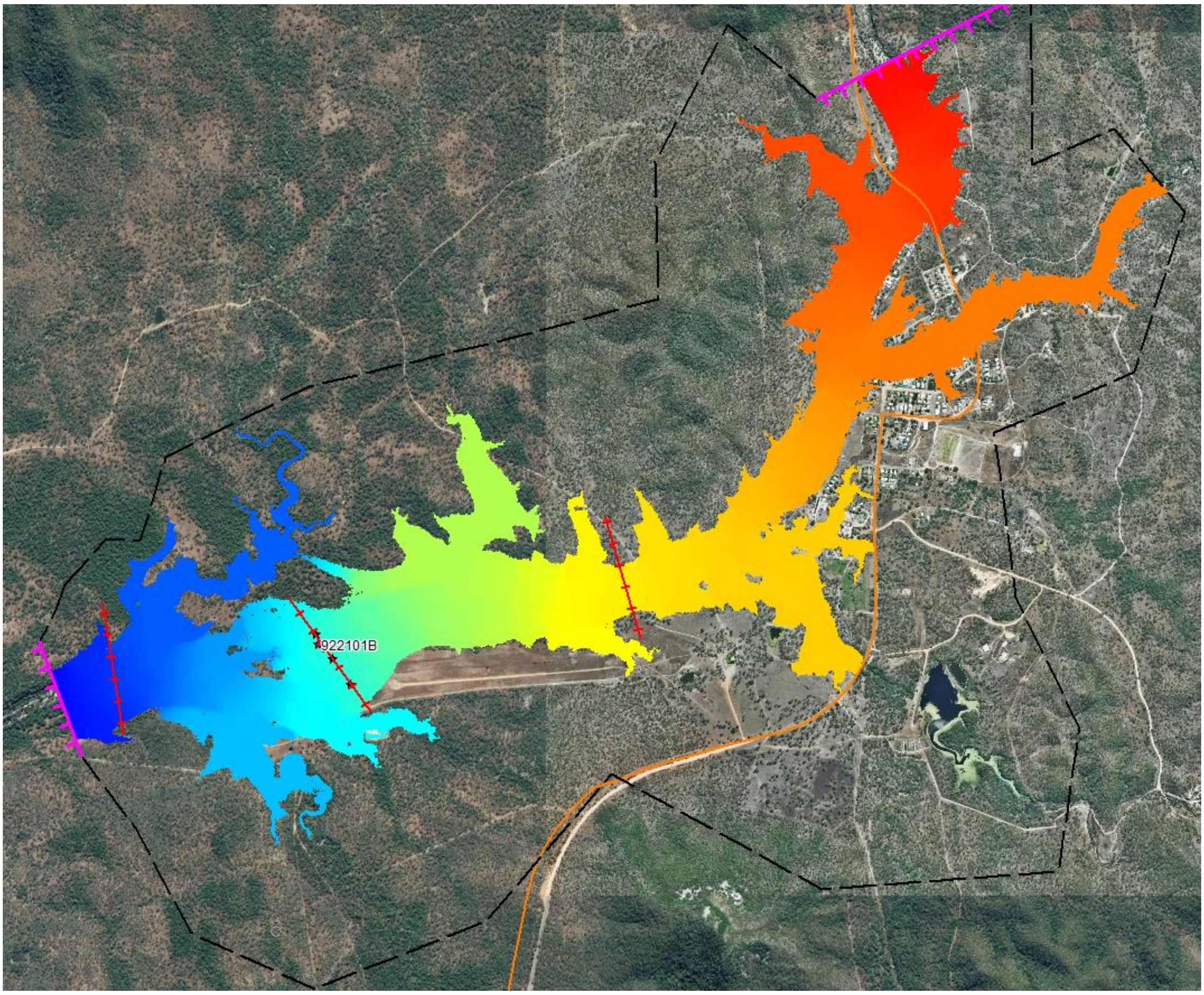
Areal Imagery with Material Polygons



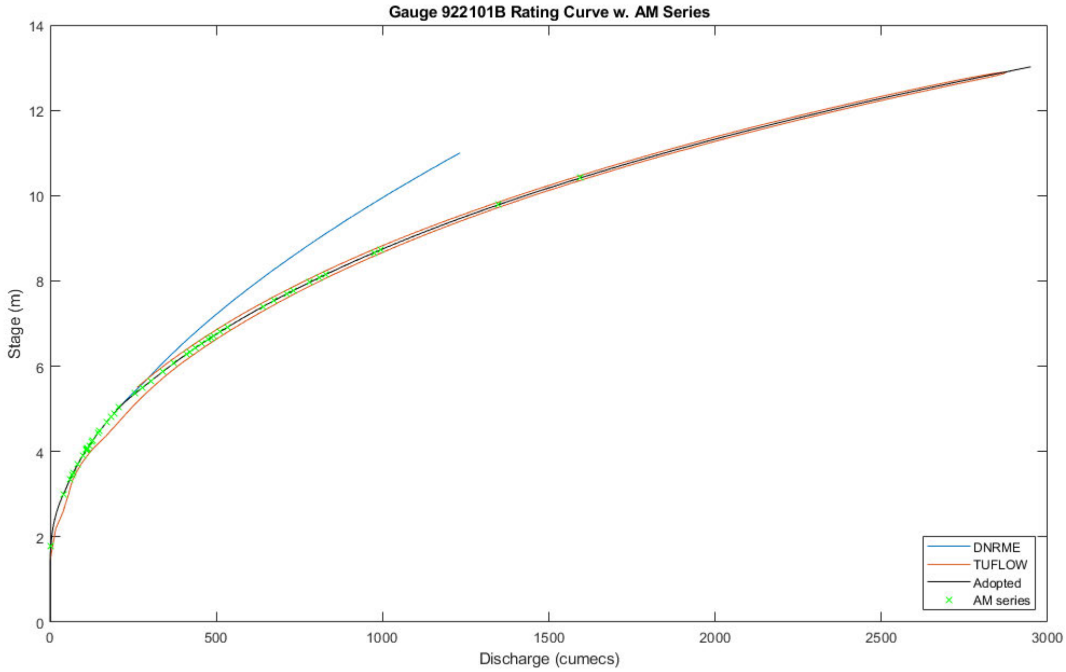
Areal Imagery with Material Polygons – Zoomed in to gauge



