Probabilistic Dynamic Rating Curves using Auxiliary Information

by

Luis Camilo Galindo Ruiz

B.S. Civil Engineering, University of Texas at Austin, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Civil Engineering)

The University of British Columbia
(Vancouver)

December 2017

Abstract

Rating curves play a vital part in hydrology for producing streamflow time-series. The derived streamflow is an integral component to any hydrological study and therefore requires proper quantification of not only a discharge point value, but also an uncertainty measure. Using multivariate Gaussian distributions as kernels, a probabilistic rating curve was developed from the conditional distribution as an alternative model for the standard deterministic rating curve. Auxiliary information from a run-of-river hydroelectric project, as well as the temporal variability from the gauging measurements, were used to study the possible reduction in the uncertainty of the developed rating curve. The temporal information was modeled using an exponential function that updated upon receiving new gaugings and the sluicing model was a continuously updated kernel distribution that assigned more weight to gaugings taken after a sluicing event. Four models of varying complexity were created and their performance was evaluated using information theory measures such as surprise and the Kullback-Leibler divergence measure. The results indicate that probabilistic rating curves are useful tools for modeling and evaluating the dynamic uncertainty of the curves. The uncertainty was shown to be reduced by up to 19% by including the temporal information of the gaugings and sluicing information. Auxiliary information can be beneficial to rating curve development and an argument is made for why probabilistic rating curves should become a norm in the hydrology field.

Lay Summary

The volume of water (streamflow) flowing through rivers has large applications in engineering design and modeling. All these applications require that the proper amount of streamflow be quantified to ensure the best possible design and usage of the water. For this reason, stations that continuously record streamflow have been placed at various locations along rivers all over the world. Directly measuring streamflow however is highly expensive and an indirect method must be used. This method is called a rating curve, but the largest flaw in the method is that it does not properly capture the uncertainty in the streamflow or give any insight into the probability of the discharge. This research addresses the flaw by developing a model that highlights the truly probabilistic nature of the streamflow. Available information impacting the streamflow uncertainty was also introduced to help reduce the uncertainty of the probabilistic dynamic rating curve.

Preface

This thesis is original, unpublished, and independent work by the author, Luis Galindo.

Table of Contents

List of Tables

List of Figures

Glossary

USGS United States Geological Survey

PDF probability distribution function

WSC Water Survey of Canada

V-SHIFT variable shift diagram

MCMC Markov Chain Monte Carlo

OU Ornstein-Uhlenbeck

GLUE Generalized Likelihood Uncertainty Estimation

ANN artificial neural network

ENNF explicit neural network formulation

ENSO El Niño Southern Oscillation

PDO Pacific Decadal Oscillation

EC electricial conductivity

IFR instream flow requirement

FKC Forrest Kerr Creek

LOOCV leave-one-out cross validation

CDF cumulative distribution function

NHC Northwest Hydraulic Consultants

ROR run-of-river

BC British Columbia

RMSE root mean square error

Acknowledgments

I would like to thank Dr. Steven Weijs for all his help throughout the past few years, and for the notorious 1+ hour long meetings in your office. It has been great fun working under your supervision. Thank you Dan Moore for being my second reader. Thank you to Akhil Kumar and Hossein Foorozand for your fruitful discussions and friendship during the past year. Valentine Arrieta and Pauline Millet, those few months that you were each in Vancouver were unforgettable moments.

Thank you to Stu Hamilton and Touraj Farahmand from Aquatic Informatics for the support on my thesis and for helping with the NSERC Engage grant. Also thank you to Andre Zimmerman from NHC for generously providing me the data to complete my thesis.

To everyone in the upstairs CEME graduate offices, thank you for all the great conversations. To Kevin, Elnaz, and Glenda, thank you for all the distractions, coffee, candy, and help.

My experience in the hydrotechnical group would not have been the same without the following people: Abhishek Agrawal, Mehretab Tadasse, Jonathan Van Groll, Andy Lee, Erica Kennedy, and Julio Portocarrero. Thank you all so much for all the great times that we have had.

To the many faces of 2116, it has been a pleasure living with all those who have come and gone; you guys have been great roommates! James Roberts, thank you for all the fun that we have had the past few months in the house—that haunted escape room was brilliant! Simone Tengattini, living with you was a beautiful experience, and the friendship that flourished will never be forgotten.

Sylvia Woolley, thank you for everything. Your companionship the past year has been an extraordinary adventure.

To all my friends from Texas that I have maintained in contact with, I am incredibly happy to have you all in my life. Thank you to all those who have traveled to Vancouver to see me, especially Juan De Loera, Drake Hernandez, Brandon Comisarenco, Anais Picarelli, and Joanne Neira.

Lastly, thank you to my family. None of this would have been possible without their support. Gracias a mi papá y mi abuela por todo tu amor y apoyo—los amo con todo mi corazón.

Chapter 1

Introduction

This thesis is focused on the one compound that everyone in the world needs for survival—water. It's both vast, powerful, and dynamic as well as finite, fragile, and lethargic, all at the same time. Human life depends on this compound, but its usability is threatened every single day by pollution and anthropogenic climate change. This makes one wonder how water will play a factor in the future and/or how much usable water will there be? x m^3? y m^3? This information is useful, but it leaves a lingering question of, "How sure are you?" This question instinctively forces the person providing the x and y estimates to give a response like—"I'm fairly sure" or, "About 67% confident." So when given a prediction, it is natural to know how certain the answer is. Think about it. What if you turned on your television and the weather spokesperson said, "Today it will rain." This statement leaves one wondering if the individual meant that it will rain all day, or if the rain will be scattered with higher probability soon after lunch. Without any further information on the probability of rain today, one is left to make the judgment of when he/she believes it will rain. So now, one can make a subjective statement on what the probability of rain will be, most likely determined by looking outside and examining the gloom in the clouds. However, one is not an expert in weather forecasting and the probability of rain determined by individual users may be different, and more likely to be incorrect, from the probability that the expert would assign. The reason for this is that the expert has all the information available to him/her and is more qualified to be the one making the judgment on the chances of rain today. The aforementioned is true when trying to predict how much water is flowing in a river channel.

Today, all around the world the task of quantifying how much water is flowing in any river is delegated to primarily government agencies. Here in North America, the two largest agencies are Water Survey of Canada (WSC) and the United States Geological Survey (USGS). Hydrographers for WSC and USGS perform field visits to rivers, and carefully measure water level and discharge values to develop an empirical relationship between the two variables. The true relationship is grounded on fundamental hydraulic principles and is known as the rating curve. However, the rating curve produces a purely deterministic result. Therefore, similar to the scenario described in the prior paragraph, users of the assembled discharge time series are tasked with the exercise of applying a

probability to the values, or in many cases, applying no probability and using the time series as is, causing an "illusion of certainty" (Krzysztofowicz, 2001). Many hydrologists have come to terms with the importance of probabilistic models and their importance (Beven, 2010; Liu & Gupta, 2007; McMillan, Freer, Pappenberger, Krueger, & Clark, 2010; Montanari & Brath, 2004; Weijs, 2011). Unfortunately, the rating curve has yet to be transformed into a truly probabilistic model and it is here, where the author argues that a probabilistic rating curve is essential for not only the hydrology field, but those that are dependent on discharge time series. The results from a rating curve are fundamental to accurate depictions of floodplain mapping, emergency and evacuation decisions, engineering design structures, and more.

The topic of rating curve uncertainty is continuously being studied to develop models that can produce more accurate values. Again, the author has not found in the literature an explicit probabilistic dynamic rating curve model. The literature does contain probabilistic rating curves in the sense that the parameters of the rating curve are described by a probability distribution, but no rating curve has been developed as a pure conditional distribution, like the one in this thesis. The author's first objective in this thesis is to develop a probabilistic rating curve model that produces similar discharge values but with the added benefit of providing a probability distribution with the discharge point value. The second objective of this thesis is to illustrate to the reader the importance of adding available information to improve the accuracy of a rating curve model. Available information in the context of rating curves means information that may help explain the behavior of rating curves throughout the temporal and spatial domains, as well as any relevant data that may have also been collected while recording water levels and discharge, such as electrical conductivity, as done in (Weijs, Mutzner, & Parlange, 2013). Information has the main benefit of possibly reducing uncertainty. In the context of rating curves, auxiliary information that can be brought into the model is dependent on what kind of other signals were collected at the hydrometric station, or nearby. Rivers with high energy, perched in the mountains, tend to have their discharge measured by electrical conductivity. Weijs et al. (2013) showed that electrical conductivity could be useful in reducing rating curve uncertainty. Other signals that have a strong correlation with the discharge may also serve as good side information to include, but they must be independent of the water level. Thinking about what local information relates to the discharge is a useful exercise that can have high impacting results on modeling rating curve uncertainty. Caution must be adhered to ensure that a complex model is not developed as a substitute for the current practical rating curve. Instead, the ideal model should balance complexity and practicality. Thus, by grounding the rating curve on hydraulic principles and identifying useful information, one can create a model that meets both complexity and practicality needs.

More specifically, this thesis will demonstrate a probabilistic rating curve that uses information available from a run-of-river hydroelectric project in northwestern BC. The thesis is organized as follows. Chapter 2 will provide the background on the current state of developing rating curves, as

well as present a literature review of the current work that has been done on modeling rating curve uncertainty. The literature review over rating curve uncertainty extrapolates some of the work presented in the literature review done by Le Coz in 2012, and expands on it by work done since that manuscript's publication. Chapter 3 describes the methodology used in producing the probabilistic rating curve, as well as a description of how the added information was incorporated into the rating curve. This chapter finishes by introducing a few information theory metrics and measures used to help evaluate the added information, and the performance of the models. Chapter 4 is a presentation of the probabilistic rating curve produced, as well as an assessment of the auxiliary information used in the model. Finally, Chapter 5 provides a discussion on the results, the importance of probabilistic models and using available information, and future work that can be done to expand the work presented herein this thesis.

Chapter 2

Background

2.1 Stage-discharge rating curve

A stage-discharge rating curve is an empirical relationship between water level and discharge along a cross-section of a river. A rating curve ideally consists of 10 or more points that are time invariant and supports the hypothesis of a stable relationship between the variables (Ministry of Environment Science and Information Branch for the Resources Information Standards Committee, 2009). A power law (Equation 2.1) is often used and forms the one-to-one relationship between the two variables.

$$Q = a \cdot (h_t - h_0)^b \tag{2.1}$$

Equation 2.1, when plotted on a logarithmic scale, creates a linear relationship (when the true cease-to-flow water level is known) that can be interpreted in terms of hydraulic principles. The scaling parameter (a) and exponent (b) have been attributed to the geometry and physical constraints of the river. In his seminal work, Chow (1959) demonstrated that the scaling parameter is related to the characteristics of the channel or section control, while the exponent parameter can approximate the shape of the cross-section. This can provide an indication as to whether the stage-discharge relationship is controlled by a section or channel control. An exponent less than two would indicate a channel control while a larger exponent is typical of a section control. This led Chow (1959) to generalize the following:

- b = 3/2 is typical of a rectangular channel
- b = 2 is typical of a rough parabolic shape
- b = 5/2 is typical of a triangular or semi-circular section

2.1.1 Hydraulic Controls

The stage-discharge relationship at a gauging station is governed by the downstream hydraulic controls of the river (Rantz, 1982). The three types of hydraulic controls are section, channel, and

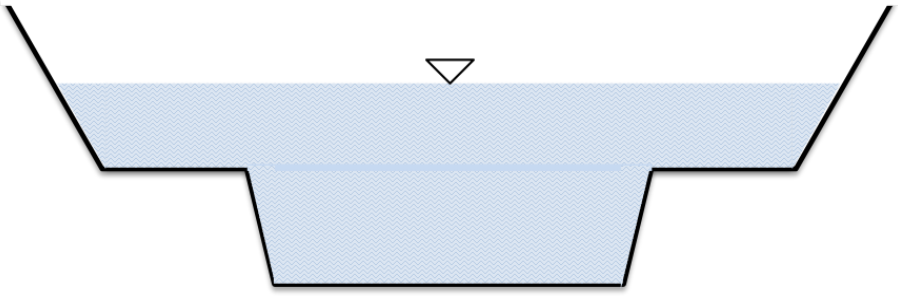

Figure 2.1: Compound channel cross-section

compound controls. A section control is either a natural or man-made structure, such as a rock outcrop or bridge/weir, while a channel control is a function of the channel geometry, slope, and roughness. Typically, a channel control is more evident in large and wide rivers, while a section control is more commonly found in small narrow rivers. The third control is the compound control, a combination of section and channel controls as shown in Figure 2.1. Here, the section control manages the low end of the rating curve and as the flow increases, the section control is drowned out and the relationship is dominated by the channel control. In such a hydraulic situation, it is common to identify a transition region between the two controls. When plotting the rating curve, this can usually be identified by changes in the slope between low and high flows (Herschy, 1995). It is also important to note that the hydraulic equation may also be segmented to account for a compound section. In this case, a segmented hydraulic equation will have unique parameters for the different sections that correspond to the dominant control. Also when considering hydraulic controls, one needs to account for the offset value, or h_0 in Equation 2.1. This offset value is the height adjustment for the control, and differs if the stage-discharge rating curve is compounded to represent the different physical characteristics of the channel.

2.2 Development of a rating curve

2.2.1 Criteria

When developing a rating curve, it is pertinent that the location of a rating curve satisfy the following criteria (when possible) as is mentioned in the *Manual of British Columbia Hydrometric Standards*. These criteria include:

- Accurate water level measurements must be available at low and high water levels
- A stable hydraulic control must be available, either natural or artificial
- The station must be accessible year long
- Straight, stable, and aligned banks
- No tributaries between gauge and metering site
- Metering section should have uniform depth and velocity flow lines

- Accountability for backwater effects
- The station should be able to withstand large peak events and intense rainfall and snow events
- The proper station should be constructed for the type of body of water one wishes to gauge

Unfortunately, not all criteria will always be met, and this could cause a less stable stage-discharge relationship. If this is the case, one may need to perform more site visits.

2.2.2 Site investigation

Site visits allow the hydrographer to develop an intuition for how the river is behaving at a given point in time; as well as give an image of changes that may have occurred since the last visit. Careful observations regarding high flowing rivers should be exercised and evidence of past floods should be sought out. These include, but are not limited to, water marks, debris deposition, and fragmented vegetation in the river's floodplain. It is also important to ensure that if a housing structure, such as a stilling well, is on site, that it remains intact and fully operational.

A careful observation of the channel cross-section shape (including banks of the river) will help the hydrographer develop a hypothesis for what a possible exponent value (b) may be in Equation 2.1. Signs of large boulders along the river may indicate high energy and temporal variations of the controls (Baker, 2009; Kieffer, 1985; Le Coz, Renard, Bonnifait, Branger, & Le Boursicaud, 2014). It is recommended that during site visits, the controls should be identified to ensure no changes to the regime that controls the stage-discharge relationship have happened. Gaugings taken during site investigations should be aggregated to the most current rating curve. This practice allows the hydrographer to identify if the data collected fall within 5% of the current rating curve, or if a shift or new curve may be required (Herschy, 1999).

2.2.3 Interpretation of the rating curve

A shift is described as a temporary deviation from the main rating curve. Shifts can be attributed to a number of reasons, primarily temporary changes in the control, such as a log/debris jam, ice, and vegetation. Backwater effects may also cause a temporary shift to be implemented. If the shift persists for a period of months, then this is may be indication that a permanent effect may have taken place in the stream and a new rating curve is required. It is sometimes possible to predict when a shift is going to occur based on the temporal patterns of the stream and basin. For example, it is possible that through a thorough ecological investigation, a hydrographer may notice that beavers always build their dams during a specific time period and/or that during spring and summer months, vegetation tends to grow and a shift in the curve is needed until that time period is over.

Determining shifts can be an art, as stage and discharge have natural scatter due to the volatility of the cross-section as impacted by erosion/deposition, vegetation, and/or ice. Gaugings that deviate from the adopted threshold (typically $\pm$ 5%) may indicate a possible shift in the stage-discharge

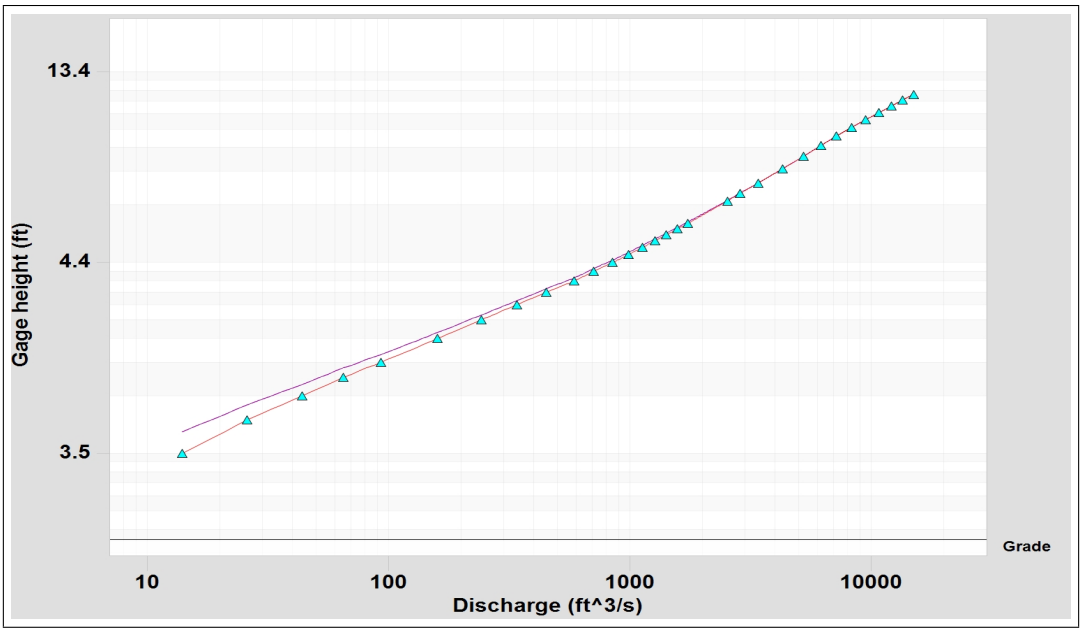

Figure 2.2: Illustrates a shift correction. The red line represents the baseline rating curve with triangular markers used to construct linear lines. The purple line is a temporary shift that occurred during the active period of the rating curve.

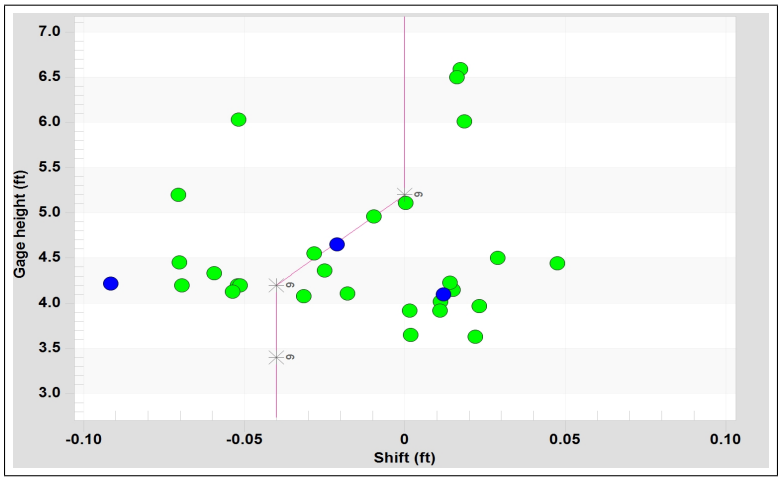

Figure 2.3: Variable-shift diagram displaying a negative shift.

relationship, but a well-grounded physical reason must justify the decision on when to apply a shift, and where the shift should occur on the rating curve. Figure 2.2 presents a shift for the active rating curve of a gauging station between September 2008 and September 2009. A variable shift diagram (V-SHIFT) diagram is often used to help identify shifts and their location. A sample shift diagram is shown in Figure 2.3 The V-shift diagram plots the water level of the gaugings on the ordinate, and the distance between the measured water level and that from the baseline rating curve on the abscissa. When the data are plotted, the gaugings should resemble a Gaussian distribution around the zero shift reference line. If not, this could indicate a bias and supports the notion that

a shift may be necessary. A negative shift will cause the temporary aberration in the rating curve to be to the left of the baseline curve, and the opposite for a positive shift. Shifts are meant to be temporary and should always yield back to the baseline curve. If over time, gaugings systematically plot above or below the baseline curve, a new rating curve should be considered.

Typically, rating curves can exhibit a pattern of moving to the left or right of the previous rating curve. A simple sensitivity analysis was performed by Rantz (1982) in which he stated that by plotting the rating curve on log paper, changes in the control are identifiable. For example, an increase in the width of the channel control causes the new rating curve to plot parallel and to the right of the original, while the opposite indicates a decrease in control width. Scouring and deposition are also possible to detect and Rantz proposes that scouring will cause a new curve to plot to the right but will be concave downward, rather than parallel, while deposition will plot to the left and will be concave upward.

2.3 Uncertainties in rating curve

As one can see, the stage-discharge relationship thus far is affected by the hydraulic controls and their variability over time. However, there are many more factors that can affect the stability of a rating curve. The following is a list outlining some of the most common factors affecting a rating curve as shown in Di Baldassarre and Montanari (2009); Hamilton and Moore (2012); Herschy (1999); Turnipseed and Sauer (2010):

- Scour and fill sand-bed channel
- Aquatic vegetation
- Ice cover
- Variable backwater
- Changing discharge

Understanding how these factors play a role in modeling the rating curve is crucial for ensuring an accurate discharge time series. For example, during the summer when vegetation is higher than the rest of the year, a shift may be probable until the end of the summer when the vegetation no longer has a strong influence on the river's discharge. Ice plays a unique role in measuring streamflow, and in many cases, forces hydrographers to have to remove water level sensors to prevent cold weather damage to the instruments. This poses a challenge in trying to recreate the discharge time series. In many cases, WSC will interpolate their winter discharge time series with as few as two gaugings (Moore, Hamilton, & Scibek, 2002). Variable backwater effects in streams are usually caused by an obstruction in the river downstream that causes the water to propagate upstream, such as when a river tributary enters the main stem. Changing discharge is typically seen during passing flood waves where the rising slope of the wave is steeper than the falling. This phenomenon creates a hysteresis effect where the discharge has two different water levels for the same discharge. In a

stage-discharge plot, this is usually identifiable by a loop around the main rating curve.

