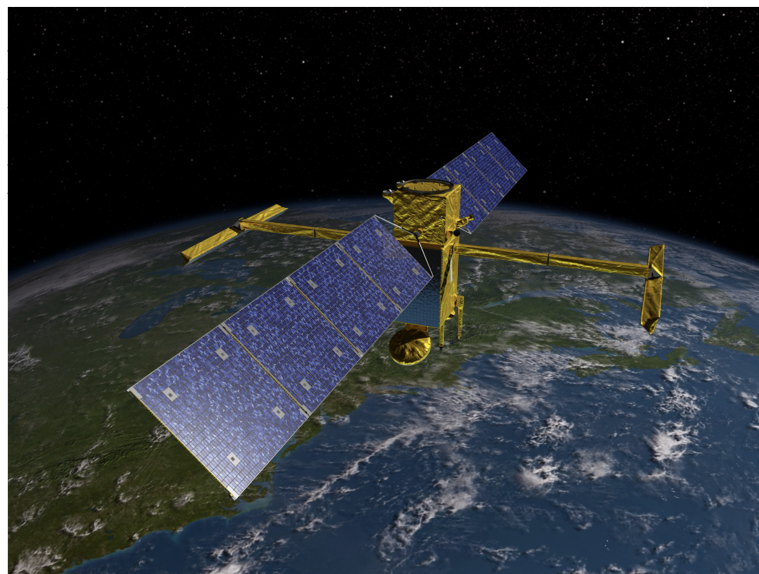


Estimation of river discharge using a Kalman Filter approach relying on simulated SWOT observations



Copyright © Courtesy NASA/JPL-Caltech

Master Thesis
Geodesy und Geoinformatics
University of Stuttgart

Siqi Ke

Stuttgart, September 2021

Supervisor: Dr.-Ing. Mohammad Tourian
University of Stuttgart

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum

Unterschrift

Abstract

Accurate estimates of river discharge are vital to quantify the amount of water resources. Hydrological monitoring through spaceborne sensors comes to a rescue for the steadily declining in-situ gauge network. As SWOT mission will be launched in November 2022, CNES has developed the Large-Scale SWOT Hydrology Simulator to identify SWOT applications, prepare for the use of SWOT data, and quantify SWOT impacts prior to launch.

In this thesis, we use this simulator to generate monthly river products for Po River from year 2002 to 2019. The input files are created with the support of satellite data, in-situ data and SWORD. The simulation process runs step by step and the final output files are river products. Besides, the changes in cross-sectional area are simulated by a Matlab program. After the simulation the post-processing is to compute river discharges. Based on Manning's equation the flow law parameters must be known. For this purpose, on the basis of mass-conservation equation the first round of Kalman Filter can be executed to obtain a priori discharges. And then using interior-point optimization with inequality constraints the flow law parameters are calculated, which will be used to recalculate discharges. These recalculated data will be reversely as additional observations for the second round of Kalman Filter and the posterior discharges could be then estimated. In the end the results are validated against in-situ gauge data.

The mean validation data of RMSE, correlation and NSE are respectively $501.87 \text{ m}^3/\text{s}$, 0.45 and 0.19. The correlation and NSE values are larger than 0, which indicates a positive linear dependence between the predicted and in-situ data and a good predictive skill of this method. Surely, the validation doesn't show the best performance. Therefore, some suggestions for improvements are also discussed in the end.

Key Words: River Discharge; SWOT Hydrology Simulator; Kalman Filter

Contents

Abstract	v
1 Introduction	1
1.1 River discharge	1
1.2 Objectives	2
1.3 Structure of this thesis	3
2 Study case and data	5
2.1 Study case	5
2.2 Data	5
2.2.1 The SWOT a priori River Database (SWORD)	6
2.2.2 In-situ data	8
2.2.3 River width time series data	8
2.2.4 Altimetry time series data	9
3 SWOT Hydrology Simulator	11
3.1 Introduction	11
3.1.1 SWOT Hydrology Toolbox	11
3.1.2 RiverObs	12
3.2 SWOT simulation products	13
3.2.1 Pixel cloud product	13
3.2.2 River product	14
3.2.3 Lake product	15
3.2.4 Raster product	16
3.3 Simulation steps	17
3.3.1 Generating time-variable polyline shapefiles	17
Time-variable width data	18
Time-variable altimetry data	20
Time-variable slope data	22
3.3.2 Creating time-variable polygon shapefiles	22
3.3.3 Downloading and setting up the simulator	26
3.3.4 Running the simulator	27

Setting up the working directory	29
Setting up your configuration files	30
Creating the orbit file	32
Running the simulator	33
4 Methodology	35
4.1 Introduction to river discharge estimation	35
4.2 Flowcharts	36
4.3 General steps	40
4.4 Collecting data from the simulator	40
4.5 Computation of the changes in cross-sectional area	45
4.5.1 Definition	45
4.5.2 Building the fitting function	45
4.5.3 Calculating A'	46
4.6 Estimation of flow law parameters	48
4.6.1 Process model	48
Space domain	48
Time domain	50
4.6.2 Observation model	53
4.6.3 Obtaining a priori discharge estimates	54
4.6.4 Computing the flow law parameters	56
4.7 Estimating posterior discharges	58
5 Results and validation	61
5.1 Results of a priori discharges	61
5.2 Results of flow law parameters	64
5.3 Results and validation of posterior discharges	67
6 Conclusion and outlook	71
6.1 Summary and conclusion	71
6.2 Outlook	72
A Estimated a priori discharges	75
B Estimated posterior discharges	85
References	95

List of Figures

1.1	Number of discharge monitoring stations according to GRDC database (Fekete & Vörösmarty 2002)	1
2.1	Po River and its corresponding area	5
2.2	SWORD networks with reach numbers per continent. Grey numbers in parentheses are reach numbers including ghost reaches (Altenau et al. 2020)	6
2.3	Divided reaches in Po River and location of gauge stations	7
2.4	Discharge time series at gauge stations	9
3.1	Pixel cloud product of a Po River section. This is the rawest product for land surfaces. The product provides longitude, latitude, height, pixel size and corrections for each point classified as water, for points on a buffer zone around these water zones as well as on systematically included areas.	14
3.2	Reach product for the Po River. Reach product provides the mean height, mean slope, etc. on each of these river sections, which will be used in the following discharge estimate.	15
3.3	Node product for the Po River. Node product is with mean values around each node for parameters such as width, height, flooded area, etc.	15
3.4	Lake product of Po River. The product for each cycle includes a median contour, and the average height of the lakes for passes when they were seen in their entirety (observations of only part of the lake are not taken into account).	16
3.5	Raster product of Po River. A raster product will be in NetCDF 2D format and will cover four point-cloud tiles (the 2 swaths and 120km along the ground track).	17
3.6	Reach 21406402131 with only sporadic data from year 2015 to 2019	18
3.7	Distribution of scale factors among nearby reaches	19
3.8	Availability of altimetry data in reaches	20
3.9	"Buffer" tool to convert a polyline to a polygon	23

3.10	Transformation from polylines to polygons in the Po River in January 2000	24
3.11	Add a field in the attribute table	25
3.12	Field types. "HEIGHT" is the height attribute, which specifies the initial WSE and is in the arithmetic type "double". "RIV_FLAG" designates polygons as river segments in "short integer". "id" is an identification field with the type "short integer"	25
3.13	Find the field calculator	26
3.14	Field calculation for "HEIGHT" attribute	27
3.15	Field calculation for attribute "RIV_FLAG"	28
3.16	Obtaining the repository of the SWOT Hydrology Toolbox and River-Obs (Elmer 2020)	28
3.17	Setting up the conda environment (Elmer 2020)	29
3.18	Setting the "PYTHONPATH" variables (Elmer 2020)	29
3.19	Modify "parameter_orbit.rdf" for January 2002	30
3.20	Modify directories in "parameter_sisimp.rdf" for January 2002	31
3.21	Modify orbit parameters in "parameter_sisimp.rdf"	31
3.22	Modify height parameters in "parameter_sisimp.rdf"	31
3.23	Modify output products in "parameter_sisimp.rdf"	32
3.24	Modify "parameter_river.rdf" in January 2002	32
3.25	Command to create orbit file	32
3.26	Orbit file "passplan.txt" in January 2002	33
3.27	Command to create the pixel cloud product	33
3.28	Naming convention of pixel cloud product	33
3.29	Command to create the river product	34
3.30	Naming convention of the river product	34
4.1	Whole process of this thesis	37
4.2	Kalman Filter steps	38
4.3	Interior-point optimization with inequality constraints	39
4.4	A part of the output data in the river product	41
4.5	Simulated vs. Landsat-observed width	42
4.6	Simulated vs. satellites-observed WSE	43
4.7	Histogram of <i>residual</i> between the simulated and Landsat width in gauge stations to verify if it is white noise	44
4.8	Composition of the cross-sectional area	45
4.9	A' with $\sigma_{A'}$ in gauge stations	48
4.10	Time domain transition matrix Φ_t	52
4.11	Time domain process noise N_t	52

4.12	Space and time domain estimates	55
4.13	Probability distribution of Q_s , Q_t and their joint prior $Q_{s,t}$	56
5.1	Comparison of a priori and in-situ discharges	62
5.2	Comparison of simulated and in-situ width	63
5.3	Comparison of simulated and in-situ WSE	64
5.4	Initial vs. optimized \bar{A}	65
5.5	Initial vs. optimized σ_z	65
5.6	Initial vs. optimized n_b	66
5.7	Calculated n for each gauge station in time series	67
5.8	Comparison of posterior and in-situ discharges	68
A.1	$Q_{est.}$ in 2002	75
A.2	$Q_{est.}$ in 2003	76
A.3	$Q_{est.}$ in 2004	76
A.4	$Q_{est.}$ in 2005	77
A.5	$Q_{est.}$ in 2006	77
A.6	$Q_{est.}$ in 2007	78
A.7	$Q_{est.}$ in 2008	78
A.8	$Q_{est.}$ in 2009	79
A.9	$Q_{est.}$ in 2010	79
A.10	$Q_{est.}$ in 2011	80
A.11	$Q_{est.}$ in 2012	80
A.12	$Q_{est.}$ in 2013	81
A.13	$Q_{est.}$ in 2014	81
A.14	$Q_{est.}$ in 2015	82
A.15	$Q_{est.}$ in 2016	82
A.16	$Q_{est.}$ in 2017	83
A.17	$Q_{est.}$ in 2018	83
A.18	$Q_{est.}$ in 2019	84
B.1	$\hat{Q}_{est.}$ in 2002	85
B.2	$\hat{Q}_{est.}$ in 2003	86
B.3	$\hat{Q}_{est.}$ in 2004	86
B.4	$\hat{Q}_{est.}$ in 2005	87
B.5	$\hat{Q}_{est.}$ in 2006	87
B.6	$\hat{Q}_{est.}$ in 2007	88
B.7	$\hat{Q}_{est.}$ in 2008	88
B.8	$\hat{Q}_{est.}$ in 2009	89

B.9	$\hat{Q}_{est.}$ in 2010	89
B.10	$\hat{Q}_{est.}$ in 2011	90
B.11	$\hat{Q}_{est.}$ in 2012	90
B.12	$\hat{Q}_{est.}$ in 2013	91
B.13	$\hat{Q}_{est.}$ in 2014	91
B.14	$\hat{Q}_{est.}$ in 2015	92
B.15	$\hat{Q}_{est.}$ in 2016	92
B.16	$\hat{Q}_{est.}$ in 2017	93
B.17	$\hat{Q}_{est.}$ in 2018	93
B.18	$\hat{Q}_{est.}$ in 2019	94

List of Tables

2.1	Location and elevation of the gauge stations	8
3.1	Comparison of output attributes of reach and node product	16
3.2	Scale factors to fill width	19
3.3	Water body identifier	26
4.1	Extracted attributes from the river product	41
4.2	Correlation of input and simulated data in gauge stations	43
5.1	Correlation of simulated and in-situ data	64
5.2	Validation in gauge stations	68

Chapter 1

Introduction

1.1 River discharge

Rivers are a link between atmospheric, terrestrial and oceanic processes, and route about two-fifths of total global rainfall over land back to the ocean (Oki & Kanae 2006). They also represent an important resource for agriculture and urban development as well a major hazard during flood events (Durand et al. 2016). Thus, accurate estimates of river discharge are vital because they quantify the amount of water available for human use, determine the amount of water during a flood event, and indicate the overall response of the watershed to atmospheric pressure. Traditionally monitoring of river discharge is obtained by using in-situ gauges. However, these in-situ gauge networks have their own limitations, especially at full catchment of water discharge observations. Globally, the available discharge database has steadily been declining over the past few years (see figure 1.1).

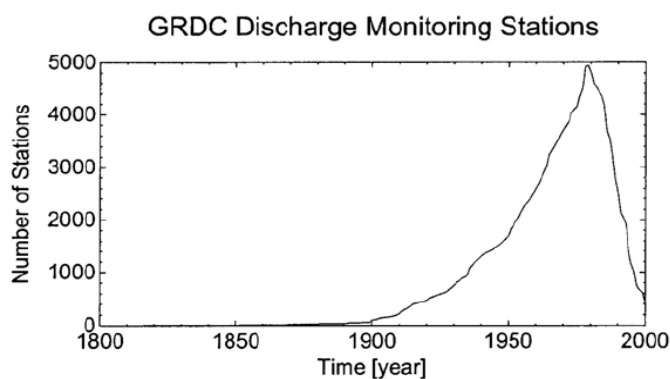


Figure 1.1: Number of discharge monitoring stations according to GRDC database (Fekete & Vörösmarty 2002)

Therefore, discharge monitoring through spaceborne sensors is a viable tool due to their large observation range, short cycle time and less restriction due to ground conditions. The Surface Water and Ocean Topography (SWOT) mission is a swath

altimeter satellite to be launched by National Aeronautics and Space Administration (NASA) and National Centre for Space Studies (CNES) in November 2022 (JPL 2021). Its main objective is to make a survey of the global Earth's surface water and ocean. It will provide measurements of water surface elevation, change in cross-sectional area, river width, slope, etc., through which discharge can be estimated indirectly for river widths down to 50–100 m (Pavelsky et al. 2014). SWOT observations (change in cross-sectional area, width, slope) are combined with flow law parameters from river database (SWORD) to estimate discharge on each pass (Durand et al. 2020). There are now five proposed algorithms for use in the SWOT mission: AMHG, GaMo, Metroman, MFG and MFCR. The median of these five is evaluated as a sixth algorithm herein. The theoretical basis of the algorithm AMHG is at-many-stations hydraulic geometry and the others' is Manning's equation. These algorithms can estimate river discharge in some ungauged rivers with approximately 35% root-mean-square error. (Durand et al. 2016)

Currently, CNES develops the Large-Scale SWOT Hydrology Simulator, which accounts for a comprehensive error budget and produces output in formats comparable to that expected from official SWOT products (Elmer et al. 2020).

1.2 Objectives

In this thesis, we will take advantage of the SWOT Hydrology Simulator to simulate monthly SWOT-like data in Po River from January 2002 to December 2019 and then use them to estimate the monthly river discharges in this time interval via Kalman Filter. To achieve the objectives, the specific tasks are as follow:

- Creating monthly input from January 2002 to December 2019 for the simulator via the available satellite data, in-situ data and a priori database of SWOT, namely SWORD.
- Doing simulation to generate monthly river products.
- Collecting simulated data in time series.
- Simulating the changes in cross-sectional area.
- Estimating a priori river discharges via first round of Kalman Filter.

- Computing flow law parameters through interior-point optimization with inequality constraints.
- Recalculating discharge and estimate posterior river discharges via second round of Kalman Filter.
- Validating the results with the support of in-situ data.

1.3 Structure of this thesis

The structure of this thesis is divided into the following chapters. In Chapter 2, the study case Po River and the utilized data in this thesis will be introduced. In Chapter 3, the simulator will be presented in detail, especially the composition of the simulator, the products of it and the specific simulation steps. In Chapter 4, post-processing after simulation to estimate flow law parameters and the river discharges through Kalman Filter will be illustrated. In addition, flowcharts of the process in this thesis will be displayed. In Chapter 5 the results of Kalman Filter will be shown and a validation via in-situ data will be executed. And some tips for improvement will be given. Chapter 6 indicates the summary, conclusion and further re- search direction.

Chapter 2

Study case and data

2.1 Study case

Po river is the case study of this thesis. Stretching 652 km, the Po River is Italy's longest river, rising in the Monte Viso group of the Cottian Alps on Italy's western frontier. It passes through the cities of Turin, Piacenza, and Cremona as well as dozens of smaller towns and discharges into the Adriatic Sea in the east. At its widest point, the opposing riverbanks are more than 500 m apart (Britannica 2019). Figure 2.1 shows the Po River and its corresponding area.



Figure 2.1: Po River and its corresponding area

2.2 Data

Three types of data were used in this study: **SWOT A Priori River Database** (SWORD), in-situ data, altimetry and Landsat data. SWORD, altimetry and Landsat data mainly played a role in the simulation step as database and input. In this thesis, the SWORD version v0.5 is used. In-situ data were utilized for the discharge estimate and validation.

2.2.1 The SWOT *a priori* River Database (SWORD)

The SWORD is a global *a priori* database of river networks and reaches (see Figure 2.2). The SWORD was created by merging multiple global hydrographic databases into one consistent product. The SWORD provides high-resolution river nodes and reaches with attached hydrologic variables (WSE, width, slope, etc.). Each river over 30 m wide is defined in the SWORD either as the centerline of the river divided into reaches approximately 10 km long or as nodes every 200 m along this line. Altenau et al. (2021)

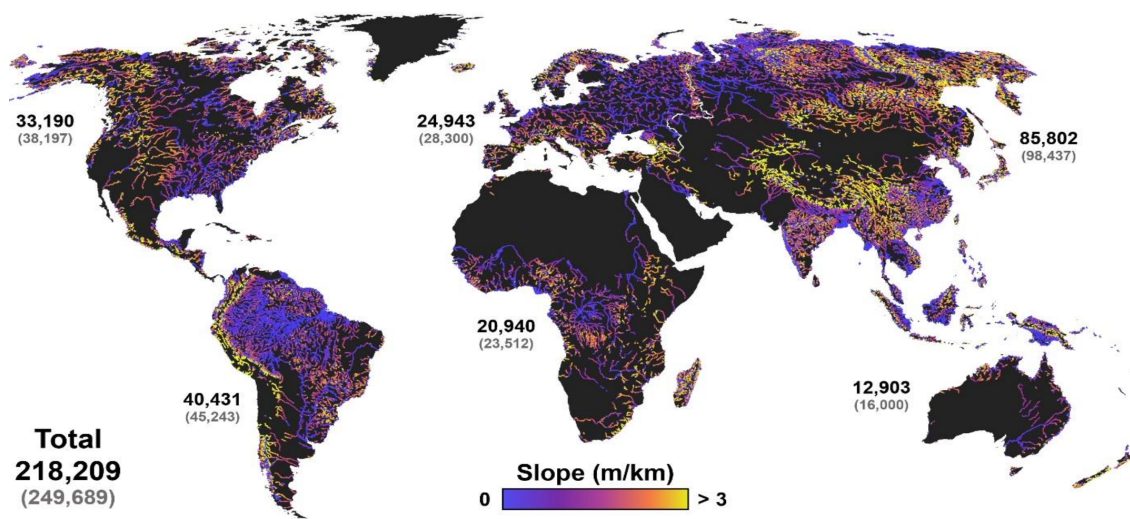


Figure 2.2: SWORD networks with reach numbers per continent. Grey numbers in parentheses are reach numbers including ghost reaches (Altenau et al. 2020)

SWORD data are saved in shapefile and NetCDF formats. The shapefile format is distributed as a set of files according to continent and the major basins on a continent, while the NetCDF format is allocated by continent. Altenau et al. (2020) In this thesis SWORD data support the execution of the SWOT Hydrology Simulator in the creation of pixel cloud products and the transformation from pixel cloud products to river products.

The shapefiles are separated for nodes and reaches, with nodes represented as spaced points and reaches represented as polylines. All shapefiles follow geographic projection with reference to datum WGS84. In this study, the shapefile of the Po River was processed to convert polylines to polygons for its monthly width, which is part of the input for simulation.

The NetCDF format is more comprehensive and is the official format required by the Jet Propulsion Laboratory (JPL) to process SWOT observations. Each NetCDF file contains a set of global attributes as well as three groups of variables in different spatial scales (`/centerlines`, `/nodes`, `/reaches`). The NetCDF format also contains an expanded set of subgroups under the reaches group (`/area_fits`, `/discharge`) that are not included in the shapefile formats. These subgroups are needed for SWOT discharge algorithms. The corresponding NetCDF file plays a role as an a priori database while running the simulator (Altenau et al. 2020).

The SWORD version v0.5 divides the Po River into 286 reaches, with 65 reaches in the mainstream. Each reach is roughly 10 kilometers long (see Figure 2.3). In preparation for the simulation, the shapefile of this area was extracted from the SWORD, and the NetCDF file in Europe was utilized.

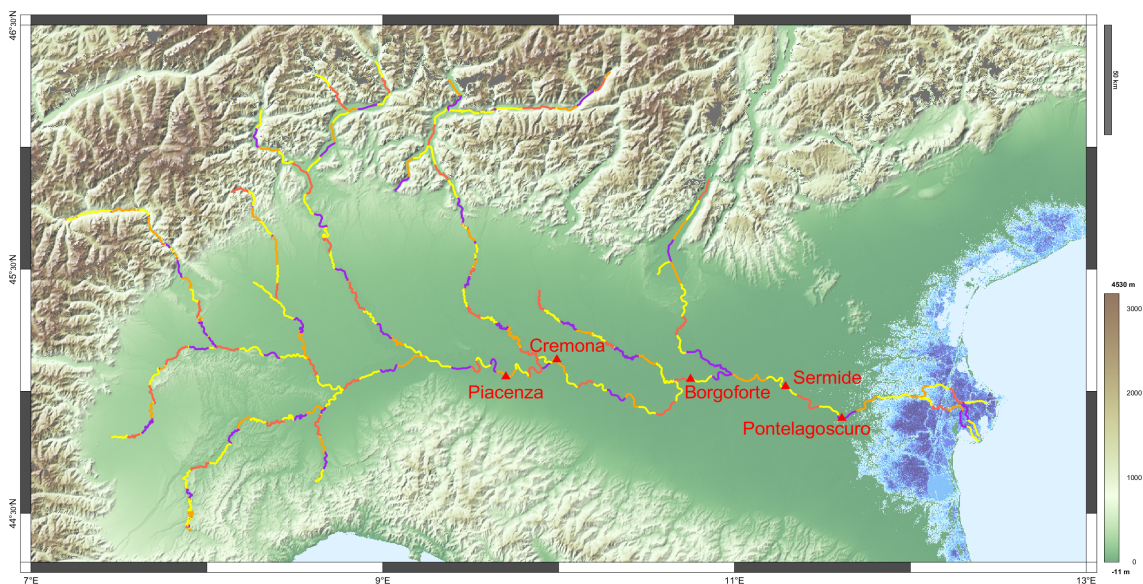


Figure 2.3: Divided reaches in Po River and location of gauge stations

Notably, the SWORD database is still under development and therefore contains some artifacts remaining. The value of these attributes was set as "not a number" (NaN). Some of the attributes were inserted with valid values via "online" SWOT processing or via the "offline" computations of the research team. More specifically, the reach subgroups `/area_fits` and `/discharge` were filled with "NaN" because several important elements for discharge estimation were based on the first year of mission operations or have not yet been defined.