Terrain with relief shading



Maximum water surface height

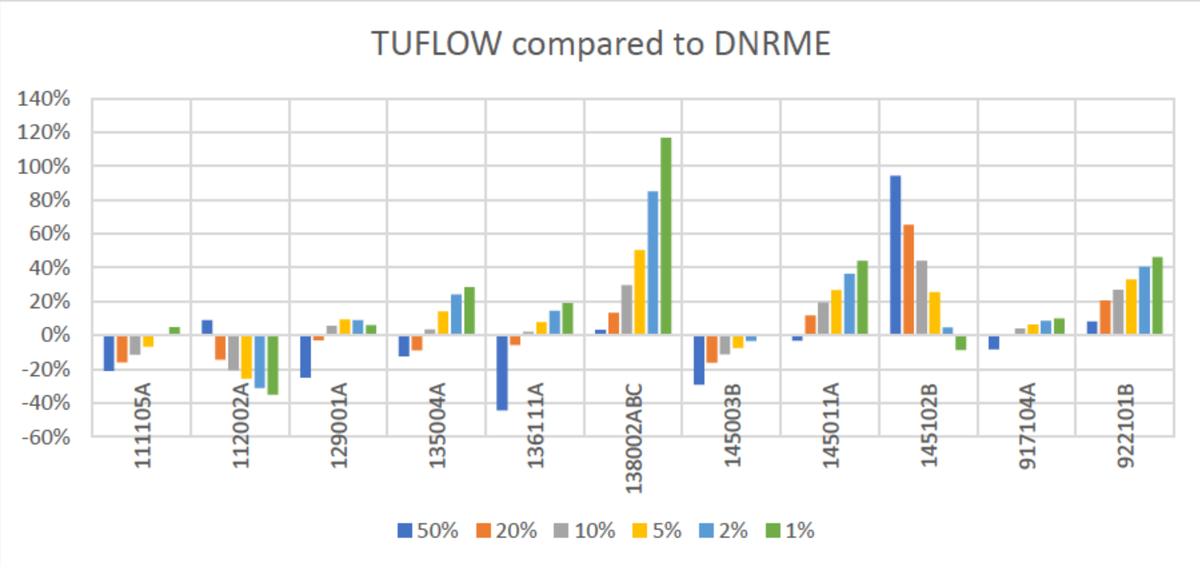


Rating Curve graph

Appendix D – Rating Curve Comparison

Quantile comparison of Revised rating curve to Existing rating curve

Percentage difference

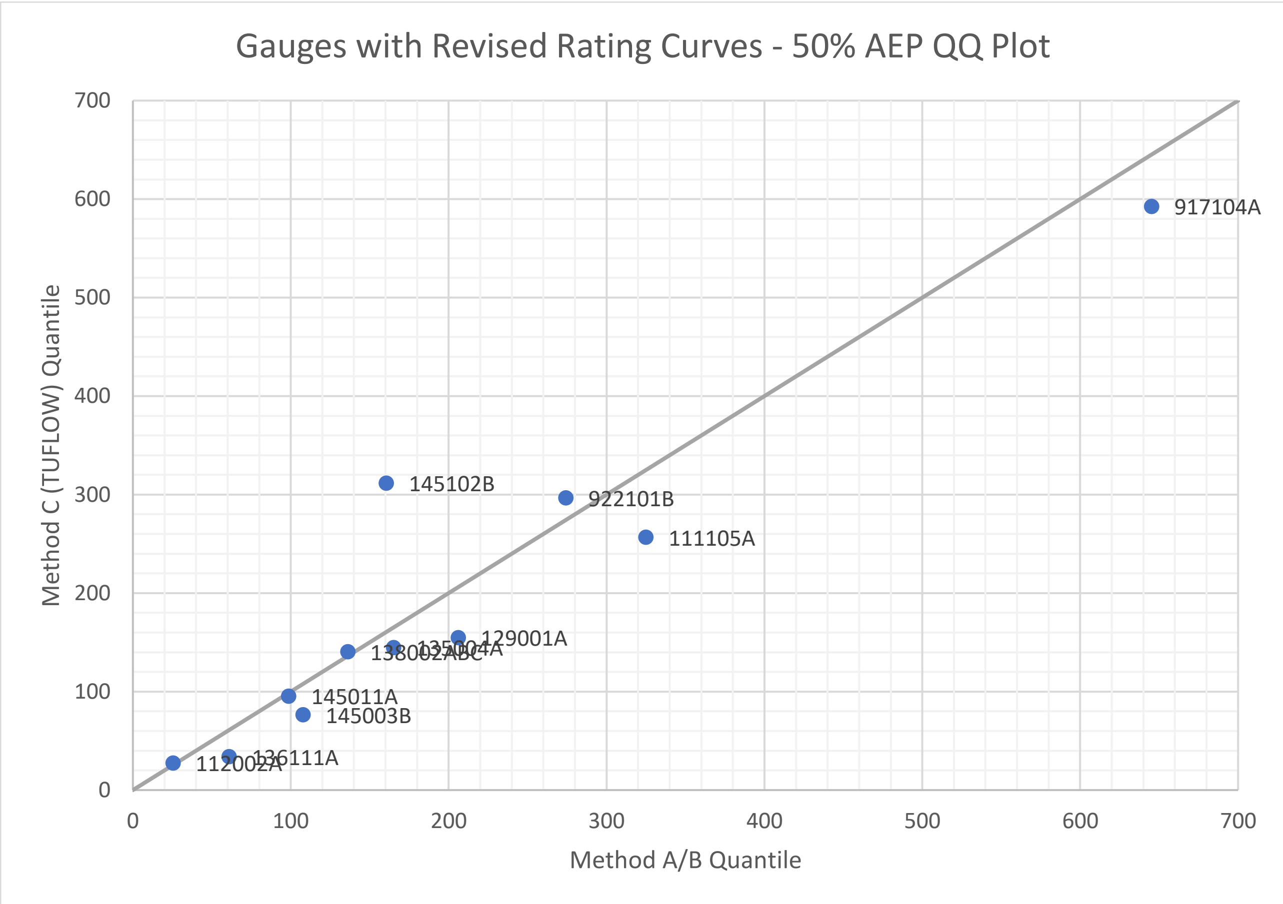


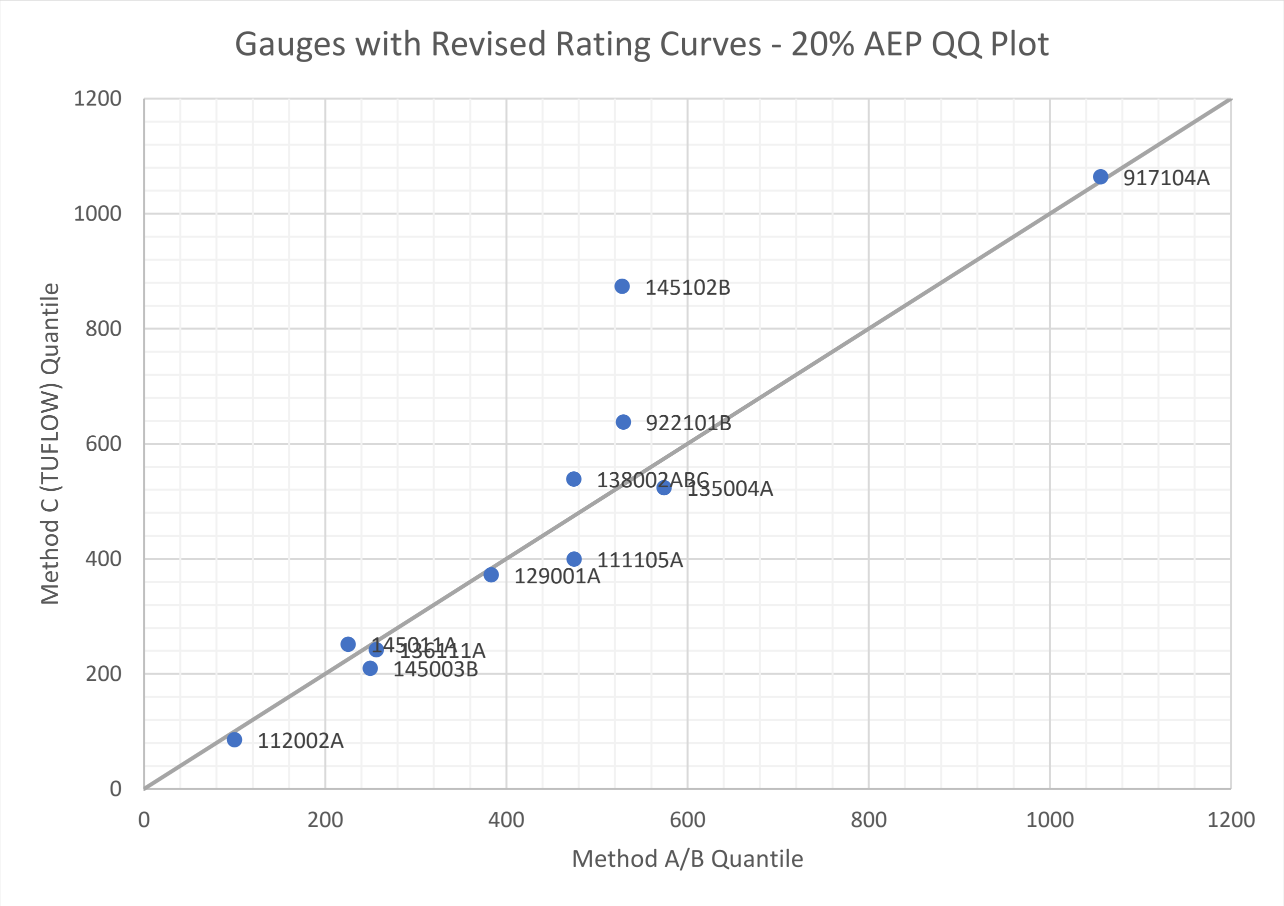
Discharge quantiles (1%, 2%, and 5% AEP)

	1% AEP		2% AEP		5% AEP	
	Method A/B	Method C	Method A/B	Method C	Method A/B	Method C
111105A	718.51	753.06	679.82	679.28	615.5	575.59
112002A	350	227.8	287	197.8	209.3	156
129001A	1631.8	1728.57	1196.2	1305.04	782	855.42
135004A	2877.2	3696.47	2141.1	2657.98	1385.7	1582.13
136111A	534.1	635.6	493.5	565	422.6	455.9
138002ABC	2195.02	4760.43	1754.97	3247.65	1198	1803.03
145003B	1200.2	1197.1	874.7	846.5	559.9	518.4
145011A	524	755.2	462.2	630.9	374.8	475.2
145102B	2043	1867.9	1620.9	1697.5	1137.5	1426.8
917104A	2330.4	2564.6	2022.6	2197.3	1631.2	1734.7
922101B	1096.3	1604.4	983.2	1382.3	819.6	1090