One other source of uncertainty in a rating curve, is the one associated with extrapolation. High flows can make it dangerous for hydrographers to visit the site, and safely record water level and discharge. At sites with permanent measuring devices, high flows can cause the instruments to be washed away. Extrapolation is required for an array of purposes, but special care should be taken to ensure that the extrapolation is confined to the physical constraints of the channel cross-section. It is best practice to plot stage vs area on arithmetic or logarithmic axes to identify the relationship of the two variables. Ideally, both variables will exhibit a linear relationship until a break in the cross-section is observed, similar to the break in a rating curve when a compound cross-section is used. Typically, one can extrapolate until the point where the stage-area curve is outside the bounds of the channel area. This however is still an assumption, and is never guaranteed to give the true relationship, but can provide a reasonable approximation in some cases.

The first published work on modeling rating curve uncertainty was done by Venetis (1970), in which residual variance was used to determine the bounds of the uncertainty in a power law rating curve. The data was log transformed and a least square regression was performed on the log-transformed discharge. Venetis however did not account for measurement uncertainty. The work done by Rantz (1982) and Herschy (1999) really grounded the hydrometry field and provided a foundation for the work that follows.

Di Baldassarre and Montanari (2009) presented a general uncertainty framework that assumes a global uncertainty accounting for the discharge measurement technique, interpolation and extrapolation errors, the presence of unsteady flow conditions, and seasonal changes. They performed a one-dimensional hydraulic model on the Po River in Italy to carry out their estimation of rating curve uncertainty. Their results found that discharge uncertainty could be well over 40% and have high implications on hydrological models such as rainfall-runoff models that are dependent on the results of a rating curve to calibrate the models. However, Di Baldassarre and Montanari also made an exhaustive list of assumptions, primarily that the cross-section of the river is time invariant and that the uncertainties affecting the discharge values are independent, implying that the measuring technique used, velocity-area method, is independent of the flow conditions and vegetation.

Heteroscedasticity is a common problem observed in rating curve modeling in which discharge uncertainty increases as the magnitude increases (Sorooshian & Dracup, 1980). Using least square regression on a power law function, however, violates this assumption, unless log-transformed. Petersen-Øverleir (2004) pointed out this common error of unaccounted heteroscedasticity uncertainty affecting rating curve parameters and estimated discharge. Petersen-Øverleir suggested a maximum likelihood method that better accounts for the natural uncertainty exhibited in the stage-discharge relationship. However, in that manuscript, only the uncertainty associated between stage

and discharge was examined, and not the *individual* gauging errors. The same author has also investigated the impact of hysteresis (Petersen-Øverleir, 2006) and overbank flow in rivers (Petersen-Øverleir, 2008). In the first paper, the Jones Formula was used to identify a change of stage rate. In the latter paper, the author discussed issues with modeling multi-segmented rating curves, in which the range of the parameter estimates might be incorrect and the location of the breakpoint may also not have strong physical meaning. The author also made simplifying assumptions about channel characteristics and hydraulic controls. Some of the numerical challenges were also not addressed by the author in his 2005 manuscript (Petersen-Øverleir & Reitan, 2005), but were in later publications.

2.3.1 Bayesian and MCMC methods

As computation power has increased in recent years, so has the number of works using Bayesian and Markov Chain Monte Carlo (MCMC) methods. In Bayesian analyses, hydraulic knowledge can be incorporated into the modeling process via prior distributions. As the model receives new data, it updates its beliefs, via a likelihood function, and creates a posterior distribution (Reitan & Petersen-Øverleir, 2008). It is most common to use Gaussian distributions as priors and posteriors because of the closed form solution. However, if the prior and posterior are not conjugate to each other, a closed form solution does not exist and MCMC methods must be used to sample from the posterior (Godsill, 2001). The usage of these methods naturally allows the modeler to sample from a probability distribution, and therefore assess the uncertainty associated with the parameters of Equation 2.1 and the predicted discharge. Incorporating hydraulic knowledge into the model via priors allows the modeler to use more information, rather than the gaugings alone.

The segmented rating curve problem by Petersen-Øverleir (2008) was improved using Bayesian and MCMC methods, and further helped with modeling the uncertainty of the rating curve (Reitan & Petersen-Øverleir, 2009). A compounded channel was evaluated using Bayesian methods to estimate the parameters in each power law segment, as well as the breakpoints in the stage-discharge relationship. The posterior distribution that resulted from the Markov Chain Monte Carlo sampling was then used to determine the uncertainty of the rating curve. The authors acknowledged that their model was time invariant and assumed that their rating curves were not affected by changes in the river geomorphology, hysteresis, or variable backwater. For backwater effects, Petersen-Øverleir and Reitan used two gauging stations to determine the slope of the reach and the correct discharge in their 2009 publication. While this solution does help with estimating a more correct discharge, it is considered impractical in many situations to include more than one gauging station. Multi-segmented rating curves were also looked at in McMillan and Westerberg (2015). Here, the authors created a new likelihood function called the "Voting Point likelihood." Their model assumed a logistic distribution for the discharge gaugings and prior distributions for the parameters of the power law function. A rating curve was then created from sampling the prior distributions of the parameters, and the curve is assigned a likelihood weight, dependent on the logistic distributions it has

intersected with, as well as the range of stage and discharge that the rating curve spans. This procedure was then repeated using a MCMC method to create a posterior distribution of possible rating curves. The results of this method allowed to explicitly model aleatory and epistemic uncertainty. The authors treated all gaugings as equal, and do not account for permanent changes that may occur in the river after larger altering rainfall events.

Le Coz et al. (2014) also showed the value of adding hydraulic knowledge to the development of the rating curve, as well as the importance of properly classifying the gauging error. The uncertainty in this study was divided into the uncertainty associated with the gaugings, the location of the gaugings, and the remnant uncertainty. Since the remnant uncertainty was unknown, a MCMC simulation was used to estimate the value. What the authors show is that when the uncertainty in the gaugings was assumed to be 5%, as suggested by (Herschy, 1995; Rantz, 1982; Turnipseed & Sauer, 2010), the remnant uncertainty increases. When the actual (true) uncertainty in the gaugings was used, the remnant uncertainty was lower. These results indicate the importance of capturing the correct discharge uncertainty, rather than assuming a standard value for all. The work done by Le Coz et al. in 2014 helped the community acknowledge the importance of Bayesian and MCMC usage in rating curve development, through their incorporation of the techniques into their user friendly developed method called, BaRatin. However, a major limitation of the work is its stationary assumption.

2.3.2 Time variance in rating curves

It is accepted that river characteristics naturally change over time and therefore the controls which affect the stage-discharge relationship are also evolving temporally (Di Baldassarre & Montanari, 2009; Hamilton & Moore, 2012; Leonard, Mietton, Najib, & Gourbesville, 2000; Schmidt, 2002). Common practice is to apply time-varying shifts and/or increase the frequency of gaugings. However, the latter option is often highly expensive. A few studies have tackled the task of time-variant rating curves. In Reitan and Petersen-Øverleir (2011), the authors transformed the assumed time invariant parameters to time dependent through the combination of a Bayesian hierarchy approach and the Ornstein-Uhlenbeck (OU) process. For those interested in the OU process, please refer to Reitan and Petersen-Øverleir (2011) for further information. The OU process serves as the continuous time-varying function for the power law and the parameters are estimated by updating the models with information as it is received. The results produced 95% credibility bands around the estimated parameters in the power law function, and the rating curve. The uncertainty bands around the parameters, however, have sharp changes in the uncertainty due to the OU process, implying no gradual change in the uncertainty over time. It can be argued this sharp increase in uncertainty is not always observable in nature. Measurement errors were also not considered by the authors. The credible bands around the rating curve are purely Bayesian and uniform in shape, rather than dynamic uncertainty bands that illustrate regions of high and low confidence with respect to the gaugings and their date of measurement. Guerrero, Westerberg, Halldin, Xu, and Lundin (2012) studied the

dynamic relationship of stage and discharge on six hydrometric stations in Honduras through the usage of topographic surveys and Manning's equation to develop their rating curve model. The uncertainty in their model was evaluated using Monte Carlo analysis on a moving window of 30 data points using the Generalized Likelihood Uncertainty Estimation (GLUE) method. Their results indicated a temporal variability in the scaling, offset, and exponent parameters of Equation 2.1. The scaling exponent appeared to be the most sensitive to time and was the focus of their conclusions. The uncertainty, however, did not account for the uncertainty in each individual gauging and the other parameters were not discussed. In Westerberg, Guerrero, Seibert, Beven, and Halldin (2011), time variability of the rating curve and its uncertainty was derived through a weighted fuzzy regression, but some of the assumptions are too subjective. For example, their inclusion of the top three largest discharge values in the time series into each 30 data point window, and the usage of what appears to be an arbitrary window does not account for anything related to changes in the geomorphology of the river or vegetation growth. McMillan et al. (2010) divided the window of gaugings used in the uncertainty analysis of rating curves as being dependent on a 0.5 year return period. This return period is used as a threshold to classify major flood events and is subjective. In this study, the authors looked into discharge errors stemming from rating curve uncertainty and the impact that they had on rainfall-runoff models. They provided a methodology to quantify the uncertainty in discharge as a probability density function conditioned on errors in stage and velocity measurements, assumptions about the stage-discharge relationship, extrapolation of the rating curve, and changes to the cross-section due to vegetation growth and bed movement. Coxon et al. (2015) showed that there are regional differences in the stage-discharge relationship of 500 gauging stations in the UK, but that local conditions such as weed growth and channel instability typically dominated the magnitude of discharge uncertainty.

2.3.3 Machine learning

In this day and age, it would not give justice to this thesis to not briefly mention machine learning and its impact on hydrology and rating curves. There are also similarities between the Gaussian kernels used in this thesis, and those in Gaussian Processes. At its core, machine learning is about learning and adapting, as most notability seen in nature (Goldberg & Holland, 1988). Machine learning reduces the burden of having to code every rule, and instead lets the machine learn the rules via the data. This however puts a large emphasis on data driven models and can cause erroneous estimates when the physical constraints of the model and system being investigated are ignored. The reader is referred to Solomatine and Ostfeld (2008) to learn more about the common pitfalls in data driven models using machine learning, as well as interesting new avenues that this exciting field can take hydrology and water resources. For a more practical overview, the reader may find Hsieh (2009) to be useful.

Bhattacharya and Solomatine (2005) used an artificial neural network (ANN) in combination with an M5 model tree to develop a stage-discharge relationship. ANN, as of 2013, is the most common used

machine learning algorithm in water resources research (Govindaraju & Rao, 2013). ANN works as a collection of nodes, or neurons, that transmit information from neuron to neuron as received from the input-ouput data. Weights are assigned, by learning, to each neuron and are used to develop a function that explains the relationship between input and outputs (Hsieh, 2009). The M5 model tree used is a type of tree like regression where the algorithm splits the parameter space (into subspaces) and estimates a local regression for each subspace. The machine learning model proved to be more effective and had a lower root mean square error in training and verification modes. However, each traditional rating curve is fundamentally connected to the physical constraints of the river, which helps in modeling and predicting values past the highest measured discharge. The extrapolation of rating curves was looked at by Sivapragasam and Muttil (2005) but again, no connection to the physical constraints of the river were investigated, although having mentioned a few of the most common methods in hydrometry. In Guven, Aytek, and Azamathulla (2013), the authors used the stage and discharge time series to teach an explicit neural network formulation (ENNF) model. The authors state the performance of their model is better than the traditional rating curve. Using the time series over the gaugings forces the uncertainty associated with the original rating curve model to propagate into their machine learning model. This therefore suggests overfitting by the ENNF.

While machine learning does provide a good alternative for models with sparse data, caution should be exercised to ensure modeling within the confinements of the natural system, and using local conditions to drive the model. This notion that local conditions impact rating curve uncertainty the most (Coxon et al., 2015) triggers an interesting question of, "What local *information* is most useful in quantifying and reducing rating curve uncertainty?"

2.3.4 Information is informative

Information can be defined as the particular arrangement of signals that can provide a change in belief about the current state (Schement & Ruben, 1993). This belief, or better yet, the current state of our knowledge, changes over time as one receives information (Nearing et al., 2016). If we can provide information to our current state of knowledge, we implicitly can increase how much more we know. This gives rise to the idea that information has value and can lead to better decisions (Weijs, 2011). But information does not arise out of nothing, it is simply a matter of extracting all the available information. A classic example in hydrometry is shown in Figure 2.4. The gaugings provide more information than the relationship between stage and discharge alone, but also describe *what* is causing the change. Therefore, providing more information to the hydrographer and an example of increasing our knowledge, by updating our belief about the state. But just like you can extract more information from the data, one can also include *new sources of information* to the model.

Hamlet, Huppert, and Lettenmaier (2002) showed that the inclusion of El Niño Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) signals into their streamflow prediction model

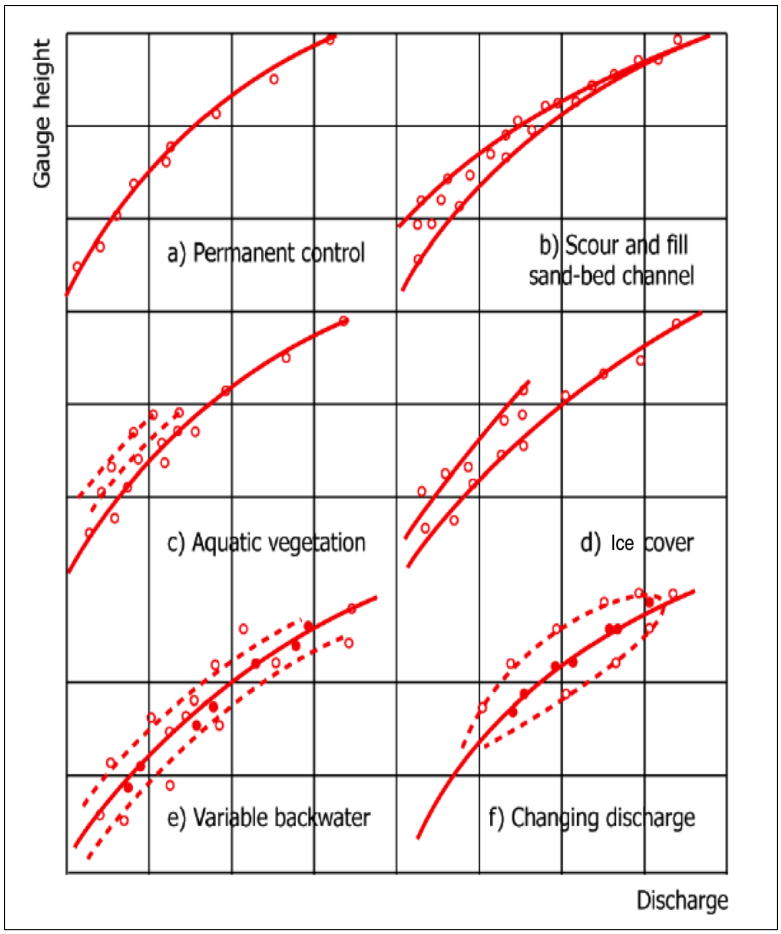

Figure 2.4: Different physical processes that are common in rating curve and an example of extracting more information than stage and discharge relationship (from Herschy (1995)).

increased the lead time of their model by six months, and improved the operating systems of a hydroelectric power plant in the Columbia River basin. It can be argued that the hydroelectric stakeholders may have found high value in including this exogenous information into the model. In terms of rating curves, Weijs et al. (2013) found that by using electricial conductivity (EC) signals, the uncertainty in the rating curve of an alpine stream in Switzerland was reduced by over 40%. The predictive power of the EC in the stream was found to be of similar magnitude to that of the water level. This manuscript is the only example known to the author, in which auxillary information has been used to improve the rating curve and its associated uncertainty, and raises an interesting question of, ”what other auxiliary information can be used to improve rating curve uncertainty?”

In British Columbia (BC) hydroelectric power plants are a common mean of producing electricity because of the advantageous geographical attributes in the province. As of 2014, there were 56 independent run-of-river (ROR) projects in BC, with 25 others expected to be operational by 2018 (Clean Energy BC, 2015). Each run-of-river hydroelectric plant must ensure compliance with the

regulatory authorities and meet an instream flow requirement (IFR) downstream of their stream intake. An IFR is placed to ensure the livelihood of aquatic and terrestrial life. To comply with IFR requirements, a rating curve may be used to predict the discharge. Depending on the location of the ROR project, a rating curve may be volatile due to the sediment load and energy of the river. These dynamics can create high uncertainty in the predicted discharge and can have high economical impact on the profit of a ROR project. This scenario sets up an interesting research landscape to identify the information available from a ROR project to be utilized in the development and improvement of a rating curve and its uncertainty.

This research will address the current gaps in the literature and develop a probabilistic dynamic rating curve that highlights regions of high and low conditional probability. The new rating curve will be described using conditional distributions that are updated as new gaugings are added to the model thus forcing the uncertainty bands to by dynamic. The new model will also allows users to produce continuous discharge time series, as well as identify the conditional probability of discharge occurring, given a water level measurement. The research will further explore the notion of identifying and utilizing available information towards uncertainty reduction as seen in the Weijs et al. (2013). In addressing these two gaps, the author hopes to convince the audience the necessity for transitioning into a more probabilistic mindset, as well as the importance of assessing, quantifying, and modeling rating curve uncertainty.

2.4 Research objectives

This thesis will focus on exploring the data/information available from a run-of-river hydroelectric project in northwestern BC to model a probabilistic dynamic rating curve. The central research questions to be addressed at the end of this journey are the following:

1. Can a rating curve be developed using probability distributions to produce similar results such as from a deterministic model?
2. Can the inclusion of auxiliary information from a run-of-river hydroelectric plant reduce the uncertainty in the IFR rating curve?

The thesis will also examine big picture ideas such as:

- The importance of utilizing available information to reduce model uncertainty
- Probabilistic versus deterministic models
- Dynamic uncertainty bands for hydrographs
- Economic considerations of modeling rating curve uncertainty

A handful of information theory measures and metrics will also be considered when evaluating the developed probabilistic rating curve and the auxiliary information used. Employing an information

theory framework becomes a natural choice when analyzing problems concerned with information and uncertainty.

Chapter 3

Data and methods for fitting rating curves

3.1 Site description

The data used in this project is from a hydrometric stations located on the Iskut River between the confluence with the Forrest Kerr Creek (FKC) and the braided reach of the Iskut River, a distance of about 20 km (Figure 3.1). The net drainage area of the watershed is approximately 9,500 km^2, but the watershed affecting the project is a fraction of the net drainage area, an estimated area of 6,978 km^2. The watershed includes a high plateau with broad valleys to the east and glaciated mountains to the west. The maximum elevation in the watershed is 2,558 m with a minimum of 244 m downstream at the FKC intake. The region around the Iskut River includes high glacial landforms that covers approximately 10.2% of the Iskut watershed (Northwest Hydraulic Consultants, 2016).

The area has a volcanic history that has helped form the geology of the region. Heavy volcanic rock is found directly upstream of the confluence on the Iskut River, that is transported through deep narrow volcanic bedrock valleys. The sediment enters the reach of the Iskut River where the project is located and is fed by small tributaries that originate from meltwater and precipitation. Due to the location of the project, the climate of the watershed is influenced by its mountainous terrain and its proximity to the Pacific Ocean. The annual precipitation of the watershed exceeds 2,500 mm/year (Northwest Hydraulic Consultants, 2016).

All the hydrometric stations have stage-discharge rating curves that are commonly altered by the unique characteristics of the Iskut Watershed. The high sediment load of the Iskut is a driving force for the production of new rating curves. The Iskut River is a highly active river that transports large amounts of sediment. It is estimated that the river caries over six million tonnes of fine sediment per year. Such high quantity of particles and boulders traversing the reach in which the six hydrometric stations are located is a cause for the volatile relationship between stage and discharge at each sta-

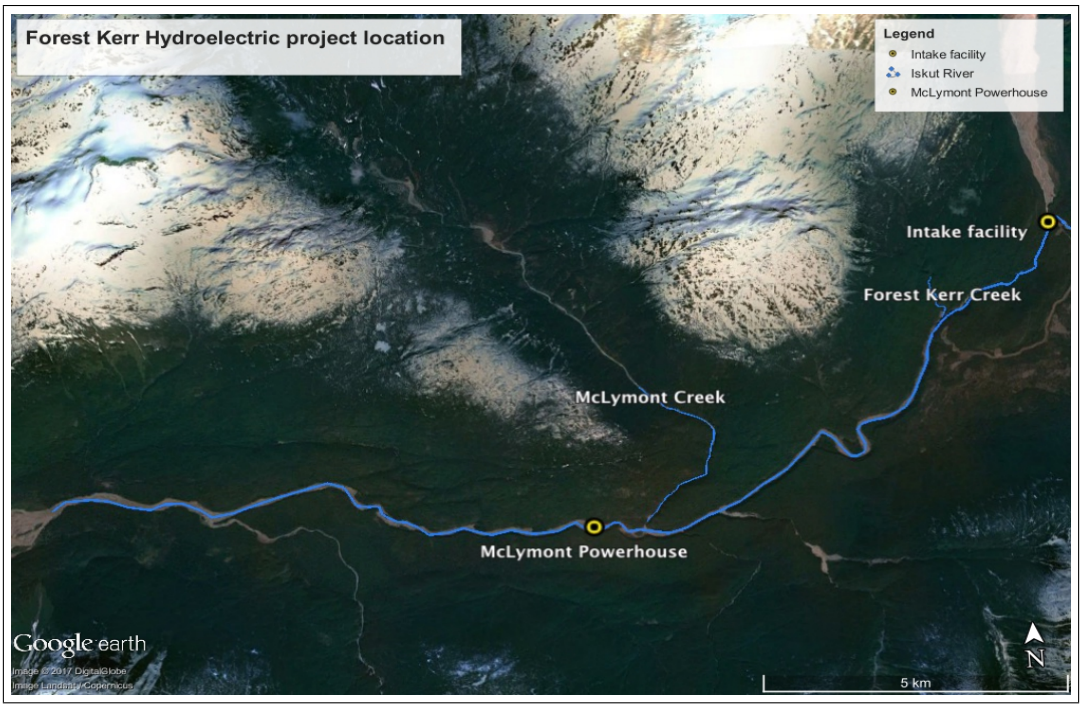

Figure 3.1: Location of Forrest Kerr Hydroelectric project. The Iskut river flows from east to west.

tion. To mitigate the impact of the sediment on the intake of the hydroelectric project, a box culvert was designed upstream of the intake, with a sluiceway and radial gate to facilitate the passage of sediment (refer to Figure 3.2). The sediment, however, does aggregate over time and as a result, the radial gate is opened to flush the sediment out (sluicing). This has an impact on the downstream hydrometric station that is used to ensure IFR compliance of 10 m^3/s since the large sediment loads typically alter the stage-discharge relationship.