2.2.2 In-situ data

Along the Po mainstream, there are five gauge stations that provide daily observation data, or so-called "in-situ data". Table 2.1 lists the location and average elevation of these gauge stations and Figure 2.3 also shows the stations on the map.

Name	Latitude [°]	Longitude [°]	Elevation [m]
Piacenza	45.06	9.70	42.37
Cremona	45.13	9.99	29.03
Borgoforte	45.05	10.75	14.05
Sermide	45.02	11.29	9.50
Pontelagoscuro	44.89	11.61	3.48

Table 2.1: Location and elevation of the gauge stations

The daily observation data includes stage, water level, water depth, discharge and flow area from the years 1995 to 2011. Moreover, the approximate width w_{gauge} in time series can also be computed as:

$$w_{gauge}(t) = \frac{A_{gauge}(t)}{d_{gauge}(t)} \quad (2.1)$$

where A_{gauge} is the flow area and d_{gauge} is the water depth. These data play important roles in simulation, estimation, and validation. In-situ discharge is a particularly important measurement tool in validating the results of the Kalman Filter. The monthly discharge of the five gauge stations is shown in Figure 2.4. Generally, the discharge of the gauge stations follows the same seasonal behavior. It is noteworthy to mention that the discharge peaked in October 2000. Actually, during the October 2000 flood more than 80 km² of rural and settled areas located in the floodplain close to the junction of the Sesia and Po rivers were inundated (Fugazza et al. 2008).

2.2.3 River width time series data

As the SWORD provides a static database with no time series, it was necessary to input the varying width and altimetry data to reflect the change in discharge in different time periods.

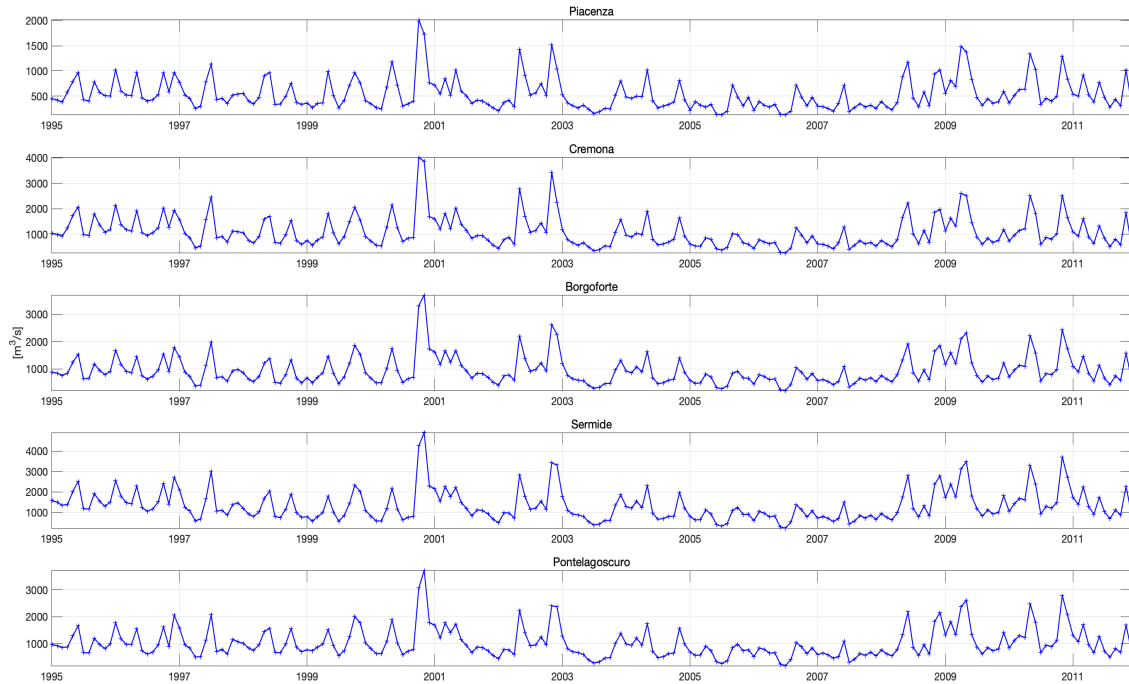


Figure 2.4: Discharge time series at gauge stations

The monthly width data were taken from cloud-free Landsat imagery from the years 2000 to 2019. Only widths from March to October of every year were available in the data set for each reach. (Elmi et al. 2016) Thus, data for the other months needed to be manually created before being input into the SWORD data, which will be described in Section 3.3.1

2.2.4 Altimetry time series data

Altimetry data included merged data produced by several satellite missions, such as Envisat, Topex, Jason, and so on (Tourian et al. 2016). The altimeters on the satellites measure the roundtrip time taken by a radar wave to reach the surface from the satellite (Durand et al. 2010). The available data varied from 2002 to 2019, which also determined the simulation period. There was no fixed time interval, and some reaches had only sporadic data or no data at all.

To overcome the above problems and provide a complete timeline of altimetry data, it is necessary to manually create appropriate data, as described in Section 3.3.1.

Chapter 3

SWOT Hydrology Simulator

3.1 Introduction

The SWOT Hydrology Simulator provides open source tools that enable end users to generate virtual SWOT products with representative characteristics; in this way, users can become familiar with product contents and formats, use the data to conduct studies, validate methodology and propose improvements. (Elmer 2020) The SWOT Hydrology Simulator is coded in high-level programming language (Python 3) for simple distribution, installation, and use. The managing environment of the SWOT Hydrology Simulator is the conda environment, which serves to create, remove, and package virtual environments as well as install software for the simulator.

The SWOT Hydrology Simulator consists of two toolboxes: the SWOT Hydrology Toolbox and RiverObs. The SWOT Hydrology Toolbox aims to easily and quickly generate L2 (level-2) HR (high-rate) pixel cloud products over huge spatial (e.g., large basin, national, or even continental scale) and time scales (e.g., months, years) with relatively representative errors. After that, RiverObs is used to run end-to-end simulation of pixel cloud to L2 HR vector based river or lake products. Conveniently, the simulator provides overall scripts that allow all steps to be run consecutively. Furthermore, both algorithms and product formats are still under development and are regularly updated. Hence, the versions of both toolboxes must be associated with each other while running the simulator. For this study, the SWOT version "release_version_11_24_2020" and the compatible RiverObs version "release_version_06_18_2020" are used.

3.1.1 SWOT Hydrology Toolbox

The SWOT Hydrology Toolbox contains the following main tools:

- **select_orbit_cnnes:** An orbit selection tool to compute SWOT orbit files (pass-plan). The orbit files include cycle, orbit number, and mission time of SWOT over the study area. Based on this information, the following tools are able to generate products for the appropriate region.
- **sisimp:** A large scale simulator to produce L2 HR pixel cloud products from an input shapefile of a water mask.
- **scripts:** This is a river processing tool that contains scripts to improve geolocation and link RiverObs to convert one or more tiles of pixel cloud into the same number of L2 HR river products.
- **processing:** This project is dedicated to the production of L2 HR lake products from pixel cloud.
- **test:** This folder saves the input, output, and configuration files (rdf.file) of our study cases.

It is noteworthy to mention that although the large scale simulator (sisimp) provides fairly representative statistical errors, several simplifications or approximations are made. For instance, the simulator utilizes spherical Earth approximate geolocation equations, which potentially leads to a loss of accuracy at high latitudes ($> 60^\circ$). In addition, topography is not taken into account, and idealized pixel cloud processing with perfect water detection is assumed. However, there is an option to introduce dark water. The simulator also allows simplified representation of water height, such as a spatially constant height for each water body or 2D height models. There are also options to insert true heights from an input shapefile or a NetCDF file.

3.1.2 RiverObs

RiverObs contains the following main packages:

- **RiverObs:** This is the main package for associating data with river reaches, and estimating hydrology parameters base on reach averaging.
- **Centerline:** This provides a class that can be used to project data or refine a river centerline.
- **GeometryDataBase:** This package aims to quickly find which reach intersects with a geometry of interest. The geometries are assumed to be stored in a shapefile.

- **SWOTRiver:** This package contains classes that use the RiverObs capabilities to produce hydrology outputs from SWOT (simulated) data.
- **GDALOGRUtilities:** This provides homegrown utilities for reading and writing various GIS files.
- **GWDLR:** This is an optional package to convert Global Width Database Large Rivers raster data into vectors that can be used as centerlines.

RiverObs must be linked to the a priori SWORD database because the latter contains fixed node locations, reach boundaries, and high-resolution reach centerlines that are necessary for data processing. The version of SWORD must be associated with the version of RiverObs. In our case, the SWORD version is v0.5.

3.2 SWOT simulation products

The SWOT Hydrology Simulator provides four main L2 HR products: pixel cloud products, river products, lake products, and raster products. By executing the corresponding scripts, they can be generated in a certain order.

3.2.1 Pixel cloud product

The pixel cloud product is the rawest product of SWOT and it is available in NetCDF and shapefile format, covering a tile corresponding to a swath over 60 km long on one of the two sides (left or right) of the satellite's ground track. The pixel cloud covers an area that is defined a priori in input shapefile and classified as water and a surrounding buffer zone. Each pixel indicates longitude, latitude, height, pixel size, metadata (date, pass number), and corrections. This product is generated for each pass of the satellite. Figure 3.1 shows an example of a pixel cloud product in a section of the Po River.

The required inputs of this product are SWOT orbit files in NetCDF format, a polygon-shapefile covering water zone, and eventually roll residual errors etc.

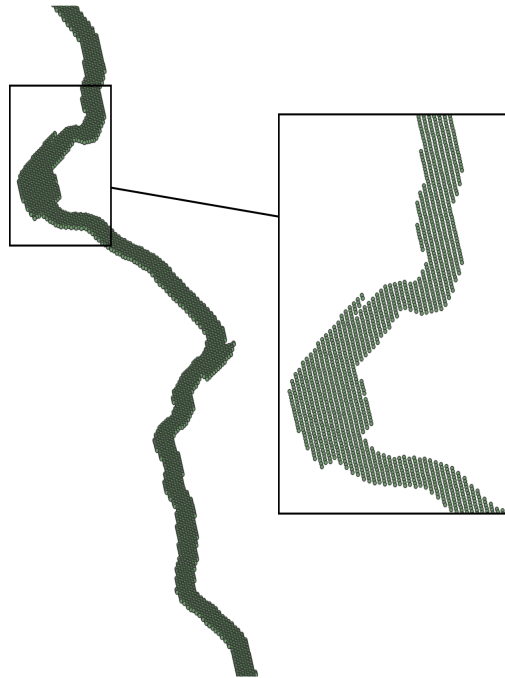


Figure 3.1: Pixel cloud product of a Po River section. This is the rawest product for land surfaces. The product provides longitude, latitude, height, pixel size and corrections for each point classified as water, for points on a buffer zone around these water zones as well as on systematically included areas.

3.2.2 River product

Like the a priori SWOT database, the river product has two types of breakdown. The reach product (see Figure 3.2) is located at the centerline of the river and divided into sections. The reach product provides mean height, mean width, mean slope, and so forth for each of the river sections. This product was the main product used to estimate river discharge in this study. The node product (see Figure 3.3) is composed of nodes placed approximately every 200 m along the centerline and provides the mean values of each node for parameters such as width, height, flooded area, and so on. These products are provided for each satellite pass.

Table 3.1 compares the attributes of the reach and node products. The differences between the products mean that they are used for different purposes in post-processing or analysis.

The required inputs of this product are SWOT v0.5, the associated pixel cloud product, and a configuration file "parameter_river.rdf" to control the output. Importantly, the river and the following lake product only consider the water pixels.

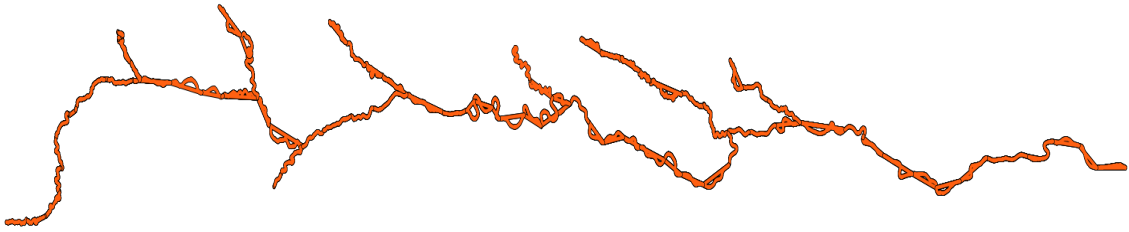


Figure 3.2: Reach product for the Po River. Reach product provides the mean height, mean slope, etc. on each of these river sections, which will be used in the following discharge estimate.



Figure 3.3: Node product for the Po River. Node product is with mean values around each node for parameters such as width, height, flooded area, etc.

3.2.3 Lake product

The lake product (see Figure 3.4) for each individual satellite pass gives information about height, surface area, contour, and so on for each observation of water surfaces that are not known rivers and that are more than one hectare in size. Like the river product, an a priori lakes database must be linked for the identification of lake elements and to provide relevant information.

The required inputs of this product are an a priori lakes database, the associated river product, and a configuration file "parameter_laketile.cfg" to control the output.

Reach Product	Node Product
- reach ID	- reach ID
- time	- node ID
- latitude	- time
- longitude	- latitude
- WSE	- longitude
- width	- WSE
- slope	- width
- total water surface area	- total water surface area
- geoid slope	- solid Earth tide height
- fractional area of dark water	- river flow direction
- number of nodes in the reach that have a valid WSE	- number of pixels that have a valid WSE
...	...

Table 3.1: Comparison of output attributes of reach and node product

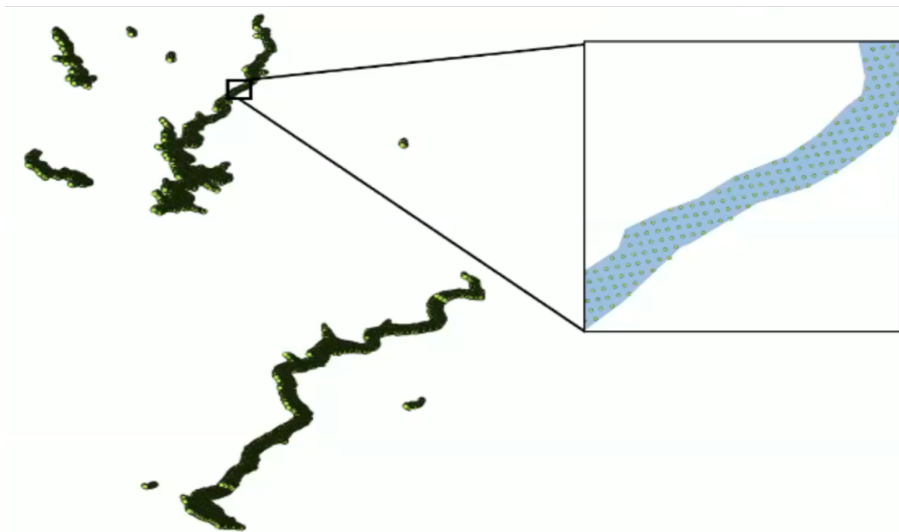


Figure 3.4: Lake product of Po River. The product for each cycle includes a median contour, and the average height of the lakes for passes when they were seen in their entirety (observations of only part of the lake are not taken into account).

3.2.4 Raster product

The computation of the raster product is still under development. It will be systematically generated from the pixel cloud product and will be in the NetCDF and GeoTiff formats with two resolutions: 100 m and 250 m. It will cover four point-cloud tiles corresponding to two swaths of 120 km along the ground track. The

points of the pixel cloud will be plotted on a regular grid according to a Universal Transverse Mercator (UTM) projection. This will be a one-per-pass product. Figure 3.5 shows an example of the raster product for the Po River.

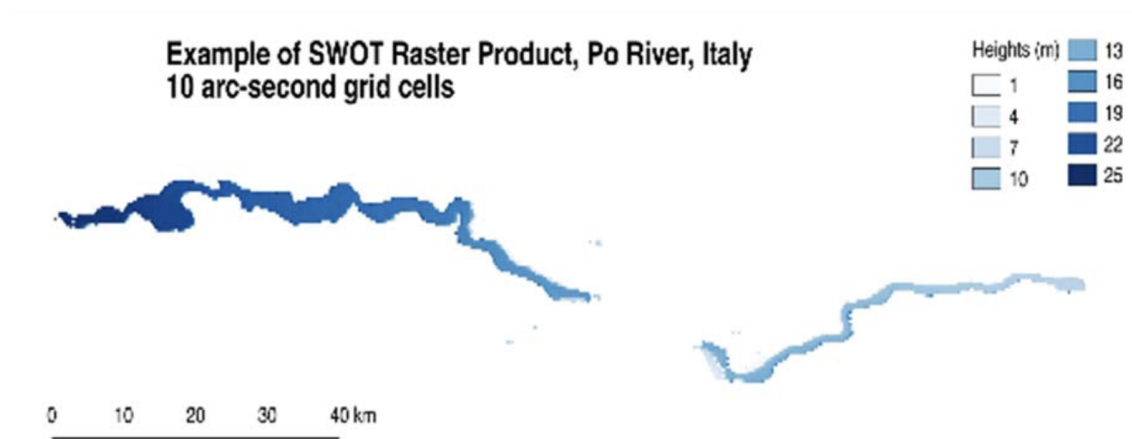


Figure 3.5: Raster product of Po River. A raster product will be in NetCDF 2D format and will cover four point-cloud tiles (the 2 swaths and 120km along the ground track).

3.3 Simulation steps

Much of the effort required to run the simulator is made upfront to create the simulator input. In our case, the SWOT Hydrology Toolbox generated monthly river products from 2002 to 2019. To reflect the monthly variation of the Po River, width data from Landsat and altimetry data (see Section 2.2.3) replaced the constant values in SWORD for each reach. Then, the time-variable polyline shapefiles from SWORD were processed to the required polygon shapefiles based on ArcGIS because the simulator uses the polygon extent to calculate river width. Finally, the time-variable input was allocated to each month's working directory, where each month has its own configuration files and output directory. The simulation was executed separately for each month.

3.3.1 Generating time-variable polyline shapefiles

As mentioned, the constant width, altimetry data, and slope in SWORD were replaced. Since SWORD data is saved in shapefile and NetCDF formats, files in both formats have to be changed at the same time with consistent information. The input of width and altimetry data is based on existing satellite data, and

slope is calculated from altimetry data and reach length. Finally, each month has its own specific shapefiles and NetCDF files with corresponding data, which is termed the "time-variable input".

Time-variable width data

Ideally, Landsat would have provided each reach's monthly width data and variance from March to October for each year from 2000 to 2019 for the Po River. However, in reality, there was no complete data for any of the reaches, and in some cases there was no data for a whole year (see Figure 3.6). As such, these gaps needed to be filled before the input. If width data was available in a given month for other years, the mean monthly data was inserted for the missing month. Additionally, if there was no data for a given month in any year, data inter- or extrapolation from other months was utilized to create an appropriate width for the month.

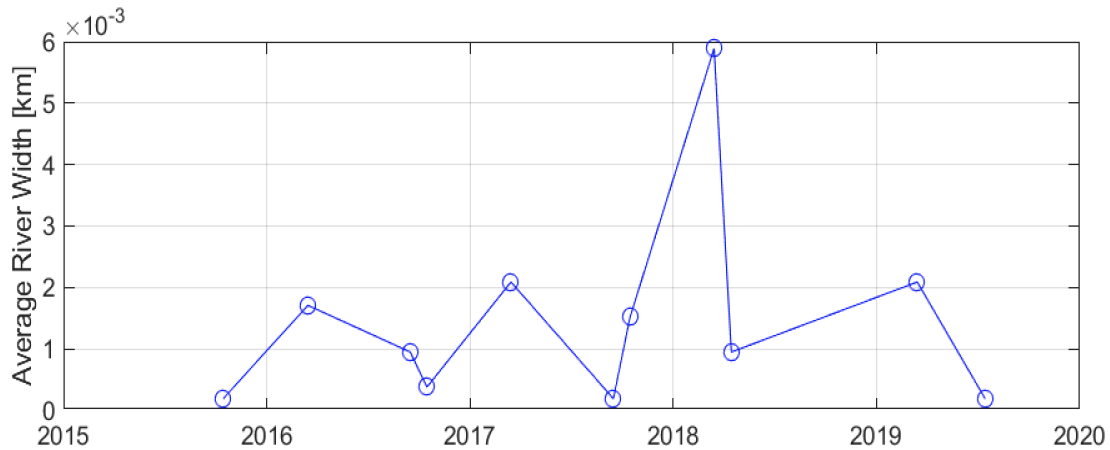


Figure 3.6: Reach 21406402131 with only sporadic data from year 2015 to 2019

Following these steps, the width data for January, February, November, and December was created manually. First, for each gauge station i , the mean width $\bar{w}_{i\text{gauge}}$, standard deviations $\sigma_{\bar{w}_{i\text{gauge}}}$, and mean monthly width $\tilde{w}_{ij\text{gauge}}$ for these months j were calculated. Then, a scale factor was computed for each month at each gauge station:

$$s_{ij} = \frac{\tilde{w}_{ij\text{gauge}} - \bar{w}_{i\text{gauge}}}{\sigma_{\bar{w}_{i\text{gauge}}}} \quad (i : \text{Piacenza} \dots \quad j : \text{January} \dots) \quad (3.1)$$

Table 3.2 shows the values obtained.

[-]	Jan.	Feb.	Nov.	Dec.
Piacenzao	-0.155	-0.222	0.489	0.086
Cremona	-0.050	-0.172	0.465	0.156
Borgoforte	0.061	-0.102	0.541	0.306
Sermide	0.053	-0.116	0.515	0.327
Pontelagoscuro	0.116	-0.054	0.515	0.358

Table 3.2: Scale factors to fill width

These scale factors were assigned to nearby reaches (see Figure 3.7). Using the distributed scale factors, widths $w_{k\text{SWORD}}$ and standard deviations $\sigma_{w_{k\text{SWORD}}}$ from SWORD, the mean monthly widths \tilde{w}_{ijk} for each reach k were computed:

$$\tilde{w}_{ijk} = w_{k\text{SWORD}} + s_{ij} \cdot \sigma_{w_{k\text{SWORD}}} \quad (3.2)$$

i : Piacenzao...

j : January...

k : reaches

\tilde{w}_{ijk} was then allocated to those months with missing data in each year. At this point, all reaches had time-variable width data and could be overwritten into their own shapefiles and NetCDF files. The corresponding variance is overwritten by the Landsat width variance data. In addition, nodes within the same reach were assigned the width of this reach.