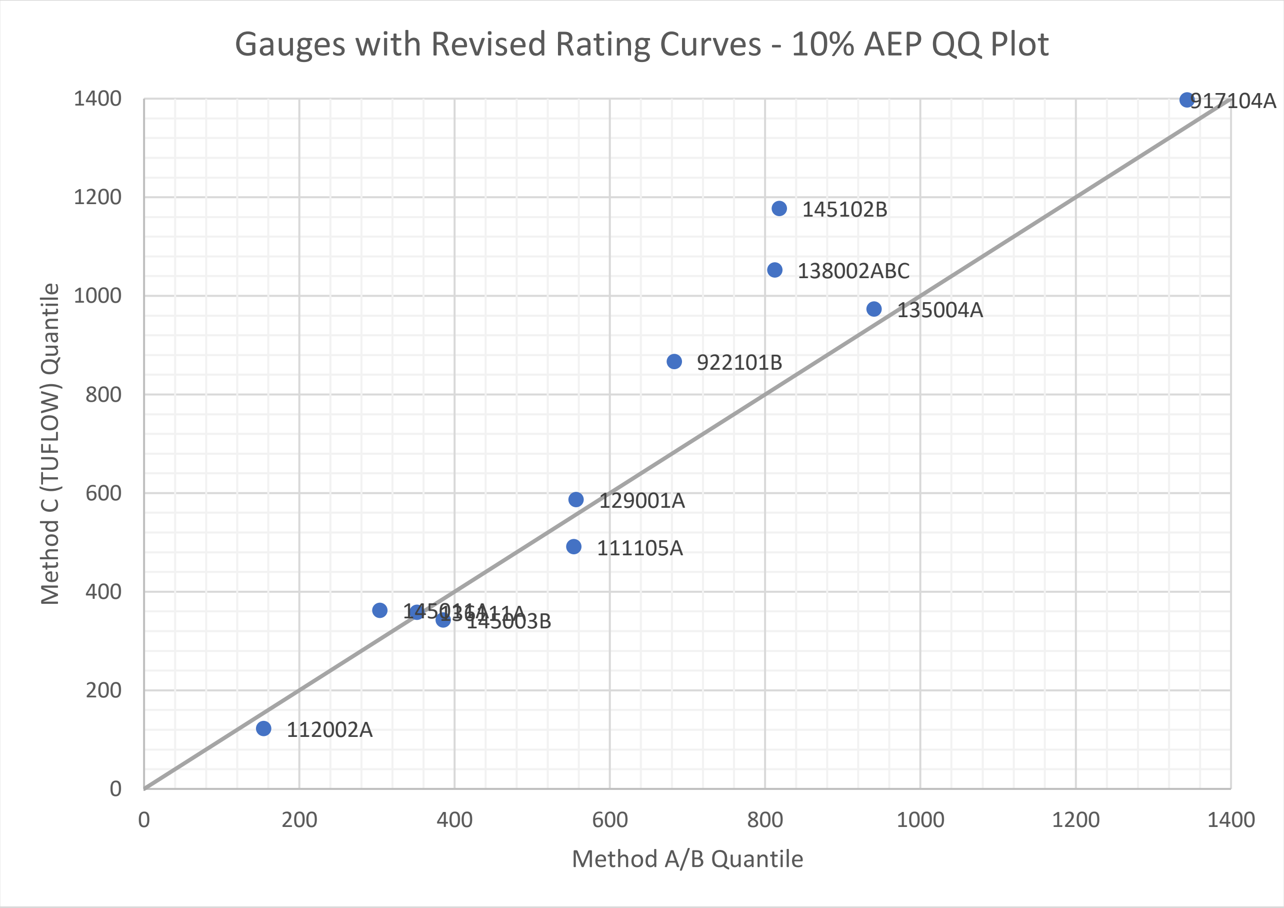
Discharge quantiles (10%, 20%, and 50% AEP)

	10% AEP		20% AEP		50% AEP	
	Method A/B	Method C	Method A/B	Method C	Method A/B	Method C
111105A	553.37	491.21	474.46	399.47	324.8	256.62
112002A	153.8	122.3	99.8	85.7	25.3	27.6
129001A	556.1	587.26	383	372.03	205.9	154.86
135004A	940	973.44	573.8	523.65	165.1	144.7
136111A	351.1	358.2	256.3	242	60.9	34
138002ABC	812.07	1052.86	474.37	538.39	135.95	140.59
145003B	385.2	342.6	249.4	209.8	107.7	76.4
145011A	303.4	362.4	225	251.3	98.5	95.5
145102B	817.9	1177.5	527.8	873.6	160.4	311.8
917104A	1343.3	1397.6	1055.8	1064	645.1	592.7
922101B	682.6	867.2	528.9	637.8	274.1	296.7

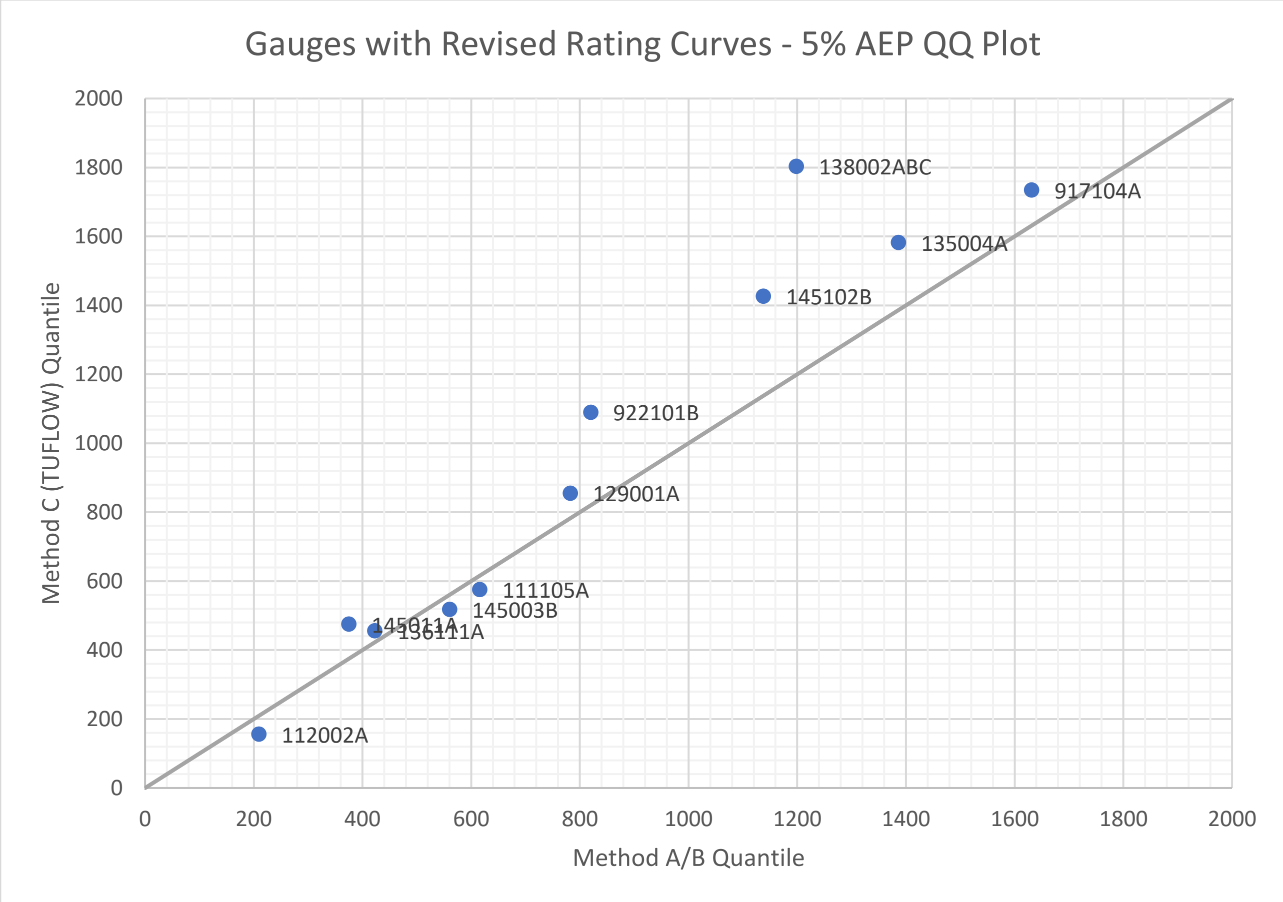




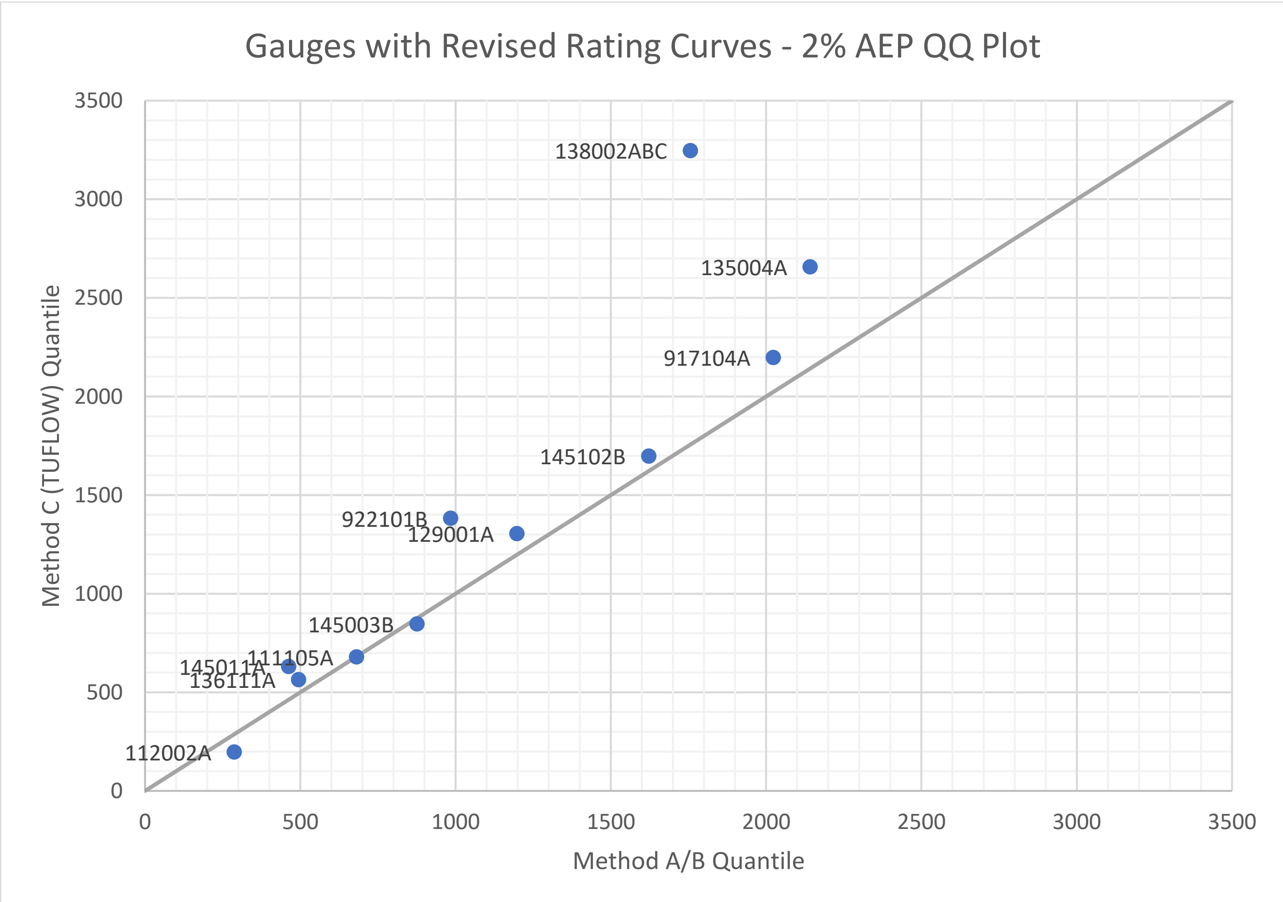
Gauges with Revised Rating Curves - 20% AEP QQ Plot
Method C (TUFLOW) Quantile
Method A/B Quantile
917104A
145102B
922101B
138002ABC
135004A
111105A
129001A
145011A
136111A
145003B
112002A