Due to the altering effects that sluicing may have on the geomorphology of the river, new gaugings must be recorded immediately after a sluicing event to ensure an accurate discharge estimate. Altering events like this also imply that prior gaugings are probably not representative of the new stage-discharge relationship and therefore, older gaugings should have less weight in the estimation of the most recent rating curve.

Current practice involves utilizing only the most recent gaugings, and sometimes a subset of older gaugings, to create a rating curve. Unfortunately, it may not always be possible to measure a new full range of stage and discharge values to form a new rating curve, and so previously recorded gaugings are selected based on how the hydrographer believes the rating curve has changed. Although the modeler may have a strong intuition for how to select the gaugings, information on how the rating curve is changing may be lost due to selecting only a subsample of gaugings. Instead, this author argues that all the gaugings should be used with a weighted coefficient that gives more

Figure 3.2: A schematic of the intake structure just upstream of the IFR hydrometric station. Courtesy of Tim Argast of NHC.

weight to the most recent values, as well as those after a sluicing event. This allows the rating curve to include all information from the gaugings, as well as the new sluicing and time information.

3.2 Data availability

The data used to establish the rating curve originate from the IFR hydrometric station, while the added information is from the intake structure upstream of the IFR station shown in Figure 3.1. The hydrometric station has a record between April 2013 to July 2017 for a total of 105 gaugings. The first 19 gaugings were used to develop the initial rating curve on which the remaining gaugings were added. The reasoning for this is that a subset of gaugings is required to develop *any* rating curve, and it would not make sense, nor is it practical, to develop a rating curve with a small set of gaugings as it is unsure what the true relationship will be. The first 19 gaugings were selected as the initial points to develop the rating curve as this was the first curve created by NHC during their analysis.

The offset value (cease-to-flow water level) was determined to be 234 m using only the initial 19 gaugings. This offset value was carried into all the rating curve analyses and was not updated to reflect changes in the controls. A discussion on the decision for selecting 234 m is presented in the discussion section of this thesis. Lastly, all the data have undergone a natural log transformation to better fit the assumptions of normality. A bias does occur when the variables are log-transformed, used in linear regression by applying least-squares regression, and then back-transformed without accounting for the error term in the regression. When working with log-transformed variables, the error term in the regression has a mean of zero in logarithmic units, but the mean parameter of the error changes when the variables are back-transformed, and the error must be properly compensated for to ensure accurate predictions of the discharge. Readers should take note of this and can further read on the issue found in Newman (1993). However, in this thesis, the author is concerned more with producing probability distributions for the discharges, rather than mean values, but mean values can be calculated, and if back-transformed, the necessary corrections should be applied.

3.2.1 Sluicing data

Although sluicing events were not explicitly recorded, they were inferred by evaluating certain patterns from four other signals at the intake structure of the FKC project (Figure 3.3). High sediment build-up necessitates sluicing, and to prevent damage to the turbines and structures downstream of the intake pipeline, the power tunnel must be turned off (or reduced significantly). Less water is now diverted through the power tunnel and the radial gate must be opened to allow the water to flow downstream. The radial gate causes the water levels at the forebay and the sluiceway to decrease—forcing a higher discharge downstream at the IFR hydrometric station. This pattern is not perfect, and the water level at the sluiceway may not always drop when sluicing occurs. The reason for this is a combination of how fast the intake to the power tunnel is closed, how fast the radial gate opens, and the discharge of the Iskut River coming in. The four time series used in the identification of the sluicing events are on a one minute interval with records for each beginning at different stages of the project. The overlap between the time series begins in 2015 with each time series having over 1,500,000 values. A manual inspection of how the four time series change in relation to each other would be a laborious task. To aid in identifying sluicing events, a MATLAB code was written to find the location of the sharp changes in three of the four time series. The time series of the sluiceway water level was not used since it was identified by NHC employees to not always follow the pattern. Each time series is constrained with a threshold that was characterized by a sluicing event in October 2016. This sluicing event and its pattern is the only sluicing event that was used as an illustration to the author for what the general pattern in the other time series are during sluicing. The threshold was relaxed slightly to ensure the algorithm did not miss events. The dates extracted from each time series when the power tunnel discharge decreased, IFR increased, and forebay level decreased were found by the findpeaks MATLAB algorithm and then analyzed. The sluiceway signal was not used in the analysis because NHC employees found that the signal did not always behave as expected. Since the changes of all time series were not expected to align perfectly, a window of

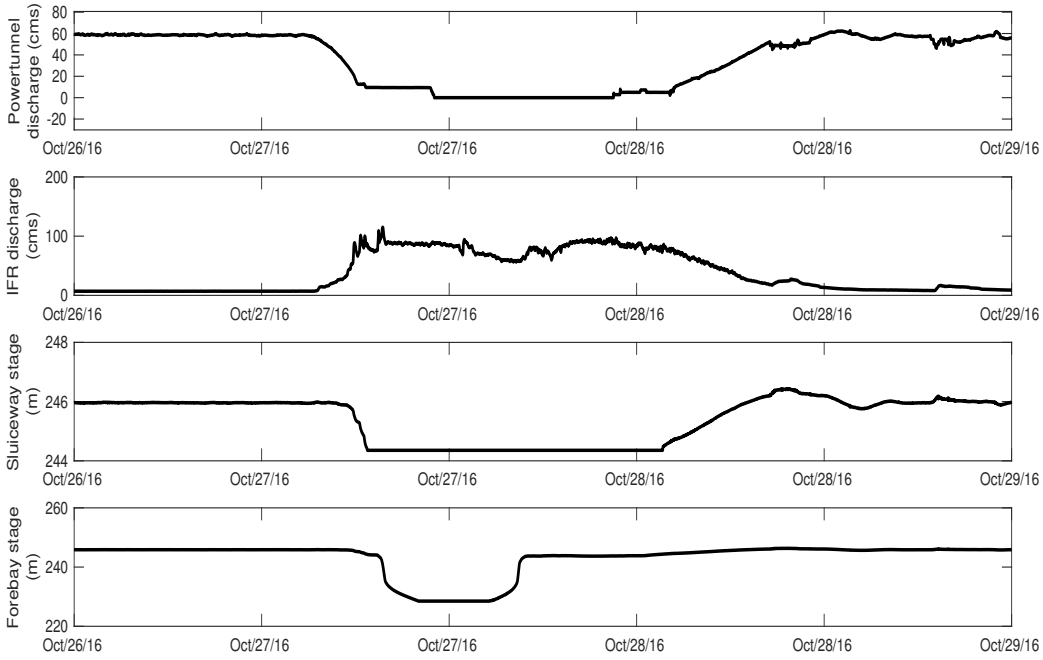

Figure 3.3: Time series of stage and discharge at two two different locations. A clear pattern for when sluicing was identified as having occurred. The forebay stage is located right before the power tunnel intake in the desander basin.

2 hours was selected as a possible period of overlap. The window was selected to try and ensure a greater overlap in dates between the time series, but the selection of the window was an arbitrary decision.

The algorithm identified over 70 possible events that shared the pattern described above. The events were then used to perform a second round of inspection by visually scanning through the time series more efficiently, and verifying if the identified events were sluicing episodes, or not. The second round of inspection done manually found a few sluicing events that the algorithm missed, as well as a few misclassified patterns. After filtering, 25 possible sluicing events were identified and verified. The reason that many of the original sluicing events identified by the algorithm were not selected as possible sluicing events is due to 1) characteristics and unpredictable behavior of the sluiceway and 2) storm events. The latter made it difficult to assess with certainty if sluicing occurred or not, since the pattern would display high variability in the signals. Although one could argue that sluicing is possible during heavy rain events, it may not always be the case, and for this reason, these dates from the original set identified by the algorithm, were not used. The estimated time for each sluicing event was also identified. This was done by measuring the average time that it took the forebay water level and discharge from the power tunnel to go back to the level prior to dropping. The forebay and powertunnel were used since they proved to be the most consistent and stable time

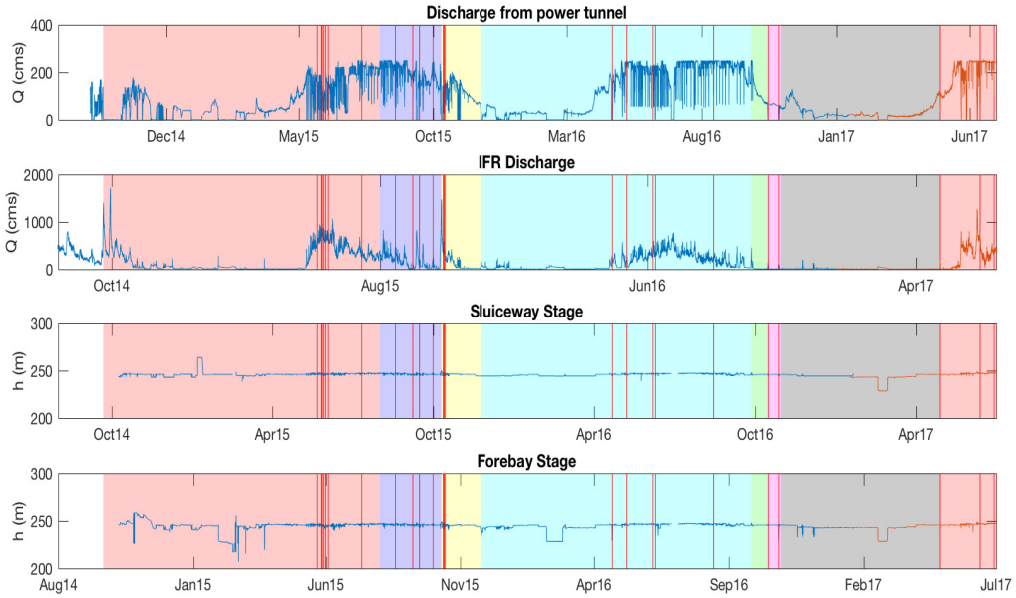

Figure 3.4: Time series of stage and discharge at two different locations. Changes in the time series with identified sluicing events detailed as red vertical lines. The colored patches represent time periods in which new NHC rating curves were developed.

series. Figure 3.4 presents the four time series with red lines indicating periods where sluicing was presumed to occur. The colored patches represent different rating curves created by NHC and the duration the curves were active. Figure 3.4 shows the relationship that the duration of some of the rating curves had with the number of sluicing events. Although not every new curve was a consequence of a sluice event, there were a few that appeared to be. To better illustrate the sluicing events, a histogram of the events was created (Figure 3.5). The months with the largest events were May, June, and October, which coincided with the behavior of the Iskut watershed. May and June are typical months when snow and ice melt and force an increase in discharge on the Iskut River (Figure 3.6). With an increased discharge and steep valleys near the confluence, large sediments and boulders can be carried and deposited at the intake structure. The reason for a high number of sluicing events in October may be due to a partial combination of the amount of precipitation that the region receives from the Pacific Ocean, the volcanic geology, and the high glacial concentration in the watershed. High rainfall can cause sediment from across the watershed, including glacial sediment, to be transported downstream through the Iskut and into the sluiceway.

3.3 Rating Curve Model

The rating curve model developed in this thesis is purely probabilistic and can operate in forecast mode, which uses real-time data to develop the most recent rating curve. This mode is desirable for operators needing to monitor IFR to ensure compliance. The model predicts the next discharge

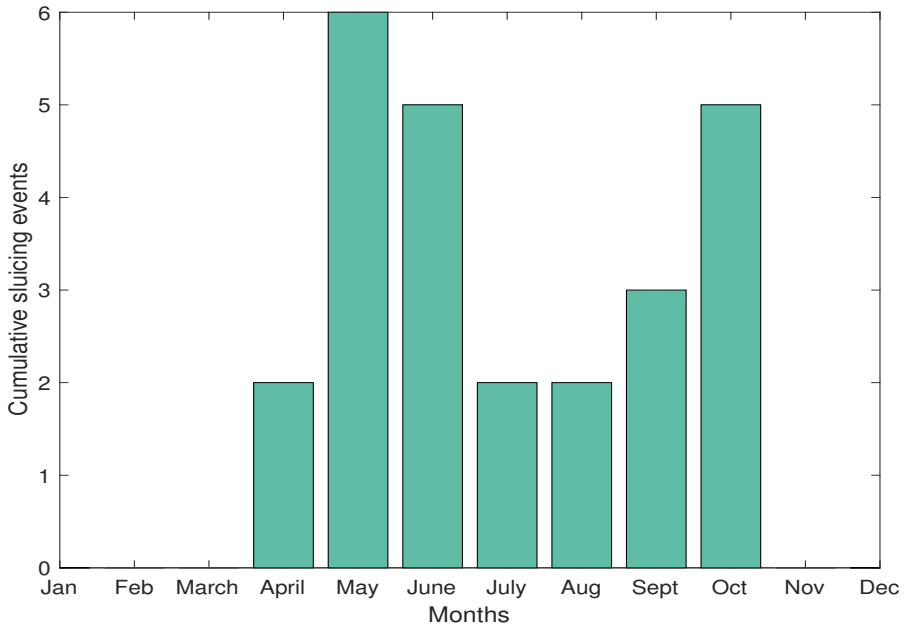

Figure 3.5: A bar graph showing which months experienced the most sluicing events between January 2015 and July 2017.

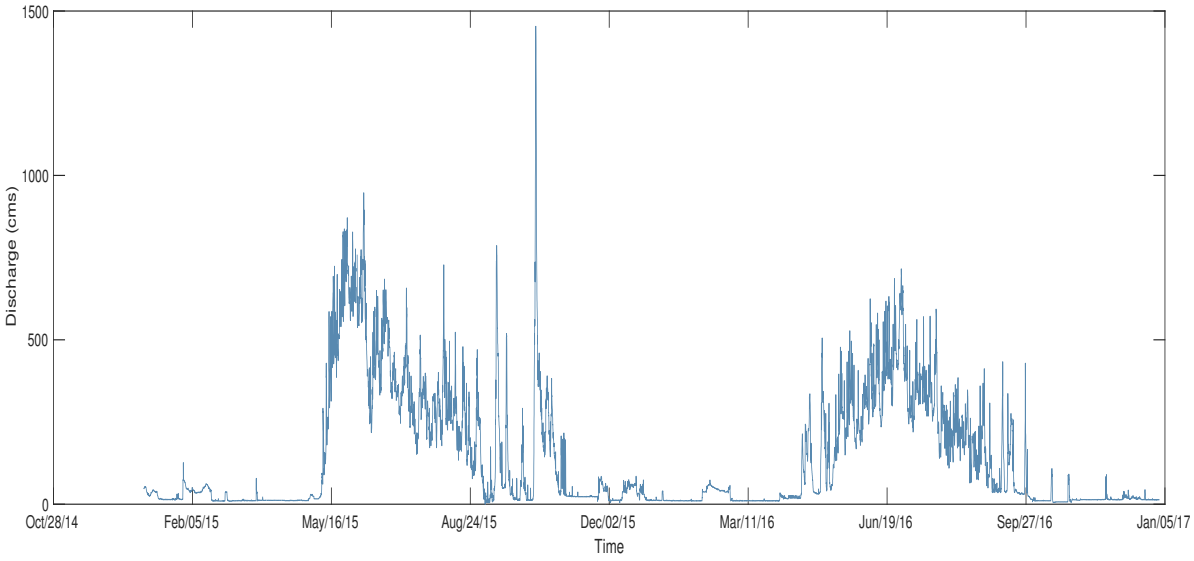

Figure 3.6: Hydrograph at IFR hydrometric station between January 2015 and July 2017.

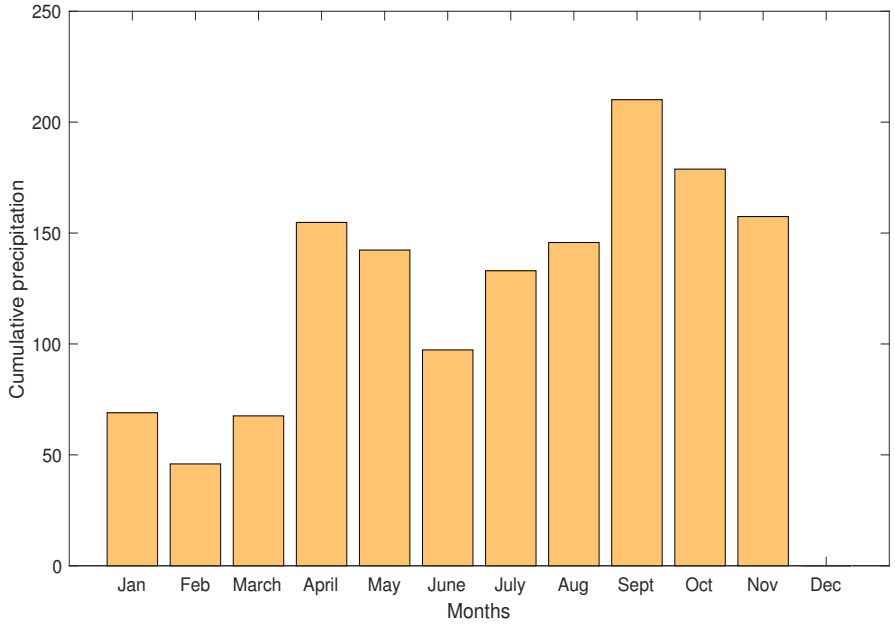

Figure 3.7: A bar graph showing which months experienced the most precipitation between January 2015 and July 2017.

gauging by propagating the information used (i.e. time/sluicing weights) to update the Gaussian kernels to the most up-to-date model, prior to the next gauging being added.

3.3.1 Forecast model

The forecast model operates by updating the rating curve with gaugings as they are observed and makes a prediction for the next gauging. The model uses a special case of the leave-one-out cross validation (LOOCV) method, in which the time and sluicing models are propagated to the period in which the next gauging is observed, and then makes a prediction for that value *without* using that gauging in the subset of data. Once the prediction is made, the observed gauging is added, and a prediction for the next gauging is made using the same process. The LOOCV method was preferred for the validation of the rating curve due to the limited dataset, and is a realistic evaluation of uncertainty in the way rating curves would be used without using data from the future. Unfortunately, dividing the dataset into training and test sets would not provide enough data in either set.

Time model

The time model used in the forecast option of the rating curve resembles an exponential distribution

$$y = f(x|\mu) = \frac{1}{\mu} e^{\frac{-x}{\mu}} \tag{3.1}$$

where y represents the weight, x is the time difference between the current time step and previously observed gaugings, and μ is the expected value of the exponentially distributed variable—its inverse, λ, is the rate at which the weights of the gaugings change. Equation 3.1 is typically used in financial models for monitoring future market prices (Longerstaey & Spencer, 1996; Pafka & Kondor, 2001). Recent data are assigned larger weights than the older data that may not necessarily represent the present. In hydrometry, rivers of high energy carrying large sediment can generate new cross-sections and invalidate prior gaugings.

The variable μ was selected to be 365 so as to make the time weights change at a rate of $\frac{1}{365}$. A sensitivity analysis was performed and results showed minimal variability between the final results and the value of μ chosen. An illustration of the process is shown in Figure 3.8. As time

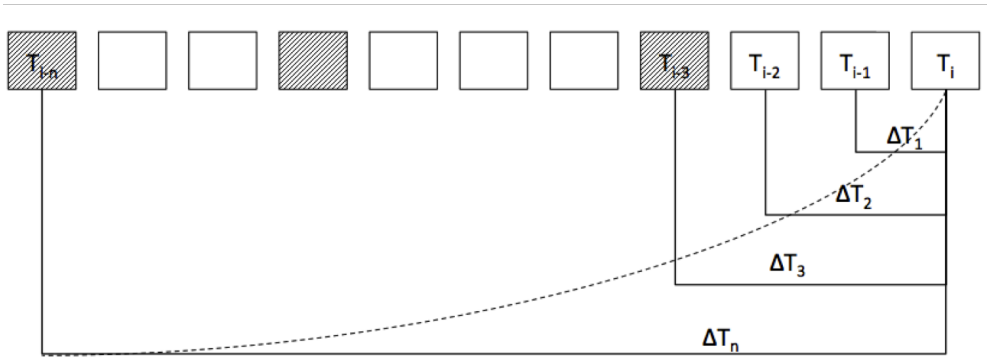

Figure 3.8: More weight from the exponential function (dotted line) is assigned to the gaugings (colored boxes) that are closer to the current time step, T_i.

progresses towards the next observed gauging, the exponential distribution is updated to reassign weights to the already observed gaugings.

Sluicing model

To create a probabilistic model of the sluicing information, a Gaussian kernel was drawn around each event and then summed to create a kernel distribution. The weights of the individual kernels were determined by the time duration of the sluicing events. As sluicing information becomes available in time, the kernel distribution is updated to eventually span between May 2015 and July 2017 (Figure 3.9). In the forecast mode, only the observed sluicing events are used to develop the kernel distribution. The cumulative distribution function (CDF) is then obtained and used to assign weights to the observed gaugings. Gaugings after the last recorded sluicing event are assigned higher weight than those that occurred prior to sluicing. Figure 3.10 shows the evolution of the normalized weights taken from the updated CDFs, while Figure 3.11 shows the final sluicing weight assignment when all the gaugings were observed. A comparison of Figures 3.10 and 3.11 show a sharp change in the value of the weights when the weight for the 52nd gauging is evaluated. This was done since sluicing information was unfortunately only available post May 2015 and there were

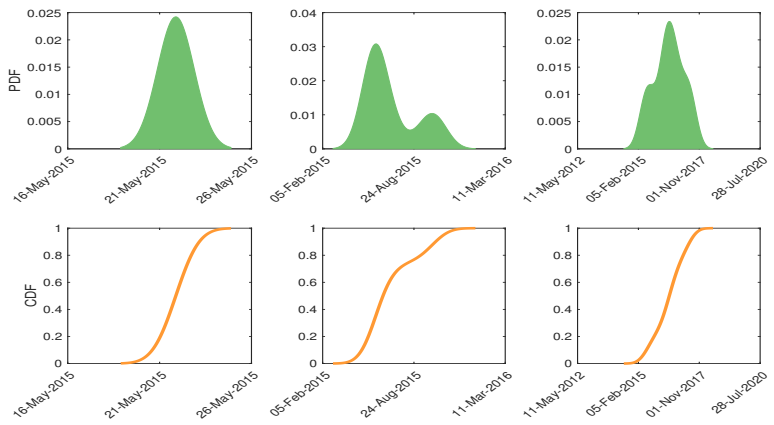

Figure 3.9: The kernel PDF and CDF are updated as more information on possible sluicing events enter the model. The first panel shows the PDF and CDF of the first sluicing event, the second panel shows and updated version, and the final panel shows the model with all the sluicing information.