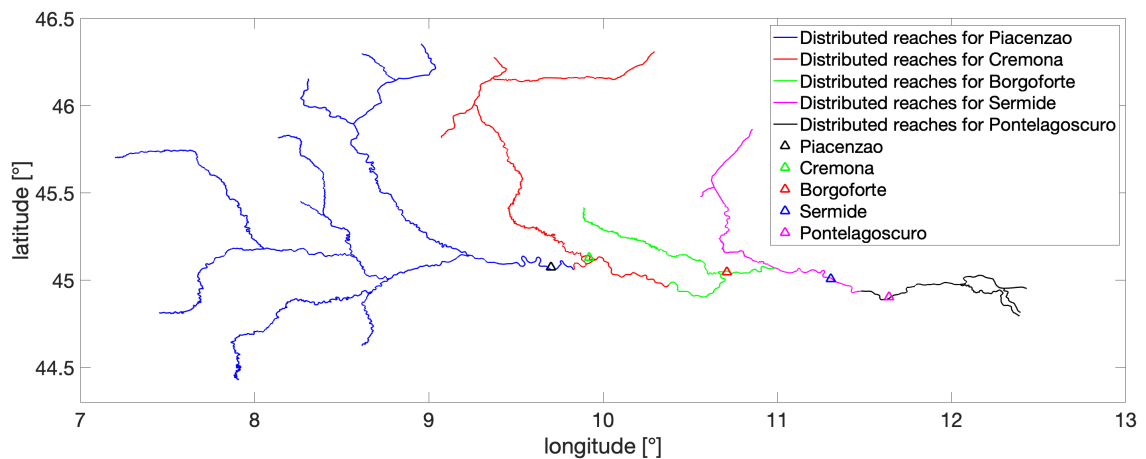


Figure 3.7: Distribution of scale factors among nearby reaches

Time-variable altimetry data

Ideally, the various satellites would have provided each Po River reach's altimetry data and standard deviation from 2002 to 2019. However, as with width data, there was no complete data for any of the reaches, and some reaches had no data at all (see Figure 3.8). Therefore, it was necessary not only to fill the gaps but also to create suitable data for the reaches without data. The process for filling the gaps in altimetry data was similar to that followed for width data. If the month missing data in one year has data from other years, the missing data was filled with mean monthly data from other years. Otherwise, inter- or extrapolation with data from other months was performed to create an appropriate value.

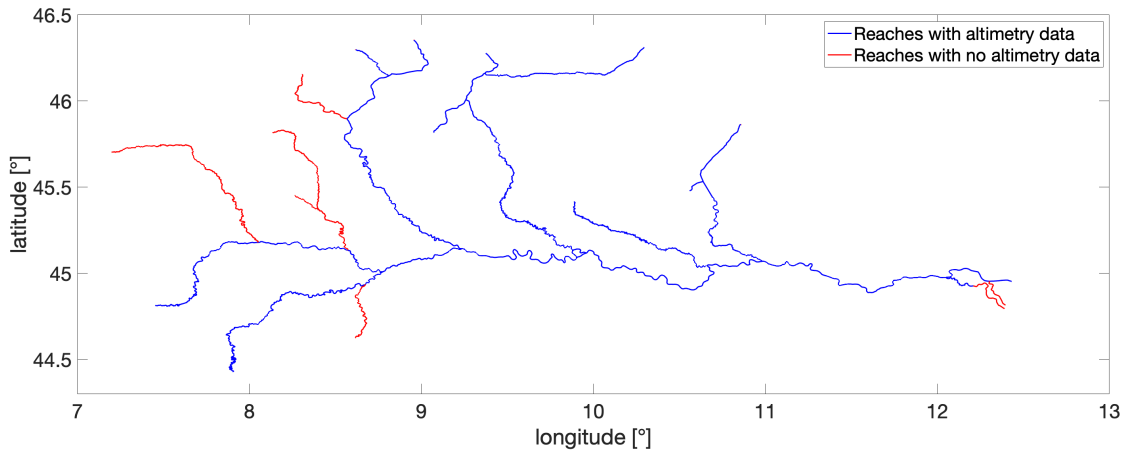


Figure 3.8: Availability of altimetry data in reaches

After this, reaches with unavailable data were processed. Neighboring reaches with valid data were utilized along with wave travel time between reaches.

To build the wave travel time, average flow velocity (Tourian et al. 2016) between reaches was computed:

$$V = 1.48 \cdot \bar{w}_{\text{Landsat}}^{0.8} \cdot s_{\text{SWORD}}^{0.6} \quad (3.3)$$

V is the average flow velocity of the reach. \bar{w}_{Landsat} is the mean width of the reach according to Landsat. s_{SWORD} is the slope of the reach according to SWORD.

Wave travel time T_L could then be estimated:

$$T_L = \frac{L}{c} \quad (3.4)$$

where L is the distance between two reaches from their center points (provided by SWORD) and c is the celerity. In the kinematic wave equation, the momentum balance can be arranged using the Manning formula; consequently, celerity c was expressed as

$$c = b \cdot V \quad (3.5)$$

where $b = \frac{5}{3}$ and V is the above computed average flow velocity.

Next, the time lag was applied to the computation of missing altimetry data based on data from neighboring reaches either up- or downstream. Since the WSE and standard deviation of all reaches are accessible in SWORD, a factor based on it was built first:

$$f = \frac{\sigma_{\text{WSE}_{\text{SWORD}_2}}}{\sigma_{\text{WSE}_{\text{SWORD}_1}}} \quad (3.6)$$

where $\sigma_{\text{WSE}_{\text{SWORD}_1}}$ is the WSE standard deviation from SWORD of the neighboring reach and $\sigma_{\text{WSE}_{\text{SWORD}_2}}$ is that to compute. If $h_1(t)$ and $\sigma_{h_1(t)}$ are the known altimetry time series data and the corresponding standard deviations of the neighboring reach, the unknown $h_2(t)$ and $\sigma_{h_2(t)}$ can be derived by

$$h_2(t) = f \cdot (h_1(t + T_L) - \text{WSE}_{\text{SWORD}_1}) + (\text{WSE}_{\text{SWORD}_2} - \text{WSE}_{\text{SWORD}_1}) \quad (3.7)$$

$$\sigma_{h_2(t)} = f \cdot \sigma_{h_1(t+T_L)} \quad (3.8)$$

When one reach, which is to be computed, has a number of neighboring reaches, the $h_2(t)$ and $\sigma_{h_2(t)}$ for each reach were first processed separately, after which their weighted mean was determined.

Next, the time-variable altimetry data and corresponding standard deviations of all reaches were obtained, which could be written into the specific shapefiles and NetCDF files. Nodes within the same reach were also assigned with data from this reach.

Time-variable slope data

Slope data were derived from altimetry data and the reach length. The slope of a reach was computed through its neighboring reach:

$$s(t) = \frac{h_{up}(t) - h_{down}(t)}{L} \quad (3.9)$$

where $h_{up}(t)$ and $h_{down}(t)$ are the altimetry data of the up- and downstream at t , respectively, and L is the distance between two reaches from their center points.

The corresponding standard deviations were obtained by the propagation of uncertainty:

$$\sigma_s(t) = \frac{\sqrt{\sigma_{h_{up}(t)}^2 + \sigma_{h_{down}(t)}^2}}{L} \quad (3.10)$$

When a reach had more than one neighbor, the weighted mean was computed.

After obtaining the time series of the slope for all reaches, this was written into the corresponding shapefiles and NetCDF files. Unlike other data, slope is specific to each reach section, and the node section therefore did not need to be changed.

Up to this point, the time series input files were already prepared. The shapefiles will continue to be processed to complete the transformation from polyline to polygon files. The details of the operation are explained in Subsection 3.3.2. Additionally, NetCDF files will be used as input directly. The main purpose of the shapefiles is to produce a pixel cloud product, while NetCDF files enable the generation of the river product.

3.3.2 Creating time-variable polygon shapefiles

There are many ways to create a polygon shapefile depending on the application used, such as ArcGIS, QGIS, or even Python (for experienced users). Furthermore, a polygon shapefile can be generated from different types of files, such as a digital elevation map, a water mask raster, or a polyline shapefile. In this study, the starting point was the polyline shapefile, and the processing platform was ArcGIS.

Before the transformation, it had to be ensured that the shapefiles had a WGS84 coordinate system, which is necessary for running the simulator. Otherwise, the "Project" tool would be used to reset it. Fortunately, all shapefiles in SWORD are referenced to datum WGS84. The coordinate system could also be verified by considering the shapefile properties.

In ArcGIS, the "Buffer" tool was used to convert a polyline to a polygon (see Figure 3.9).

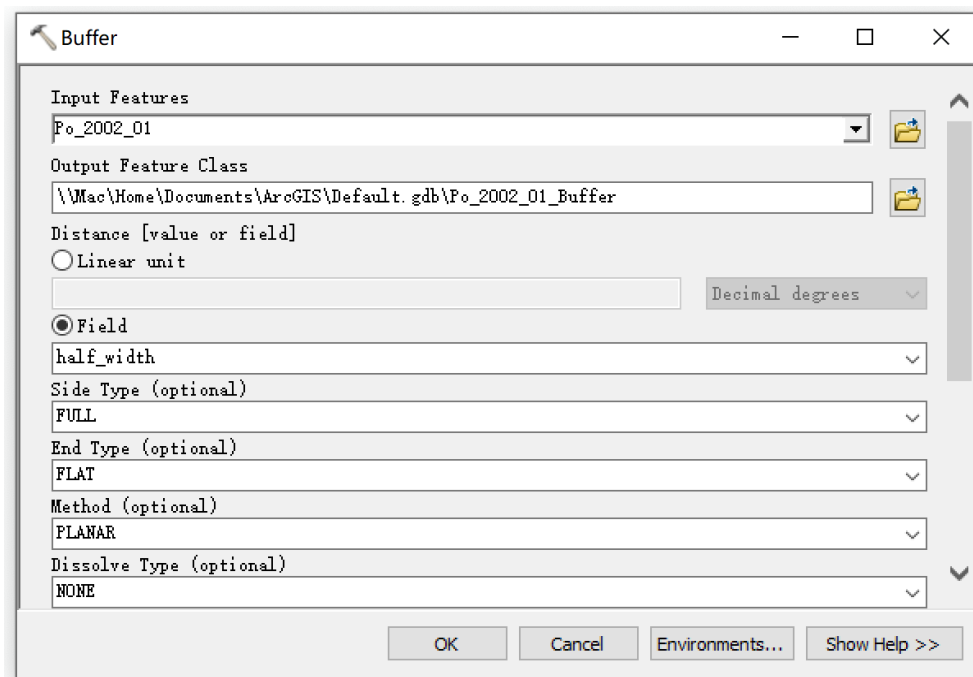


Figure 3.9: "Buffer" tool to convert a polyline to a polygon

The "Distance" field can be specified using a constant value ("Linear Unit") or a width field ("Field"). In this study, the polyline shapefiles already had the width attribute, but the distance was one-sided; thus, setting a value of 50 meters resulted in a polygon width of 100 meters. As such, half of the width of all reaches was calculated and saved as a new attribute of the polyline shapefiles ("half_width"), or the "Distance" field in the "Buffer" tool. Other options like "Side type" and "End type" could also be selected. A comparison between polylines and polygons in the Po River is given in Figure 3.10.

Once the polygon shapefile was generated, three attributes were added:

- **HEIGHT:** A height attribute specifies the initial WSE.
- **RIV_FLAG:** A river flag designates polygons as river segments.

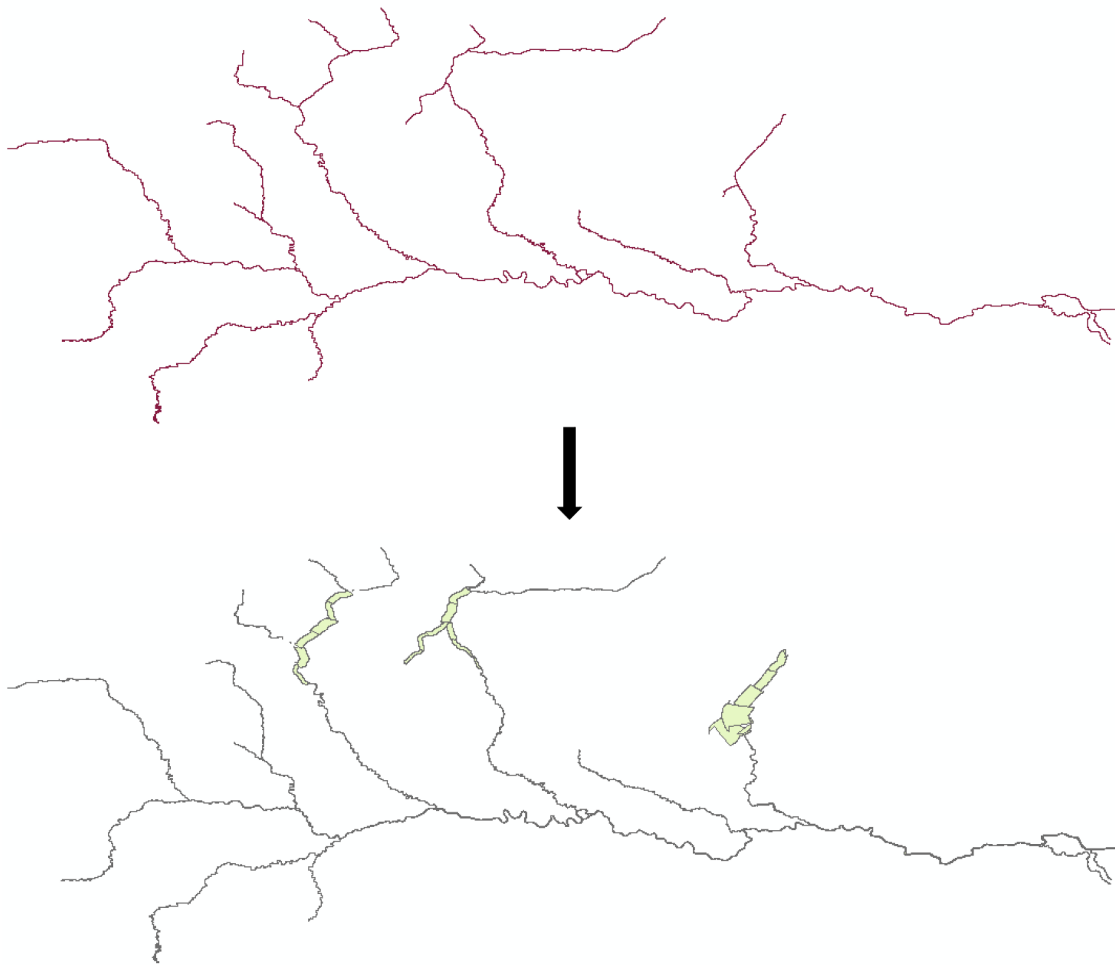


Figure 3.10: Transformation from polylines to polygons in the Po River in January 2000

- **id:** This is an identification field.

This operation can be done by opening the attribute table of the polygon shapefile and adding the above-mentioned attributes (see Figure 3.11). The "HEIGHT" attribute is specified as the type "double," while the other two are "integer" type (see Figure 3.12).

Afterwards, values were assigned to these attributes through the field calculator (see Figure 3.13). Input altimetry data was duplicated to "HEIGHT" and saved in the field "wse" (see Figure 3.14). "RIV_FLAG" is set to 1 for river segments and 0 for lakes or reservoirs along the network. To distinguish, the shapefile attribute "lakeflag" can be used, which is a water body identifier for each reach (see Table 3.3). The "lakeflag" attribute allows a simple Python code to be written in the field calculator to identify the composition of water bodies for "RIV_FLAG" (see

reach_len	n_nodes	wse	wse_var
10787.853335	54	6.640813	2.465317
19662.185722	98	9.944157	2.459238
12938.718634	65	10.543656	2.46878
10505.810595	53	11.44092	2.468302
8133.742722	41	12.440259	2.472027
15591.273752	78	-0.156322	2.463915
13401.510699	67	0.943572	2.464254
		1.940626	2.468095
		2.0416	2.467365
		4.838973	2.489658
		6.341193	2.464248
415.85009	2	6.640687	2.465673
12626.822412	63	43.749826	0
12408.196366	62	56.752393	0
13037.013123	65	10.495556	0.000057
19189.072141	96	192.238604	0
11165.850342	56	193.598757	0
15573.677366	78	68.004324	0.007372
9168.266088	46	193.598757	0
10518.110983	53	103.057515	0.03349
10768.011616	54	103.057515	0.03349
10744.46668	54	193.598757	0
10767.732816	54	193.598757	0
10334.869463	52	-1.46366	2.473066
9834.78574	49	-1.463707	2.47332
9859.881812	49	-1.063613	2.472813
14031.245829	70	-0.66366	2.473066
195.187532	1	12.540218	2.475699
210.452911	1	103.057515	0.03349
163.634926	1	63.299999	0
30 Polygon	21406100566	215.802481	1 -1.46366 2.473066
31 Polygon	21406200161	10622.808409	53 22.57181 0.176091
32 Polygon	21406200171	18493.318577	92 25.773791 0.177771
33 Polygon	21406200184	392.45836	2 29.327247 2.191891
34 Polygon	21406200191	9257.362827	46 30.56493 0.177135
35 Polygon	21406200201	9316.638519	47 34.27255 0.018415
36 Polygon	21406200211	9312.294402	47 36.271414 0.175626
37 Polygon	21406200221	11463.276461	57 39.372809 0.17568

Figure 3.11: Add a field in the attribute table

Figure 3.12: Field types. "HEIGHT" is the height attribute, which specifies the initial WSE and is in the arithmetic type "double". "RIV_FLAG" designates polygons as river segments in "short integer". "id" is an identification field with the type "short integer"

Figure 3.15). The attribute "id" also utilizes the values in "FID".

Afterwards we can assign values to these attribute through the field calculator (see figure 3.13). For "HEIGHT" we can directly copy the input altimetry data, which is saved in the field "wse" (see figure 3.14). "RIV_FLAG" should be set to

1 for river segments, but 0 for an lakes or reservoirs along the network. To distinguish we can refer to the attribute "lakeflag" in the shapefile, which is a water body identifier for each reach (see Table 3.3). According to it we can write a simple python code in field calculator to identify the composition of water bodies for "RIV_FLAG" (see figure 3.15). And the attribute "id" also directly utilizes the values in "FID".

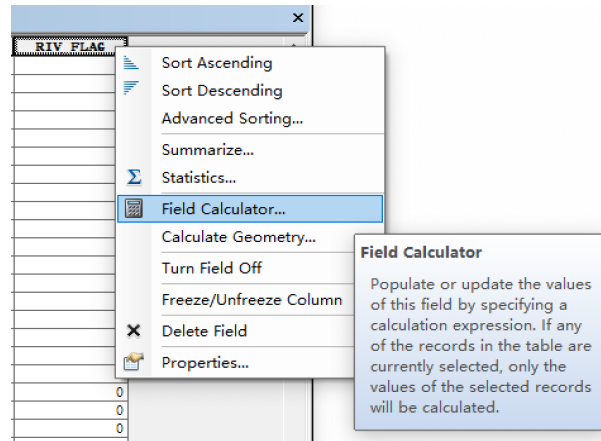


Figure 3.13: Find the field calculator

Identification number	Type
0	river
1	lake/reservoir
2	tidally influenced river
3	canal

Table 3.3: Water body identifier

3.3.3 Downloading and setting up the simulator

The simulator can be run on Windows or Linux/Mac OS. The following approach is based on Linux/Mac OS.

The SWOT Hydrology Toolbox and RiverObs can be downloaded from GitHub. Figure 3.16 shows the steps to obtain the repository of them. To run the simulator, a conda environment must be created; this can be done by visiting the Anaconda website and downloading the product "Anaconda3".

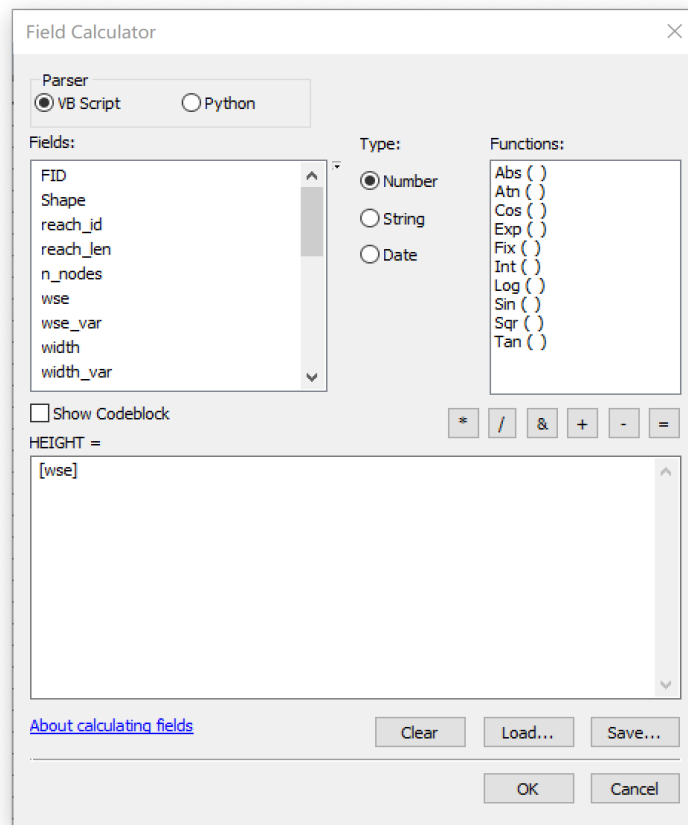


Figure 3.14: Field calculation for "HEIGHT" attribute

After setting up the conda environment, the file "environment.yml" in the SWOT Hydrology Toolbox lists the dependencies for the execution of the simulator. Figure 3.17 indicates the commands to control the environment.

If the conda environment is activated, Python must be able to find all the modules needed for the simulator (see Figure 3.18).

3.3.4 Running the simulator

Once the simulation was ready for implementation, a separate working directory was established for each observation time (i.e., per month). The input in time series was consequentially allocated to the appropriate directory based on month.

After this, the configuration files in each directory were modified. Finally, commands were issued sequentially in the terminal to create the output. The detailed process that followed is described below using January 2002 as an example. Other

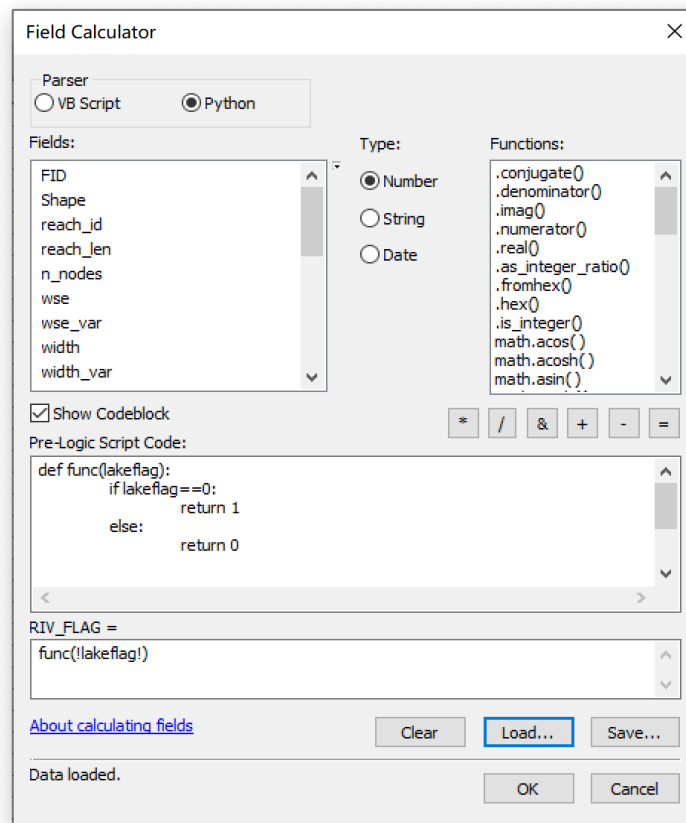


Figure 3.15: Field calculation for attribute "RIV_FLAG"

1. Clone `swot_hydrology_toolbox` repo

```
% git clone https://github.com/cnes/swot-hydrology-toolbox.git
```

The repository `swot_hydrology_toolbox` should be assigned to the `SWOT_HYDROLOGY_TOOLBOX` variable.

```
% export SWOT_HYDROLOGY_TOOLBOX=your_installation_path/swot-hydrology-toolbox
```

2. Clone `RiverObs` repo

```
% git clone https://github.com/SWOTAlgorithms/RiverObs.git
```

The repository `RiverObs` should be assigned to the `RIVEROBS` variable.

```
% export RIVEROBS=your_installation_path/RiverObs
```

Figure 3.16: Obtaining the repository of the SWOT Hydrology Toolbox and RiverObs (Elmer 2020)

months were approached in a similar manner.