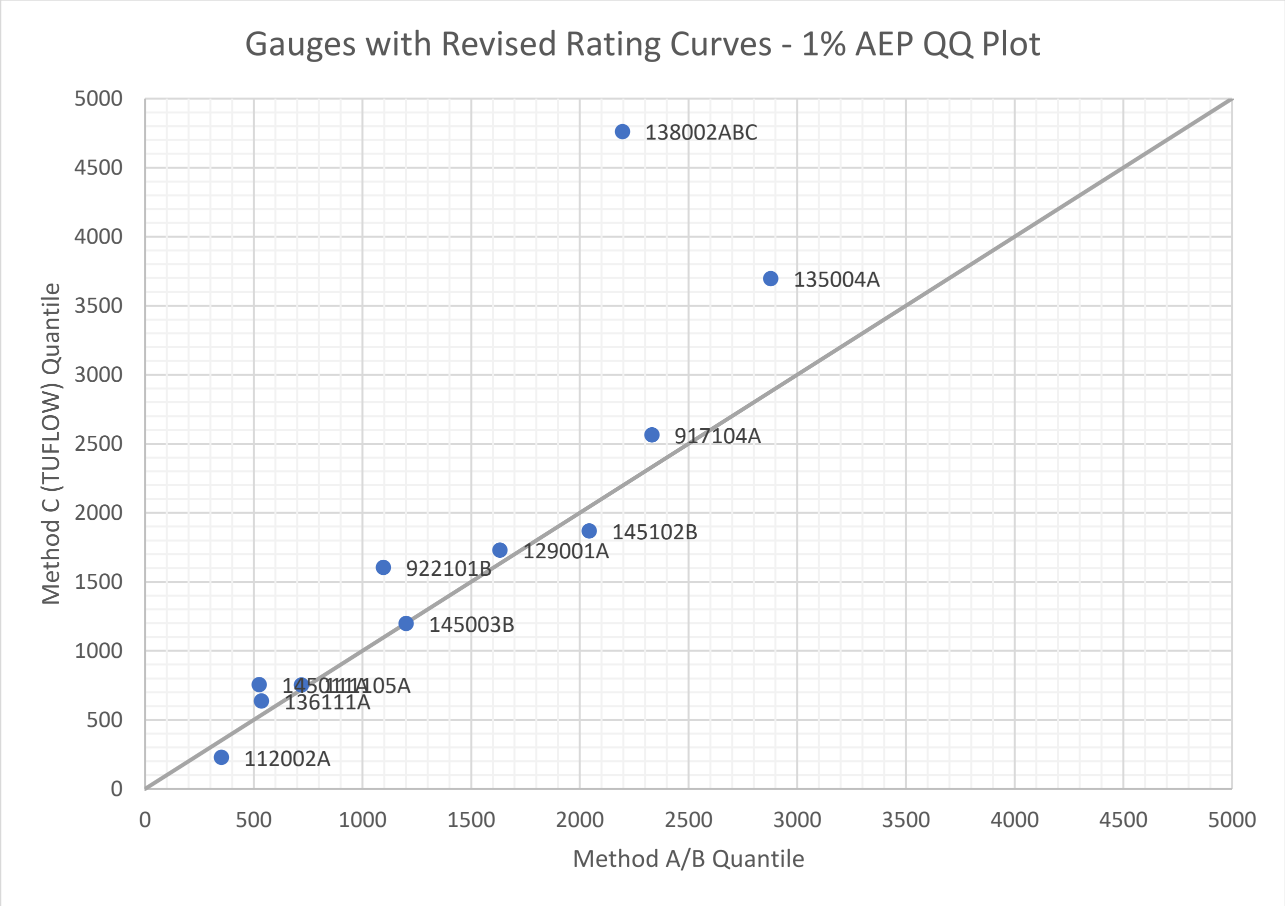
Gauges with Revised Rating Curves - 10% AEP QQ Plot
Method C (TUFLOW) Quantile
Method A/B Quantile
917104A
145102B
138002ABC
135004A
922101B
129001A
111105A
145011A
136111A
145003B
112002A



Gauges with Revised Rating Curves - 5% AEP QQ Plot
Method C (TUFLOW) Quantile
Method A/B Quantile
0
200
400
600
800
1000
1200
1400
1600
1800
2000
138002ABC
917104A
135004A
145102B
922101B
129001A
111105A
145003B
145011A
136111A
112002A



Gauges with Revised Rating Curves - 2% AEP QQ Plot
Method C (TUFLOW) Quantile
Method A/B Quantile
0
500
1000
1500
2000
2500
3000
3500
138002ABC
135004A
917104A
145102B
922101B
129001A
145003B
145011A
111105A
136111A
112002A


Gauges with Revised Rating Curves - 1% AEP QQ Plot
Method C (TUFLOW) Quantile
Method A/B Quantile
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
138002ABC
135004A
917104A
145102B
129001A
922101B
145003B
145011A
111105A
136111A
112002A

Appendix E – Flood Frequency Analysis Results

FFA and Distribution Selection Data

Station	AM Series Method	Selected Distribution	Years of Fitting Data	Years Censored	Poor Years	RFFE skew	Manual Input skew
102101A	A	GEV - Shift 2	44	0	7	-0.128	-0.45685
105105A	A	GEV - Shift 4	42	6	2	-0.122	-1.0485
107001AB	A	LP3 - uncensored	53	2	5	-0.127	-0.45471
108002A	A	GEV - Shift 0	47	1	2	-0.123	-0.58363
110002A	A	GEV - Shift 4	69	29	5	-0.119	-0.79164
111005A	A	LP3 - uncensored	46	0	6	-0.115	-0.52165
111105A	C	LP3 - uncensored	42	0	10	-0.03	-0.4952
112002A	C	GEV - Shift 4	44	43	3	-0.028	-1.09158
112003A	A	GEV - Shift 4	60	0	0	-0.027	-0.23249
112004A	A	LP3 - uncensored	50	0	2	-0.025	-0.74224
113004A	A	LP3 - uncensored	46	0	6	-0.029	-0.14919
116008B	A	GEV - Shift 4	31	14	20	-0.03	-1.3286
116010A	A	GEV - Shift 4	56	0	2	-0.025	-0.69404
116011A	A	GEV - Shift 4	55	0	3	-0.027	-0.93169
116013A	A	LP3 - RFFE	43	13	1	-0.04	-0.88689
116014A	A	GEV - Shift 1	43	10	4	-0.145	-0.54699
120102A	A	GEV - Shift 4	42	9	0	-0.011	-1.14744
120307A	A	GEV - Shift 4	32	17	1	-0.01	-1.19685
121001A	A	LP3 - censored	58	3	1	-0.034	-0.15458
125002C	A	GEV - Shift 3	27	28	6	-0.041	-1.38476
129001A	C	LP3 - uncensored	64	0	3	0.098	-0.01687
130207A	A	GEV - Shift 3	44	9	1	0.064	-1.10047
130407A	B	GEV - Shift 2	22	9	22	0.039	-1.47474
135004A	C	LP3 - uncensored	52	0	1	0.131	-0.36722
136006A	B	LP3 - censored	51	1	1	0.076	-0.82062
136108A	A	LP3 - censored	31	24	1	0.101	-0.98387
136111A	C	GEV - Shift 4	42	11	1	0.112	-1.00412
136202D	A	LP3 - uncensored	51	2	1	0.14	-0.80021
136301B	A	LP3 - uncensored	40	10	3	0.119	-0.75926
137001AB	A	LP3 - censored	33	33	5	0.128	-1.35505
137101A	A	GEV - Shift 4	52	0	1	0.127	-0.71867
137201A	A	GEV - Shift 4	45	7	1	0.085	-1.03519
138002ABC	C	LP3 - uncensored	68	41	2	0.125	-0.71035
138110A	A	GEV - Shift 4	59	0	0	0.109	-0.69603
138111A	A	GEV - Shift 4	55	0	0	0.121	-0.89302
142202A	A	GEV - Shift 4	41	9	3	0.097	-1.49232
143107A	A	GEV - Shift 1	37	14	6	0.104	-0.99813
143108A	A	GEV - Shift 4	44	11	2	0.106	-0.77966

143212A	A	LP3 - censored	35	11	5	0.103	-1.09158
145003B	C	GEV - Shift 4	55	5	5	0.138	-0.31907
145011A	C	GEV - Shift 4	50	2	1	0.073	-1.07356
145102B	C	GEV - Shift 4	47	43	1	0.103	-1.07925
145103A	A	GEV - Shift 0	32	10	15	0.115	-1.10863
146010A	A	LP3 - censored	41	12	3	0.106	-0.95203
422211A	A	GEV - Shift 4	25	25	1		-1.62798
422306A	A	LP3 - uncensored	77	21	2	0.086	-0.90211
915011A	A	GEV - Shift 2	37	9	1	-0.028	-1.10315
917104A	C	GEV - Shift 1	37	10	5	-0.04	-1.11967
917107A	A	LP3 - censored	42	7	1	-0.024	-0.78161
917114A	A	GEV - Shift 0	34	3	9	-0.04	-1.07212
922101B	C	GEV - Shift 2	29	21	1	-0.131	-0.86612

The following quantiles include those completed for checking also.