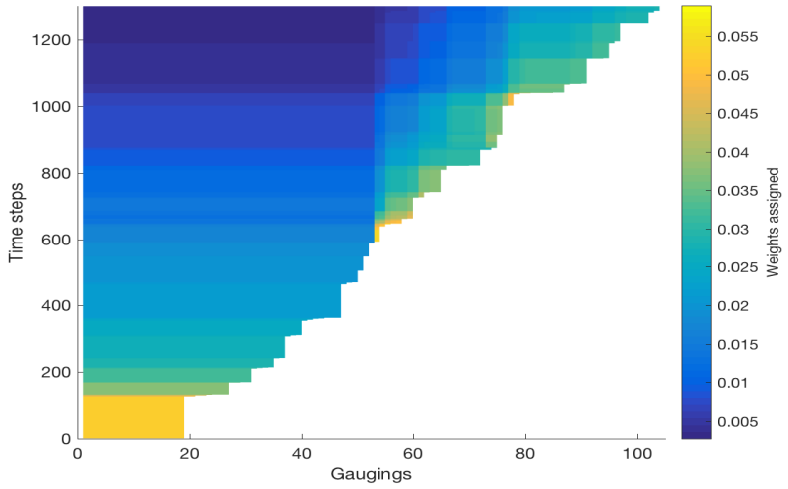

Figure 3.10: An evolution of how the sluicing weights change as new gaugings are observed (x-axis) in time. Each row is a vector of the sluicing weights

51 gaugings that were recorded prior to the first sluicing event available to the model. Using the current distribution model would assign a significantly smaller weight to those values and would not be a fair weight assignment. Instead, gaugings prior to the first sluicing event were assigned equal weight of one, which after normalization the weights would be lowered as more sluicing events enter the model.

3.3.2 Hindcast model

In hindcast mode, the probabilistic rating curve can be tested to evaluate how accurate the predicted discharge output is with the known value, given all the information. The predicted gauging is re-

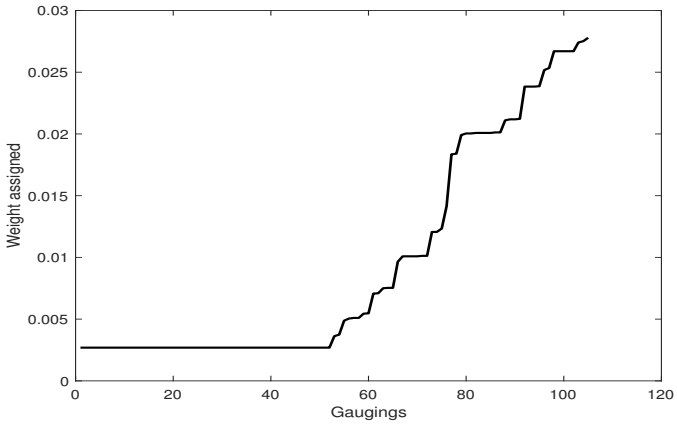

Figure 3.11: An equal weight assignment for the gaugings prior to sluicing information and then the weights increase in magnitude for values later in time. The figure is a snapshot if Figure 3.10 at the time of the last added gauging.

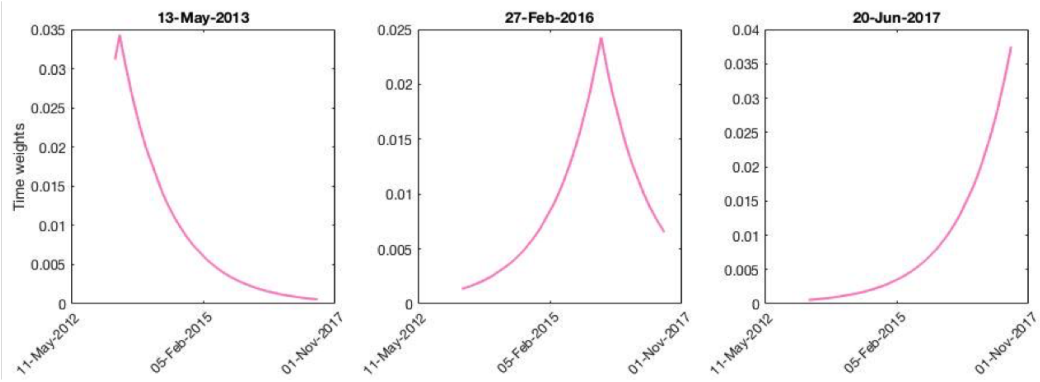

Figure 3.12: Evolution of the time model at different time period.

moved from the dataset so as to eliminate any bias, prevent overfitting, and keep testing independent of predictors. This procedure is repeated for every gauging, and therein utilizes a LOOCV method.

Time model

To perform the time model in hindcast mode, the exponential function was modified into a Laplace function so as to assign the greatest weight to the gaugings closest in time to the one being predicted, and less to those further out (Figure 3.12. The weights were computed as follows:

$$y = f(x|\zeta, b) = \frac{1}{2b} \begin{cases} e^{-\frac{\zeta - x}{b}}, & \text{if } x < \zeta \\ e^{\frac{x-\zeta}{b}}, & \text{if } x > \zeta \end{cases} \tag{3.2}$$

where the parameter ζ is the location parameter which is set as the current timestamp of the gauging under evaluation and the parameter b was set to 365, equivalent to the μ parameter in Equation 3.1. An example of what the distribution looks like is shown in Figure 3.12.

Sluicing model

The sluicing model was modified slightly to more accurately echo the effect of sluicing events. The model follows a similar pattern as the time model, where the effect of sluicing events decreases the further the gaugings are from the gauging that the hindcast model is evaluating. However, since the first 51 gaugings did not have any sluicing information available, those gaugings remained with a uniform distribution. Then, once sluicing information became available for the remaining gaugings, the weighted CDF described earlier was used to produce the function necessary to assign the remaining weights. The negative of the function was taken so as to have the function vary from zero to negative one and then a value of one was added to the function to align it with the equal weights of the first 51 gaugings. This was necessary to again ensure that the later gaugings had less influence on the first 51 gaugings. The function was then normalized and the sluicing model for the first 51 gaugings in hindcast mode is shown in the top left panel of Figure 3.13. As the hindcast model propagated towards the later gaugings, the model had to be altered to ensure highest weight is assigned to the gaugings closest to the gauging being predicted. For this, three discrete functions had to be appended together. The first function described the uniform weight for the first 51 gaugings, and the second and third functions were CDFs. The first CDF used all the gaugings starting at the 52^{nd} gauging, up to the gauging being evaluated, while the second CDF used the remaining gaugings. The second CDF was negatively transformed, and a value of one was added to translate it up and align it with the first CDF. This then produced a function as shown in the bottom left panel of Figure 3.13. The CDF was chosen because of its natural curve, and because of its feasibility to add the time duration of the sluicing events.

A few changes had to be made to the sluicing model when the hindcast model was evaluating the 52^{nd}, 53^{rd}, and the last two gaugings since producing a CDF using the *ksdensity* function in MATLAB produced undesirable results when appending all three functions. Instead, the 52^{nd} and 53^{rd} gaugings were assigned the highest weight of the CDF using the remaining gaugings (aka a value of one), and then normalized. This produced a function that looks like the top right panel in Figure 3.13. For the last two gaugings, the same was done in which the 104^{th} and 105^{th} gauging were assigned values of one (bottom right panel in Figure 3.13).

When the functions were all appended, it produced a discrete function of length L, that was 151 (100 from the *ksdensity* function and 51 from the first gaugings) when the hindcast model was evaluating the first 51 gaugings, and then had a max of 251 when evaluating the other gaugings. Using the function and the timestamps of the gaugings, the proper weights were then selected.

3.3.3 Development of rating curve

Using the information provided above, four different models per rating curve mode (forecast and hindsight) can be developed and evaluated. The four models are:

1. Rating curve that utilizes both sluicing and time information

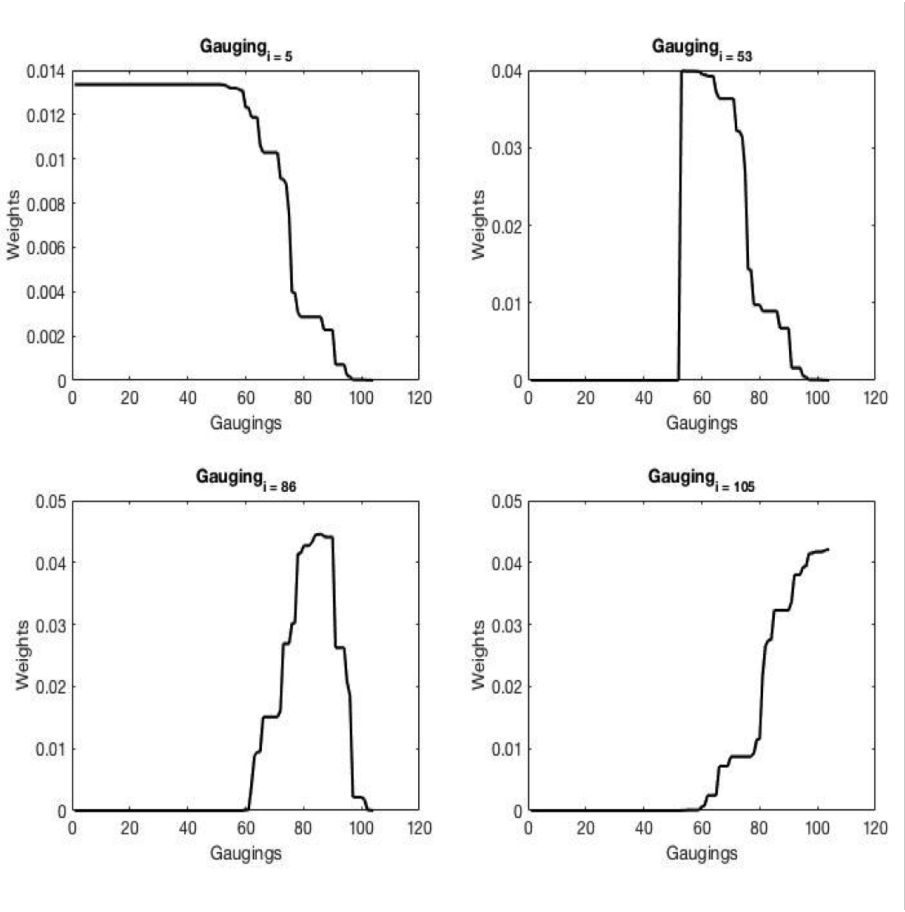

Figure 3.13: Evolution of the sluicing model when applied in hindcast mode. An evolution of how the sluicing model changes in the hindcast mode of the rating curve. The top left panel shows the function typical for the first 51 gaugings. The top right panel shows the function for the 53^{rd} gauging. The bottom left is the typical function for the gaugings between 53^{rd} and 104^{th} gauging, with the peak of the function centered around the gauging being evaluated during the hindcast model. The bottom right function is the function for the last gauging.

2. Rating curve that utilizes only time information
3. Rating curve that utilizes only sluicing information
4. Rating curve with no extra information added

Note that the model where no extra information is used assumes that all the gaugings have equal weight. The foundation of how the model operates is based on a weighted joint distribution between all the gaugings. The joint distribution for the rating curve was calculated using two dimensional kernel density estimations for the gaugings, with weights that give more, or less, importance to some gaugings than others. The kernel for each gauging was a multivariate normal probability

distribution:

$$y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'} \tag{3.3}$$

where y is the probability distribution, μ in the multivariate distribution equation is a 1-by-d vector that describes the location, or center, of the observed gauging and Σ is a d-by-d symmetric positive definite covariance matrix that describes the spread and scale between the two variables. The value d in Equation 3.3 is two since only stage and discharge are evaluated. The summation of each Gaussian distribution, and the weights, create the joint distribution for the stage and discharge relationship.

Covariance function

The covariance matrix in Equation 3.3 helps describe the geometry of the Gaussian kernel. The covariance function is written as such:

$$\Sigma = \begin{bmatrix} \sigma_H^2 & \sigma_{HQ} \\ \sigma_{QH} & \sigma_Q^2 \end{bmatrix} \tag{3.4}$$

Here, the parameters σ_H and σ_Q explain the standard deviation, or spread, in the x and y direction of the kernel, which are related to the marginal distributions, while σ_{HQ} represent how the two parameters change with respect to each other. Note that H represents h-h_0.

In Equation 3.4, the parameters are unknown but can be solved for using the measurement uncertainty. In Figure 3.14, the measurement uncertainties are used to represent the uncertainty in the gaugings. These uncertainties are related to the conditional probability of the kernel at the gaugings, and can be used to solve for the parameters in Equation 3.4. The relationship connecting the uncertainties in the conditional and marginal distributions are shown in Equation 3.5 and Equation 3.6.

The parameters σ_{H^*} and σ_{Q^*} in Figure 3.14 represent the standard deviations of the conditional distributions at the gaugings, and are related to the marginal distributions via Equations 3.5 and 3.6.

$$\sigma_H = \sqrt{\frac{\sigma_{H^*}^2}{(1-\rho^2)}} \tag{3.5}$$

$$\sigma_Q = \sqrt{\frac{\sigma_{Q^*}^2}{(1-\rho^2)}} \tag{3.6}$$

$$\sigma_{H^*} = \frac{\rho * \sigma_{Q^*}}{S} \tag{3.7}$$

The parameter ρ, a weighted correlation coefficient, was found by using only the gaugings observed thus far by the rating curve model and their assigned weights, and every time a new gauging was observed, a new correlation coefficient was calculated. However, as shown in Equation 3.7, there

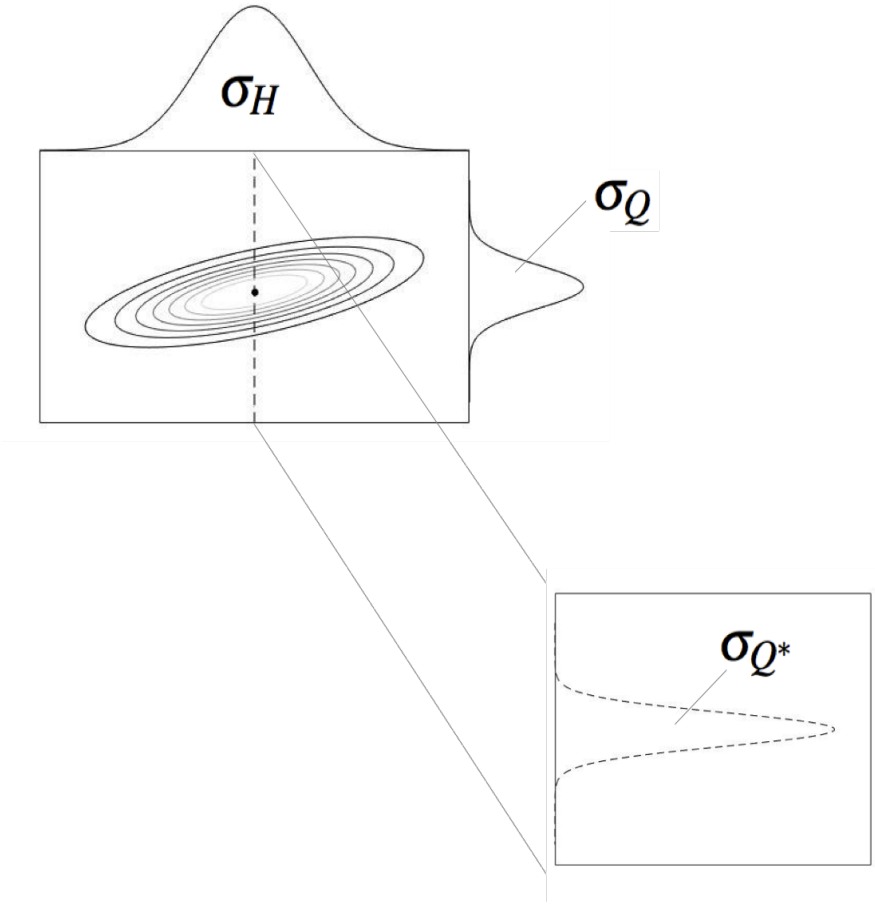

Figure 3.14: A multivariate Gaussian kernel with its marginal distributions and a conditional distribution. Note that only the scaling parameters are shown for illustrative purposes. The location parameters per kernel are fixed to the value of the gaugings.

exists a relationship between the uncertainty parameters of a Gaussian multivariate distribution, its slope, and the correlation coefficient, so therefore, all four parameters cannot be fixed without ensuring Equation 3.7 is satisfied. In this thesis, the slope (S) of each kernel was fixed to a value θ that was found by using the same subset of data as described in the correlation analysis, and performing a weighted linear regression. Ensuring that all the slopes of the kernels are aligned aids in interpolation and extrapolation of discharge values, as well as produces a model that resembles the shape of a deterministic rating curve. To ensure that Equation 3.7 was met, the slope and correlation factors were fixed by the described methods, and σ_{Q^*} was assumed to be 5% of the magnitude of the discharge gaugings. This left σ_{H^*}, the conditional uncertainty of Q given $H_{gauging}$ as the unknown parameter. This was done because it was assumed that the original assumption of 2 mm for σ_{H^*} would limit the ability to correctly identify σ_{Q^*}, and might underestimate the uncertainty in the discharge. Given that if real measurement uncertainty data were available, the correct numbers could be used to determine the shortcomings in the parameters. The assumption of 5% of the discharge magnitude as Σ_{Q^*} was shown to be a reasonable value in the literature (Di Baldassarre & Montanari,

2009; Herschy, 1995; Rantz, 1982). The author does recognize that this assumption could have an influence on evaluating the uncertainty of a rating curve, but due to the lack of data on how discharge was measured, this assumption had to be made, but can be improved on, if the true discharge measurement uncertainties are available.

Solving for all the parameters in Equation 3.4, the individual covariance matrices per gauging were calculated and Equation 3.3 was used to develop the Gaussian kernels. The summation of all the individual kernels produced a joint distribution for the gaugings numerically evaluated on a discrete grid of size 500x500 with a resolution of 0.0123 logarithm units that is based on the minimum and maximum values for the gaugings. From the joint distribution, the conditional distribution was calculated by normalizing every column in the joint distribution to one—therefore producing the probabilistic dynamic rating curve. The conditional distribution is then useful for extracting the conditional distributions of Q given a water level measurement from the continuous logger. To obtain a point value estimate, the mean or mode of the distribution can be taken. For this thesis, the weighted mean of the distributions were used. From the rating curve, the conditional distribution of Q given the continuous water level measurement can be extracted and appended to create a hydrograph with dynamic uncertainty bands, as shown in the results.

3.4 Metrics for evaluating rating curve

To understand the accuracy of the probabilistic rating curve model, an evaluation of its performance must be undertaken. In hydraulic and hydrologic models, one of the most commonly used methods is the root mean square error (RMSE) where O represents observations, P predicted values, and n is the sample size:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \tag{3.8}$$

RMSE spans between 0 and ∞, and a value of 0 indicates a perfect fit. Many local and federal authorities designated with standardizing and ensuring quality projects grade the nature of the rating curve based on the RMSE. In this thesis, the RMSE was performed on the logarithmic discharge. When in forecast mode, the RMSE is calculated prior to adding the gauging and after. The results only show the RMSE before the gauging is added to the model, only otherwise stated. The RMSE after the gauging was added always decreased.

Since the model developed is centralized around probability, it makes sense to use metrics that do justice to the probabilistic nature.

3.5 Information theory metrics

In 1948, Claude Shannon published his seminal paper on the quantification of information from messages and signals. In it, he introduces the concept of entropy as a measure of uncertainty and

developed a mathematical form to quantify the uncertainty in bits as shown in Equation 3.9:

$$He(x) = -\sum_{i=1}^{n} P(x_i) log_2(P(x_i)) \tag{3.9}$$

where He is the entropy of the discrete probability distribution and P is the probability of the event x_i. Uncertainty is defined as the information that is unknown and is a function of the current state of knowledge as defined by a probability distribution. So as one receives more information, the probability distribution can change to reflect a better outcome of the model. When evaluating the surprise of the models, the conditional probability for a given water level and expected discharge was used.

3.5.1 The model that is less surprised is better

Take the example of a coin toss. It is known that a fair coin has a 50-50 chance of landing on either heads or tails. However, after many tosses, the coin flies through the air and instead of landing on one of the faces, it lands on the edge of the coin. An event like this has such a small probability of occurring that it causes a huge surprise for you. Using information theory, surprise can also be quantified:

$$S_x = -log_2(P(x_i)) \tag{3.10}$$

where S represents surprise. What Equation 3.10 says is that an event with a small probability of occurrence will have a much larger surprise. This principle of surprise can be extremely useful when evaluating how good a predictive model is. When thinking about surprise in terms of the rating curve, a predicted discharge significantly higher than the largest gauging in the model, will tend to surprise the modeler. But in current applications, rating curves only use a subset of the data, rather than the historical set. So although a higher gauging may exist in the set of gaugings, the rating curve may not be using that gauging and the higher discharge value then becomes more of a surprise than if having used the gauging from the historical dataset.

There is of course a trade-off between robustness and accuracy when considering using all the gaugings, and the decision for this is dependent on the purpose for which the model serves. In this thesis, the robustness of the rating curve is improved by the inclusion of all the gaugings, rather than selecting a subset of the data, and as will be shown in the results section, the accuracy can also improve by including more information.

The surprise of all four models will be evaluated and compared for every predicted gauging in forecast and hindsight mode. The results will provide an indication as to which model was surprised the least, and help identify which out of the four, might be a better predictive model. The surprise will also be measured before and after an observation (when a gauging is added) to compare how the gauging impacts the surprise of the model.