To create a conda environment, execute

```
cd $SWOT_HYDROLOGY_TOOLBOX
conda env create -f environment.yml
```

To activate this environment, if the first option was used, type

```
conda activate swot-env
```

To deactivate this environment, type

```
conda deactivate
```

Figure 3.17: Setting up the conda environment (Elmer 2020)

```
export PYTHONPATH=$SWOT_HYDROLOGY_TOOLBOX/processing/src/:$RIVEROBS/src:$PYTHONPATH
```

Figure 3.18: Setting the "PYTHONPATH" variables (Elmer 2020)

Setting up the working directory

The working directory of the simulator is in the "/test" folder in the SWOT Hydrology Toolbox. An original path "/Po_2002_01" was set up. This folder already contains example datasets showing how to configure the data, run the simulation, and process the results. There are three sub-folders under the main path: "/data", "/output," and "/rdf." These three folders were also created under "/Po_2002_01" for the following purposes:

- **"/data"** saves the input polygon shapefile, and the corresponding shapefile in January 2002 was moved to it.
- **"/output"** can store all output files, or all products of the simulator. Two folders needed to be established in "/output": "/orbit" and "/simu." "/orbit" contains all orbit files. The cycle number, mission time, and so on along with the pixel cloud product were created in "/simu." Note that the river product was located in "/river," which was also under "/output," and the path was automatically generated by running RiverObs.
- **"/rdf"** includes the three configuration files "parameter_orbit.rdf," "parameter_sisimp.rdf," and "parameter_river.rdf." These files played different roles as control input in the simulation and are detailed in the next sub-section. The rdf files in the example case "/river_and_lake" under "/test" were copied with appropriate modifications.

Setting up your configuration files

First, "parameter_orbit.rdf" was opened in the text editor. This file sets parameters describing the SWOT orbit, which was used to determine the observation times for the Po River. The mission start time was specified as April 1, 2001; this was used to determine SWOT overpass times. The study area bounding box defines the area where the Po River is located. The simulation start and end time corresponded to January 2002. The simulation time for other periods was adjusted according to the simulation month. Figure 3.19 exhibits the above parameters. Other parameters can be modified as needed, although the defaults will likely work for most applications.

```

!== Mission specific parameters ==
Mission name = SWOT
Mission start time = 2001-04-01

!== Orbit parameters ==

!Directory of theoretical orbit files
Orbit repository = $SWOT_HYDROLOGY_TOOLBOX/select_orbit_cnes/
swot_science_orbit_june_2015

!Studied area bounding box
DEM south latitude (deg) = 44
DEM north latitude (deg) = 46.5
DEM west longitude (deg) = 7
DEM east longitude (deg) = 12.5

!PixC mapping parameters
Azimuth spacing (m) = 21.875000
Swath width (m) = 120000.000000
NR cross track (m) = 10000.000

!== Pass plan parameters ==
passplan = yes
simulation_start_time = 2002-01-01
simulation_stop_time = 2002-01-31

!== Output parameters ==
GDEM Orbit prefix = test.gdem_orbit

```

Figure 3.19: Modify "parameter_orbit.rdf" for January 2002

"parameter_sisimp.rdf" was then specified. This file contains the control parameters to generate pixel cloud products.

The directory for orbit files, polygon shapefiles, and output were set to the corresponding paths under "/Po_2002_01" (see Figure 3.20).

Then, orbit files were chosen for processing. There are three options (see Figure 3.21). In this simulation, "Multiple orbit = passplan" was chosen so that the orbit files would be processed according to the "passplan.txt" file, which is located in


```

Run directory for orbits = $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01/
output/orbit
Shapefile path = $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01/data/
Po_2002_01_Buffer
Output directory = $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01/output/
simu

```

Figure 3.20: Modify directories in "parameter_sisimp.rdf" for January 2002

"/orbit" and creates orbit files.

```

!### Orbit parameters
!3 options =
!Multiple orbit = yes (default) => all orbit files in Orbit
directory will be processed
!Multiple orbit = no => set the Orbit to a correct number found in
the Orbit directory; only this orbit file will be processed
!Multiple orbit = passplan => orbit files will be processed
according to passplan.txt file (generated if passplan = yes in
select_orbit.rdf)
Multiple orbit = passplan
!Orbit = 1
!Cycle number = 6

```

Figure 3.21: Modify orbit parameters in "parameter_sisimp.rdf"

Next, the height parameters were specified. There are also three options (see Figure 3.22). In this case, height model option 2 was used. Height is given as the attribute "HEIGHT" in the input shapefile, which matches "Height shp attribute name" in this option.

```

!== Height ==

!### Option 1 - Constant height model, uniform over each water body;
height varies sinusoidally with time
!Specific option: [Height model A = 0.] & [!Height model] => no
height applied, height in output onlyt contains errors
!### Constant height model parameters (same height for each water
body)
!Constant height model A = 0. !=0 to disable
!Constant height model t0 = 47076
!Constant height model period (days) = 365.25
!### Complex 2D height model parameters (2D variations added for
lakes > [Height model min area])
!Height model = polynomial !=polynomial or gaussian; if disabled,
only constant height model
!Height 2d model min area = 100. !(ha) min area of water bodies
on which to add complex 2D height model (default=100.)
!Height 2d model stdv = 1. !stdv for gaussian model (ie
Height model = gaussian)

!### Option 2 - Height is given from a specific attribute in the
shapefile of water bodies
Height model = reference_height
Height shp attribute name = HEIGHT !Name of the attribute
(default=HEIGHT)

!### Option 3 - Height is given in a NetCDF file
!Height model = reference_file
!True height file = $SWOT_HYDROLOGY_TOOLBOX/test/river_and_lake/
data/true_height_model_river_and_lake.nc

```

Figure 3.22: Modify height parameters in "parameter_sisimp.rdf"

Finally, to quicken the simulation, intermediate output files such as pixel cloud shapefiles that will not be utilized later can be excluded from the exported file

(see Figure 3.23).

```

#### Files in output
Create shapefile = no !Produce output files also as shapefiles
Create dummy pixc vec river file = no !Produce L2_HR_PIXCVecRiver
product associated to PixC files

```

Figure 3.23: Modify output products in "parameter_sisimp.rdf"

The last rdf file to be edited was "parameter_river.rdf." This file contributes to converting the pixel cloud product into a river product. The sole modification needed was updating the "reach_db_path," which references the input NetCDF file of this month (see Figure 3.24).

```

width_db_file          (-) = None
reach_db_path          (-) = $SWOT_HYDROLOGY_TOOLBOX/test/
eu_apriori_rivers_v05_2002_01.nc
l2_file                (-) = 'REPLACE_ME'
fout_reach             (-) = 'REPLACE_ME'
fout_node              (-) = 'REPLACE_ME'
fout_index             (-) = 'REPLACE_ME'
lonmin                 (-) = 'REPLACE_ME'
lonmax                 (-) = 'REPLACE_ME'
latmin                 (-) = 'REPLACE_ME'
latmax                 (-) = 'REPLACE_ME'
bounding_box           (-) = 'lonmin, latmin, lonmax, latmax'
lat_kwd                (-) = 'latitude'
lon_kwd                (-) = 'longitude'
class_kwd              (-) = 'classification'
height_kwd             (-) = 'height'
xtrack_kwd             (-) = 'cross_track'
class_list             (-) = [2, 3, 4, 22, 23, 24]
fractional_inundation_kwd (-) = 'water_frac'
use_fractional_inundation (-) = [True, True, False, False, False,
False]
use_segmentation       (-) = [False, True, True, False, True,
True]

```

Figure 3.24: Modify "parameter_river.rdf" in January 2002

Creating the orbit file

When commands were issued utilizing "parameter_orbit.rdf" (see Figure 3.25), a "passplan.txt" file and several other files were created under "/output/orbit" with information about orbit, cycle number, and mission time (see Figure 3.26). These files were selected by "parameter_sisimp.rdf" in the orbit parameters (mentioned above) to process orbit and cycle of interest in the following steps.

```

>>cd $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01
>>python $SWOT_HYDROLOGY_TOOLBOX/select_orbit_cnes/select_orbit_cnes.py rdf/parameter_orbit.rdf output/orbit

```

Figure 3.25: Command to create orbit file

```

# Mission start: 2001-04-01 00:00:00
# Simulation start: 2002-01-01 00:00:00
# Simulation stop: 2002-01-31 00:00:00
#
#run      cycle orbit MissionTime year DayOfYear      date      time
c014_t335 14 335 24468380 2002 9.19884 2002-01-09 04:46:20
c014_t251 14 251 24209074 2002 6.19762 2002-01-06 04:44:34
c014_t180 14 180 23988312 2002 3.64250 2002-01-03 15:25:12
c014_t264 14 264 24247617 2002 6.64372 2002-01-06 15:26:57
c014_t486 14 486 24932876 2002 14.57495 2002-01-14 13:47:56
c014_t514 14 514 25019307 2002 15.57531 2002-01-15 13:48:27
c014_t279 14 279 24295504 2002 7.19796 2002-01-07 04:45:04
c014_t542 14 542 25105745 2002 16.57575 2002-01-16 13:49:05
c014_t208 14 208 24074743 2002 4.64286 2002-01-04 15:25:43
c014_t557 14 557 25153637 2002 17.13006 2002-01-17 03:07:17
c014_t307 14 307 24381938 2002 8.19836 2002-01-08 04:45:38
c014_t236 14 236 24161176 2002 5.64324 2002-01-05 15:26:16
c015_t335 15 335 26271077 2002 30.06339 2002-01-30 01:31:17
c015_t251 15 251 26011771 2002 27.06216 2002-01-27 01:29:31
c015_t180 15 180 25791009 2002 24.50705 2002-01-24 12:10:09
c015_t264 15 264 26050314 2002 27.50826 2002-01-27 12:11:54
c015_t029 15 29 25326507 2002 19.13087 2002-01-19 03:08:27
c015_t279 15 279 26098201 2002 28.06251 2002-01-28 01:30:01
c015_t208 15 208 25877440 2002 25.50741 2002-01-25 12:10:40
c015_t307 15 307 26184635 2002 29.06291 2002-01-29 01:30:35
c015_t236 15 236 25963873 2002 26.50779 2002-01-26 12:11:13
c015_t001 15 1 25240068 2002 18.13042 2002-01-18 03:07:48

```

Figure 3.26: Orbit file "passplan.txt" in January 2002

Running the simulator

With the support of "parameter_sisimp.rdf," the pixel cloud product was created by issuing the command in Figure 3.27. Figure 3.28 presents the naming conventions of the pixel cloud product.

```

>>cd $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01
>>python $SWOT_HYDROLOGY_TOOLBOX/sisimp/proc_sisimp.py rdf/parameter_sisimp.rdf

```

Figure 3.27: Command to create the pixel cloud product

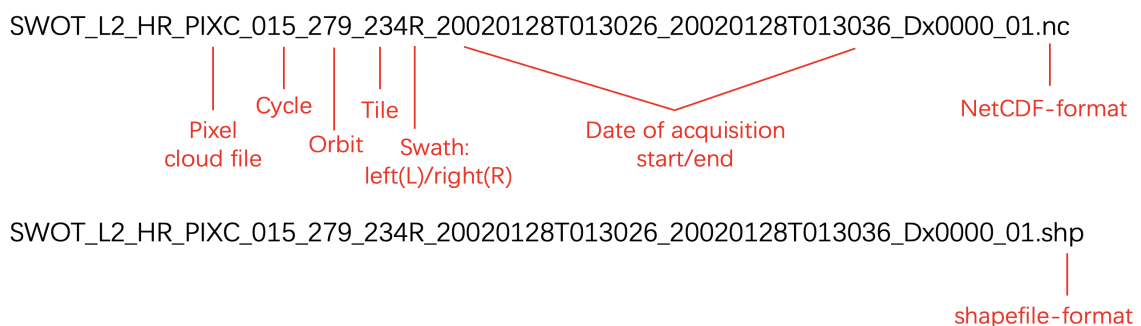


Figure 3.28: Naming convention of pixel cloud product

Further process to create the river product was completed using the "parameter_river.rdf" and the RiverObs. The RiverObs was executed through the script in the SWOT Hydrology Toolbox. Therefore, it was sufficient to issue commands for script implementation under the toolbox (see Figure 3.29). The three product types were stored under "/output/river/rivertile." Figure 3.30 indicates the naming conventions; these are similar to the pixel cloud product, but the prefix was

changed from "PixC" to "RiverTile" and the suffix of the shapefiles added "_reach" and "_node" to distinguish between the two output products. In this study, the reach product was the target product, which is collected and processed in the next chapter.

```
>>cd $SWOT_HYDROLOGY_TOOLBOX/test/Po_2002_01
>>python $SWOT_HYDROLOGY_TOOLBOX/scripts/l2pixc_to_rivertile.py output/simu output/river rdf/parameter_river.rdf
```

Figure 3.29: Command to create the river product

➤ **NetCDF-format:**

SWOT_L2_HR_RiverTile_015_279_234R_20020128T013026_20020128T013036_Dx0000_01.nc

➤ **Shapefile-format:**

• **Node product:**

SWOT_L2_HR_RiverTile_015_279_234R_20020128T013026_20020128T013036_Dx0000_01_nodes.shp

• **Reach product:**

SWOT_L2_HR_RiverTile_015_279_234R_20020128T013026_20020128T013036_Dx0000_01_reaches.shp

Figure 3.30: Naming convention of the river product

Chapter 4

Methodology

4.1 Introduction to river discharge estimation

In this study, Manning's equation (Manning et al. 1891) was used to estimate river discharge:

$$Q = \frac{1}{n}(\bar{A} + A')^{5/3}w^{-2/3}s^{1/2} \quad (4.1)$$

where n is the reach-averaged Manning's roughness coefficient, \bar{A} is the reach-averaged cross-sectional area at the time of the lowest observed river elevation, A' is the change in cross-sectional area with respect to \bar{A} , w is the observed channel top width, and s is the slope of the water surface elevation.

n and \bar{A} are known as "flow law parameters" and are normally provided by SWOT. However, these data are based on the first year of mission operations and were not defined yet. Thus, before the estimation of river discharge, an approach was needed to specify the flow law parameters for each reach.

For more precise description of n , it was modeled based on Durand et al. (2021):

$$n = n_b \left(1 + \frac{5}{6} \left[\frac{w\sigma_z}{\bar{A} + A'} \right]^2 \right) \quad (4.2)$$

where n_b is the roughness coefficient at a high flow, such as bankfull, and σ_z is the spatial variation of river bed elevation. \bar{A} , n_b and σ_z denote time-invariant parameters that must be estimated for each reach.

In the SWOT mission, A' , w and s were observed for each overpass. Until now, the SWOT Hydrology Simulator can provide simulated w , s and standard deviations in time series. A' can be calculated via integration over a function relating

water surface elevation h to width $w(h)$, which is a simulation implemented by the Matlab code "SWOTAprimeCalcs" (Durand et al. 2014).

4.2 Flowcharts

This section summarizes the process of this thesis through flowcharts. Figure 4.1 shows the whole process of simulation and post-processing until the posterior discharges $\hat{Q}_{est.}$ are calculated.

Figure 4.2 describes the details of two rounds of Kalman Filter. According to the Manning's equation (see Equation 4.1) the flow law parameters must be known before the computation of discharge. Thus, the first round of Kalman Filter is set on the basis of mass conservation to compute the prior discharge $Q_{est.}(t)$, through which the flow law parameters could be estimated by the interior-point optimization with inequality constraints. After that, using the flow law parameters discharge could be recalculated, which will be the observations of the second round of Kalman Filter for a more accurate estimation of posterior discharges $\hat{Q}_{est.}$.

Figure 4.3 depicts the concrete steps of interior-point optimization with inequality constraints.