Method A Quantile Data

Gauge	50% AEP	20% AEP	10% AEP	5% AEP	2% AEP	1% AEP
102101A	879.8	1549.9	1988.6	2405.9	2940.9	3338.1
105105A	261.3	452.9	549.5	624.6	701.4	747
107001AB	254.77	532.5	754.12	986.59	1309.05	1563
108002A	1156.2	1859.5	2349.4	2838.3	3499.9	4018
110002A	254.8	594.9	794.7	969.4	1173.2	1310.9
111005A	768.38	1485.6	2017.63	2548.76	3250	3778.47
111105A	324.8	474.46	553.37	615.5	679.82	718.51
112003A	83.7	300.5	479.6	684	1006.1	1298.5
112004A	1752.24	2949.47	3705.17	4378.77	5170.55	5706.8
113004A	156.3	289.34	395.13	508.41	671.41	805.52
116008B	431.3	832	1026.7	1173.8	1319.6	1403.3
116010A	151.1	430.6	613	786	1007.2	1171
116011A	67.5	288.1	423.3	545.4	693.2	796.8
116013A	285.13	633.58	955.98	1338.56	1948.63	2498.11
116014A	350	788.9	1167.5	1615	2348	3039.4
120102A	376.7	1154.3	1649.5	2110.3	2686.9	3104.6
120307A	246.7	453.4	599.9	748.2	951.9	1113.9
121001A	395.09	984.87	1665.97	2638.33	4553.37	6663.06
125002C	1172.4	2831.6	3625.6	4218.2	4798	5126.5
129001A	205.9	383	556.1	782	1196.2	1631.8
130207A	86.1	205.2	308.7	431.6	634.3	826.5
136108A	26.4	125.76	266.66	480.26	899.89	1341.03
136202D	73.61	340.52	658.54	1061.01	1694.45	2228.94
136301B	55.96	176.22	290.58	418.57	601.02	744.44
137001AB	80.6	383.33	634.18	846.42	1051.29	1156.55
137101A	80.5	335.6	493.9	638.3	814.8	939.8

137201A	132.2	606.3	928.5	1244.1	1662.1	1982.6
138002ABC	135.95	474.37	812.07	1198	1754.97	2195.02
138110A	369.7	1299	1898	2460.7	3172.2	3693.2
138111A	308.5	1606.6	2331.7	2941.3	3622	4062.5
142202A	217.5	466.3	645.8	830.1	1087.4	1295.2
143107A	249.9	565.7	810.7	1076.8	1473	1813.7
143108A	162.4	445.5	692.7	987.3	1475.1	1939.7
143212A	30.42	315.39	681.18	1067.29	1505.43	1757.49
145011A	98.5	225	303.4	374.8	462.2	524
145103A	62.5	147.2	228.7	333.7	523.5	720.2
146010A	128.55	344.07	534.38	741.28	1032.05	1260.06
422211A	17.5	126.7	195	257.7	334.8	389.8
422306A	30.27	99.01	162.21	229.87	320.86	388.12
915011A	91.1	154.5	207.6	268.8	366.3	455.9
917104A	645.1	1055.8	1343.3	1631.2	2022.6	2330.4
917107A	92.94	207.52	296.1	384.9	500.73	586.11
917114A	183.5	292.3	366.3	438.8	534.8	608.5
922101B	274.1	528.9	682.6	819.6	983.2	1096.3

Method B Quantile Data

Gauge	50% AEP	20% AEP	10% AEP	5% AEP	2% AEP	1% AEP
112002A	25.3	99.8	153.8	209.3	287	350
130407A	310.9	846.1	1245.6	1666.5	2272	2775.3
135004A	165.1	573.8	940	1385.7	2141.1	2877.2
136006A	70.69	240.89	453.5	761.48	1358.34	1992.62
136111A	60.9	256.3	351.1	422.6	493.5	534.1
145003B	107.7	249.4	385.2	559.9	874.7	1200.2
145102B	160.4	527.8	817.9	1137.5	1620.9	2043

Method C Quantile Data

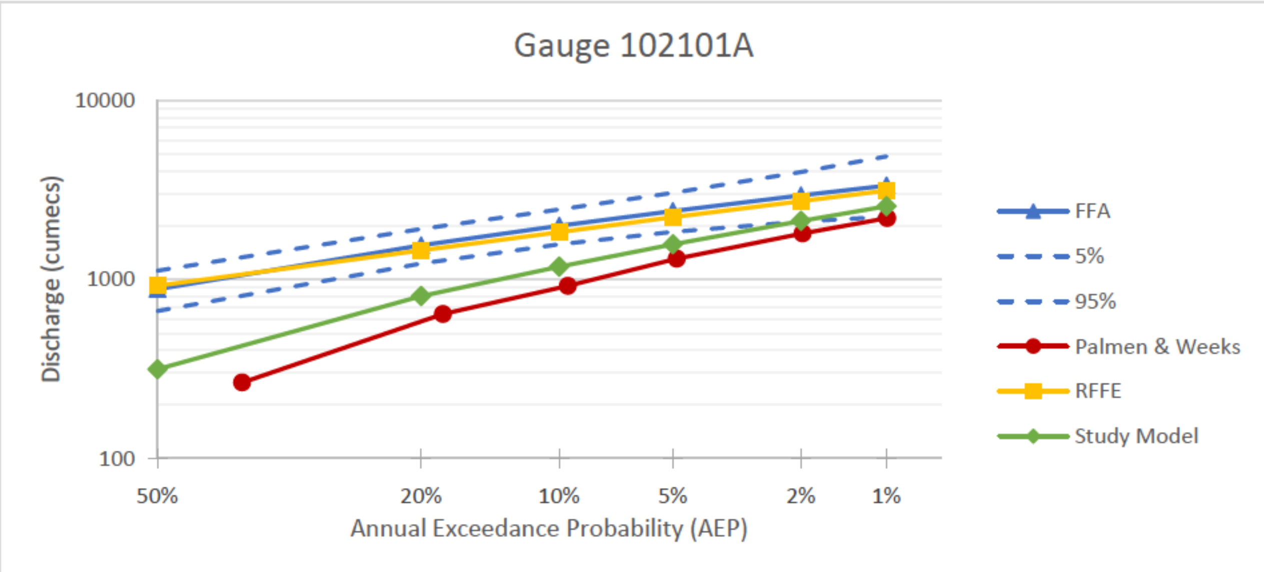
Gauge	50% AEP	20% AEP	10% AEP	5% AEP	2% AEP	1% AEP
111105A	256.62	399.47	491.21	575.59	679.28	753.06
112002A	27.6	85.7	122.3	156	197.8	227.8
129001A	154.86	372.03	587.26	855.42	1305.04	1728.57
135004A	144.7	523.65	973.44	1582.13	2657.98	3696.47
136111A	34	242	358.2	455.9	565	635.6
138002ABC	140.59	538.39	1052.86	1803.03	3247.65	4760.43
145003B	76.4	209.8	342.6	518.4	846.5	1197.1
145011A	95.5	251.3	362.4	475.2	630.9	755.2
145102B	311.8	873.6	1177.5	1426.8	1697.5	1867.9
917104A	592.7	1064	1397.6	1734.7	2197.3	2564.6
922101B	296.7	637.8	867.2	1090	1382.3	1604.4

Appendix F – Case Studies

Summary of superior model comparing to FFA

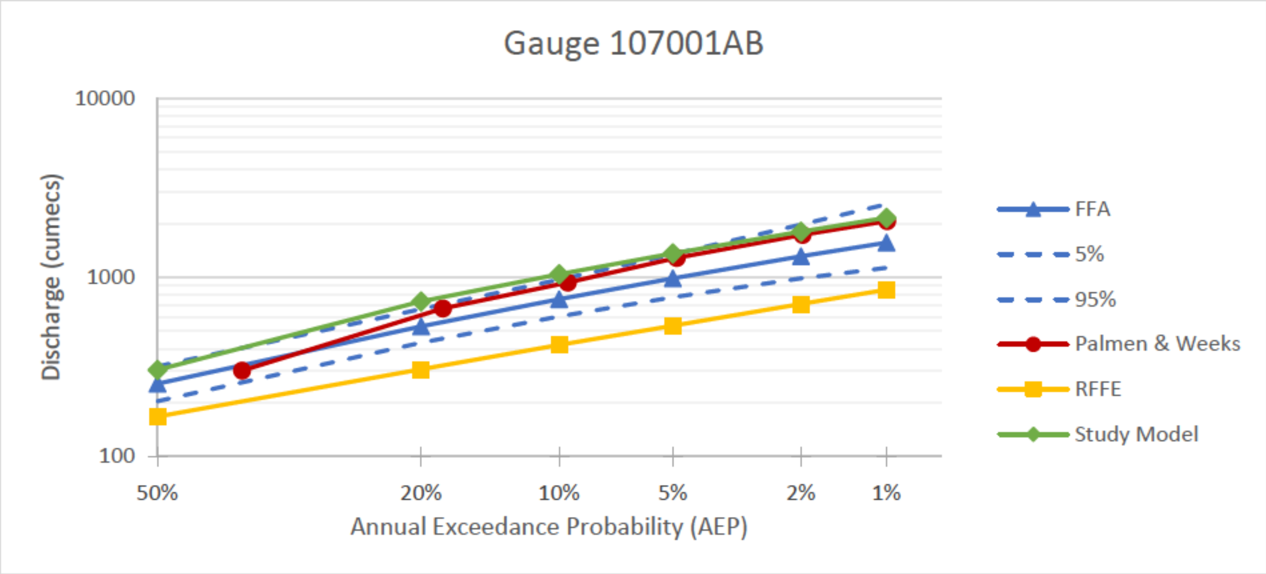
Gauge ID	Study Model	RFFE Model
102101A		✓
107001AB	✓	
111005A	✓	
116013A		✓
120307A		✓
136108A	✓	
136202D	✓	
138110A	✓	
143107A		✓
143108A		✓
146010A	✓	