3.5.2 Information gain

Information gain of individual gaugings can also be quantified through the usage of information theory by evaluating the divergence of the prior PDF and the posterior PDF when a gauging is added. The gaugings with the highest information gain, could prove beneficial to the modeler, and might provide insights to any shortcomings of the rating curve. The measure of information is determined by using the Kullback-Leibler divergence measure:

$$D_{KL}(P||Q) = \sum_i P(i) log_2(\frac{P(i)}{Q(i)}) \tag{3.11}$$

Equation 3.11 measures how far the discrete probability distribution Q is from the true distribution P. A D_{KL} of 0 would indicate that the two distributions are equal and that no information was gained, while a larger value means a difference is observed. There are two considerations to using the Kullback-Leibler divergence that should always be accounted for. The first is that the Kullback-Leibler divergence is not symmetric and so $D_{KL}(P||Q) \neq D_{KL}(Q||P)$ and the second is that although information gain was observed, the $D_{KL}(P||Q)$ will not provide details on the quality of the information (i.e. good or bad). Inferences must be made using other available data to determine, if possible, that the information was beneficial.

If the true probability distributions for the gaugings were available, rather than the 5% assumption, the information gain from the forecast to the gaugings, could be evaluated. The results would help indicate which model most closely resembles the true probability distribution of the gauging ($D_{KL}(Obs||Forecast_{model})$). Given that the true data were not available, this analysis was not performed since the results may not necessarily portray what is actually happening.

In this thesis, the information gain for each gauging will be evaluated while in forecast mode where the "true" distribution will be the probability distribution function given to the gauging, *after* the gauging was added to the model. The prior distribution will be the probability function assigned by the models prior to seeing the gauging. In other words $D_{KL}(Gauging_{after})||(Gauging_{before})$. The information between models will also be examined, where it is assumed that the true distribution is equal to the models using weights, and the prior distribution comes from the Equal Weights model. This is done by taking the probability distributions of each gauging during forecast mode for the All Weights, Only Time, and Only Sluicing models, and comparing it with the probability distribution from the Equal Weights.

Chapter 4

Results of probabilistic rating curves

4.1 Forecast mode

4.1.1 Probabilistic dynamic rating curve

To compare all four models, the probabilistic dynamic rating curves are plotted together in Figure 4.1. Note that a rating curve was produced every day during the simulation and Figure 4.1 is only showing the rating curve for May 15, 2016. Immediately, a few observations become apparent, one of them being what appears to be a linear relationship between stage and discharge, and the other being a fragment in the model around a logarithmic discharge of -1.5. The linear relationship is as expected due to the properties of log transforming a power law. Also, since the slope of all the kernels were fixed to be identical, it was expected that all the kernels would be oriented in the same direction. The fragment in the linear relationship is a possible consequence of how the slopes for the kernels were calculated, and assumed to be equal to the slope of the all the gaugings seen up to the time period at which the rating curve was produced. This implies that if a new observed gauging does not reside near the existing linear relationship, its slope will be parallel to all the other gaugings, and therefore create a parallel kernel as seen in Figure 4.1.

A closer inspection of Figure 4.1 demonstrates the differences in the models. For example, looking at the high end of the probabilistic dynamic rating curve for all four models shows that the Equal Weights and Only Sluicing model have more similar conditional distributions than the Only Time and All Weights models. The latter two models appear to have more regions of higher probability which are caused by the influence of the assigned weights decreasing the surface area of the kernels on the two dimensional space, and increasing it in the third dimension (probability) centered at the gauging. From Figure 4.1 it can be seen that the Only Sluicing model resembles the Equal Weights rating curve more than the Only Time rating curve. Prior to May 15, 2016 (date of the rating curve in Figure 4.1), there were 14 sluicing events of varying magnitudes that influenced the shape of the rating curve kernels and are the reason for the distinctions between the Equal Weights and the Only

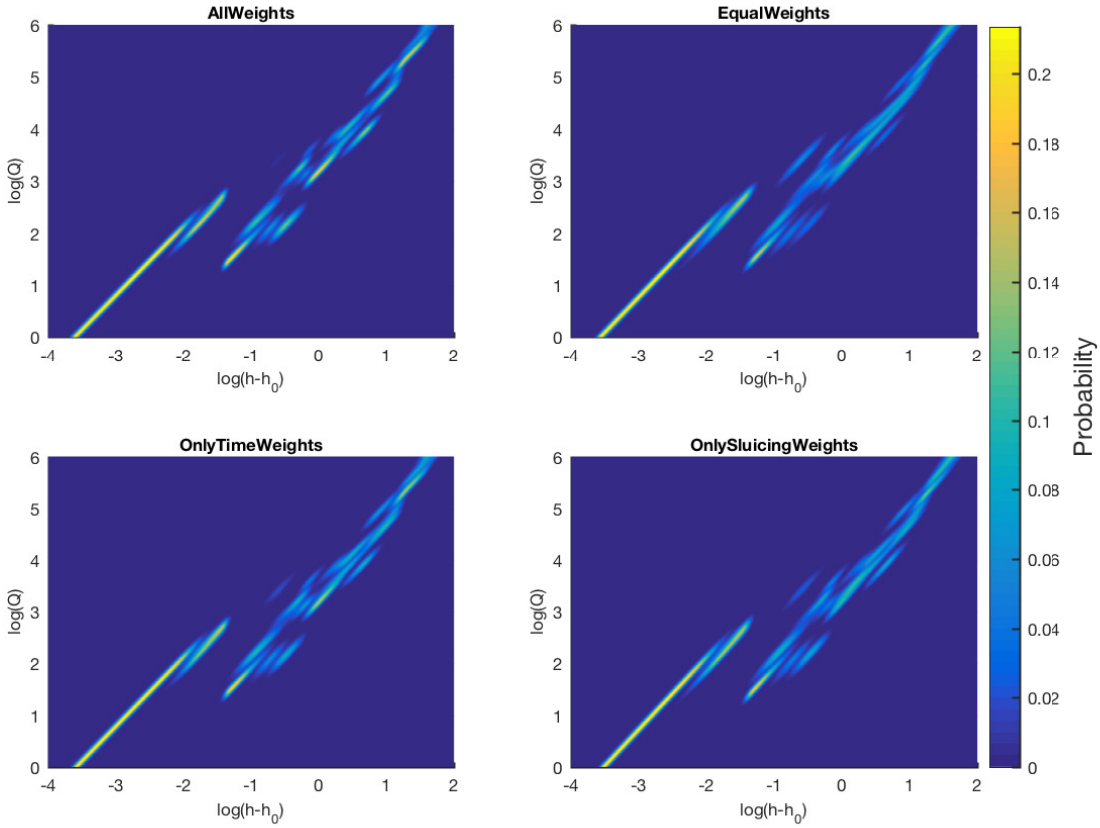

Figure 4.1: The conditional probabilistic rating curve for all four models at a snapshot in time. May 15, 2016.

Sluicing rating curves. Finally, the All Weights rating curve, which is a product of the time and sluicing weights, seems to have more concentrations of higher conditional distribution peaks than the other models. This is possible in the event where gaugings were taken after a sluicing event and were also the latest observed gaugings.

To better understand how the models differ, the conditional distributions of Q given an arbitrary logarithmic water level measurement of 1.3686 are compared in Figure 4.2. The results display that all four conditional distributions appear to have two large peaks which indicates two unique points of high conditional probability. The reason for this is due to the influence of two discharge gaugings with values centered around the peaks of the conditional distributions and their assigned weight. The results of Figure 4.2 are taken from the rating curve shown in Figure 4.1. Zooming into the rating curves, a better inspection of the kernels can be performed. The kernels from Figure 4.3 align with what is seen in Figure 4.2, where two regions of high conditional distribution are influencing the conditional distributions. The fact that the conditional distributions can be multi-modal, is the reason why a weighted mean is used to calculate the expected point value discharge rather than

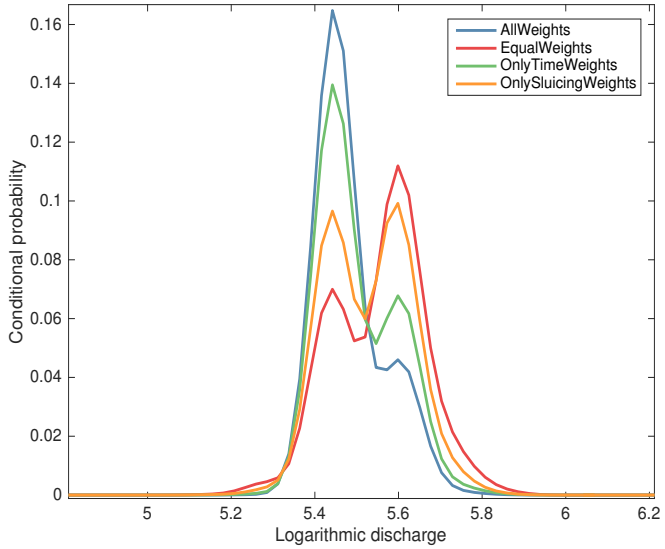

Figure 4.2: A comparison of the conditional distributions for a give logarithmic water level measurement.

using the mode.

4.1.2 Dynamic conditional uncertainty bands for hydrograph

Using the conditional distributions for a given water level measurement, a hydrograph with conditional uncertainty bands can be produced (Figure 4.4). The figure shows how the uncertainty in the hydrograph varies in time and also illustrates areas where the conditional density is highest. Also shown in Figure 4.4, are the predicted continuous discharges from the NHC deterministic rating curves, as well as the predicted discharges from the probabilistic dynamic rating curves. Taking a closer examination at the hydrograph, one can get a better understanding at how the conditional uncertainty is varying in time. Figure 4.5 also compares the predicted continuous discharge values from the deterministic and the probabilistic rating curves. Figure 4.5 shows the area of highest probability in the dynamic uncertainty band follows the same pattern as the predicted discharge values from the deterministic and probabilistic rating curves for the All Weights rating curve. The hydrographs for the other three models are attached in the Appendix.

An important take-away from the probabilistic dynamic rating curves is that the answer to an important question that could not have been asked in a deterministic rating curve is now possible. Prior to a probabilistic rating curve, one could not receive a response to, ”What is the probability distribution of a discharge, given a specific water level measurement?” This central question helped motivate the research undergone thus far and is to be considered by the author, an important question that presumably does not get asked often when dealing with deterministic rating curves. However, it is a question that should be considered highly, by not only those who produce rating curves, but anyone

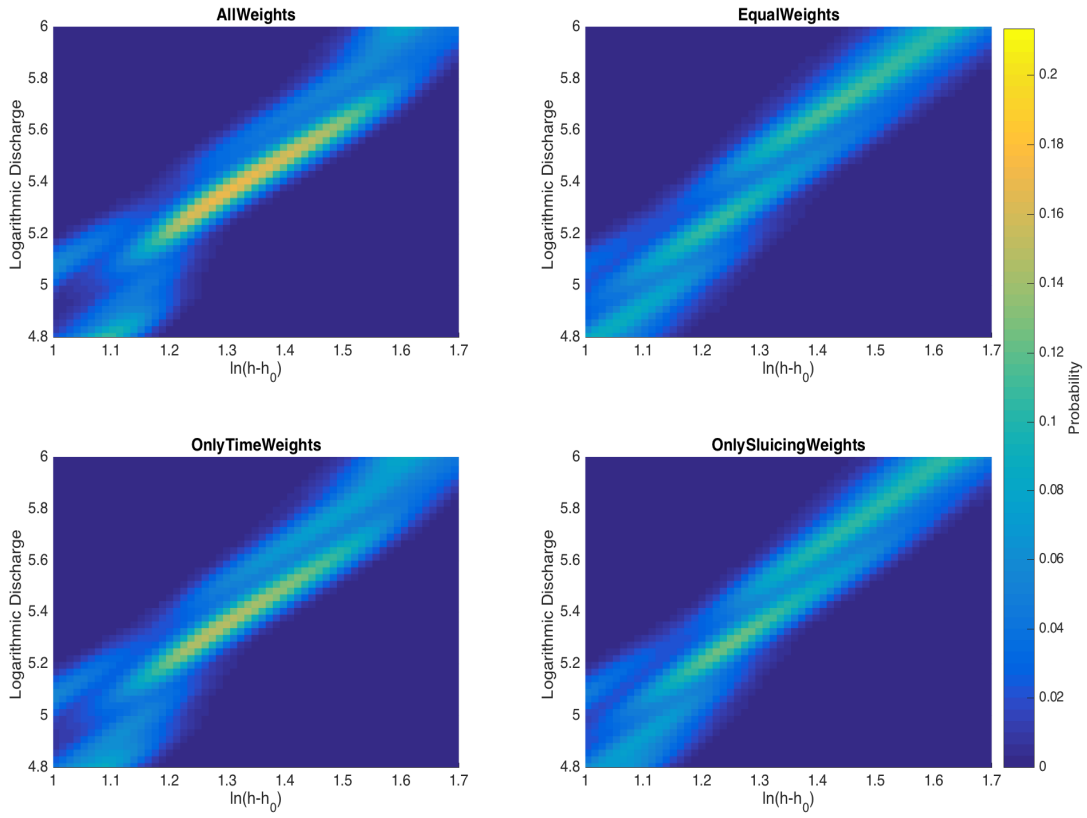

Figure 4.3: A zoomed in version of Figure 4.1 used to understand the bimodal conditional distribution shown in Figure 4.2.

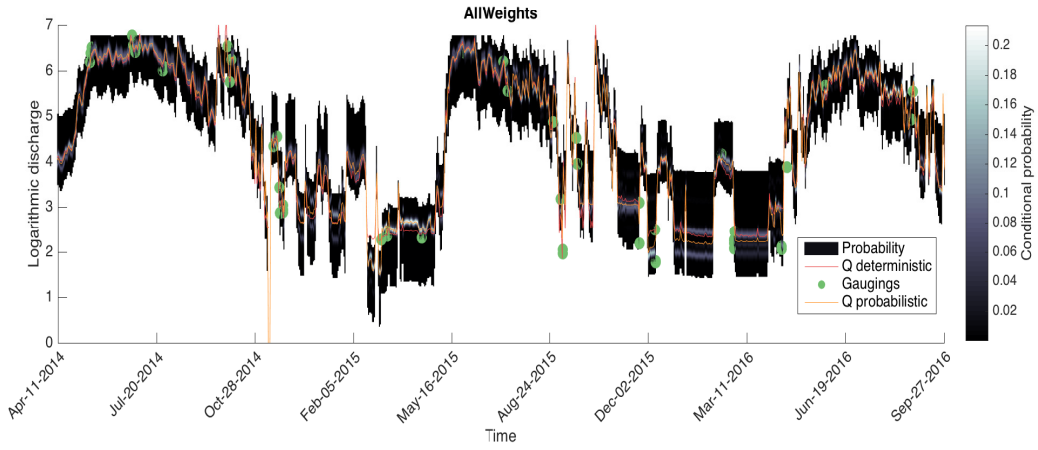

Figure 4.4: A hydrograph with dynamic conditional uncertainty bands for the All Weights model.

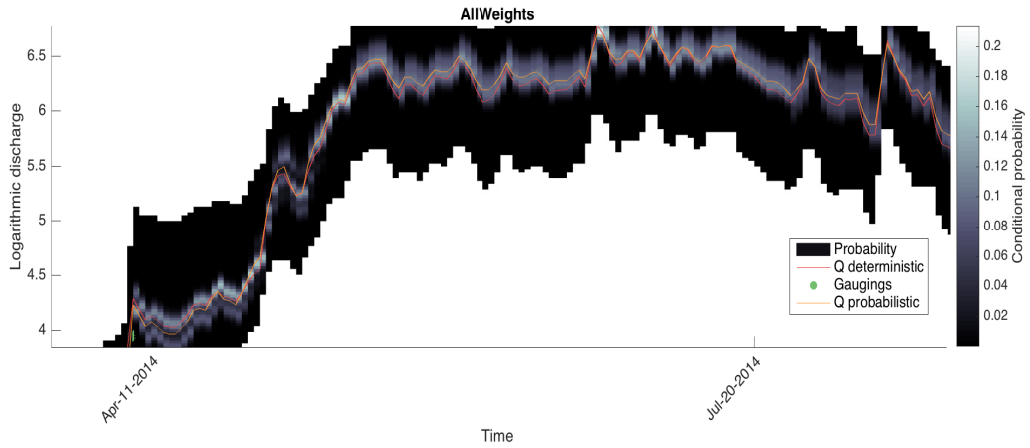

Figure 4.5: A zoomed in version of Figure 4.4

using a discharge time series.

4.1.3 Evaluation of rating curves

RMSE

When comparing the continuous discharge measurements in Figure 4.4 for the All Weights model and the deterministic rating curve predictions, it was seen that both values were closely aligned. To better understand how well the probabilistic model is performing, a scatter plot was created to compare the predicted discharge from the probabilistic model with the *measured* discharge gaugings (Figure 4.6). Immediately, one can identify the lower discharge values were both over and under estimated, and that higher discharge values tended to be well predicted. It was expected that gaugings taken later in time (darker colored markers) would reside closer to the one-to-one line in Figure 4.6 as the model would learn from previously observed gaugings. A possible reason for not seeing this result could be due to the small width of the kernels, but this makes sense since the influence of gaugings should only be felt along the slope of the kernel and not necessarily along the width. Figure 4.6 also has the RMSE values for the four models. Both the Only Sluicing and Only Time models had lower RMSE values than the Equal Weights model. These results serve as an indication that the added information was useful in reducing the uncertainty. When using both signals, the RMSE decreased to 0.42, and helped reduce the uncertainty by a percent difference of 19% when compared to the Equal Weights model.

Surprise

An analysis of the surprise about the observed value before, and after, a gauging is added to the rating curve is helpful to identify gaugings that can improve the performance of the model. To better investigate and compare the results of surprise per model, the data were categorized into the total surprise per month per model (Figure 4.7). The light colored bars represent the surprise before

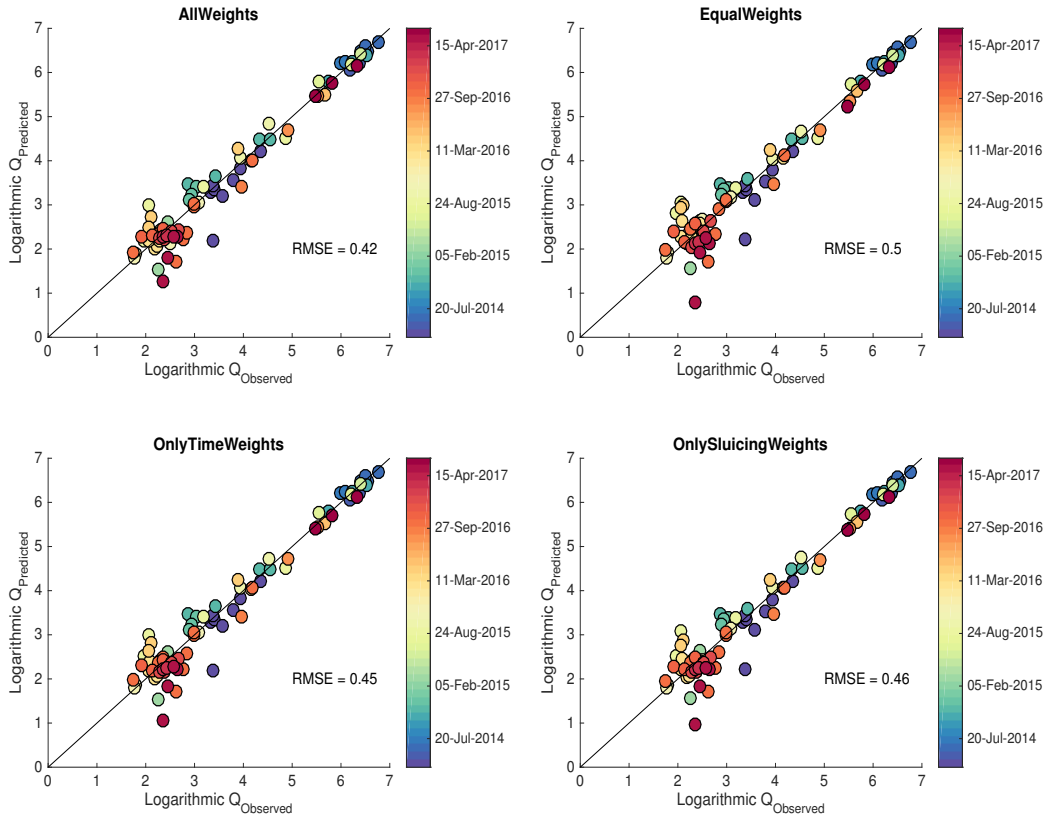

Figure 4.6: Scatter plot describing the differences between the predicted discharge from the four models and the observed values.

the gauging was added, while the darker colored bars are the surprise after adding the gauging. As expected, the surprise decreased for all when the gauging was added to the rating curve. No clear pattern is shown in the figure as to which model is best. What can be seen is that surprise results for April before the gaugings are added are not visible in Figure 4.7—this is due to two surprise values equal to infinity. The magnitude of the discharge gaugings were identified as logarithmic discharge 2.17 and 2.35, but because these gaugings fell significantly outside the linear relationship observed thus far, the rating curve models had assigned a small conditional probability. This small probability is also a consequence of the numerical limitations by the computer used in which the probability is small enough to round to zero, and therefore produces a default value of infinity. If the computation precision was not limiting, a real number would be achievable, but nonetheless, the surprise would still be of a large magnitude.