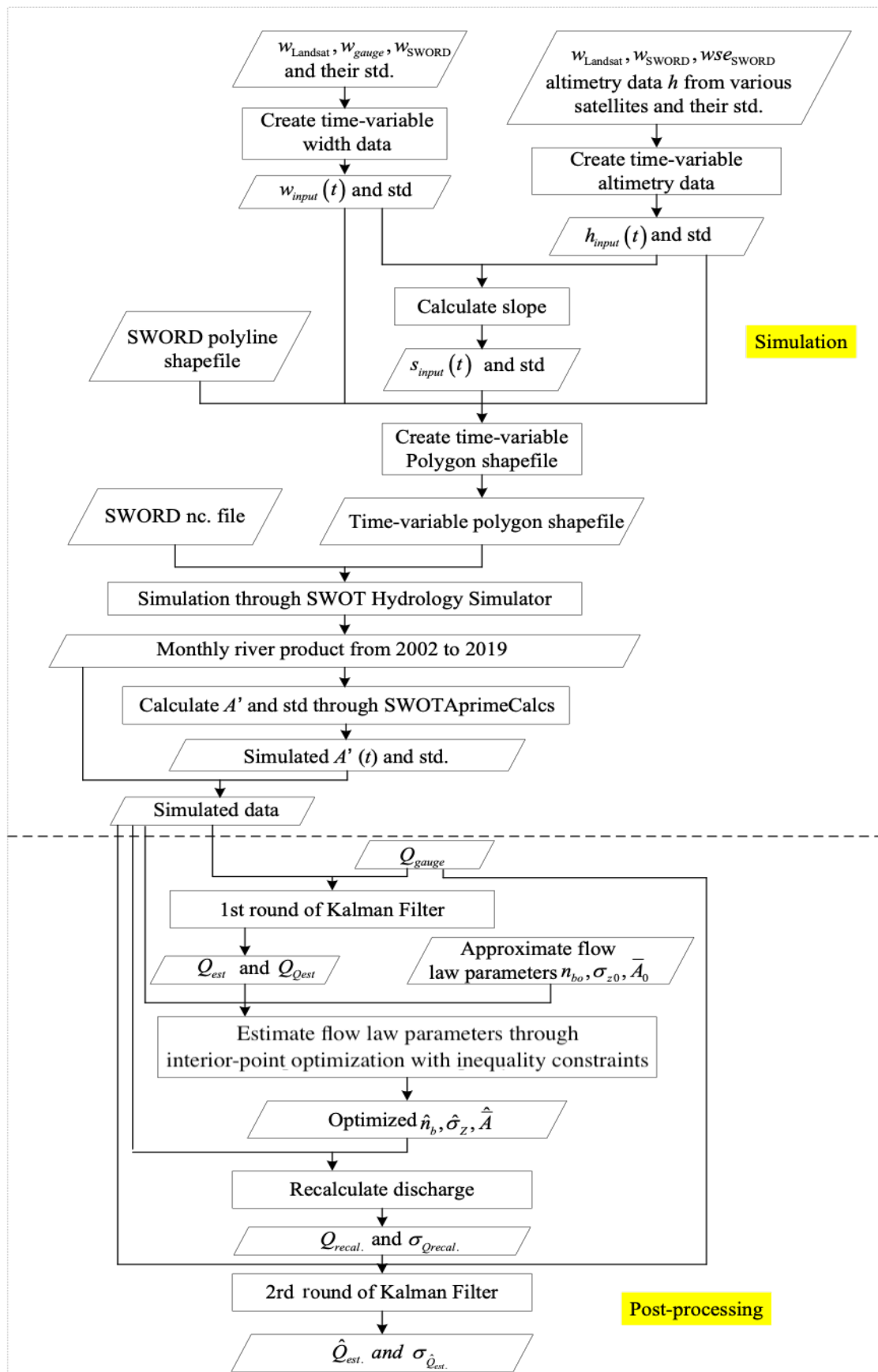


Figure 4.1: Whole process of this thesis

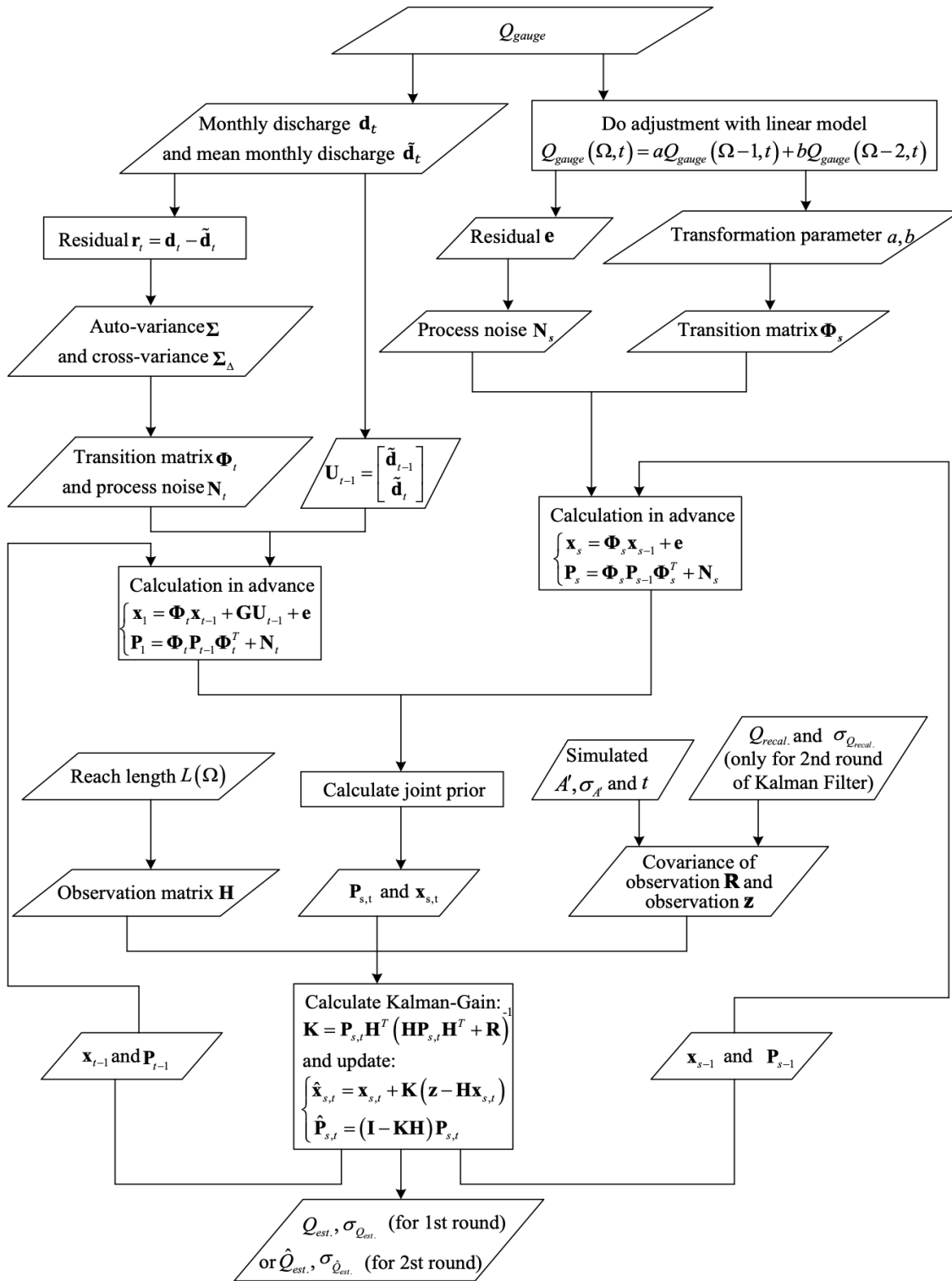


Figure 4.2: Kalman Filter steps

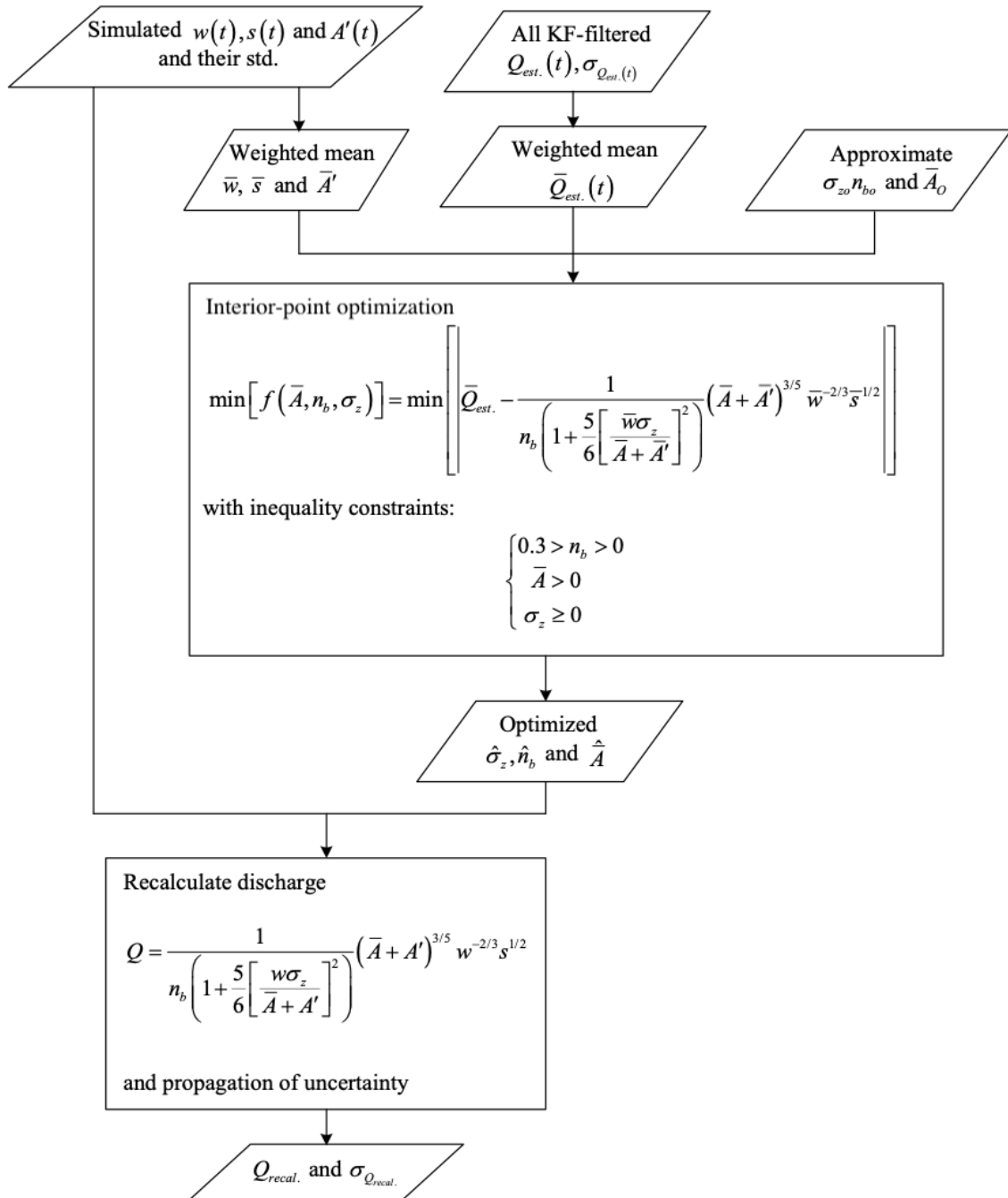


Figure 4.3: Interior-point optimization with inequality constraints

4.3 General steps

In general, four steps could be summarized from the flowcharts above to obtain time series data for monthly river discharge:

1. The required data for the subsequent process: $w(t)$, $s(t)$, $h(t)$ with their standard deviations was extracted from the simulated river product. Over time, monthly time series data was collected for each reach.
2. $A'(t)$ and $\sigma_{A'}(t)$ were computed through "SWOTAprimeCalcs" with simulated $w(t)$, $\sigma_w(t)$, h and $\sigma_h(t)$ as the input.
3. The first Kalman Filter estimated the prior discharge $Q_{est.}(t)$ for each reach in the mainstream. Flow law parameters were calculated through interior-point optimization with inequality constraints. The initial values of the Kalman Filter were taken from in situ data. The process model includes a model in the space and time domain. Before updating via the observation model, the joint prior of the two process models was computed. The mass conservation condition served as an observation equation.
4. Discharge $Q_{recal.}(t)$ was recalculated through the estimated flow law parameters. The second Kalman Filter was implemented using the results of the first Kalman Filter as starting values. The space and time process model were also the same as the first Kalman Filter. The observation model contains the mass conservation condition and $Q_{recal.}(t)$.

4.4 Collecting data from the simulator

Section 3.3.4 notes that a separate working directory was created for each month featuring output files in NetCDF format and divided by orbit, cycle number, and mission time under "/output/river/rivertile." Each file contained various attributes, and there were many invalid data points that were assigned the invalid value of "-9999999999" or "no_data" (see Figure 4.4). Hence, it was necessary to extract required and valid information from these files.

Table 4.1 lists the attributes that were collected from each output file. When all attributes of an object were assigned valid data, they were extracted from the river

reach_id	node_id	time	time_tai	time_str	lat	lon	lat_u	lon_u
21406600463	2140660046...	-999999999...	-999999999...	no_data	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	-999999999...	-999999999...	no_data	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	70991167.312	70991135.312	2002-04-01T...	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	-999999999...	-999999999...	no_data	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	-999999999...	-999999999...	no_data	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	-999999999...	-999999999...	no_data	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	70991167.307	70991135.307	2002-04-01T...	-999999999...	-999999999...	-999999999...	-999999999...
21406600463	2140660046...	70991167.157	70991135.157	2002-04-01T...	45.92391211	8.5677002	0.016991224...	0.090799596...
21406600463	2140660046...	70991167.094	70991135.094	2002-04-01T...	45.92044284	8.5672171	0.429967949...	2.297789959...
21406600463	2140660046...	70991167.056	70991135.056	2002-04-01T...	45.91854254	8.5654417	0.078044818...	0.41707207372
21406600463	2140660046...	70991167.016	70991135.016	2002-04-01T...	45.91614380	8.5653612	0.123467467...	0.659881030...
21406600463	2140660046...	70991166.984	70991134.984	2002-04-01T...	45.91404738	8.5664646	0.455428778...	2.43502811131
21406600463	2140660046...	70991166.961	70991134.961	2002-04-01T...	45.91251830	8.5661492	0.032409389...	0.17317848213

Figure 4.4: A part of the output data in the river product

product.

Notation	Attribute Name	Unit
reach_id	Reach id	[-]
time	Time (UTC)	[s]
WSE	Water surface elevation	[m]
WSE_u	Uncertainty in WSE	[m]
slope	Slope	[m/m]
slope_u	Uncertainty in slope	[m/m]
width	Reach width	[m]
width_u	Uncertainty in width	[m]

Table 4.1: Extracted attributes from the river product

After this, the data was organized into a chronological order for each river reach. Additionally, the monthly weighted mean of slope, WSE, and width were calculated for use in the Kalman Filter. Figure 4.5 compares simulated and Landsat-observed widths in gauge stations. Figure 4.6 compares the simulated WSE with the WSE observed by various satellite missions in gauge stations. As shown, the input data are highly correlated to the simulated data. Table 4.2 shows the correlations in each gauge station. We can then compute the *residual* between the simulated and Landsat width:

$$residual = width_{simu} - width_{Landsat} \quad (4.3)$$

The histogram of the *residual* could also be built and be compared with the normal distribution fit line (see Figure 4.7). We can see that the bars of residual approximately fit the normal distribution and the noise between the simulated and Landsat width is white noise.

There are always offsets for width because the simulator uses polygon extent to calculate river width. The river product is created based on the pixel cloud product, which covers the area classified as water and the buffer zone around these water zones (see Section 3.2.1). Each pixel has its own area. If a pixel happens to be at the edge of the river, it is counted as a water pixel in the RiverObs process, which offsets width because of its pixel area.

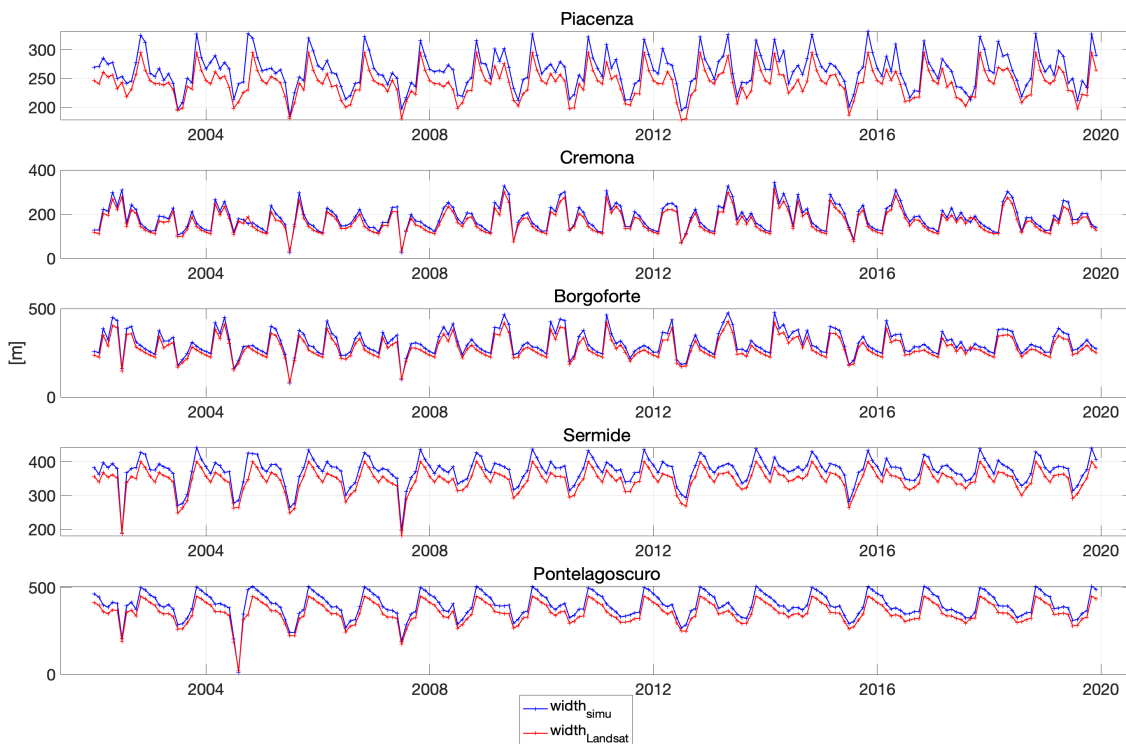


Figure 4.5: Simulated vs. Landsat-observed width

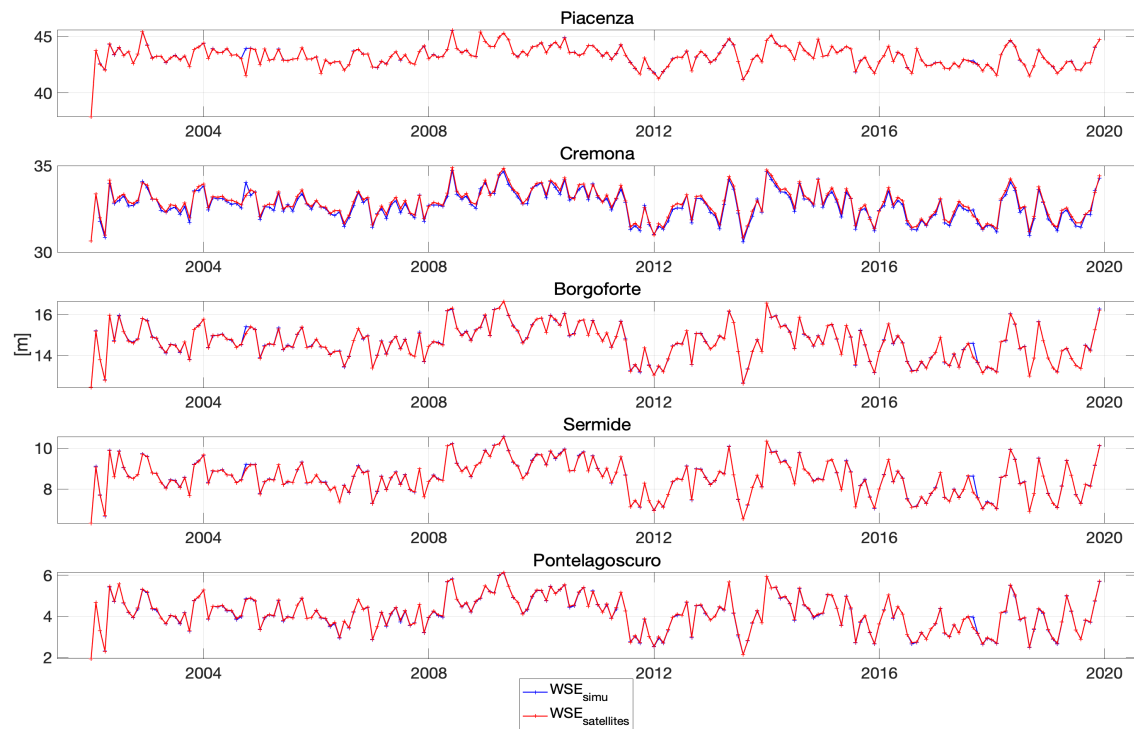


Figure 4.6: Simulated vs. satellites-observed WSE

	Corr. of width [-]	Corr. of WSE [-]
Piacenza	0.9525	0.9844
Cremona	0.9937	0.9902
Borgoforte	0.9967	0.9979
Sermide	0.9887	0.9977
Pontelagoscuro	0.9932	0.9988

Table 4.2: Correlation of input and simulated data in gauge stations

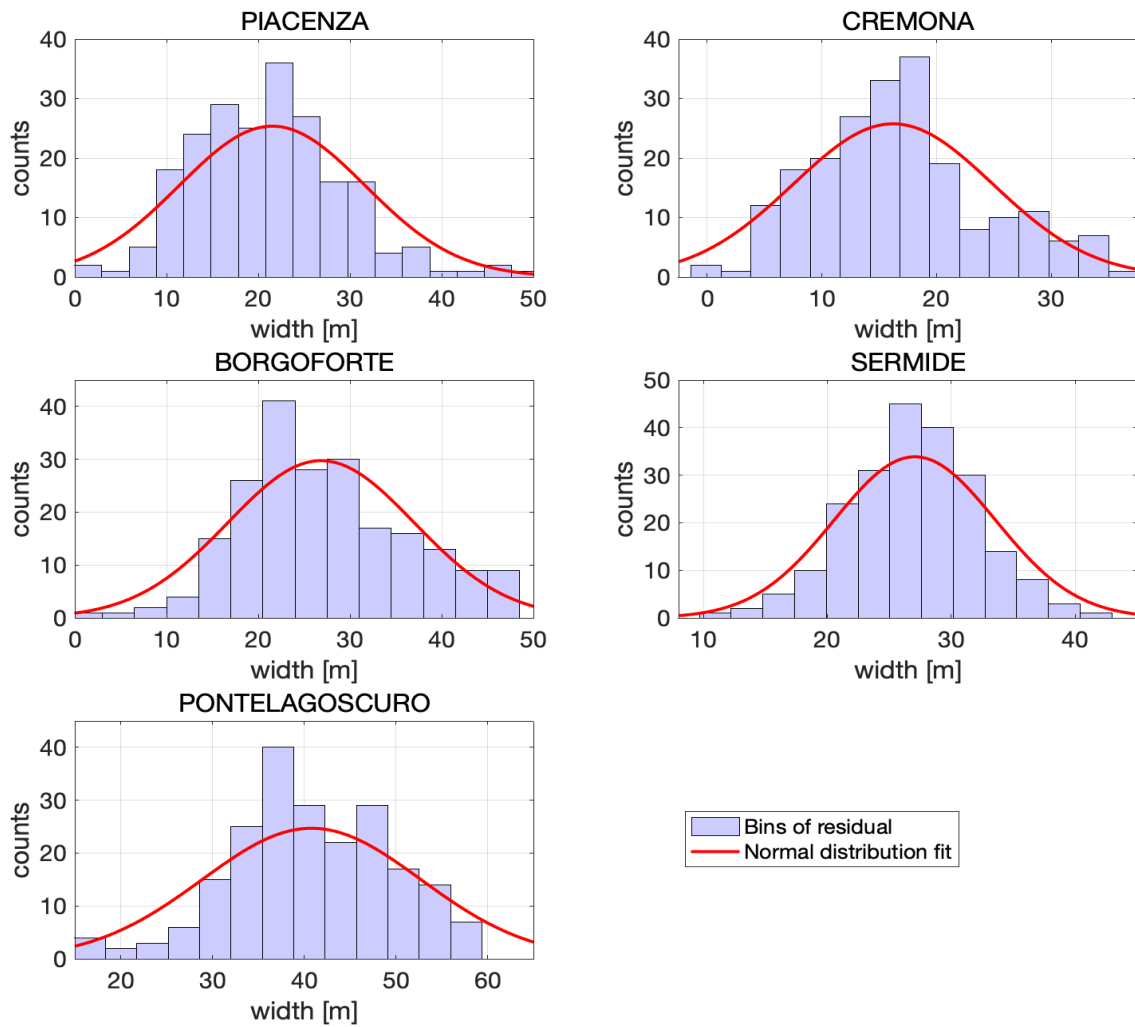


Figure 4.7: Histogram of residual between the simulated and Landsat width in gauge stations to verify if it is white noise

4.5 Computation of the changes in cross-sectional area

Changes in cross-sectional area A' are a SWOT data product. In the Manning equation for discharge algorithms (see Equation 4.1), the sum of A' and reach-averaged cross-sectional area \bar{A} determine the cross-sectional area A , which is of primary importance in river hydraulics. This section describes the estimation of A' and $\sigma_{A'}$ with the help of simulated width $w(t)$ and height $h(t)$ based on "SWOTAprimeCalcs".

4.5.1 Definition

The computation of the cross-sectional area can be written as (Durand et al. 2014):

$$A(h(t)) = \bar{A} + A'(h(t)) = \bar{A} + \epsilon_{\bar{A}} + \int_{\bar{h}}^{h(t)} f_w(h') dh' \quad (4.4)$$

where $h(t)$ is the observed WSE at time t , $\epsilon_{\bar{A}}$ is a slight mismatch that does not vary with time. As shown in Figure 4.8, A' is computed by integrating the WSE at \bar{A} (\bar{h}) to $h(t)$ through the fitting function $f_w(h')$. The challenges of this approach are finding an appropriate f_w based on height-width data and calculating A' using f_w . This were done "offline" by the research team and "online" as part of the SWOT processing chain.

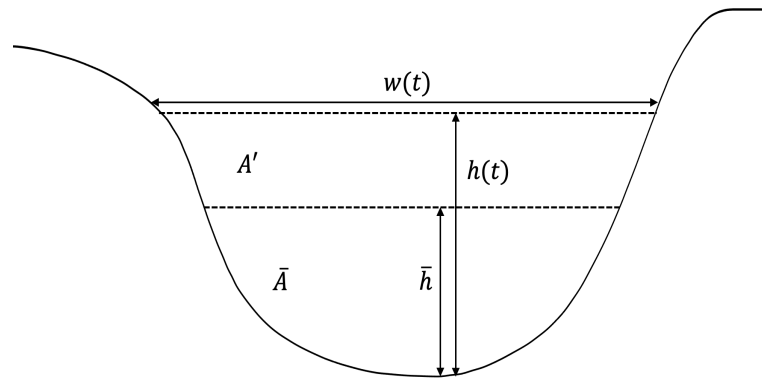


Figure 4.8: Composition of the cross-sectional area

4.5.2 Building the fitting function

To better express the height-width variation, the cross-section was divided into sub-domains. Empirically, three sub-domains are adequate for this estimation.

A continuous piecewise-linear function met these requirements, and the boundary definition of the sub-domains was done via an optimization routine of minimizing square errors by choosing different sub-domain boundaries. Therefore, an errors-in-variables model (Carroll et al. 2006) was utilized with simultaneous consideration of errors in height and width data (Durand 2018):

$$\min \sum_{s=1}^{n_s} (\sigma_w^2 + p_{s,1}^2 \sigma_h^2)^{-1} \sum_j (p_{s,1} h_j + p_{s,2} - w_j)^2 \quad (4.5)$$

where n_s is the number of sub-domains, σ_w and σ_h are the standard deviations of width and height, $p_{s,1}$ and $p_{s,2}$ are the slopes and intercepts for each sub-domain, and h_j and w_j indicate the time-various height-width observations. Furthermore, this optimization problem requires a constraint:

$$p_{s,1} h_{b,s+1} + p_{s,2} = p_{s+1,1} h_{b,s+2} + p_{s+1,2} \quad (4.6)$$

where $h_{b,s}$ are the breakpoints or sub-domain boundaries defined in terms of their river height including start and end points. This constraint equation indicates that the fitting function f_w meets at breakpoints.

After the minimization, the fit parameters $p_{s,1}$ and $p_{s,2}$ and breakpoints $h_{b,s}$ of each reach were obtained. It is important to note that these parameters were not yet defined because they would be included in the SWORD based on the first year of mission operations. Thus, in this study, the calculation of fit parameters and breakpoints was based on simulated height-width observations. In "SWOTAprimeCalcs," the height and width data with their errors were input into the "FitData" function, which solved the optimization problem.

4.5.3 Calculating A'

For the computation of A' , the optimal height point needed be found. The errors-in-variables approach finds the optimal height point by combining height and width observations:

$$\hat{h}(t) = h(t) + (\sigma_w^2 + p_1^2 \sigma_h^2)^{-1} (p_1^2 \sigma_h^2) (w(t) - p_2 - p_1 h(t)) \quad (4.7)$$

With this information, the fitting functions can be integrated. If the optimal height point $\hat{h}(t)$ falls within the sub-domain d (Durand 2018),

$$A'(\hat{h}(t)) = \sum_{s=1}^{d-1} \int_{h_{b,s}}^{h_{b,s+1}} (p_{s,1}h + p_{s,2})dh + \int_{h_{b,d}}^{\hat{h}(t)} (p_{d,1}h + p_{d,2})dh - A'(\bar{h}) \quad (4.8)$$

where $A'(\bar{h})$ is the difference with the median cross-sectional area at \bar{h} , which can be calculated through:

$$A'(\bar{h}) = \sum_{s=1}^{d-1} \int_{h_{b,s}}^{h_{b,s+1}} (p_{s,1}h + p_{s,2})dh + \int_{h_{b,d}}^{\bar{h}} (p_{d,1}h + p_{d,2})dh \quad (4.9)$$

When a height observation falls outside all sub-domains, a simple trapezoidal approximation between the height point and the nearest sub-domain boundary can be used.