102101A



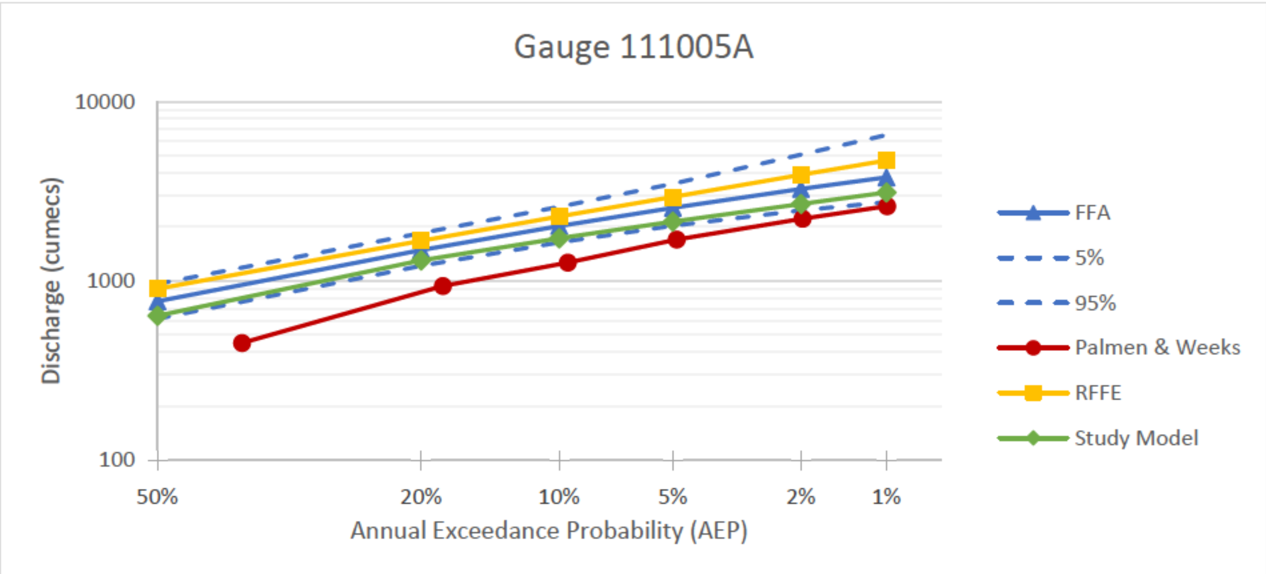
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	879.8	664.8	1116.8	920	314.2956		39.4%	264.602
20%	1549.9	1225.6	1906.8	1450	806.3086		18.1%	638.7411
10%	1988.6	1566.2	2452	1830	1177.491		9.52%	917.6872
5%	2405.9	1838	3038.6	2220	1565.302		4.88%	1304.524
2%	2940.9	2097.3	3970.5	2730	2112.909		1.98%	1803.103
1%	3338.1	2239	4859.7	3130	2559.765		1%	2197.377

107001AB



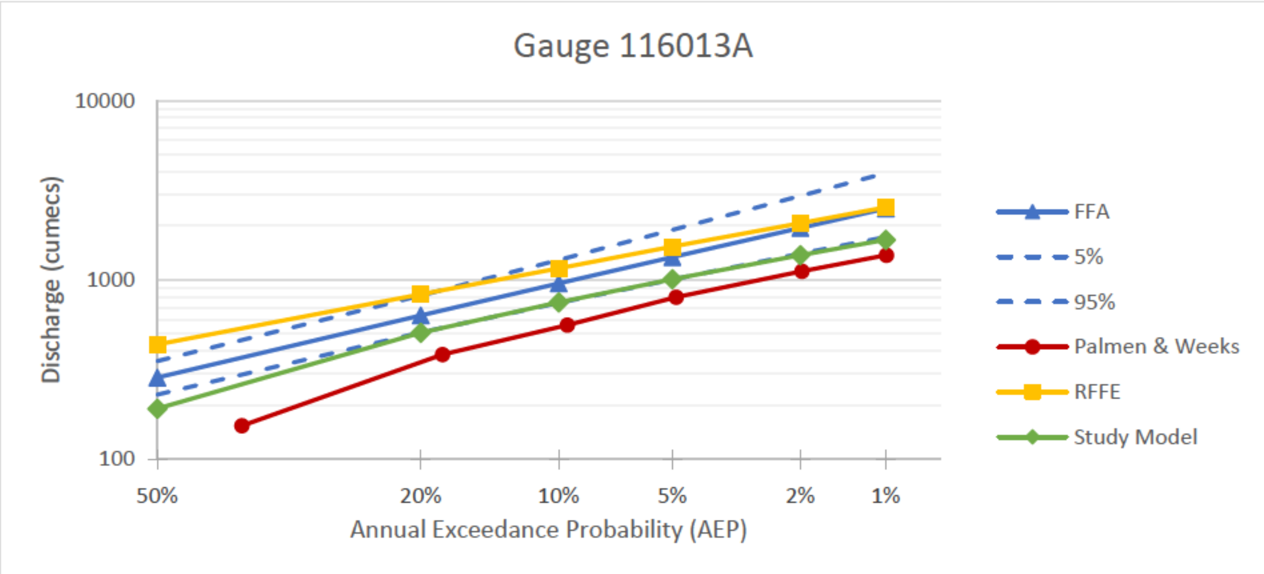
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	254.77	203.49	318.2	167	304.2259		**39.4%**	301.4385
20%	532.5	431.41	666.2	307	733.1557		**18.1%**	671.5786
10%	754.12	605.38	972.8	419	1042.844		**9.52%**	934.648
5%	986.59	774.97	1346.3	538	1359.498		**4.88%**	1285.667
2%	1309.05	986.54	1971.9	711	1798.94		**1.98%**	1724.367
1%	1563	1130.44	2558.4	853	2152.699		**1%**	2061.896

111005A



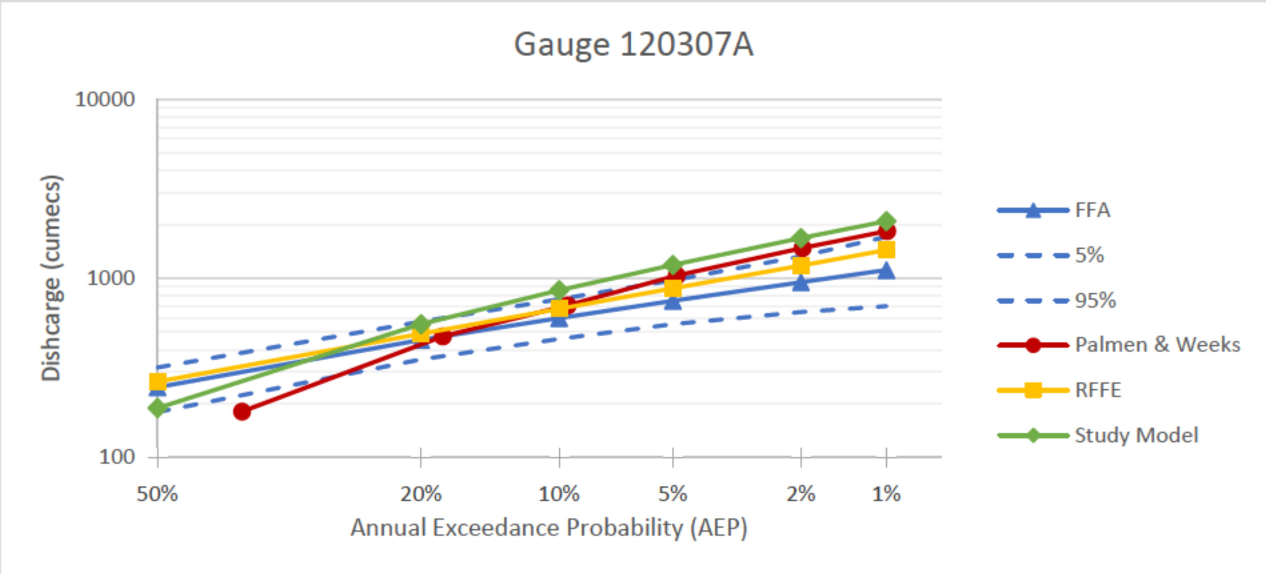
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	768.38	614.11	960.6	902	638.0962		39.4%	450.1851
20%	1485.6	1210.01	1844.7	1670	1300.793		18.1%	935.9187
10%	2017.63	1634.12	2586.8	2280	1728.269		9.52%	1262.986
5%	2548.76	2023.78	3482.8	2940	2141.064		4.88%	1696.418
2%	3250	2469.12	5051.3	3900	2687.194		1.98%	2217.925
1%	3778.47	2748.84	6505.4	4690	3110.221		1%	2608.263