Since the infinity values prevented the remainder of the surprise for April to be seen, they were removed, and the total surprise per model were recalculated (Table 4.1). The total surprise prior to adding the gauging was highest for Equal Weights, followed by the Only Sluicing, Only Time, and

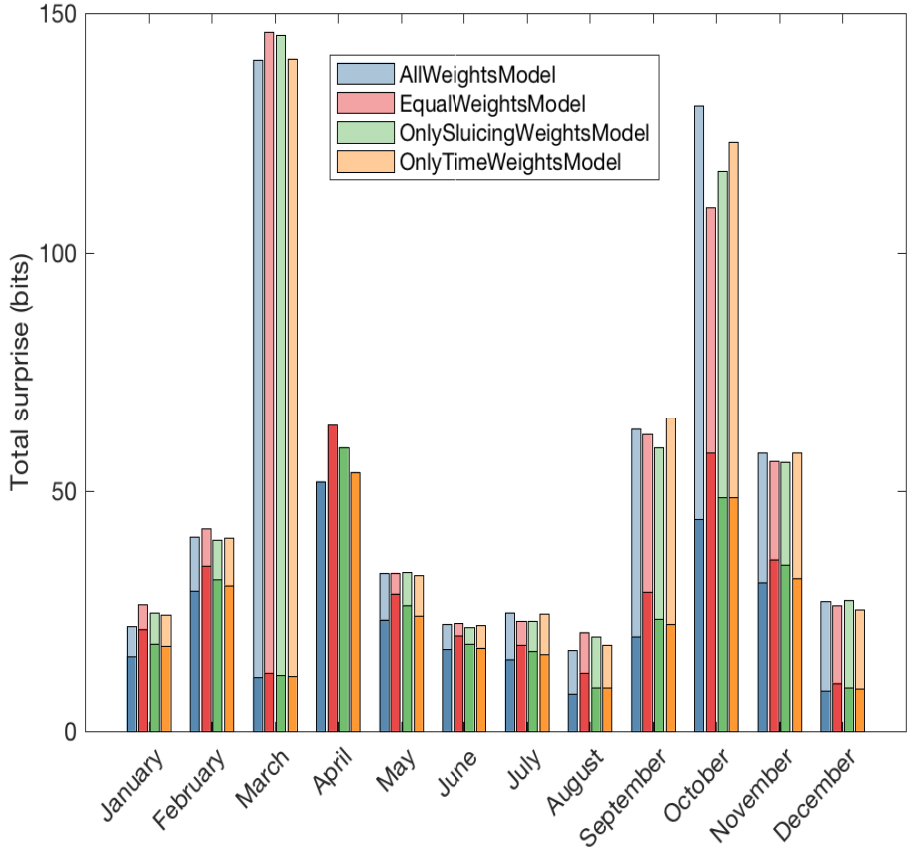

Figure 4.7: Total surprise per month. The translucent bars represent the surprise before the gauging is added (during prediction) and the opaque bars are the surprise after.

Table 4.1: Total surprise per model

Model name	Total surprise before	Total surprise after
All Weights	1194	274
Equal Weights	1414	344
Only Sluicing	1306	307
Only Time	1260	292

the All Weights model. Another observation from Figure 4.7 is that the total surprise per month appears highest in March and October. March had a total of four gaugings, with the first three of four taken in 2015, all within a week of each other. October had a total of 14 gaugings, 10 of which were taken in 2016. When the infinity values were removed for April, the results showed that the month had a total of 534 bits of surprise. The heatmap in Figure 4.8 shows that all months recorded predominantly low valued discharges, and as Figure 4.9 displays, low values tended to have a higher surprise since low probability was assigned by the rating curve. The first three discharges recorded in March 2015 were values that had not been observed by the rating curve and as the results show,

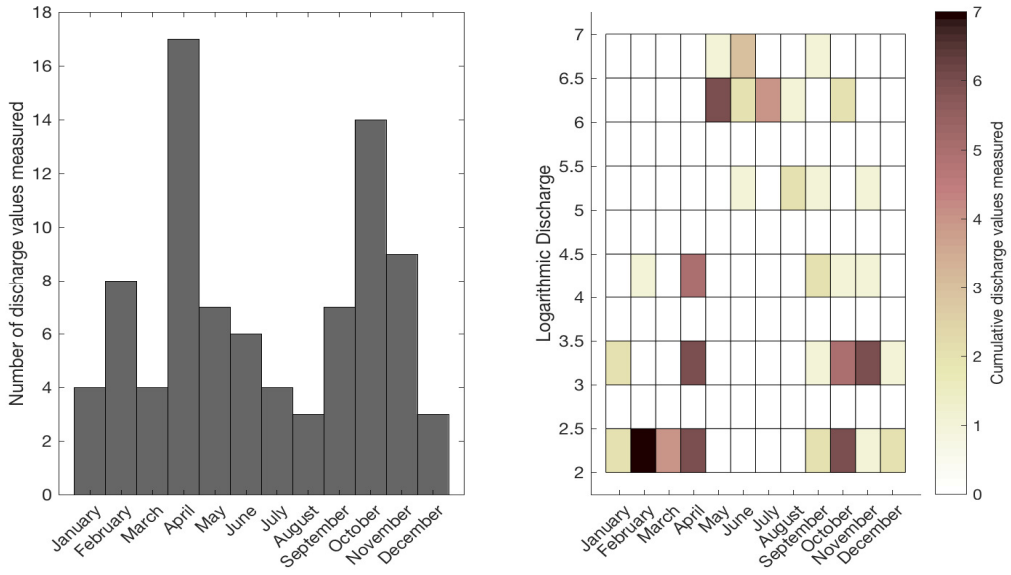

Figure 4.8: Cumulative gaugings per month are shown in the figure to the left and the range of values recorded per month and their frequency are shown in the heatmap to the right.

high surprise was achieved. In October 2016, the relationship between discharge and water level had changed, and therefore a handful of gaugings were designated a low probability. From Figure 4.9, a decrease in surprise and magnitude can be observed. A possible hypothesis for this could be that lower discharge values are more sensitive to minor changes in the channel characteristics, which is why more scatter is observed at the lower discharge gaugings. This scatter in the gaugings can have an influence on the probability that the rating curve assigns to the localized region.

A reason for why the surprise of large discharge values was low may be due to the influence that the first 19 gaugings had on the rating curve. A look at the first 19 gaugings show that NHC had done a good job at collecting a wide range of discharge values for their first rating curve, especially in collecting large discharges. Aside from the probabilistic rating curve having already observed these large discharges, Figure 4.10 shows that the large discharges have minimal scatter and all tend to follow a linear path. Therefore, the rating curve can do a decent job at extrapolating, and the conditional distribution of gaugings along this line tends to be high—meaning low surprise. However, in the case where a new gauging is recorded, is much larger than the rest, and falls outside the expected linear path, the model will assign a large surprise value to the gauging. If the gauging falls on the linear path of the previous gaugings, the conditional distribution at that point is influenced by the other gaugings and so the probabilistic rating curve should not be as surprised.

Kullback-Leibler Divergence

Probability distributions can be extracted and used to analyze the information gain from incorporating the gaugings, for the different models. Figure 4.11 shows the Kullback-Leibler divergence

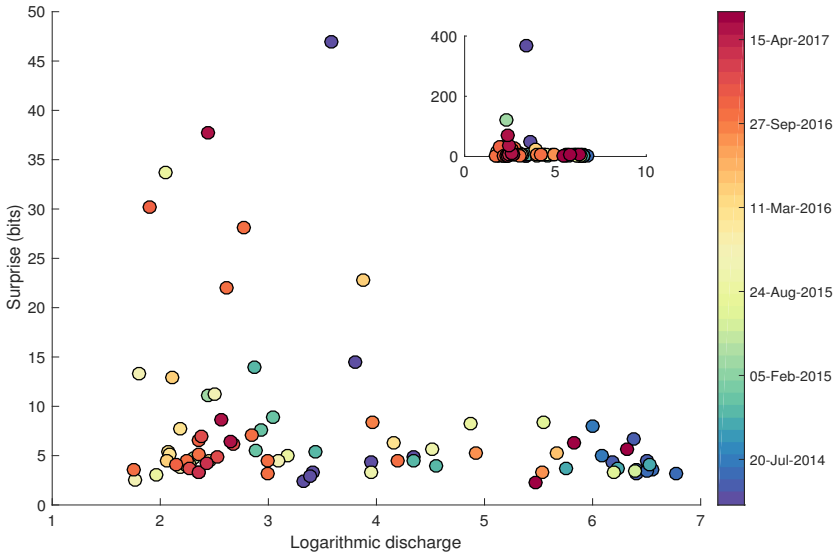

Figure 4.9: Discharge versus surprise and color coded with date in which gauging was recorded. The three largest values were removed from the main plot to better show the trend. The inset shows what the original date looked like.

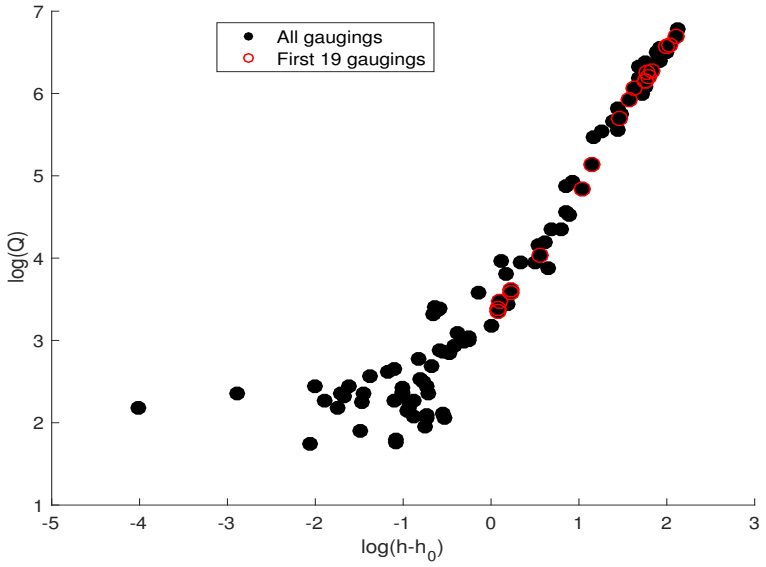

Figure 4.10: Trend in the gaugings. The reason for the nonlinear shape is due to the initially assumed h_0 of 234 m.

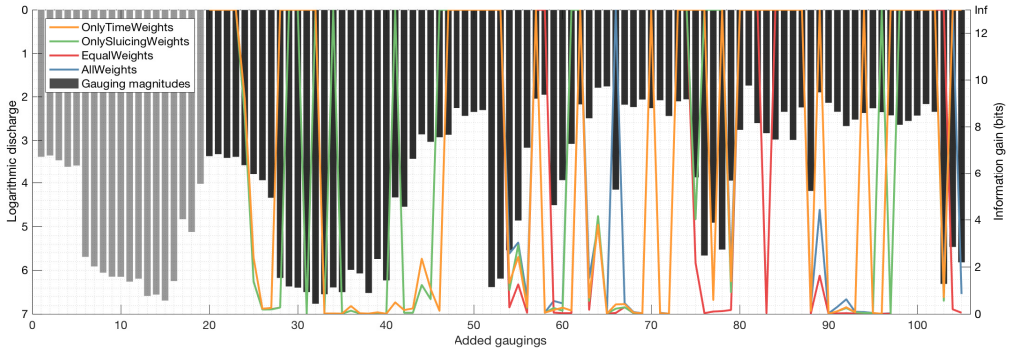

Figure 4.11: The lines represent the information gain from the Kullback-Leibler divergence measure between the probability distribution prior to a prediction being made for the gauging and the probability distribution after adding the gauging ("true" distribution). The black bars represent the discharge values of the gaugings used in the model. The light gray bars are the initial 19 gaugings used to develop the rating curve.

(colored time series) between the probability distributions of the predicted gauging before it was added to the model, and after. The black bars represent all the added gaugings for all four models. Immediately one notices numerous large spikes in the information gain for all the models. These large spikes represent an infinite amount of information gain and implies that having observed these gaugings was valuable to the probabilistic rating curve. However, what it also means is that prior to observing the gaugings, none of the models had assigned a large enough nonzero probability to that gauging (combination of Q and H) occurring. This can be attributed to several reasons. One of the reasons is that since the widths of the kernels are small, the influence of one kernel does not increase the probability space in the direction of its second principal component and so therefore, even if a gauging of similar value does occur, if it is parallel to a previous gauging, it is possible for the model to assign it a low conditional probability. The effect can be seen in Figure 4.5 where the kernels are stretched primarily along its first principal component and the width of the kernel (second principal component) is marginally small and until a new kernel is added, that probability space will remain with low values close to zero. A second possible reason for the large spikes could be due to the grid size chosen for the analysis. A 500x500 mesh was used to define the probability space. Such a fine grid would cause more areas of low probability than if a coarser mesh (i.e. 50x50) was used. Although the size of the mesh does affect the probabilities, as long as the grid size remains constant throughout the analysis, the results are meaningful. While Figure 4.11 showed the information gain of adding a gauging to a model, the Kullback-Leibler divergence can also be used to examine the information added going from one model to another. For example, it may be desirable to compare the probability distributions before a gauging is added to the All Weights model and the Equal Weights model to examine how much information was gained by using one model over the other. For this analysis, the models that are assumed to be the correct, or true, are the All Weights, the Only Time, and Only Sluicing models. These models were selected as such since the auxiliary information being brought into the rating curves better describes what is actually happening. The

results of the information gain between models showed similar infinity values as those previously described in Figure 4.11. The All Weights and Only Time models had many more infinity values (47 and 50, respectively) than the Only Sluicing model, which had 24, when calculating the information gain assuming the Equal Weights model was not the true model. Due to the large difference in the number of infinity values, it did not seem appropriate to neglect the values and do a summation of the information gain. Instead, the number of infinity values are being interpreted as an indication of how far the models were to the Equal Weights model. Since the Only Sluicing model had the least amount of infinity models, it was concluded that in general, this model was closer than the other two. This conclusion is similar to what has been seen so far in the results.

4.2 Hindcast mode

The RMSE and surprise are calculated for each gauging when the model operates in hindcast mode. Rating curves at each gauging in hindcast can also be produced, but are not shown since they all look and behave as the ones in Figures 4.1. The average of all 105 RMSE values and the total surprise are presented in Table 4.2. During hindcast mode, the surprise of the Only Sluicing model

Table 4.2: RMSE and surprise during hindcast evaluation of rating curve model

Model name	RMSE	Total surprise (bits)
All Weights	0.27	453
Equal Weights	0.34	456
Only Sluicing	0.29	433
Only Time	0.30	442

was the lowest, followed by the Only Time, All Weights, and Equal Weights. When comparing the results with the RMSE, the lowest value is from the All Weights model, preceded by the Only Sluicing, Only Time, and the Equal Weights model. The pattern of the surprise results are different from those reported in Table 4.1, while the RMSE values continue to follow the pattern found in the forecast mode of the probabilistic dynamic rating curve. What the RMSE values show is that the auxiliary information does help with improving the predictive capabilities of the rating curve. However, the model with the best RMSE, had the second largest total surprise. What this could imply is that the All Weights model is over confident about the precision of the discharge estimates, and so the surface area of the kernel decreases (since over confident implies a more ”peakier” PDF) and therefore, the probability decreases in the surrounding area. Individually, the Only Sluicing and Only Time models are not as over confident as the All Weights model, and the surface of their PDFs is larger (more spread out), hence a lower surprise. The Equal Weights model behaved as expected when examining the results of both metrics.

4.3 Sensitivity analysis

There were a few assumptions used in the development of the probabilistic dynamic rating curve. The most notable one is σ_{Q^*} in Equation 3.5 where 5% was assumed for the uncertainty in discharge gaugings. However, it is possible for gauging measurements to have more than the assumed error, and in fact, it is expected that the extreme ends of the discharge gaugings will have more uncertainty. The data used in the analysis unfortunately did not have any uncertainty values attached to it and therefore the 5% assumption had to be used. The author notes that this is a bit hypocritical to what was stated in earlier sections of the manuscript where the author argues that uncertainty assignment should not be done by the end-users, but rather the experts. Yet this contradiction serves to ignite fuel to the idea that the experts should be transparent with the uncertainty in the data, and that this important detail should be available to anyone who is interested in using the data for further modeling. If the uncertainty had been provided for each of the gaugings, this study would have used those uncertainty readings as direct inputs into the geometry of each individual multivariate Gaussian distribution. This information would have created kernels that better represent the actual uncertainty, thus forcing the global uncertainty of the rating curve and its dynamic probability bands to be more accurate. A sensitivity analysis using the forecast mode of the rating curve on this assumed uncertainty value is provided to better examine the results of the rating curve and identify the robustness in the model.

4.3.1 Discharge uncertainty

In this analysis, the water level uncertainty continued to be solved for using Equation 3.5 but the discharge uncertainty was varied between 10% and 20%, with the results from the original 5% assumption also shown for comparison purposes. The higher uncertainty values are more representative in stream gauging methods than anything lower than the originally assumed 5%. It was expected that as the discharge uncertainty increased to 10%, the kernels would also increase in width, in the ordinate direction. The results from Table 4.3 show that the RMSE stayed roughly the same, and that the surprise explodes. An example of the probabilistic dynamic rating curve, and its joint distribution, with an assumed 10% discharge measurement uncertainty is shown in Figure 4.12. By causing larger kernels, the rating curve becomes less confident and the multivariate distribution increases its surface area in the x and y direction to accommodate for the fixed slope and correlation, and decreases in the z direction. As the discharge uncertainty continues to increase, the kernels expand, and the rating curve continues to lose it confidence. Appendix C compares the joint and conditional distributions for kernels assuming a 20% discharge measurement uncertainty. The results from Table 4.3 show that in general, the RMSE stays relatively the same as the uncertainty in the discharge changes. A constant RMSE means that although the geometry of the kernels change, the mean of the kernels remains roughly the same.

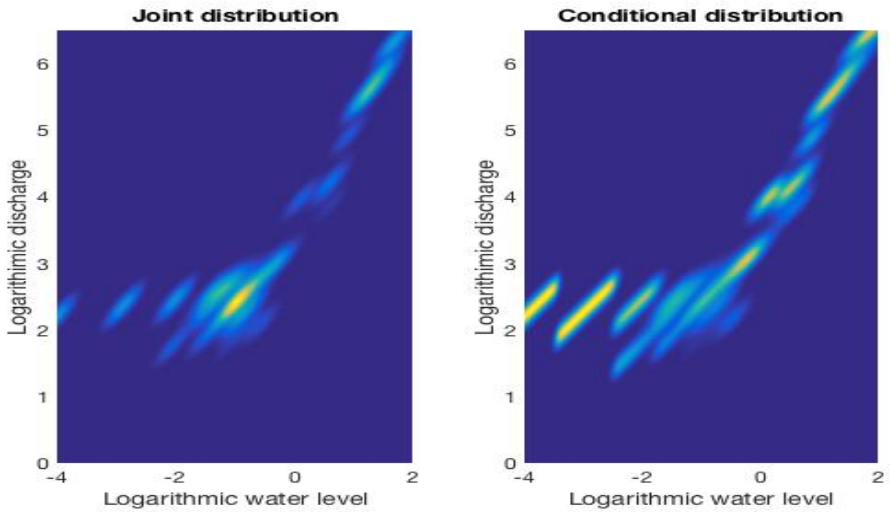

Figure 4.12: A comparison of how the joint and conditional distributions change when using a 10% discharge uncertainty.

Table 4.3: Sensitivity analysis of discharge uncertainty used in the conditional distribution of the kernels.

StdQ (%Q)	Metrics	AllWeights	EqualWeights	OnlySluicing	OnlyTime
5	RMSE	0.42	0.50	0.46	0.45
	Surprise (bits)	1194	1414	1306	1260
10	RMSE	0.40	0.47	0.44	0.43
	Surprise (bits)	858	1120	1005	959
20	RMSE	0.40	0.50	0.45	0.43
	Surprise (bits)	575	653	618	605

4.3.2 Time model parameter (μ)

The other assumed parameter in the probabilistic rating curve was the rate parameter of the exponential function (Equation 3.1) used for assigning the time weights. In the equation, a rate parameter of $\frac{1}{365}$ was used. Various other values were examined to understand the impact of the parameter on the model and the results are shown in Table 4.4. The results of varying the parameter show that the model's surprise is more sensitive than the RMSE. However, there is not a general pattern that emerges when varying μ. What is occurring is that as the rate of the exponential function increases, the most recent gauging is receiving a larger weight than when assuming a value of 365 days, while previous gaugings are losing their weights at a faster rate. This can create a push-pull effect where different combinations of kernel influence can occur depending on the weights. A further analysis should be performed to better understand the impact of μ on the rating curve.

Table 4.4: Sensitivity analysis of μ parameter in the exponential function used for the time weights while in forecast mode. Note that the Equal Weights and the Only Sluicing models are not shown since they are not effected by the time weights. AW = All Weights and OT = Only Time.

	μ = 365 days		μ = 180 days		μ = 90 days		μ = 30 days	
Metrics	AW	OT	AW	OT	AW	OT	AW	OT
RMSE	0.42	0.45	0.39	0.41	0.35	0.36	0.44	0.42
Surprise (bits)	1194	1260	1830	1193	1465	1559	1201	1191

The exponential function used in the development of the time weights has a memoryless property. This property states that the probability between events stays constant, and only when a new event is observed, will the weights update. This property is desirable, because it states that the weights of the gaugings should only change when a new gauging is observed. If sufficient time from the last measured gauging has occurred, an idea could be to let the exponential distribution converge to a theoretical equilibrium that could potentially relate to the river's natural equilibrium. Leopold, Wolman, and Miller (2012) presented the idea that rivers are always seeking to be in a state of lowest energy, and so after an event has occurred that disrupts the current state, the river will slowly transition back to its natural equilibrium. Applying this similar idea to the exponential function will imply that the weights of the gaugings that previously had little weight will now be set to an equilibrium. In a river like the Iskut that is constantly changing because of the upstream intake, this idea may not work, but could be interesting to explore in other river systems.

The parameter μ in this thesis was assumed, but it is possible for the performance of the probabilistic dynamic rating curve to be improved if the parameter learns from the data. Using a Bayesian approach is a possible alternative to optimizing the parameter. A prior probability on the parameter could be selected based on the characteristics of the river and/or watershed. As gaugings are observed, and added to the rating curve model, the posterior distribution that best describes the parameter can be determined.

Chapter 5

Discussion

5.1 Cease-to-flow water level (h_0)

The h_0 parameter is important because it helps establish the relationship between discharge and the river's water level, but also because it helps identify the height of the control. In practice, sometimes the height of the control is not known, and so this parameter is typically calibrated in hindsight, in combination with field knowledge, such as possible physical limitations to the exponential parameter in Equation 2.1. For calibration, the gaugings are often log transformed, as done in this thesis, to ensure a straight line. The value of 234 m for the h_0 water level was selected by plotting all the first 19 gaugings used in the development of the first rating curve, independent of the simulations performed on the remaining gaugings, and identifying how well the data fell on a linear relationship. The value was then fixed for the duration of all the simulations in producing the new rating curves. The h_0 parameter could be fixed by producing a moving parameter, similar to μ, that is optimized every time a new gauging is observed. More accurate cease-to-flow water levels were available to the author from NHC, but were not used, because again, the value of their $h_0 s$ was calculated in hindsight, and the author felt that using their values would be more unfair than fixing the data to the assumed value.