So for now, the computation of the error of A' is a very simple approach, assuming half of the error variance in the estimator comes from the height error, and it is approximated as a response of height error and width using the Durand et al. (2014) way of computing:

$$\sigma_{A'}(t) = \sqrt{2}w(t)\sigma_h(t) \quad (4.10)$$

Furthermore, there are several required data elements that will be considered for computation in the future, such as the width error, covariance matrix of the height and width observations, the sub-domain boundary values for WSE and width, and others.

In "SWOTAprimeCalcs," $A'(\bar{h})$ could initially be computed according to Equation 4.9 through the "CalculatedAEIV" function using the median of the height observations, the calculated fit parameters, and the breakpoints from the previous step. After this, $A'(\bar{h})$ could be inversely used as an input of the "CalculatedAEIV" function to calculate $A'(\hat{h})$ (see Equation 4.8). At the same time, $\sigma_{A'}(t)$ could be computed using the width and height error (see Equation 4.10). For further processes, the weighted mean of A' for each month must be computed. Figure 4.9 shows the monthly time series for A' with the $\sigma_{A'}$ of the gauge stations. For all gauge stations, $A'(t)$ varies around 0 m^2 . $\sigma_{A'}(t)$ has the largest value in Cremona because Cremona has the largest product of $w(t)$ and $\sigma_h(t)$.

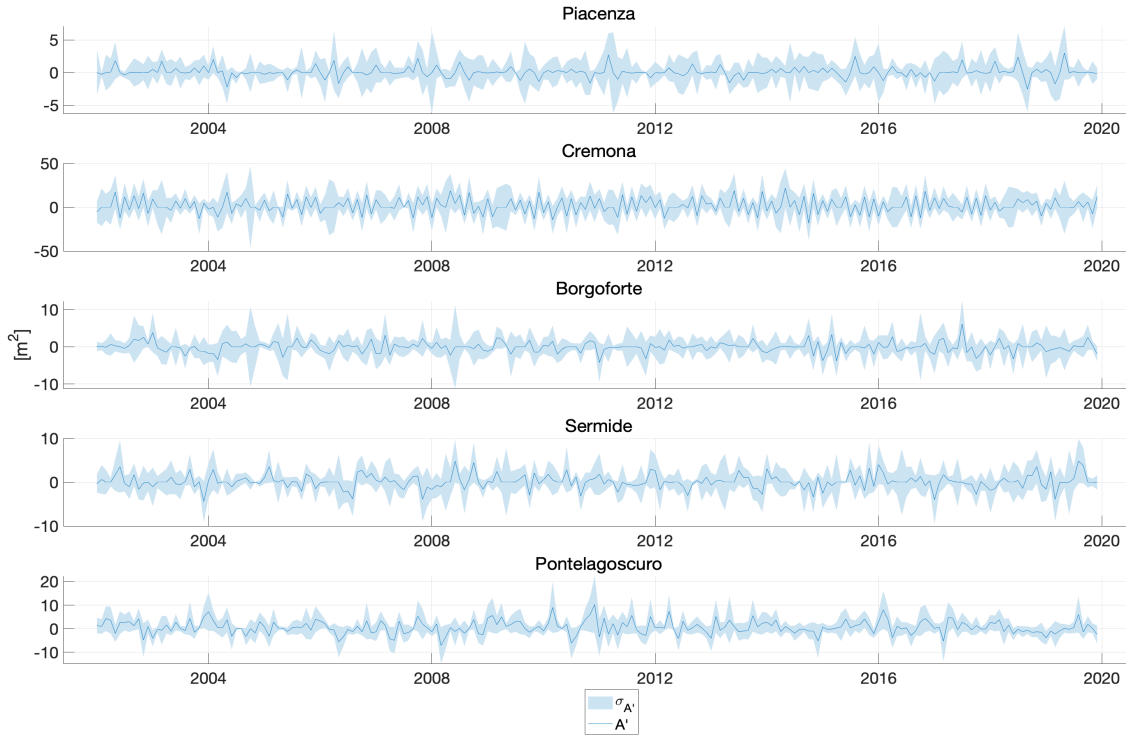


Figure 4.9: A' with $\sigma_{A'}$ in gauge stations

4.6 Estimation of flow law parameters

With these data, the first Kalman Filter was executed. The aim of this step was to derive the flow law parameters through the Kalman-Filtered a priori discharges in the mainstream. This section outlines how to build the process model, observation equation, estimate of a priori discharge, and the computation of flow law parameters.

4.6.1 Process model

First, the process model was set up using in situ data. As mentioned in Section 4.3, the process model was divided into the space and time domain. For each domain, the transition matrix and process noise must be established.

Space domain

To build the transition matrix Φ_s , the spatial transformation parameters a and b for each month must be estimated via adjustment of this linear model:

$$Q_{gauge}(\Omega, t) = aQ_{gauge}(\Omega - 1, t) + bQ_{gauge}(\Omega - 2, t) \quad (4.11)$$

where t indicates a month, Ω refers to a gauge station, and $(\Omega - 1)$ is the station before Ω . Hence, $Q_{gauge}(\Omega, t)$ indicates the in situ discharge of reach Ω in month t . Because the available in situ discharge was observed daily from 1995 to 2011, the mean value for each month in this period needed to be calculated.

To solve this model, the discharges of a given month were collected, and Equation 4.11 was extended as follows using data from January as an example:

$$\underbrace{\begin{bmatrix} Q_{gauge}(\Omega_3, Jan.1995) \\ Q_{gauge}(\Omega_4, Jan.1995) \\ Q_{gauge}(\Omega_5, Jan.1995) \\ Q_{gauge}(\Omega_3, Jan.1996) \\ \vdots \end{bmatrix}}_{\mathbf{y}_{m \times 1}} = \underbrace{\begin{bmatrix} Q_{gauge}(\Omega_2, Jan.1995) & Q_{gauge}(\Omega_1, Jan.1995) \\ Q_{gauge}(\Omega_3, Jan.1995) & Q_{gauge}(\Omega_2, Jan.1995) \\ Q_{gauge}(\Omega_4, Jan.1995) & Q_{gauge}(\Omega_3, Jan.1995) \\ Q_{gauge}(\Omega_2, Jan.1996) & Q_{gauge}(\Omega_1, Jan.1996) \\ \vdots & \vdots \end{bmatrix}}_{\mathbf{A}_{m \times 2}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{x}_{2 \times 1}} \quad (4.12)$$

This equation was solved through the following adjustment ($\mathbf{P} = \mathbf{I}$):

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \quad (4.13)$$

$$\mathbf{e} = \mathbf{y} - \mathbf{A} \hat{\mathbf{x}} \quad (4.14)$$

$$\sigma_0^2 = \frac{\mathbf{e}^T \mathbf{P} \mathbf{e}}{m - 2} \quad (4.15)$$

$$\hat{\Sigma}_{\mathbf{x}} = \sigma_0^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \quad (4.16)$$

$$\hat{\Sigma}_{\mathbf{y}} = \mathbf{A} \hat{\Sigma}_{\mathbf{x}} \mathbf{A}^T \quad (4.17)$$

$$(4.18)$$

The adjustment was executed separately for each month and yielded the monthly a , b and residuals \mathbf{e} . At this point, prediction of the state vector was possible:

$$\underbrace{\begin{bmatrix} Q_s(\Omega, t) \\ Q_s(\Omega - 1, t) \\ Q_s(\Omega, t - 1) \\ Q_s(\Omega - 1, t - 1) \end{bmatrix}}_{\hat{\mathbf{x}}_s} = \underbrace{\begin{bmatrix} a(t) & b(t) & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & a(t - 1) & b(t - 1) \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Phi_s} \underbrace{\begin{bmatrix} Q_s(\Omega - 1, t) \\ Q_s(\Omega - 2, t) \\ Q_s(\Omega - 1, t - 1) \\ Q_s(\Omega - 2, t - 1) \end{bmatrix}}_{\hat{\mathbf{x}}_{s-1}} \quad (4.19)$$

where Ω here denotes the reach and t still indicated the month. -1 refers to the previous reach or month. The corresponding prediction of the covariance matrix is:

$$\mathbf{P}_s = \Phi_s \mathbf{P}_{s-1} \Phi_s^T + \mathbf{N}_s \quad (4.20)$$

where \mathbf{P}_s and \mathbf{P}_{s-1} are the covariance matrix of state $\hat{\mathbf{x}}_s$ and $\hat{\mathbf{x}}_{s-1}$. \mathbf{N}_s is process noise given by residuals \mathbf{e} :

$$\mathbf{N}_s = \begin{bmatrix} var(\mathbf{e}) & 0 & 0 & 0 \\ 0 & var(\mathbf{e}) & 0 & 0 \\ 0 & 0 & var(\mathbf{e}) & 0 \\ 0 & 0 & 0 & var(\mathbf{e}) \end{bmatrix}$$

Notably, the starting values of the space domain process model for the first period (i.e., January 2002) were taken from the monthly discharge of two neighboring gauge stations, namely Cremona ($\Omega - 1$) and Borgoforte (Ω) in January ($t - 1$) and February (t) 2002. The corresponding covariance matrix of state was built using 10% of $\hat{\sigma}_y$ estimated in adjustment (see Equation 4.16).

Time domain

The creation of the transition matrix Φ_t was more complicated than that of Φ_s . The first residual of monthly discharge required calculating (Tourian et al. 2017)

$$\mathbf{r}_t = \mathbf{d}_t - \tilde{\mathbf{d}}_t \quad (4.21)$$

with \mathbf{d}_t as the in-situ monthly time series discharge and $\tilde{\mathbf{d}}_t$ as the corresponding in situ mean monthly discharge, which represents the cyclostationary behavior of monthly river discharge. Because there are $17 \cdot 12 = 204$ months from 1995 to 2011 and five gauge stations, the size of \mathbf{r}_t is 204×5 . Hence, Φ_t needed to be found to build the linear process model

$$\mathbf{r}_t = \Phi_t \mathbf{r}_{t-1} + \mathbf{e} \quad (4.22)$$

with minimized process error:

$$tr[cov\{\mathbf{e}\}] = min \quad (4.23)$$

The outer product of the error vector can be calculated first:

$$\begin{aligned} \mathbf{e}\mathbf{e}^T &= (\Phi_t \mathbf{r}_{t-1} - \mathbf{r}_t)(\Phi_t \mathbf{r}_{t-1} - \mathbf{r}_t)^T \\ &= \Phi_t \mathbf{r}_{t-1} \mathbf{r}_{t-1}^T \Phi_t^T - \mathbf{r}_t \mathbf{r}_{t-1}^T - \Phi_t \mathbf{r}_{t-1} \mathbf{r}_t^T + \mathbf{r}_t \mathbf{r}_t^T \end{aligned} \quad (4.24)$$

The auto-covariance and cross-covariance can be defined as:

$$\begin{aligned} \text{Auto-covariance: } \quad \Sigma &= cov\{\mathbf{r}_t\} = E\{\mathbf{r}_t \mathbf{r}_t^T\} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t^T \mathbf{r}_t \\ \text{Cross-covariance: } \quad \Sigma_{\Delta} &= cov\{\mathbf{r}_t, \mathbf{r}_{t-1}\} = E\{\mathbf{r}_t \mathbf{r}_{t-1}^T\} = \frac{1}{T-1} \sum_{t=2}^T \mathbf{r}_{t-1}^T \mathbf{r}_t \end{aligned} \quad (4.25)$$

where T refers to the total of 204 months.

Then, the error covariance matrix of the process model is as follows:

$$\begin{aligned} cov\{\mathbf{e}\} &= E\{\mathbf{e}\mathbf{e}^T\} = \Phi_t \Sigma \Phi_t^T - \Sigma_{\Delta} \Phi_t^T - \Phi_t \Sigma_{\Delta}^T + \Sigma \\ &= \Sigma - \Sigma_{\Delta} \Sigma^{-1} \Sigma_{\Delta}^T + (\Phi_t - \Sigma_{\Delta} \Sigma^{-1}) \Sigma (\Phi_t - \Sigma_{\Delta} \Sigma^{-1})^T \end{aligned} \quad (4.26)$$

Because the term $\Sigma - \Sigma_{\Delta} \Sigma^{-1} \Sigma_{\Delta}^T$ is independent of Φ_t , the minimum trace can be determined when

$$\Phi_t = \Sigma_{\Delta} \Sigma^{-1} \quad (4.27)$$

The corresponding covariance matrix is:

$$cov\{\mathbf{e}\} = \Sigma - \Sigma_{\Delta} \Sigma^{-1} \Sigma_{\Delta}^T \quad (4.28)$$

which was treated as process noise \mathbf{N}_t in the process model. Φ_t and \mathbf{N}_t are 5×5 -matrices which indicating the change in five gauge stations. The values in the matrices were evenly distributed to the reaches around each gauge station. Figures 4.10 and 4.11 show the distributed Φ_t and \mathbf{N}_t respectively.

At this point, the time domain process model needed to be adjusted accordingly. Combining Equations 4.21 and 4.22 yielded

$$\begin{aligned} \mathbf{d}_t - \tilde{\mathbf{d}}_t &= \Phi_t (\mathbf{d}_{t-1} - \tilde{\mathbf{d}}_{t-1}) + \mathbf{e} \\ \mathbf{d}_t &= \Phi_t \mathbf{d}_{t-1} - \Phi_t \tilde{\mathbf{d}}_{t-1} + \tilde{\mathbf{d}}_t + \mathbf{e} \end{aligned} \quad (4.29)$$

which could be rewritten as:

$$\mathbf{d}_t = \Phi_t \mathbf{d}_{t-1} + \mathbf{G} \mathbf{U}_{t-1} + \mathbf{e} \quad (4.30)$$

where $\mathbf{G} = \begin{bmatrix} -\Phi_t & \mathbf{I} \end{bmatrix}$ and $\mathbf{U}_{t-1} = \begin{bmatrix} \tilde{\mathbf{d}}_{t-1} \\ \tilde{\mathbf{d}}_t \end{bmatrix}$.

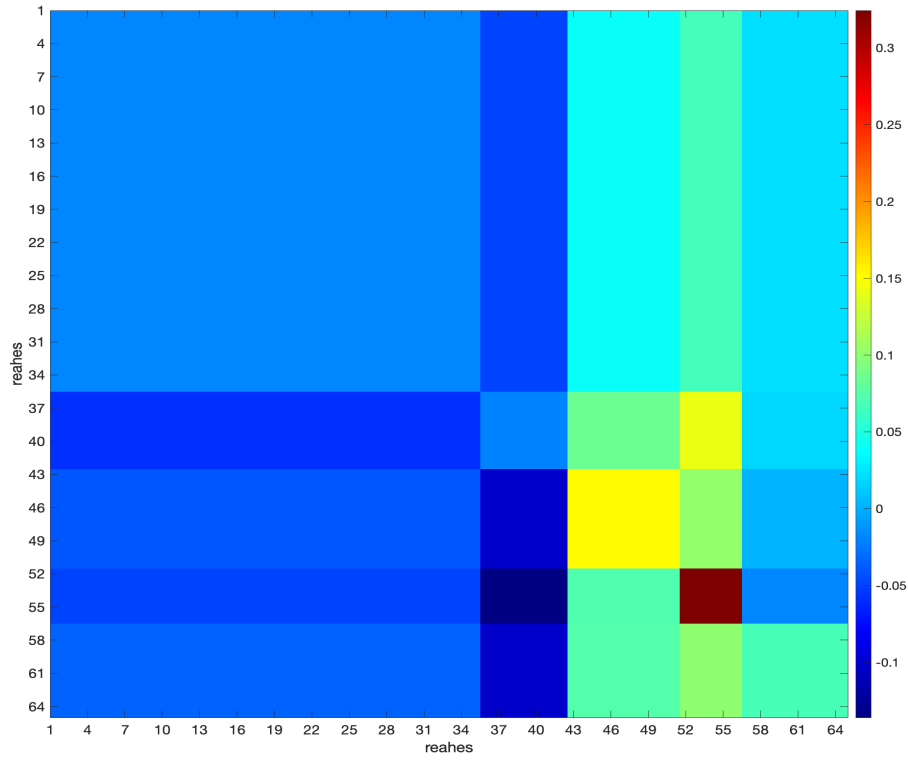


Figure 4.10: Time domain transition matrix Φ_t

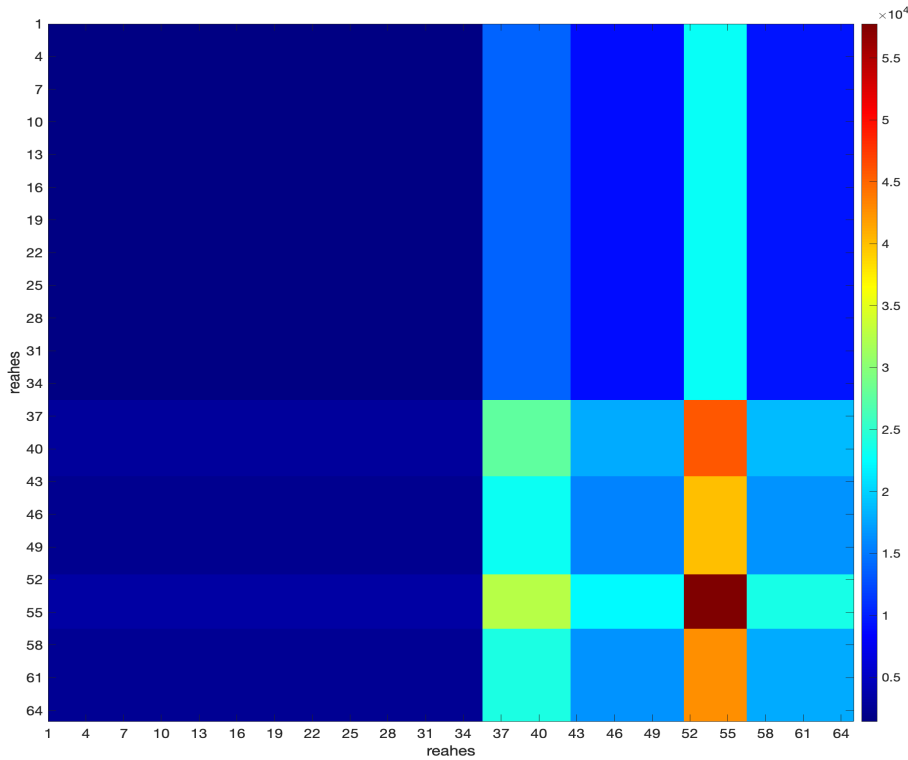


Figure 4.11: Time domain process noise N_t

Thus the time domain process model could be understood as

$$\begin{cases} \mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \mathbf{G} \mathbf{U}_{t-1} + \mathbf{e} \\ \mathbf{P}_t = \Phi_t \mathbf{P}_{t-1} \Phi_t^T + \mathbf{N}_t \end{cases} \quad (4.31)$$

where vector \mathbf{x}_t refers to the discharges of 65 reaches in the Po mainstream during

the period t : $\mathbf{x}_t = \begin{bmatrix} Q_t(\Omega_1, t) \\ Q_t(\Omega_2, t) \\ \vdots \\ Q_t(\Omega_{65}, t) \end{bmatrix}$. The initial values of \mathbf{x}_t and \mathbf{P}_t were informed by

the Kalman-Filtered results of the previous period.

4.6.2 Observation model

The observation equation benefited from mass conservation for the channel (Durand et al. 2014):

$$\frac{\partial \bar{Q}}{\partial x}(\Omega, t) + \frac{\partial \bar{A}}{\partial t}(\Omega, t) = \bar{q}(\Omega, t) \quad (4.32)$$

where the overbars indicate the average reach of each quantity. On the right side, $\bar{q}(\Omega, t)$ is the reach's average floodplain-channel exchange, which was set as $0 \text{ m}^2/\text{s}$ for a rough estimation. On the left side, the first and second terms are the "flow imbalance" and "mass imbalance," respectively. Given that the flow imbalance term represents the reach's average discharge spatial derivative, Equation 4.32 can be understood as

$$\frac{1}{2} \left(\frac{\partial \bar{Q}}{\partial x} \Big|_t + \frac{\partial \bar{Q}}{\partial x} \Big|_{t-1} \right) + \frac{\partial \bar{A}}{\partial t}(\Omega, t) = \bar{q}(\Omega, t) \quad (4.33)$$

which can be continued to be extended as:

$$\frac{1}{2} \left[\frac{Q(\Omega, t) - Q(\Omega - 1, t)}{L(\Omega)} + \frac{Q(\Omega, t - 1) - Q(\Omega - 1, t - 1)}{L(\Omega)} \right] + \frac{A'(\Omega, t) - A'(\Omega, t - 1)}{\Delta t} = \bar{q}(\Omega, t) \quad (4.34)$$

This can be rewritten as

$$\bar{q}(\Omega, t) - \frac{A'(\Omega, t) - A'(\Omega, t - 1)}{\Delta t} = \frac{1}{2L(\Omega)} [Q(\Omega, t) - Q(\Omega - 1, t) + Q(\Omega, t - 1) - Q(\Omega - 1, t - 1)] \quad (4.35)$$

Using Equation 4.35, the observation model can be established as

$$\underbrace{\bar{q}(\Omega, t) - \frac{A'(\Omega, t) - A'(\Omega, t-1)}{\Delta t}}_{\mathbf{z}} = \underbrace{\begin{bmatrix} \frac{1}{2L(\Omega)} & -\frac{1}{2L(\Omega)} & \frac{1}{2L(\Omega)} & -\frac{1}{2L(\Omega)} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} Q(\Omega, t) \\ Q(\Omega-1, t) \\ Q(\Omega, t-1) \\ Q(\Omega-1, t-1) \end{bmatrix}}_{\mathbf{x}} \quad (4.36)$$

which illustrates that if there are $A'(\Omega)$ in two adjacent months (and $\bar{q}(\Omega, t) = 0 \text{ m}^2/\text{s}$), the state estimates can be updated via this equation. In addition, the corresponding covariance of observation \mathbf{R} can be obtained by propagation of uncertainty:

$$\mathbf{R} = \frac{\sigma_{A'(\Omega, t)}^2 + \sigma_{A'(\Omega, t-1)}^2}{\Delta t^2} \quad (4.37)$$

4.6.3 Obtaining a priori discharge estimates

The process and observation models can offer a priori discharge estimates Q_{est} through Kalman Filter.

Firstly, the Kalman Filter was executed in the space domain only. In other words, starting from the neighboring gauge stations in Cremona and Borgoforte in January 2002, only the space domain process models (see Equations 4.19 and 4.20) were utilized for the estimate and the observation models (see Equation 4.36); this allowed discharge values for each reach in January 2002 to be obtained.

Then, the normal Kalman Filter was implemented for other epochs t . Using the

time domain process model (see Equation 4.31), the state vector $\mathbf{x}_t = \begin{bmatrix} Q_t(\Omega_1, t) \\ Q_t(\Omega_2, t) \\ \vdots \\ Q_t(\Omega_{65}, t) \end{bmatrix}$

can be predicted from discharges in the previous \mathbf{x}_{t-1} . Secondly, using the space

domain process model the state vector $\mathbf{x}_s = \begin{bmatrix} Q_s(\Omega, t) \\ Q_s(\Omega-1, t) \\ Q_s(\Omega, t-1) \\ Q_s(\Omega-1, t-1) \end{bmatrix}$ can be predicted.