116013A



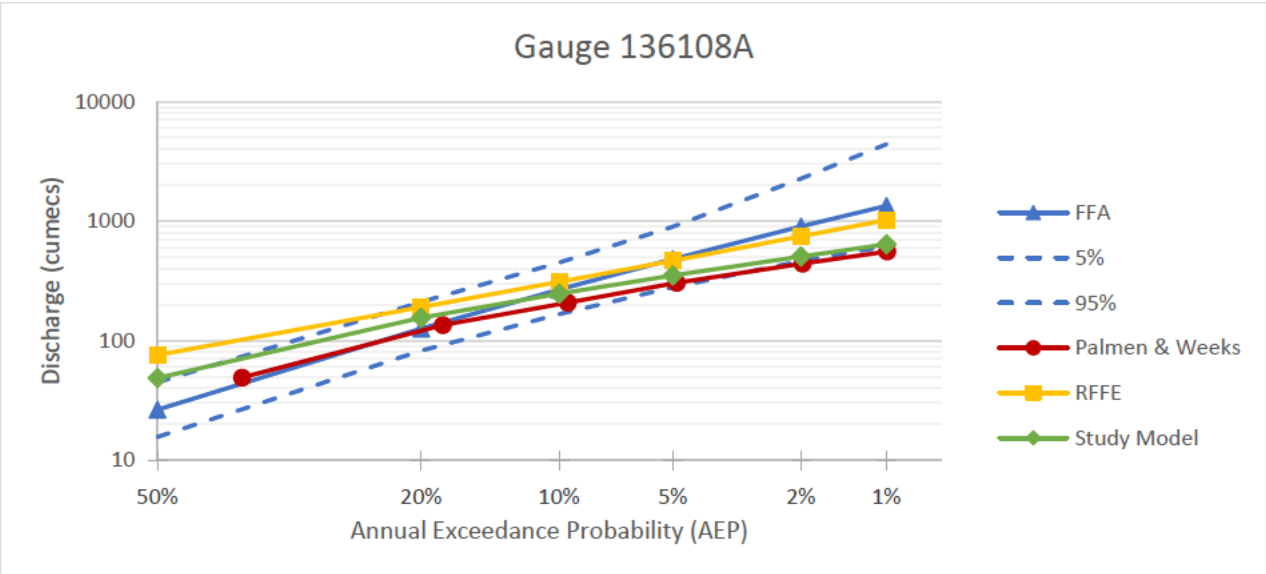
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	285.13	228.94	352.1	436	191.2937		39.4%	153.488
20%	633.58	508.47	815.3	831	507.0584		18.1%	382.5475
10%	955.98	742.38	1291.3	1160	749.6985		9.52%	559.8464
5%	1338.56	1003.47	1901.5	1530	1006.043		4.88%	800.1135
2%	1948.63	1397.98	2933.7	2070	1371.882		1.98%	1118.431
1%	2498.11	1732.93	3941.6	2540	1673.122		1%	1373.978

120307A



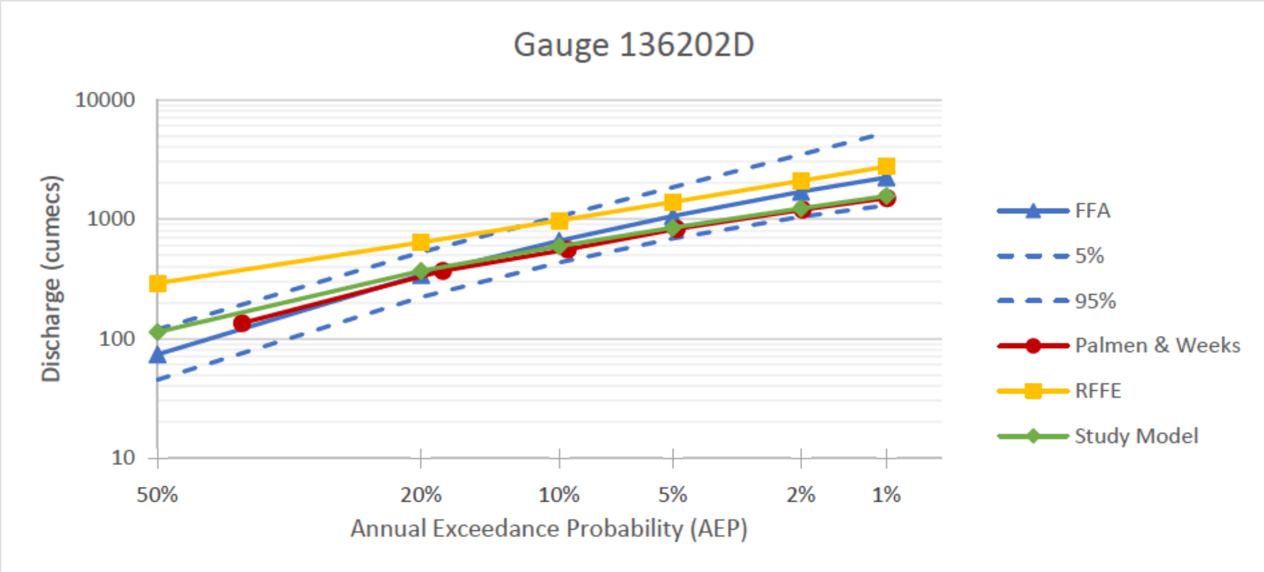
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	246.7	179.2	318	265	188.5433		39.4%	180.5849
20%	453.4	353.2	573.3	490	554.9496		18.1%	474.3768
10%	599.9	458.9	763.6	675	857.3264		9.52%	706.8084
5%	748.2	554.1	974.2	878	1188.591		4.88%	1036.18
2%	951.9	647.6	1330.1	1180	1675.694		1.98%	1477.805
1%	1113.9	699.5	1706.4	1440	2086.324		1%	1837.775

136108A



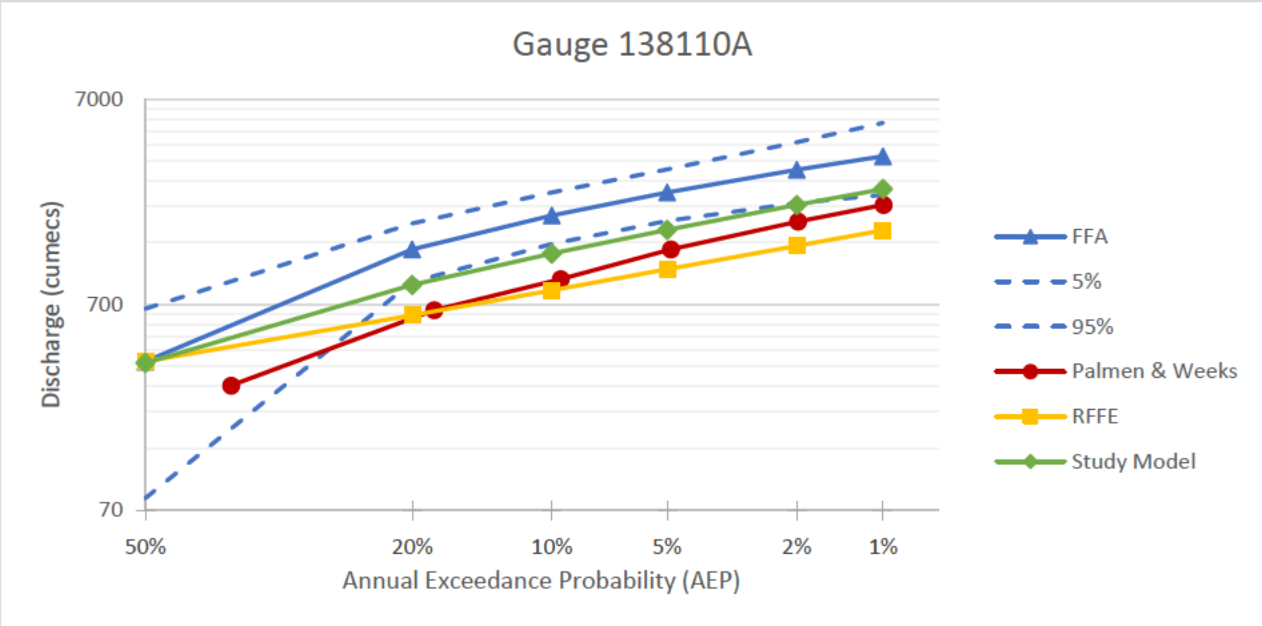
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	26.4	15.6	44.7	75.7	48.62055		39.4%	49.10642
20%	125.76	82.41	209.5	190	154.9511		18.1%	135.0808
10%	266.66	166.32	448.1	310	246.5954		9.52%	207.9497
5%	480.26	279.95	894.3	467	349.7205		4.88%	304.7304
2%	899.89	463.6	2258.3	745	505.3621		1.98%	441.3829
1%	1341.03	615.83	4393.8	1020	639.5165		1%	555.5559