A consequence of the assumed 234 m for h_0 can be seen in the final probabilistic rating curve (Figure 5.1). It is known that the log transformation of data following a power law function produces a linear relationship. However, Figure 5.1 shows otherwise. The reason for this is attributed to two gaugings that occurred on April 29th and 30th of 2017 with low water level measurements that did not conform to a linear relationship with the other data. This error was discovered when deciding which value for h_0 to assume, and since the gaugings occurred near the end of the dataset (4th and 5th from the last gauging), it was decided that 234 m would work for the other gaugings. A much better calibration could have been done, but again, this would produce an even more unfair forecast.

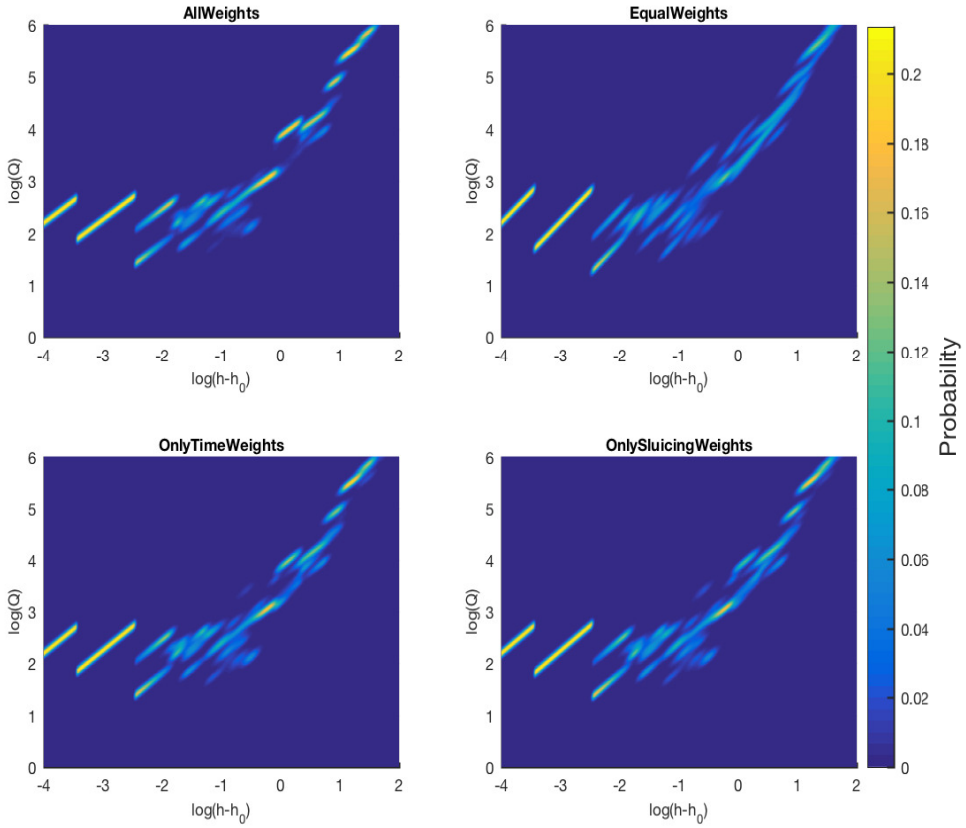

Figure 5.1: Last dynamic probabilistic rating curve produced in simulations.

5.2 The effect of fixing the slope of all kernels to be equal

As described in the data and methods section, the slopes of the kernels were all fixed to be the same by using the observed gaugings and performing a weighted linear regression to obtain the slope. A consequence of such method is that there is a chance that gaugings that do not fall exactly on the linear function, risk forming parallel lines as shown in Figure 5.2, and producing a fragmented rating curve. The example shown is from May 2016 in which only five gaugings with low water level measurements had been recorded up to that point. These five gaugings were on the lower end of the original linear regression and slightly away from all the other data points and so therefore, a parallel line was formed when their slopes were fixed. This causes a similar bimodal distribution, as seen in Figure 4.2 and is a reason for why the weighted mean of the conditional PDF was used, rather than the mode. A possible fix to this problem is to use localized regression on the 'k' nearest gaugings. In doing so, the original gauging of the five gaugings, highlighted within the red box in Figure 5.2, would have had its slope fixed with gaugings that fell closer to the linear regression, and as more gaugings entered the model, the slopes would update locally. For a proper linear function to be produced, it would be important to ensure that the 'k' nearest gaugings selected would produce a monotonic function.

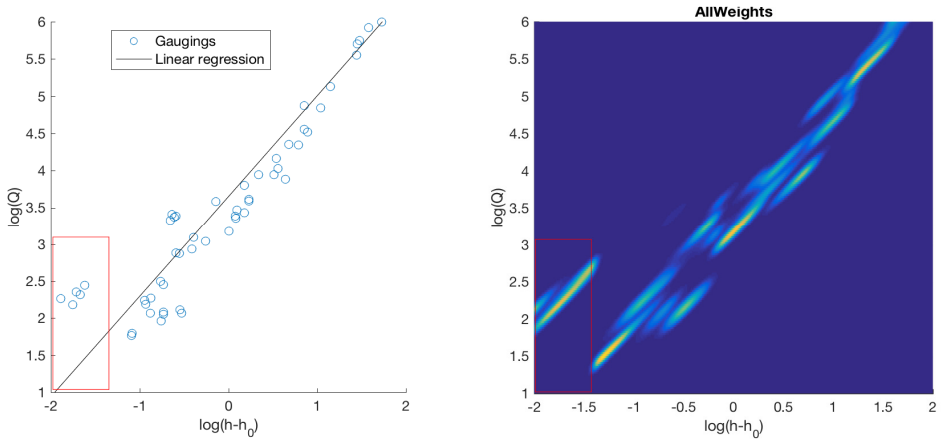

Figure 5.2: A display of how parallel rating curves can emerge using the described linear regression for fixing the slope of all the kernels.

5.3 Extrapolation of the rating curve

Extrapolation of the probabilistic rating curve works by the influence of the kernels and the orientation at which they are facing. Since the kernels in theory ”stretch-out” to infinity along their primary principal component, there will always be a probability assigned to the extreme ends of the rating curve where extrapolation might be desired. The conditional would then be taken, and the PDF would be renormalized to a sum of one. An exception to this could be that if the extrapolation region is far enough, computational limitations might take over and assign values of zeros for the probability due to limited computer precision. In this case, a correct conditional PDF might not be guaranteed.

Similar to extrapolating in a deterministic rating curve plotted on log-log space, caution should always be adhered to ensure that extrapolation is only occurring within the bounds of the linear relationship between the physical constraints of the river (i.e. stage and cross-sectional area). If a shift in the stage-area relationship occurs, then a shift in the orientation of the kernels should also occur to ensure proper extrapolation. This is one way to ensure that the physical characteristics of the river are also observed in the probabilistic model. These shifts in the stage-area relationship are also easily adoptable in the probabilistic rating curve if a multi-segmented rating curve in the deterministic domain is known to exist. The weighted linear regression would have to be altered into possibly a piecewise function with the nodes being influenced by the height of the corresponding controls.

5.4 Maximizing what is known to help solve the unknown

In the development of the probabilistic rating curve, time and sluicing information were brought in to help reduce model uncertainty. It was argued that gaugings after sluicing, and those most recent in time, were more representative of the current stage-discharge relationship than those prior

to sluicing, and taken earlier. The results showed that both the time and sluicing models did help reduce the uncertainty of a probabilistic rating curve when compared to a model using no auxiliary information, by up to a maximum of 19%, when in forecast mode. This shows that the information used to building the rating curves can become of practical importance for hydroelectric companies that rely on a rating curve downstream of their intake to ensure an accurate estimate of the IFR compliance. Operators of the intake plant can use the probabilistic results to make better informed decisions on whether more water should be released downstream to help ensure IFR compliance, within a degree of certainty, or, to stay business-as-usual because enough confidence is assigned to the predicted discharges. What can also be done is the inclusion of operator knowledge on expected sluicing events. By giving an approximate date and duration of the event, the model can evaluate how a near future sluicing event could affect the probability of ensuring the IFR after the expected sluicing event.

The model can also be used to help plan gaugings. For example, it could be possible to optimize the dynamic probabilistic rating curve to ensure that a minimum conditional uncertainty is always met. In doing so, it could help identify when to best take gaugings. Similar to how in the financial sector different portfolios are built, evaluated, and then the best one is selected, the same can be done with gaugings. Different scenarios can be used to train the dynamic probabilistic rating curve, and once calibrated and validated, an optimal plan for when to take gaugings may be possible.

Studies evaluating the usefulness of adding/collecting different information signals in conjunction with a cost-benefit analysis, may be performed. This analysis can also be used to define the value of added information to the rating curve to making better informed decisions. This idea was explored by Raso, Weijs, and Werner (2017) in their recent manuscript, in which they used the Optimal Design (OD) problem to evaluate the value of adding extra gaugings to the rating curve.

5.5 Dynamic uncertainty bands

The dynamic uncertainty bands are a beneficial product of the produced rating curve and could potentially serve as a good incentive for producing probabilistic dynamic rating curves. Figure 5.3 shows a closer examination at how the conditional probabilities change when a gauging is added while in forecast mode. The figure illustrates that the column where the gauging is found (green dot) has a flatter conditional PDF than after the gauging is added, column where the vertex in the predicted continuous discharge value begins. The conditional PDF becomes ”peakier” and is shown by the change in color in the column only *after* the gauging. This result is important because it provides an answer to a question that maybe does not get frequently asked when looking at discharge time series because of how the deterministic rating curve shows its results. The fact that the conditional PDF is changing when the data are added, helps answer the question of, ”What is the probability distribution of the discharge, given a water level measurement of h_i when a gauging is recorded?” Or another question that could be interesting, ”What is the probaiblity distribution

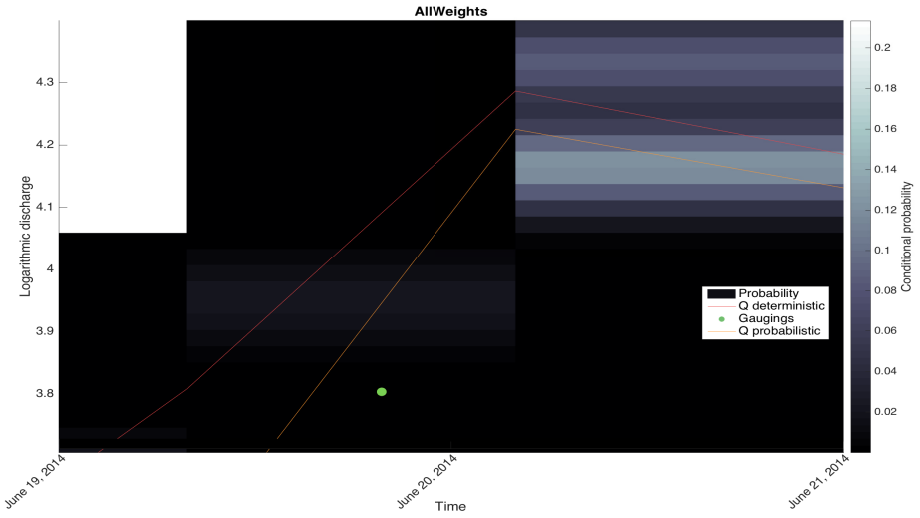

Figure 5.3: A closeup as to how the conditional uncertainty bands change when a gauging is added.

of the discharge, given a water level, *prior* to adding a gauging to the rating curve?" Both these questions help set up a canvas where the data are now more useful than they were when analyzed in a deterministic domain. When working outside a stochastic framework, 95% confidence intervals are typically appended to rating curves. However, the answers from these intervals are to different questions because the interval does not imply that the result has a 95% confidence of occurring within the interval, but rather, it states that if the samples were to be repeated, and the 95% confidence intervals recalculated, that 95% of the intervals would capture the population mean. So really, the 95% confidence interval is useful, but it answers a question that is not necessarily the immediate question that someone using predicted discharges is truly concerned with. This method for displaying the uncertainty around a rating curve gives no valuation to knowing prior information that may be useful in the design of rating curve uncertainty. For this reason, Bayesian methods have gained traction as a valuable tool in modeling rating curves (Le Coz et al., 2014; Moyeed & Clarke, 2005; Petersen-Øverleir & Reitan, 2009). Within a Bayesian framework, prior knowledge may be used to bound the limits of the possible parameters in the power law function (Equation 2.1), and as new data enter the framework, a likelihood function is used to determine the posterior that best describes the parameters. The method used in the development of the probabilistic rating curve in this thesis incorporates a similar approach where the probabilities are continuously updated as new data is observed. The method used for updating the priors is not formally defined as Bayesian, but implicitly, it is of a similar nature.

5.6 The need for probabilistic rating curves

In many cases when one downloads discharge time series data from WSC or USGS, there is no information on the probability distribution of the values. WSC does provide the percent deviation that

the last gauging measured is from the current rating curve. WSC and USGS will also sometimes provide categorical uncertainty measures that inform the users about any possible anomalies with the data. A categorical value of "B" for WSC indicates that the data were collected under the influence of ice coverage (Hamilton & Moore, 2012). However, this is not a true measure of uncertainty and is not a pragmatic approach to data uncertainty.

With the probabilistic rating curve developed in this thesis, a conditional probability value, or function, can easily be included with the predicted discharge and give a much more transparent, and informative, result. This is shown in Figure 4.4 where the conditional PDFs are extracted and a dynamic uncertainty band showcasing the conditional uncertainty of the discharge given the recorded water level is produced. For decision makers who depend on the discharge time series, having this extra information can help in making more informed decisions. For example, operators who use the predicted discharge from a hydrometric station to ensure IFR compliance can use the conditional distributions to better assess the decision to release more water downstream to ensure that the IFR is met, or to take in more water for energy production/profit. The scenario described earlier about optimizing the rating curve to better plan out when to take gaugings can also be applied for IFR rating curves. The results of such a scenario could imply a larger up front cost by having to send people out to the field, but the decision could potentially save money in the long run by preventing the streamflow to fall under the IFR, and force a hydroelectric company to pay a large monetary sum due to infraction fees.

Probabilistic rating curves allow for better decision-making processes in which decision makers can have more, or less, confidence in their actions as a result of the probability from the rating curve. In this thesis, an example of decision makers from a hydroelectric ROR project has been used, but what has not been discussed are the other factors influencing the decision on river withdrawals (i.e. electricity market demands) as well as the other stakeholders involved. In British Columbia, salmon have economic, ecological, and cultural value. To ensure that salmon keep coming back to their spawning grounds, sufficient water and the proper conditions must be available for the salmon to swim upstream. The amount of water flowing downstream can fluctuate due to the local climate and this can have an impact on water licenses. One now enters a multiobjective decision making problem in which enough water must be supplied downstream so that salmon can swim upstream, while trying to maximize economic benefit for the hydroelectric company. Decisions as these can be improved by having a sense of the uncertainty in the predicted discharge and can aid in taking either more conservative or liberal actions, as well as reduce the risk associated with these decisions.

What is also beneficial from probabilistic rating curves is that one can condition the probabilities on all available information. This is to say that the probability of a discharge value is conditioned on not only the recorded water level, but also time and sluicing, and possibly electrical conductivity. These conditional probabilities are derived from a probability function—in our case a mixture model of

multivariate Gaussians. Therefore, one can present much more information to the end users of a discharge series by not only including the point value conditional probability, but also the distribution function from which it stems. Krzysztofowicz (2001) identified four potential benefits for probabilistic predictions/forecasts. These equally apply to measurements of discharge, which are in fact also predictions. They have all have been discussed in this section, but are worth summarizing below:

1. Probabilistic models tend to be more scientifically honest and eliminate the illusion of certainty
2. Risk-based criteria can be established for example in flood forecasting
3. Rational decisions can be enhanced by knowing the true uncertainty of the data
4. There is a potential for economic benefit by using probabilistic models

These four points help promote accountability for the research, and decisions, that are taken by experts and leaders and have large benefits to society. Engineers are entrusted with the lives of thousands, and even millions, of people every single day and better ethical decisions and designs can be developed by knowing the uncertainty in the data that is used. However, for this notion of uncertainty transparency to kickstart, a shift in how data is collected and viewed must occur. Those collecting the data (i.e. streamflow) must be conscientious of how the data is recorded and an effort to calculate/estimate the uncertainty should be made so that the collecting authorities (water agencies like WSC and USGS) can publicly publish these results. Water agencies may be hesitant to do such a move, because it may show flaws in their data, but again, only through this can honest science progress and rational decisions be made.

Chapter 6

Conclusion and future work

The primary objectives laid out at the beginning of this thesis were to develop a probabilistic dynamic rating curve, and to explore the usability of auxiliary information from a run-of-river hydroelectric project to reduce model uncertainty. The data and methods section of this thesis highlight the development of the probabilistic dynamic rating curve. For a probabilistic model to be developed, the gaugings were converted into multivariate Gaussian distributions (kernels). The geometry of the kernel was shown to be important to control to ensure that all the kernels were aligned in the proper orientation so that interpolation and extrapolation of discharge values could be accurately predicted. Had the slope of the kernels not been fixed, they would be free to align themselves according to the variables used in Equation 3.6. While although the gaugings would align along a straight line as expected because of the log transformation, the kernels would not have followed that line. This would cause less accurate conditional distribution functions, and using a weighted mean of the function would have produced erroneous predictions. Although the method used for the kernel orientation was, per say, not the best (Figure 5.2), it is possible to improve it by using a local regression approach, which would be expected to produce better results.

The conditional distribution of the kernel was used to assign the measurement uncertainty of the discharge, and hence control the spread of the $(Q|h)$ distribution. The combination of using the marginal and conditional distributions served as the primary method to fully controlling the kernels. A value of 5% of the discharge magnitude was assumed for every kernel's discharge conditional distribution uncertainty. Under this assumption, reasonable results were produced and it is expected that in events where the discharge uncertainty is known, those values could be substituted in. The effect would be that the produced kernels would vary more than what are seen in this thesis, but would be a better representation of the actual uncertainty.

With the kernels constructed, their summation was taken to develop a mixture model of multivariate Gaussian distributions that represented the joint distribution. Using this finely discretized density grid, the conditional distributions ($P(Q|h = h_i)$) were calculated and normalized to one. These conditional distributionsm when appended together as shown in Figure 4.1, represent the probabilistic

dynamic rating curve. Using this model, the continuous discharge predictions could be extracted by using the continuous water level time series. What is more is that the conditional distributions can also be displayed in conjunction with the discharge time series to show a continuous discharge time series with dynamic uncertainty bands (Figure 4.5). These uncertainty bands were discussed and an argument was provided as to why these bands are more relevant to users than 95% confidence intervals. Using these conditional uncertainty bands also helps in improving models that use the discharge time series as inputs since the uncertainty in the time series is now transparent to the user. This helps reduce the ambiguity that modelers use to assign the uncertainty to the discharge and helps improve the overall performance of hydrological models and promotes better decision-making.

The probabilistic dynamic rating curve model developed is also simple enough to be adopted by engineering practitioners who develop rating curves for clients. In combination with current methods for developing deterministic rating curves, the slopes could be extracted after manual calibration of a linear regression through the gaugings. A simple graphic user interface could be created where the user inputs the measurement uncertainty and grid size, and the probabilistic rating curve could be produced instantly. Calibration of the curve could be done by changing a few of the assumed parameters measured to minimize the RMSE. Predictions from the deterministic and probabilistic results could be easily compared and dynamic uncertainty bands can be extracted for the probabilistic model.

An exploration into reducing the uncertainty was undergone by assessing the usage of auxiliary information. The information used in this thesis were sluicing signals from a ROR hydroelectric project and the timestamps of gaugings. The sluicing signal was inferred by referencing four unique signals in the intake structure of the hydroelectric project. Weighted normal CDFs were used to derive a function that was believed to help best describe the change in sluicing weights. The weights used were taken from the duration of the individual sluicing events, where large events were given more weight than short sluicing events. For the time weights, an exponential model was used to assign more weight to the recent gaugings and less to those that were taken in the past. Three unique models were created that utilized different combinations of the weights and one model was produced that used no weights, which implies that all the gaugings had an equal weight assignment of $\frac{1}{n}$ where n is the number of gaugings observed. The results show that using the auxiliary information helped reduce the RMSE of the model that used an equal weight distribution by up to 19%. The surprise of the model that used both the time and sluicing weights (All Weights model) was also shown to be that smallest of the four. This implies that its ability to predict future gaugings was the best. For hindcast mode, the results of the surprise varied and it showed that the All Weights model was over confident, which is possibly why that model had the largest total surprise. In all, it was shown that when used in forecast mode, the probabilistic dynamic rating curve proved to be a useful model with auxiliary information being capable of reducing the model's uncertainty.

There are many opportunities in which the work presented in this thesis can be taken further. The first would be an improvement on the assumed value for the exponential distribution parameter. As described, it may be possible to train this parameter using a Bayesian approach. In doing so, the model will adapt to the data and theoretically provide even better results then what were shown. It would also be interesting to explore how the model behaves when working with segmented rating curves. It is anticipated that as long as the individual slopes are extracted from the deterministic weighted linear regression, that a reasonable rating curve should be produced. The probabilistic dynamic rating curve can also be optimized to minimize RMSE, or any other performance metric, like maximizing entropy. In doing so, different scenarios can be studied to better understand when to best take gaugings. Lastly, in this thesis, multivariate Gaussian distributions were used. These same distributions are also used in Gaussian Processes and a further exploration of how to tie the results of this thesis with machine learning could be of high value, especially at a time when machine learning is currently undergoing a lot of development.

References

Baker, V. (2009). *High-energy megafloods: planetary settings and sedimentary dynamics* (Vol. 32). Special Publication.
6

Beven, K. (2010). *Environmental modelling: An uncertain future?* CRC Press.
2

Bhattacharya, B., & Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing*, *63*, 381–396.
12

Chow, V. T. (1959). *Open channel hydraulics*. McGraw-Hill Book Company, Inc; New York.
4

Clean Energy BC, B. C. (2015). *Run-of-river hydro power.* Retrieved 2017-10-13, from `https://www.cleanenergybc.org/about/clean-energy-sectors/run-of-river`
14

Coxon, G., Freer, J., Westerberg, I., Wagener, T., Woods, R., & Smith, P. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *51*(7), 5531–5546.
12, 13

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences*, *13*(6), 913.
8, 9, 11, 31

Godsill, S. J. (2001). On the relationship between Markov Chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics*, *10*(2), 230–248.
10

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, *3*(2), 95–99.
12

Govindaraju, R. S., & Rao, A. R. (2013). *Artificial neural networks in hydrology* (Vol. 36). Springer Science & Business Media.
13

Guerrero, J.-L., Westerberg, I. K., Halldin, S., Xu, C.-Y., & Lundin, L.-C. (2012). Temporal variability in stage–discharge relationships. *Journal of Hydrology*, *446*, 90–102.
11

Guven, A., Aytek, A., & Azamathulla, H. M. (2013). A practical approach to formulate stage–discharge relationship in natural rivers. *Neural Computing and Applications*, *23*(3-4), 873–880.
13

Hamilton, A. S., & Moore, R. D. (2012). Quantifying uncertainty in streamflow records.