Notably, the starting values of $Q_s(\Omega, t)$ and $Q_s(\Omega - 1, t)$ represented timely predicted discharges of the Borgoforte gauge station and its neighboring reach, while the starting values of $Q_s(\Omega, t - 1)$ and $Q_s(\Omega - 1, t - 1)$ were extracted from the Kalman-Filtered results of the previous period:

$$\mathbf{x}_{s0} = \begin{bmatrix} Q_t(\text{Borgoforte}, t) \\ Q_t(\text{Borgoforte} - 1, t) \\ Q_{est.}(\text{Borgoforte}, t - 1) \\ Q_{est.}(\text{Borgoforte} - 1, t - 1) \end{bmatrix} \quad (4.38)$$

Figure 4.12 indicates the space and time domain estimates.

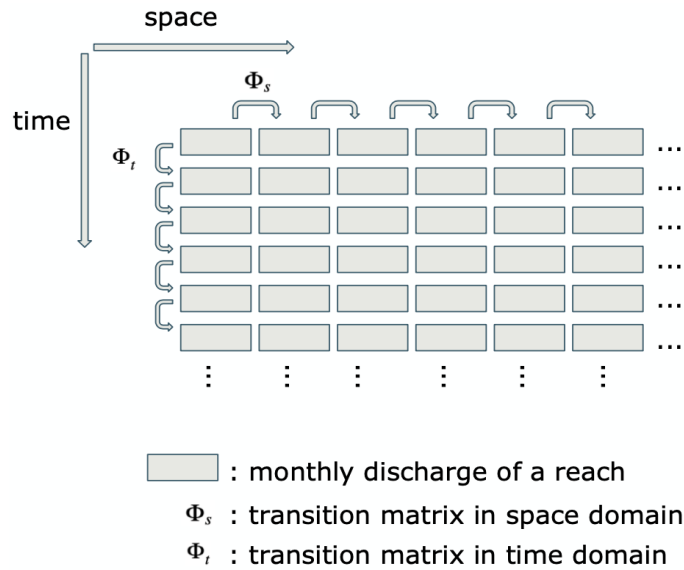


Figure 4.12: Space and time domain estimates

Thirdly, a joint prior $\mathbf{x}_{s,t}$ was calculated that combined the predictions of the two process models \mathbf{x}_s and \mathbf{x}_t with the corresponding covariance matrix Σ_s and Σ_t :

$$\underbrace{\begin{bmatrix} Q_s(\Omega, t) \\ Q_s(\Omega - 1, t) \\ Q_s(\Omega, t - 1) \\ Q_s(\Omega - 1, t - 1) \\ Q_t(\Omega, t) \end{bmatrix}}_{\mathbf{x}_s \text{ and } \mathbf{x}_t(\Omega, t) \text{ as } \mathbf{y}_{5 \times 1}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{A}_{5 \times 4}} \underbrace{\begin{bmatrix} Q_{s,t}(\Omega, t) \\ Q_s(\Omega - 1, t) \\ Q_s(\Omega, t - 1) \\ Q_s(\Omega - 1, t - 1) \end{bmatrix}}_{\mathbf{x}_{s,t} \ 4 \times 4} \quad (4.39)$$

The weight matrix of this model is:

$$\mathbf{P} = \mathbf{Q}_y^{-1} = \begin{bmatrix} \Sigma_s & 0 \\ 0 & \Sigma_t \end{bmatrix}^{-1} \quad (4.40)$$

Then the joint prior $\hat{\mathbf{x}}_{s,t}$ could be obtained by adjustment (see Equation 4.13). The corresponding covariance of $\mathbf{x}_{s,t}$, namely $\hat{\Sigma}_x$, can be computed as:

$$\mathbf{e} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{s,t} \quad (4.41)$$

$$\sigma_0^2 = \frac{\mathbf{e}^T \mathbf{P} \mathbf{e}}{5 - 4} \quad (4.42)$$

$$\hat{\Sigma}_x = \sigma_0^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \quad (4.43)$$

Figure 4.13 shows the probability distribution of Q_s , Q_t and their joint prior $Q_{s,t}$. We can see that joint prior $Q_{s,t}$ is the combination of Q_s and Q_t with improved standard deviation.

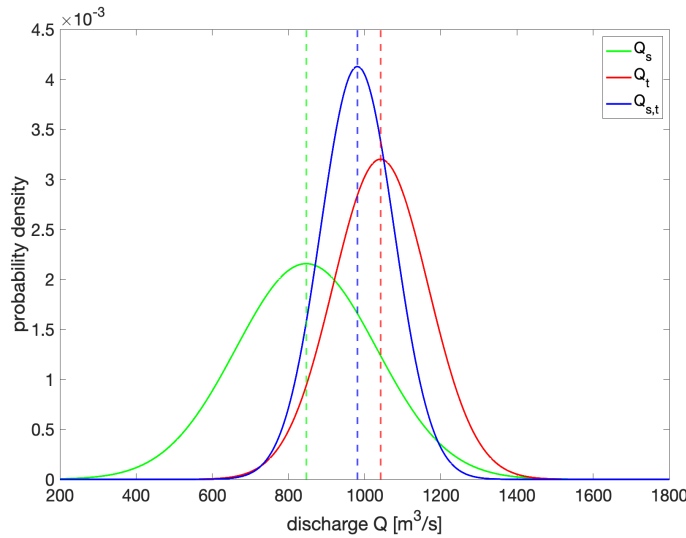


Figure 4.13: Probability distribution of Q_s , Q_t and their joint prior $Q_{s,t}$

Next, the joint prior was updated if possible; otherwise, only predictions were available for this reach. After performing these steps for all reaches in the main-stream, a priori discharge estimates $Q_{est.}$ were obtained for this period.

4.6.4 Computing the flow law parameters

Using the obtained a priori discharges $Q_{est.}$, the flow law parameters for each reach were estimated using interior-point optimization with inequality constraints.

The function to optimize was based on the combination of Manning's equation (see Equation 4.1) and the extended description of the roughness coefficient n (see Equation 4.2):

$$Q = \frac{1}{n_b(1 + \frac{5}{6}[\frac{w\sigma_z}{\bar{A}+\bar{A}'}]^2)}(\bar{A} + \bar{A}')^{5/3}w^{-2/3}s^{1/2} \quad (4.44)$$

With the support of $Q_{est.}$ and simulated w , s and A' , the time-invariant parameters \bar{A} , n_b and σ_z for each reach can be estimated by minimizing the following function:

$$\min[f(\bar{A}, n_b, \sigma_z)] = \min \left[\left| \bar{Q}_{est.} - \frac{1}{n_b(1 + \frac{5}{6}[\frac{\bar{w}\sigma_z}{\bar{A}+\bar{A}'}]^2)}(\bar{A} + \bar{A}')^{5/3}\bar{w}^{-2/3}\bar{s}^{1/2} \right| \right] \quad (4.45)$$

where the overbars of $Q_{est.}$, w , s and A' indicate their weighted mean over all periods in each reach given that the parameters to be computed are time-invariant (constant).

In addition, inequality constraints can be obtained by minimizing the function

$$\begin{cases} 0.3 > n_b > 0 \\ \bar{A} > 0 \\ \sigma_z \geq 0 \end{cases} \quad (4.46)$$

It was observed that the interior-point optimization required initial parameters values to be computed.

- σ_{z0} was set as 0 for all reaches.
- n_{b0} was derived from gauge and SWORD data. For each gauge station, a mean value was calculated ($\sigma_{z0} = 0$):

$$n_{b0} = \text{mean} \left[\frac{1}{Q_{gauge}(t)} A_{gauge}(t)^{5/3} w_{gauge}(t)^{-2/3} s_{SWORD}^{1/2} \right] \quad (4.47)$$

where $Q_{gauge}(t)$, $A_{gauge}(t)$ and $w_{gauge}(t)$ are time series gauge data, while s_{SWORD} is extracted from SWORD. Then, each n_{b0} was allocated to the reaches neighboring the gauge stations as per Figure 3.7.

- For \bar{A}_0 , temporary $\bar{A}(t)$ was first computed for each epoch with the help of n_{b0} , $Q_{est.}(t)$, simulated $w(t)$, $s(t)$ and $A'(t)$:

$$\bar{A}(t) = \sqrt[3]{n_{b0} Q_{est.}(t) w(t)^{2/3} s(t)^{-1/2}} - A'(t) \quad (4.48)$$

After this, the weighted mean of $\bar{A}(t)$ was used as the initial \bar{A}_0 .

The whole procedure of interior-point optimization with inequality constraints is presented in Figure 4.3. After that we can obtain optimized $\hat{\sigma}_z$, \hat{n}_b and \hat{A} for each reach were obtained and utilized to recalculate the discharges. These recalculated discharges were treated as observations for a new round of Kalman Filtering with the aim of estimating posterior discharges.

4.7 Estimating posterior discharges

As noted above, a new round of Kalman Filtering was executed. The time and space domain process models (see Section 4.6.1) and the computation of joint prior (see Section 4.6.3) followed the same steps as above, with one difference in the observation model and its use of the mass conservation equation (see Equation 4.36) and recalculated discharges. With the aid of estimated flow law parameters and simulated data, the discharges were recalculated as follows:

$$Q_{recal.} = \frac{1}{\hat{n}_b \left(1 + \frac{5}{6} \left[\frac{w\hat{\sigma}_z}{\hat{A} + A'}\right]^2\right)} (\hat{A} + A')^{5/3} w^{-2/3} s^{1/2} \quad (4.49)$$

And the corresponding variance can be obtained by propagation of uncertainty.

The new observation model is:

$$\underbrace{\begin{bmatrix} \bar{q}(\Omega, t) - \frac{A'(\Omega, t) - A'(\Omega, t-1)}{\Delta t} \\ Q_{recal.}(\Omega, t) \\ Q_{recal.}(\Omega - 1, t) \\ Q_{recal.}(\Omega, t - 1) \\ Q_{recal.}(\Omega - 1, t - 1) \end{bmatrix}}_{\mathbf{z}} = \underbrace{\begin{bmatrix} \frac{1}{2L(\Omega)} & -\frac{1}{2L(\Omega)} & \frac{1}{2L(\Omega)} & -\frac{1}{2L(\Omega)} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} Q(\Omega, t) \\ Q(\Omega - 1, t) \\ Q(\Omega, t - 1) \\ Q(\Omega - 1, t - 1) \end{bmatrix}}_{\mathbf{x}} \quad (4.50)$$

In addition, the corresponding covariance of observation \mathbf{R} is:

$$\mathbf{R} = \begin{bmatrix} \frac{\sigma_{A'(\Omega, t)}^2 + \sigma_{A'(\Omega, t-1)}^2}{\Delta t^2} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{Q_{recal.}(\Omega, t)} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{Q_{recal.}(\Omega-1, t)} & 0 & 0 \\ 0 & 0 & 0 & \sigma_{Q_{recal.}(\Omega, t-1)} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{Q_{recal.}(\Omega-1, t-1)} \end{bmatrix} \quad (4.51)$$

Finally, after the new round of Kalman Filter, the posterior time series discharges $\hat{Q}_{est.}(t)$ for each station were collected. Note that, the steps of Kalman Filter are shown in the Figure 4.2. The results of the flow law parameters and prior and posterior discharges are presented in the Chapter 5. In addition, the data is validated via comparison with gauge data.

Chapter 5

Results and validation

This chapter will show the results of post-processing after simulation. For the validation in-situ data is an important measurement tool. The Table 4.2 has shown that the simulated data are highly correlated to the input data derived from satellite data. But the measurement positions of gauge station and satellites could be different. Hence, for the better comparison and validation with estimated discharges, the in-situ discharges will be fitted to the satellites discharges through their width data.

According to the Manning's equation (see Equation 4.1) for each gauge station a scale factor s can be built:

$$s = \text{mean} \left[\left(\frac{w_{\text{Landsat}}(t)}{w_{\text{gauge}}(t)} \right)^{-2/3} \right] \quad (5.1)$$

where $w_{\text{Landsat}}(t)$ and $w_{\text{gauge}}(t)$ are respectively the monthly Landsat and in-situ width from year 2000 to 2011. s is the mean value of the monthly scale factors. After that the fitted in-situ discharges in each gauge station can be computed:

$$Q_{\text{fit}} = Q_{\text{gauge}} \cdot s \quad (5.2)$$

Notably in this chapter all mentioned or utilized in-situ discharges Q_{gauge} refer to Q_{fit} , which will not be specifically declared in the following.

5.1 Results of a priori discharges

Through the first round of Kalman Filter the prior discharges from year 2002 to 2019 could be computed. The monthly results for each reach and the comparison with in-situ data are shown in Appendix A. It is noteworthy that the comparison

with in-situ data is only available from year 2002 to 2011. Furthermore for a better comparison, the data in gauge stations were extracted separately and shown in time series in Figure 5.1.

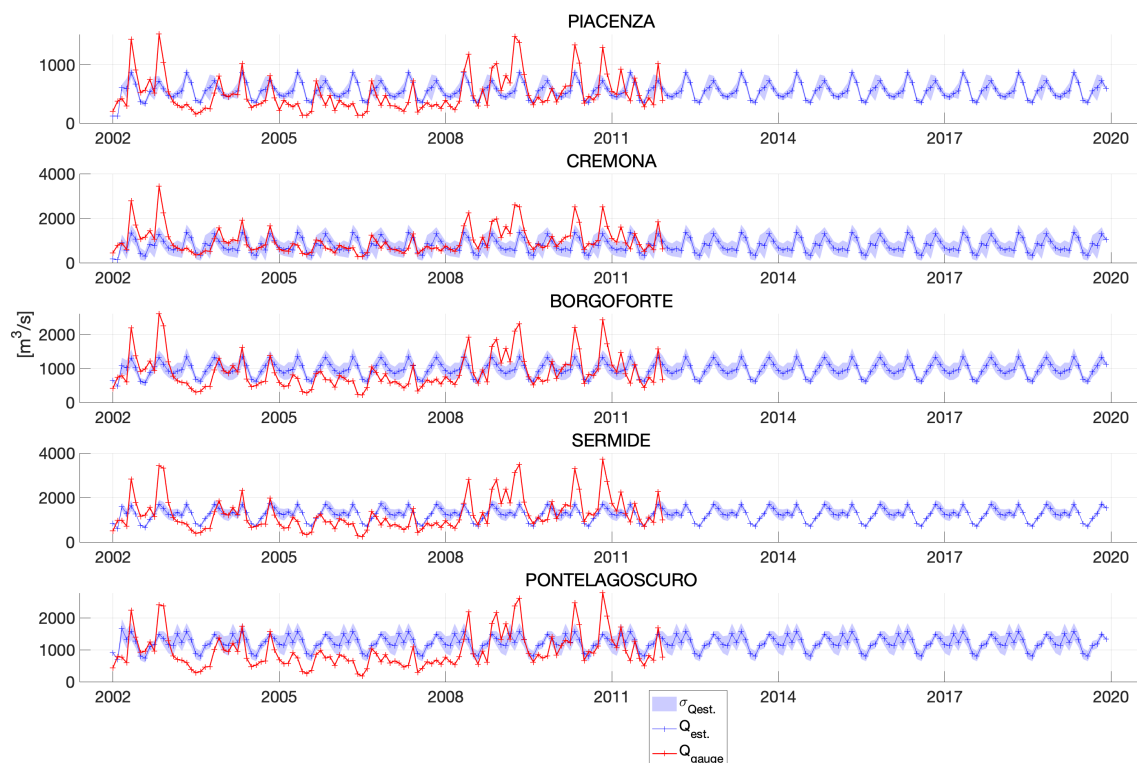


Figure 5.1: Comparison of a priori and in-situ discharges

It can be seen that, the trend of the curves in Appendix A is similar for the same month in each year. Figure 5.1 also reflect that, the prior discharges are almost periodic. In addition, there are obvious differences between a priori and in-situ discharges.

One of the possible reasons for this is that in the process model, Φ_s is a mean monthly matrix and Φ_t is even constant. The linear model may not be well adapted to predict discharges, which could be influenced by different factors, such as weather conditions, human factors, etc. On the other hand in the first round of Kalman Filter, only the mass conservation equation (see Equation 4.36) was used as the update. In this thesis, the reach's average floodplain-channel exchange $\bar{q}(\Omega, t)$ was roughly set as $0 \text{ m}^2/\text{s}$. And by computing of A' through "SWOTAprimeCalcs" (see Section 4.5), the fit parameters and breakpoints were not yet defined. Thus, the calculation of them was based on simulated height-width observations. This requires high demands on the accuracy of simulated height-width observations. Worse still, the computation of the error of A' is still

being elaborated. All the above mentioned factors have a negative impact on the computation of mass conservation equation, which also leads to less accurate updates of the predicted estimates.

To verify the influence of simulated height-width observations, the simulated width and height were separately compared with them in gauge stations in Figure 5.3 and 5.2. Table 5.1 also lists the correlations in each gauge station. All these information reveal that there are big differences between the in-situ and the simulated height-width observations, which could result in the differences between prior and in-situ discharges by computing A' , and even have a negative influence on the following estimation of flow law parameters and posterior discharges.

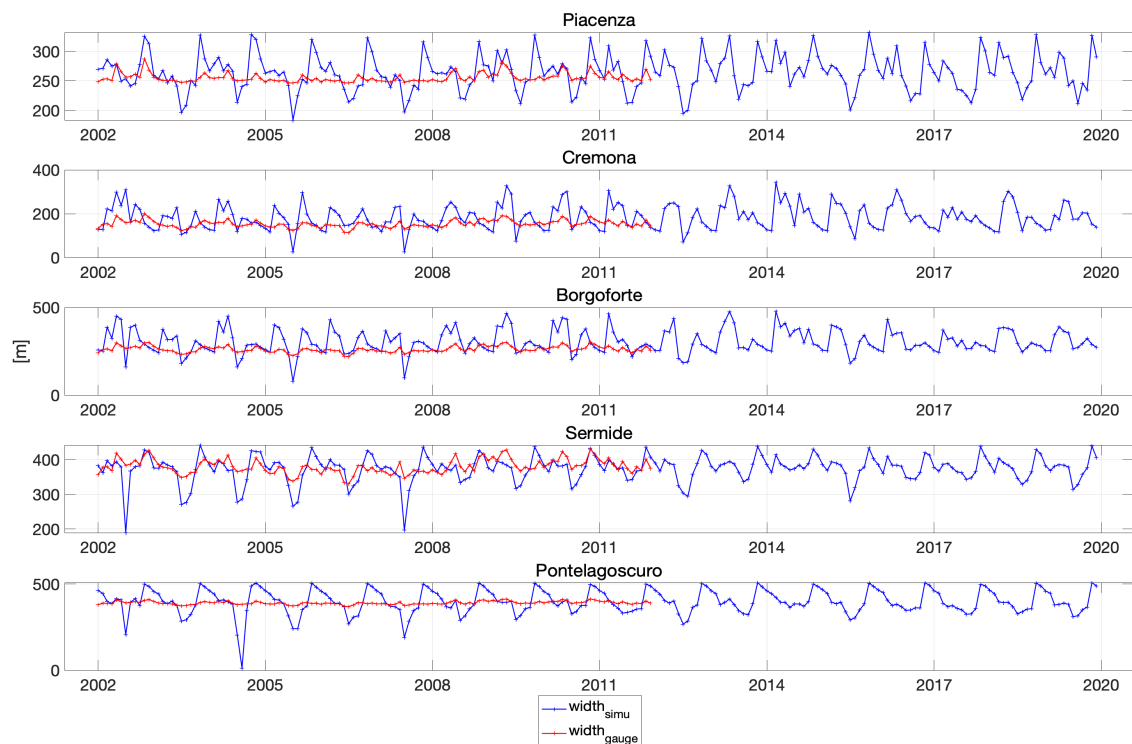


Figure 5.2: Comparison of simulated and in-situ width

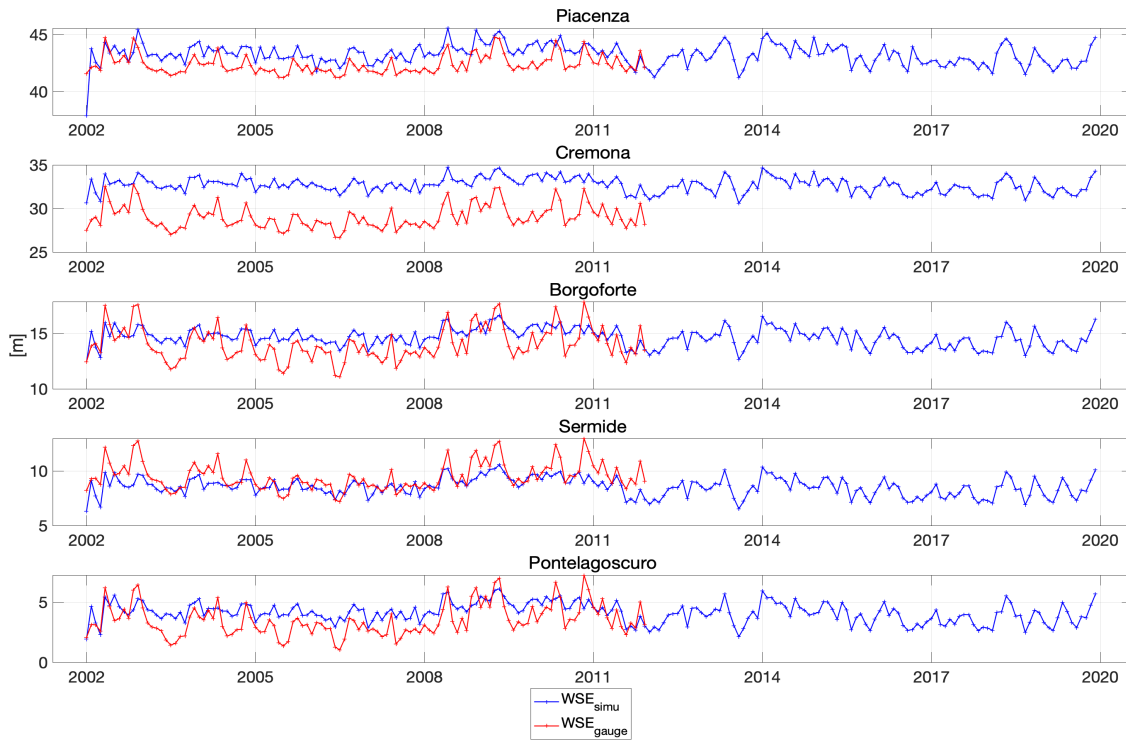


Figure 5.3: Comparison of simulated and in-situ WSE

	Corr. of width [-]	Corr. of WSE [-]
Piacenza	0.4646	0.5789
Cremona	0.4470	0.6200
Borgoforte	0.5261	0.6293
Sermide	0.4685	0.6426
Pontelagoscuro	0.4440	0.6598

Table 5.1: Correlation of simulated and in-situ data

5.2 Results of flow law parameters

After estimating the prior discharges, the interior-point optimization with inequality constraints could be executed. The respective comparison of initial and optimized parameters \bar{A} , σ_z and n_b for each reach are shown in Figures 5.4, 5.5 and 5.6. It can be seen that the data has been improved essentially. It is noteworthy that not every reach can obtain optimized flow law parameters, because in these reaches the simulated w , s or A' are not available to execute the interior-point optimization.