136202D



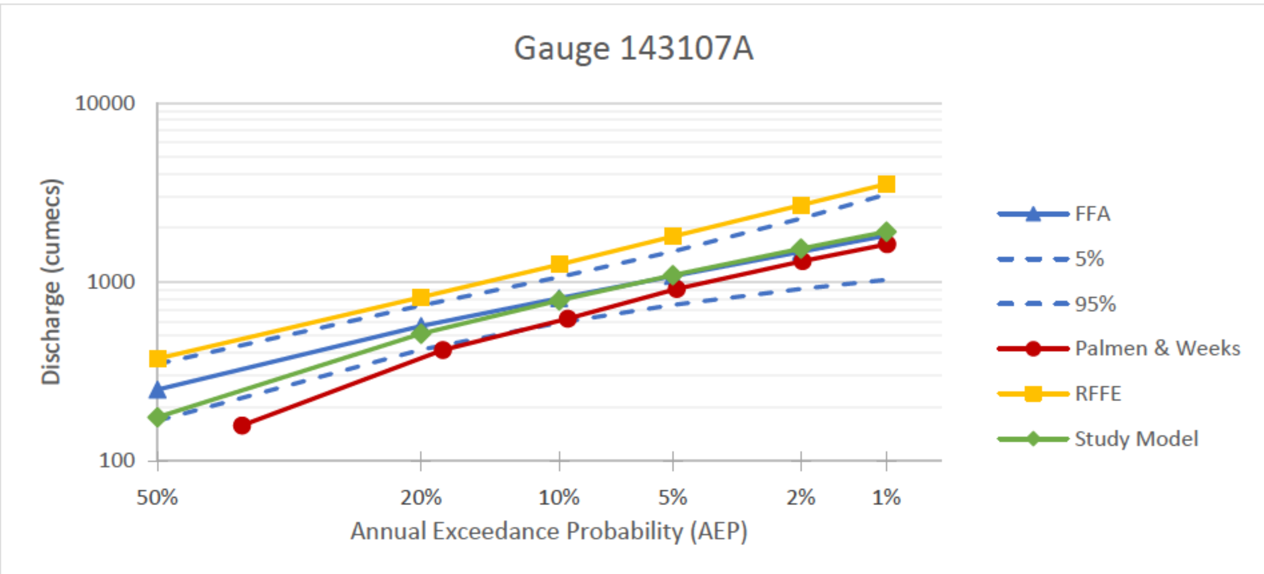
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	73.61	45.19	119.5	291	113.3139		39.4%	134.7842
20%	340.52	221.67	527.7	637	368.1133		18.1%	367.55
10%	658.54	433.41	1051.6	973	591.9094		9.52%	557.3335
5%	1061.01	685.73	1847.2	1390	845.5683		4.88%	826.6481
2%	1694.45	1044	3467.8	2090	1229.819		1.98%	1195.209
1%	2228.94	1302.82	5288	2750	1561.624		1%	1499.882

138110A



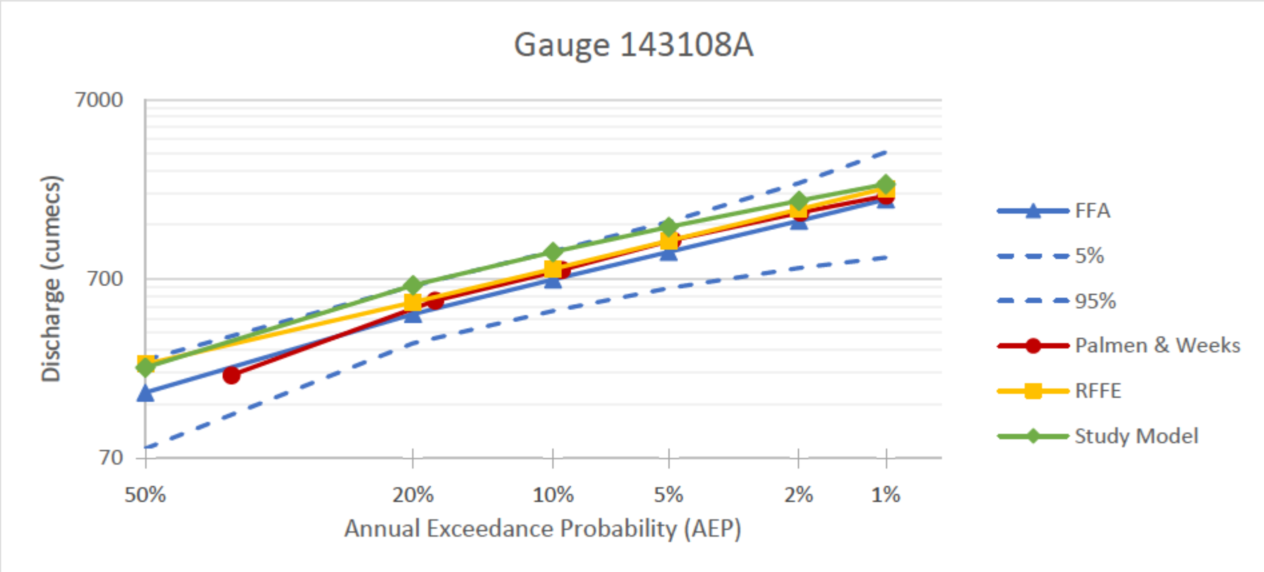
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	369.7	79.8	667.6	370	363.7		**39.4%**	282.6
20%	1299	899.9	1740.1	622	873.5		**18.1%**	656.9
10%	1898	1381	2457.3	821	1241		**9.52%**	929.7
5%	2460.7	1786.7	3184	1040	1617		**4.88%**	1301
2%	3172.2	2186.6	4320	1350	2138		**1.98%**	1773
1%	3693.2	2394.2	5365.5	1610	2556		**1%**	2142

143107A



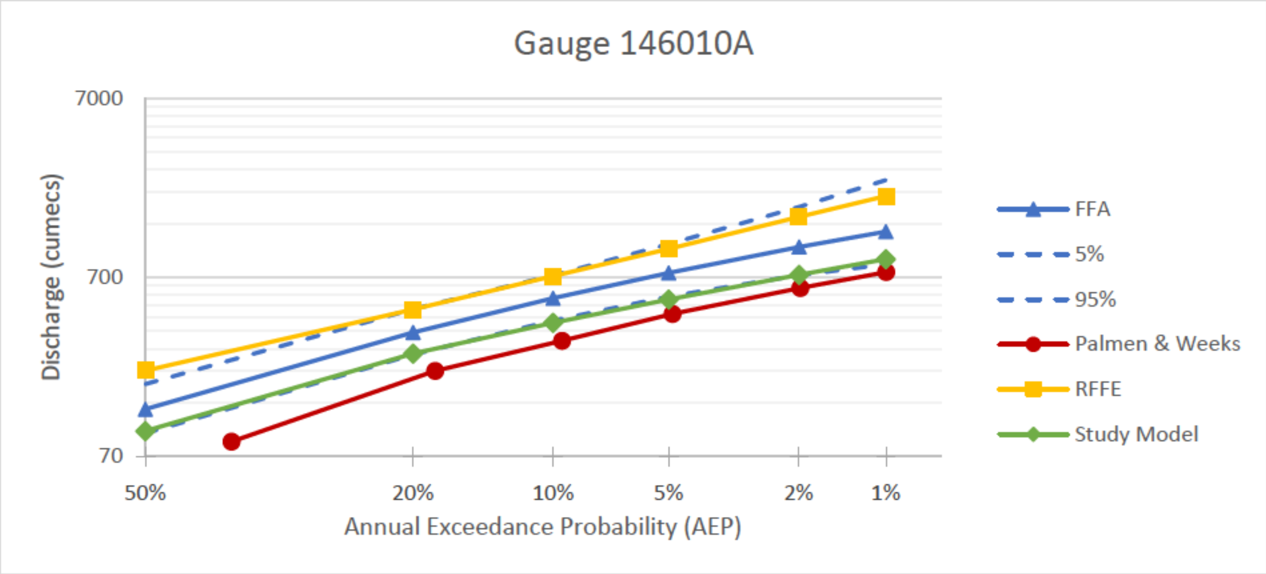
AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	249.9	167.5	347.4	372	175.2432		39.4%	157.5855
20%	565.7	417.8	737.1	821	512.0855		18.1%	416.072
10%	810.7	589.8	1064	1250	788.5544		9.52%	622.1211
5%	1076.8	741.7	1479.7	1790	1090.663		4.88%	912.0965
2%	1473	911.5	2252.4	2680	1534.052		1.98%	1303.076
1%	1813.7	1026.5	3080.3	3520	1907.308		1%	1622.632

143108A



AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	162.4	79	247.2	234	223.9901		39.4%	203.0452
20%	445.5	306	638.5	519	645.1009		18.1%	529.903
10%	692.7	464	994.8	795	988.0086		9.52%	786.5134
5%	987.3	620.4	1453	1140	1360.883		4.88%	1151.872
2%	1475.1	803.3	2392.3	1710	1905.52		1.98%	1639.042
1%	1939.7	919	3555.5	2240	2362.14		1%	2034.931

146010A



AEP	FFA quantiles	5% Confidence Limit	95% Confidence Limit	RFFE	Study Model		AEP	Palmen & Weeks
50%	128.55	94.29	177.4	211	97.14915		**39.4%**	84.60535
20%	344.07	259.72	465.9	464	262.4951		**18.1%**	210.1064
10%	534.38	401.31	718.1	710	390.5241		**9.52%**	309.2159
5%	741.28	546.56	1071	1010	526.6449		**4.88%**	437.1142
2%	1032.05	720.68	1732.8	1520	722.3281		**1.98%**	609.6446
1%	1260.06	826.73	2441.2	1990	884.53		**1%**	748.7211