Canadian Water Resources Journal, *37*(1), 3–21.
8, 11, 54

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management*, *128*(2), 91–101.
13

Herschy, R. W. (1995). *Streamflow measurement*. CRC Press.
viii, 5, 11, 14, 32

Herschy, R. W. (1999). *Hydrometry: principles and practices*. CRC Press.
6, 8, 9

Hsieh, W. W. (2009). *Machine learning methods in the environmental sciences: Neural networks and kernels*. Cambridge university press.
12, 13

Kieffer, S. W. (1985). The 1983 hydraulic jump in Crystal Rapid: Implications for river-running and geomorphic evolution in the Grand Canyon. *The Journal of Geology*, *93*(4), 385–406.
6

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, *249*(1), 2–9.
2, 55

Le Coz, J. (2012). *A literature review of methods for estimating the uncertainty associated with stage-discharge relations* (Tech. Rep. No. 6).
3

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Le Boursicaud, R. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, *509*, 573–587.
6, 11, 53

Leonard, J., Mietton, M., Najib, H., & Gourbesville, P. (2000). Rating curve modelling with Manning's equation to manage instability and improve extrapolation. *Hydrological Sciences Journal*, *45*(5), 739–750.
11

Leopold, L. B., Wolman, M. G., & Miller, J. P. (2012). *Fluvial processes in geomorphology*. Courier Corporation.
48

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7).
2

Longerstaey, J., & Spencer, M. (1996). Riskmetricstmtechnical document. *Morgan Guaranty Trust Company of New York: New York*.
25

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., & Clark, M. (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, *24*(10), 1270–1284.
2, 12

McMillan, H., & Westerberg, I. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, *29*(7), 1873–1882.
10

Ministry of Environment Science and Information Branch for the Resources Information Standards Committee. (2009). *Manual of British Columbia hydrometric standards.* Resources

Information Standards Committee.
4
Montanari, A., & Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, *40*(1).
2
Moore, R. D., Hamilton, A. S., & Scibek, J. (2002). Winter streamflow variability, Yukon Territory, Canada. *Hydrological Processes*, *16*(4), 763–778.
8
Moyeed, R. A., & Clarke, R. T. (2005). The use of Bayesian methods for fitting rating curves, with case studies. *Advances in Water Resources*, *28*(8), 807–818.
53
Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, *61*(9), 1666–1678.
13
Newman, M. C. (1993). Regression analysis of log-transformed data: Statistical bias and its correction. *Environmental Toxicology and Chemistry*, *12*(6), 1129–1133.
20
Northwest Hydraulic Consultants, N. (2016). *Forrest Kerr Hydroelectric Project Hydrometric Report 2016* (Tech. Rep.). Northwest Hydraulic Consultants. (unpublished report)
17
Pafka, S., & Kondor, I. (2001). Evaluating the riskmetrics methodology in measuring volatility and value-at-risk in financial markets. *Physica A: Statistical Mechanics and its Applications*, *299*(1), 305–310.
25
Petersen-Øverleir, A. (2004). Accounting for heteroscedasticity in rating curve estimates. *Journal of Hydrology*, *292*(1), 173–181.
9
Petersen-Øverleir, A. (2006). Modelling stage-discharge relationships affected by hysteresis using the Jones formula and nonlinear regression. *Hydrological Sciences Journal*, *51*(3), 365–388.
10
Petersen-Øverleir, A. (2008). Fitting depth–discharge relationships in rivers with floodplains. *Hydrology Research*, *39*(5-6), 369–384.
10
Petersen-Øverleir, A., & Reitan, T. (2005). Objective segmentation in compound rating curves. *Journal of Hydrology*, *311*(1), 188–201.
10
Petersen-Øverleir, A., & Reitan, T. (2009). Bayesian analysis of stage–fall–discharge models for gauging stations affected by variable backwater. *Hydrological Processes*, *23*(21), 3057–3074.
10, 53
Rantz, S. E. (1982). *Measurement and computation of streamflow: volume 2, computation of discharge* (Tech. Rep.). USGPO,.
4, 8, 9, 11, 32
Raso, L., Weijs, S., & Werner, M. (2017). Balancing costs and benefits in selecting new information: Efficient monitoring using deterministic hydro-economic models. *Water Resources Management*, 1–19.
52

Reitan, T., & Petersen-Øverleir, A. (2008). Bayesian power-law regression with a location parameter, with applications for construction of discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, *22*(3), 351–365.
10

Reitan, T., & Petersen-Øverleir, A. (2009). Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, *23*(5), 627–642.
10

Reitan, T., & Petersen-Øverleir, A. (2011). Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes. *Water Resources Research*, *47*(2).
11

Schement, J. R., & Ruben, B. D. (1993). *Between communication and information* (Vol. 4). Transaction Publishers.
13

Schmidt, A. R. (2002). *Analysis of stage-discharge relations for open-channel flows and their associated uncertainties* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
11

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical J.*, *27*(3), 379–423.
32

Sivapragasam, C., & Muttil, N. (2005). Discharge rating curve extension–a new approach. *Water Resources Management*, *19*(5), 505–520.
13

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3–22.
12

Sorooshian, S., & Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, *16*(2), 430–442.
9

Turnipseed, D. P., & Sauer, V. B. (2010). Discharge measurements at gaging stations. In *U.s. geological survey techniques and methods book 3* (chap. A8). United States Geological Survey.
8, 11

Venetis, C. (1970). A note on the estimation of the parameters in logarithmic stage-discharge relationships with estimates of their error. *Hydrological Sciences Journal*, *15*(2), 105–111.
9

Weijs, S. V. (2011). Information theory for risk-based water system operation. 2, 13

Weijs, S. V., Mutzner, R., & Parlange, M. B. (2013). Could electrical conductivity replace water level in rating curves for alpine streams? *Water Resources Research*, *49*(1), 343–351.
2, 14, 15

Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K., & Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Processes*, *25*(4), 603–613.
12

Appendix A

Snapshots of the rating curve in time

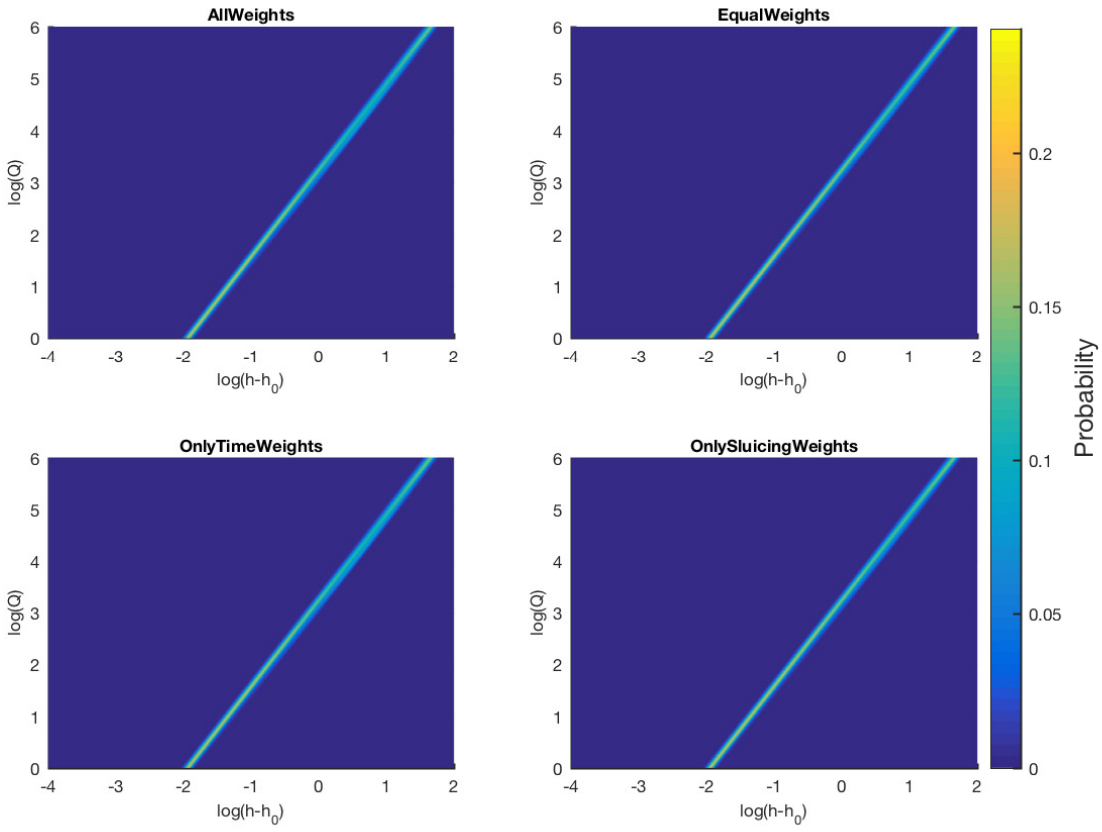

Figure A.1: November 28, 2013

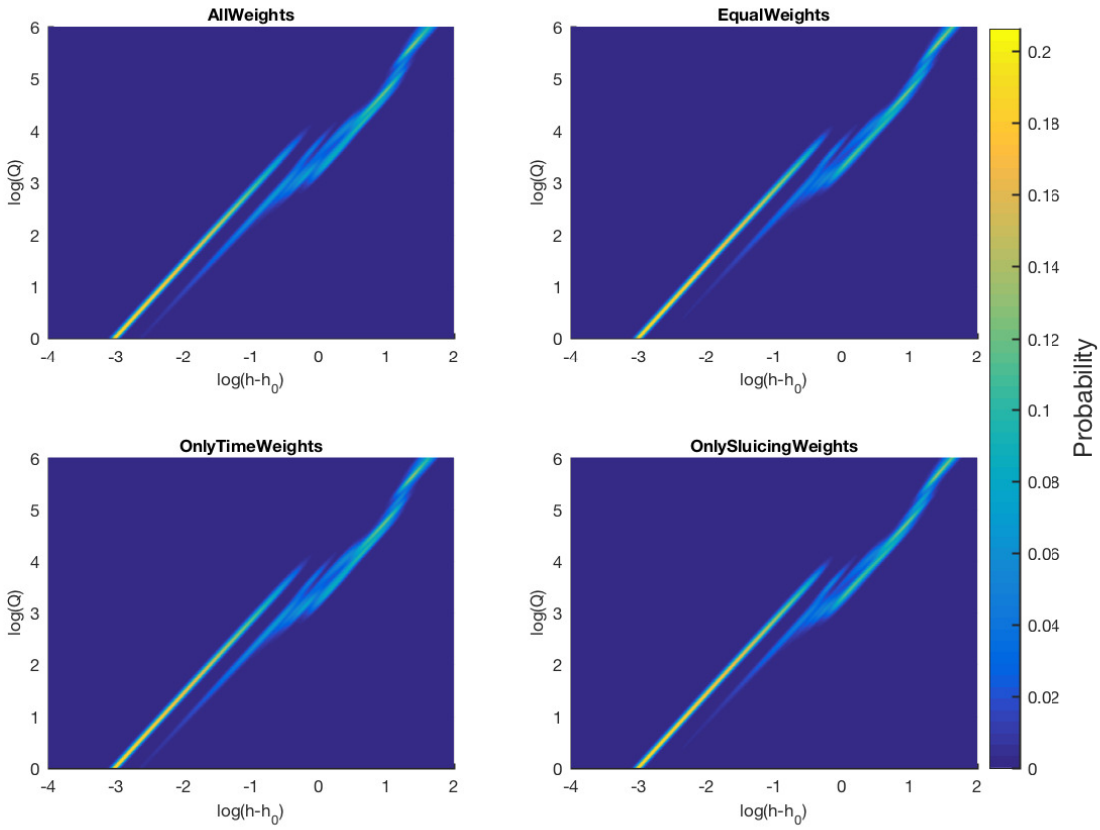

Figure A.2: January 1, 2015

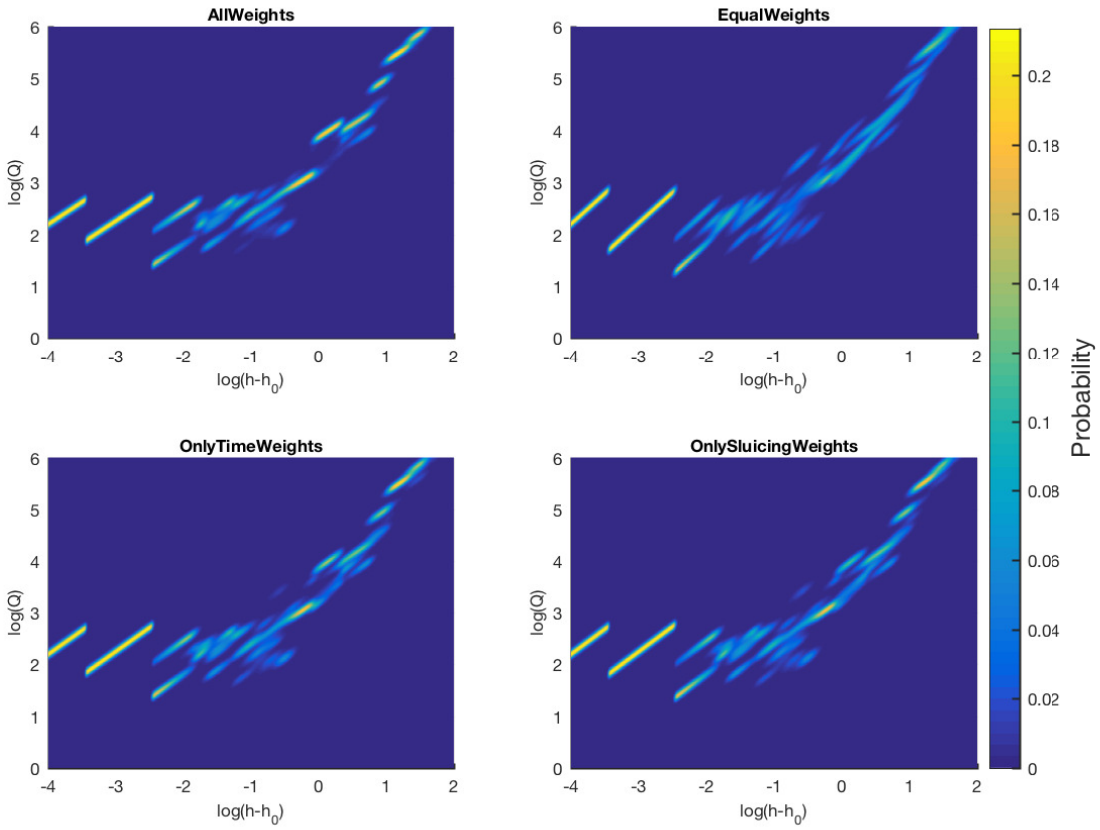

Figure A.3: June 20, 2017

Appendix B

Hydrographs

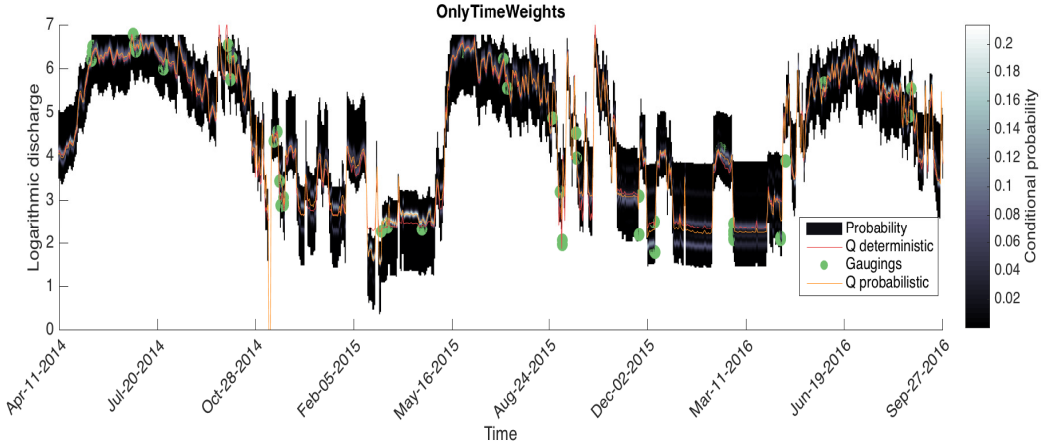

Figure B.1: Hydrograph for Only Time model

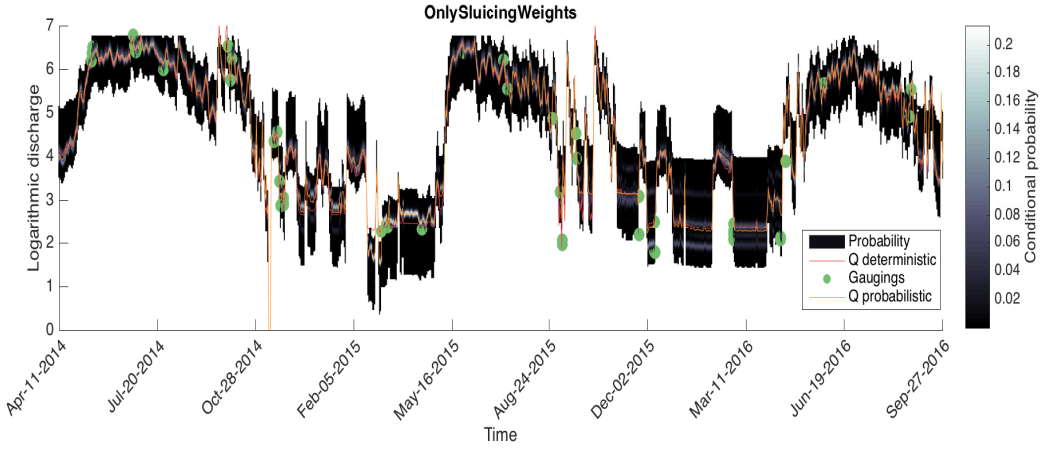

Figure B.2: Hydrograph for Only Sluicing model

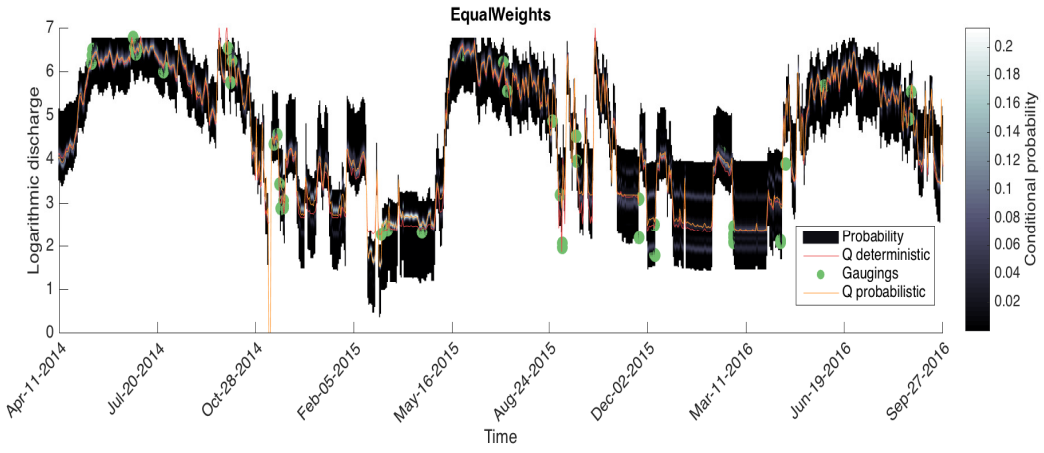

Figure B.3: Hydrograph for Equal Weights model

Appendix C

20% assumed discharge uncertainty

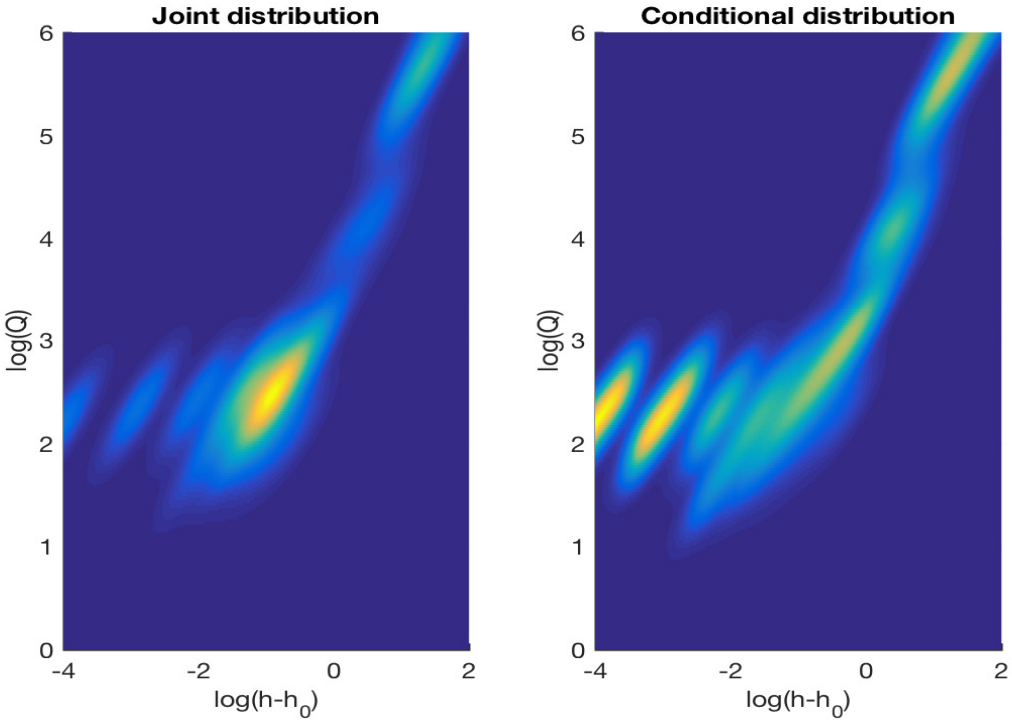

Figure C.1: A comparison of how the joint and conditional distributions change when using a 20% discharge uncertainty.