Moreover, the reach-averaged Manning's roughness coefficient n could be computed through the Equation 4.2 with the aid of optimized flow law parameters

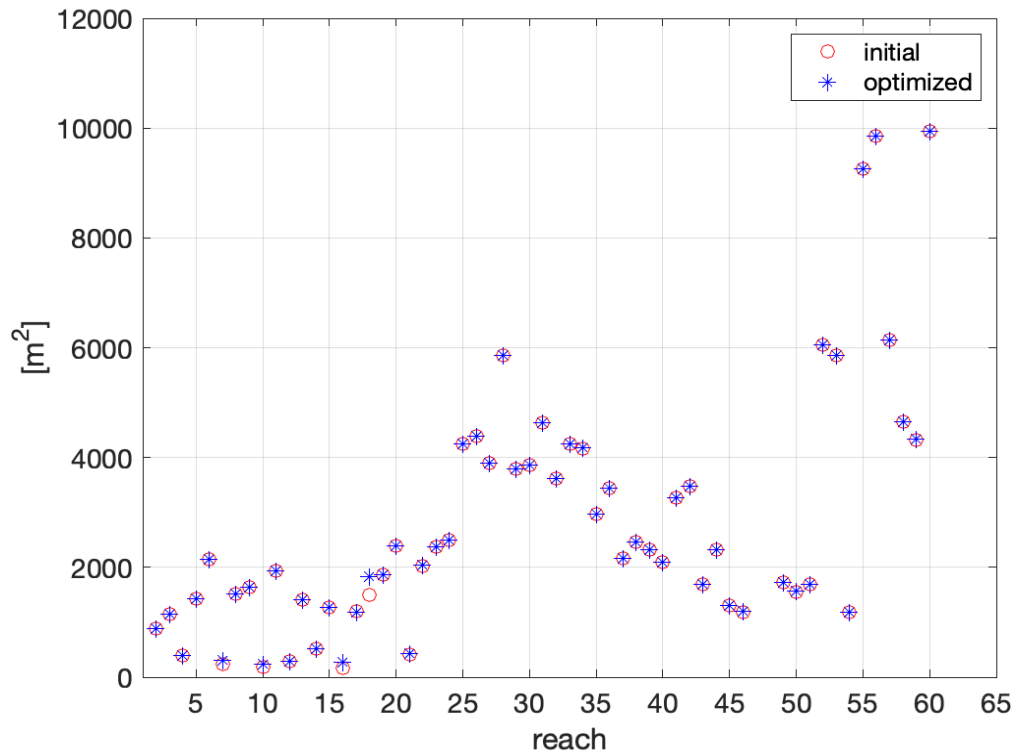


Figure 5.4: Initial vs. optimized \bar{A}

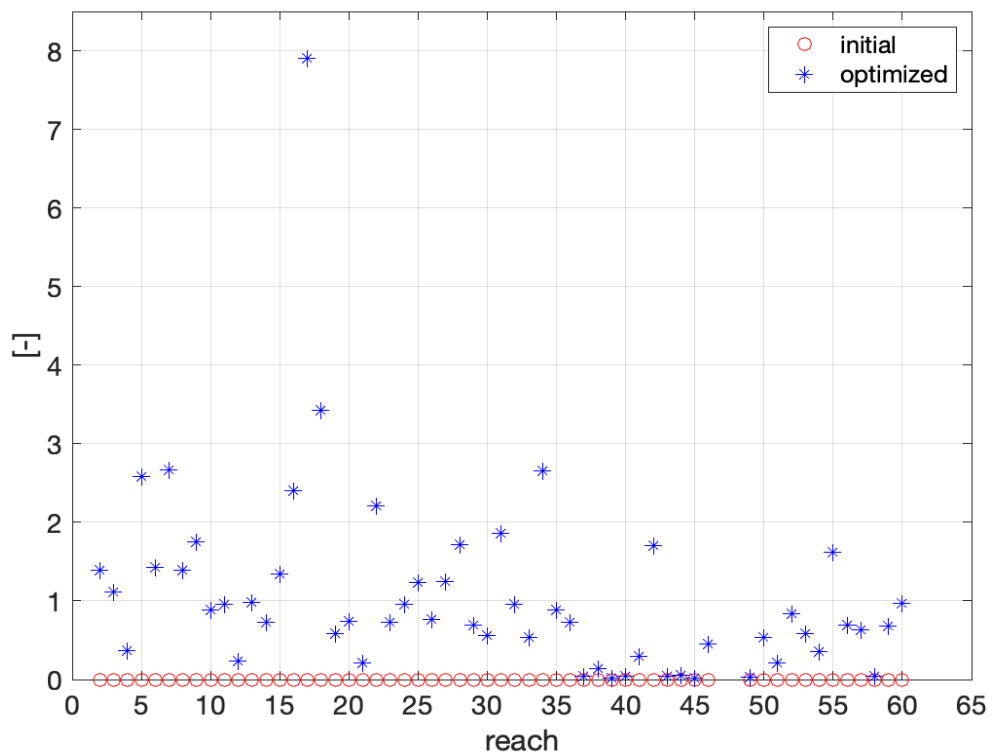


Figure 5.5: Initial vs. optimized σ_z

and simulated data. Figure 5.7 shows calculated n for each gauge station in time

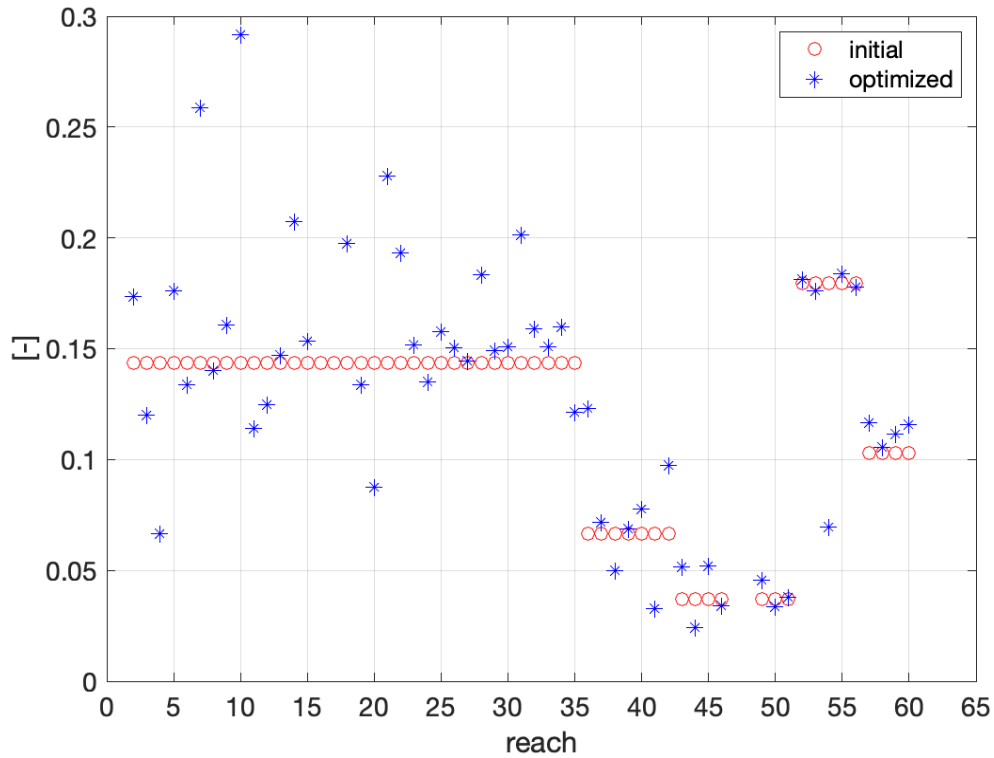


Figure 5.6: Initial vs. optimized n_b

series. In comparison with the time-invariant n_b , n changes in time series accordingly due to the variation of simulated data, but the change is stable in a relatively small interval.

The interior-point optimization (see Equation 4.45) has some potential risks. On the one hand, it uses $Q_{est,r}$ that doesn't perform well according to Section 5.1. On the other hand, the optimization equation utilizes mean values of Q_{est} and simulated data, which leads to the risk that the results are less accurate, since the particular circumstances of each period are not taken into account. Besides, as the inequality constraints are manually established, it may not contribute to the best optimization.

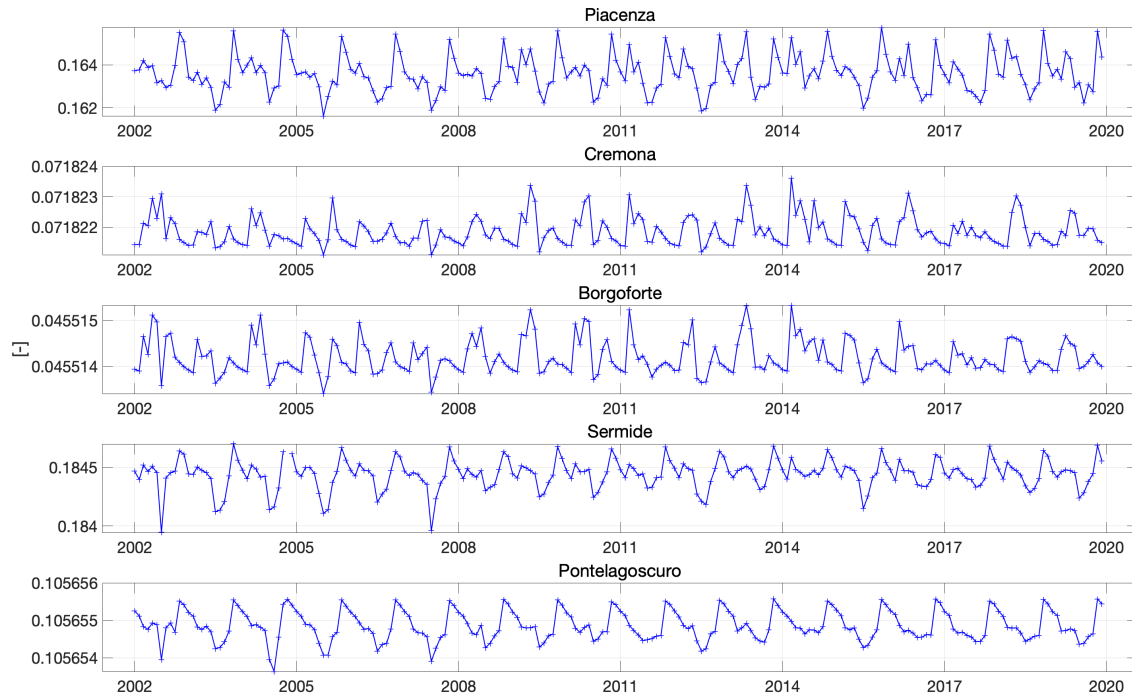


Figure 5.7: Calculated n for each gauge station in time series

5.3 Results and validation of posterior discharges

After the second round of Kalman Filter the posterior discharges can be obtained. The monthly results for each reach and the comparison with in-situ data are shown in Appendix B. The curve trend for the same month each year now reveals a difference. The standard deviation has also been improved. Differences from in-situ discharges have been also slightly reduced in some periods.

Figure 5.8 displays the comparison with in-situ discharges in time series. To further validate the results, the Root-Mean-Square Error (RMSE), correlation, and Nash-Sutcliffe Efficiency (NSE) values have been calculated between in-situ and posterior discharges (see Table 5.2).

It can be observed that there are still considerable differences between posterior and the in-situ data. The RMSE is a little bit large (approximately 30% of the mean in-situ discharge) and is maximum in the gauge station Sermide with the largest mean discharge. The correlation is larger than 0, which indicates a positive linear dependence between the estimated and in-situ discharge. The NSE is also greater as 0, which reflects that this model is a better predictor than the mean of in-situ discharge. The best correlation and NSE values are in the gauge station Borgoforte, which is also the starting station of the spatial process model

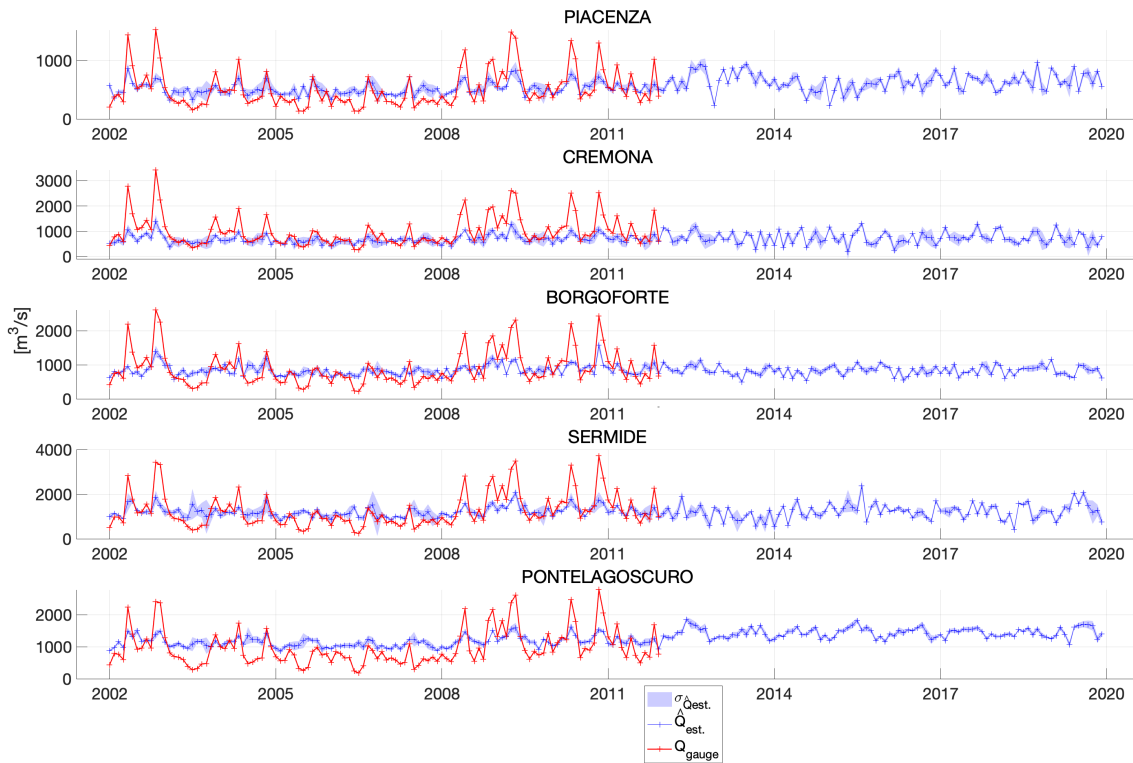


Figure 5.8: Comparison of posterior and in-situ discharges

	RMSE [m^3/s]	Correlation [-]	NSE [-]
Piacenza	283.3072	0.4434	0.1910
Cremona	554.8363	0.4490	0.1816
Borgoforte	452.7548	0.4532	0.2028
Sermide	670.1480	0.4497	0.2015
Pontelagoscuro	548.3018	0.4472	0.1731

Table 5.2: Validation in gauge stations

in Kalman Filter.

In general, validation results do not show good performance. In Section 5.1 and 5.2, the disadvantages of first round of Kalman Filter and risks of interior-point optimization with inequality constraints have been analyzed, which have negative chain effects to the estimation of posterior discharges. Additionally, in the second round of Kalman Filter the discharges were recalculated with the support of simulated $w(t)$, $s(t)$ and $A'(t)$. Table 5.1 has shown that, the simulated and in-situ width and WSE are not well correlated. Since the computation of slope is dependent on WSE, it can be presumed that the simulated $s(t)$ and in-situ slope

do not correlate very well either. Besides, the calculation of $A'(t)$ is based on simulated height-width observations. Therefore it can be inferred that, $Q_{recal.}$ have a poor performance and have a low correlation with $Q_{gauge.}$. The combination of the above reasons led to the unsatisfactory results.

Chapter 6

Conclusion and outlook

6.1 Summary and conclusion

This article consists of two main steps: simulation and post-processing:

In the simulation, first the SWOT-like data in time series over the Po River are simulated by SWOT Hydrology Simulator. As input Landsat provide width data, various satellites such as Envisat, Topex etc. provide altimetry data, and via altimetry data slope could be created. Then the Matlab code SWOTAprimeCalcs could compute the change in cross-sectional area with the support of the simulated height-width observations.

In the post-processing, at first the prior discharge is obtained by a Kalman Filter estimation with a spatio-temporal process model and mass conservation condition as the observation equation. Next using the obtained prior discharge flow law parameters are estimated through interior-point optimization with inequality constraints. Finally, posterior discharge estimates are obtained by a second round of Kalman Filter, with adding discharge observations derived from simulated measurements and estimated flow law parameters.

The validation through comparison with in-situ discharge shows a poor performance with the average RMSE of $501.87 \text{ m}^3/\text{s}$, correlation of 0.45 and NSE of 0.19. RMSE is a little bit large. The correlation and NSE values are larger than 0, which indicates a positive linear dependence between the predicted and in-situ data and a good predictive skill of this method. The validation doesn't show the best performance. Such low performance may be due to the following reasons:

- The input width and altimetry data for simulation reveals a low correlation with in-situ data, but the simulated data are highly dependent on the width data.

- The computation of change in cross-sectional area is still being elaborated and some parameters have not yet been optimally defined.
- The reach's average floodplain-channel exchange in mass conservation condition was only roughly set as $0 \text{ m}^2/s$.
- The interior-point optimization method uses the not satisfactory prior discharge. And the mean value of prior discharge and simulated data were applied for the computation, so that the specific scenarios in each period cannot be considered.
- The inequality constraints of interior-point optimization are set manually but there is no error estimate to evaluate the quality of flow law parameters.
- As additional observations to estimate posterior discharge, the recalculated discharges uses the simulated data and optimized flow law parameters, which have been proved to have potentially poor performance.

6.2 Outlook

The following tips could probably improve the results:

1. The input data of simulation needs to be improved. As mentioned in Section 3.3.1, many of the input width or height data are manually created because they are not available in some periods or reaches. Hence, to obtain more information, data from more various satellites should be collected and compared.
2. Since the SWOT Hydrology Toolbox and SWOTAprimeCalcs are still in development, we can use the latest version to check if the output data have been improved.
3. Instead of the traditional linear Kalman Filter, extended or other non-linear Kalman Filter could be tried.
4. The real reach's average floodplain-channel exchange in mass conservation condition should be taken into account.
5. A better interior-point optimization method should be derived, which could consider the variation of simulated data in time series and combine all the scenarios to give an optimal result. And the errors of the flow law parameters also

need to be estimated in some way, so that the inequality constraints could be accordingly modified.

Appendix A

Estimated a priori discharges

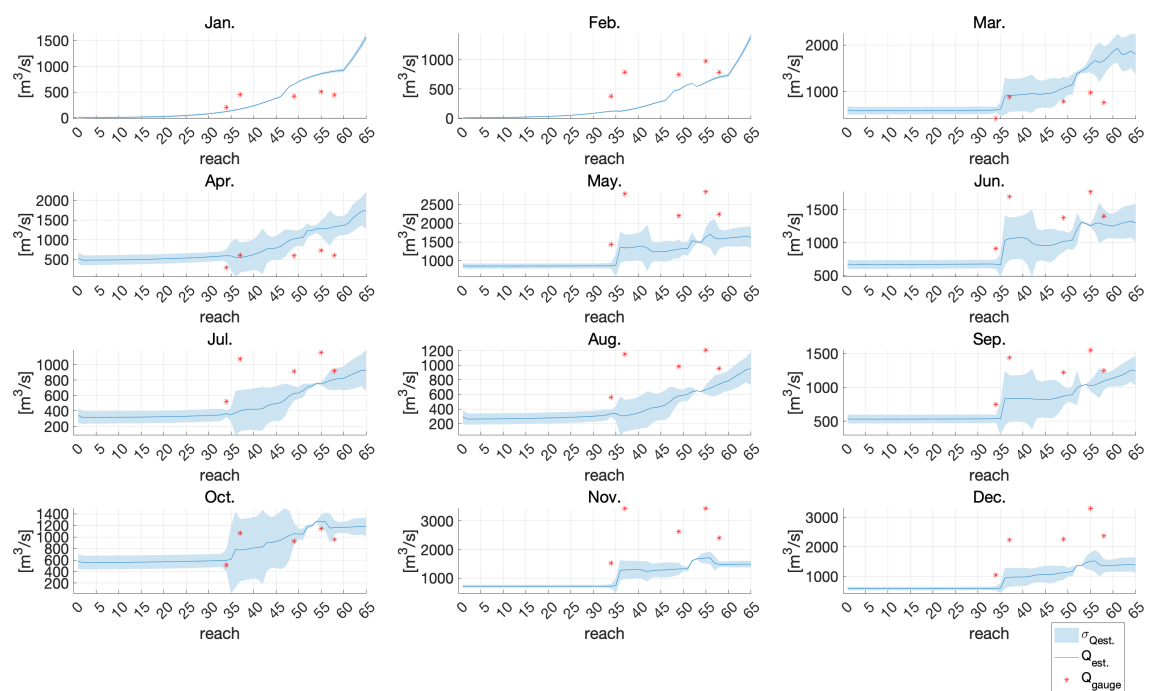
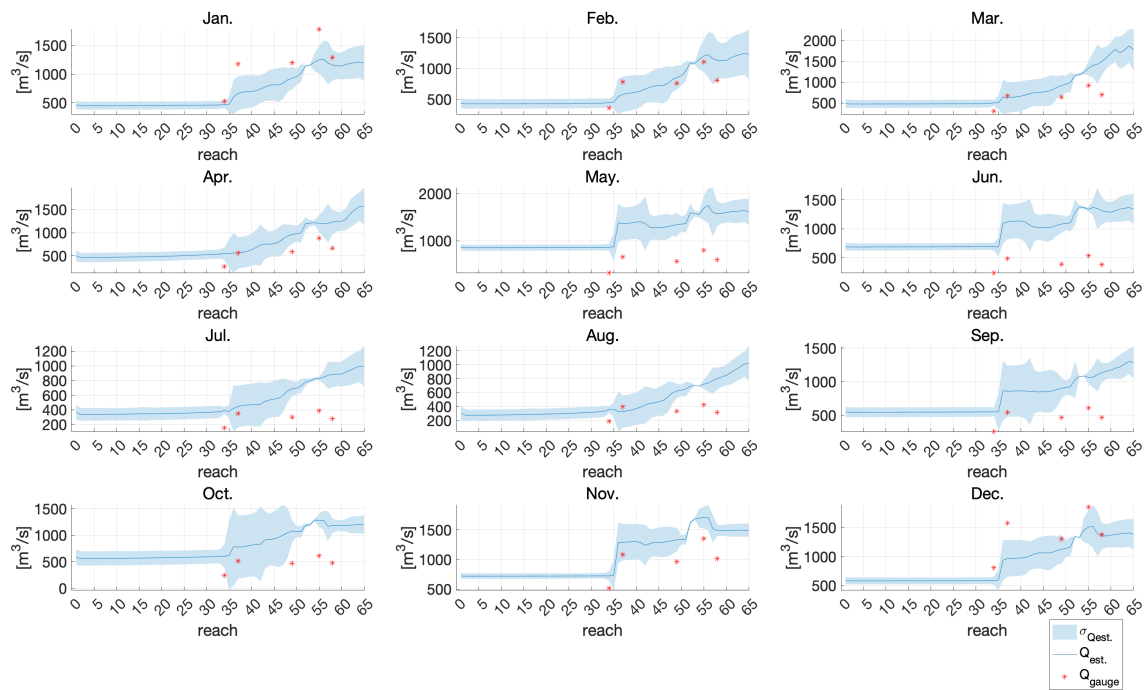
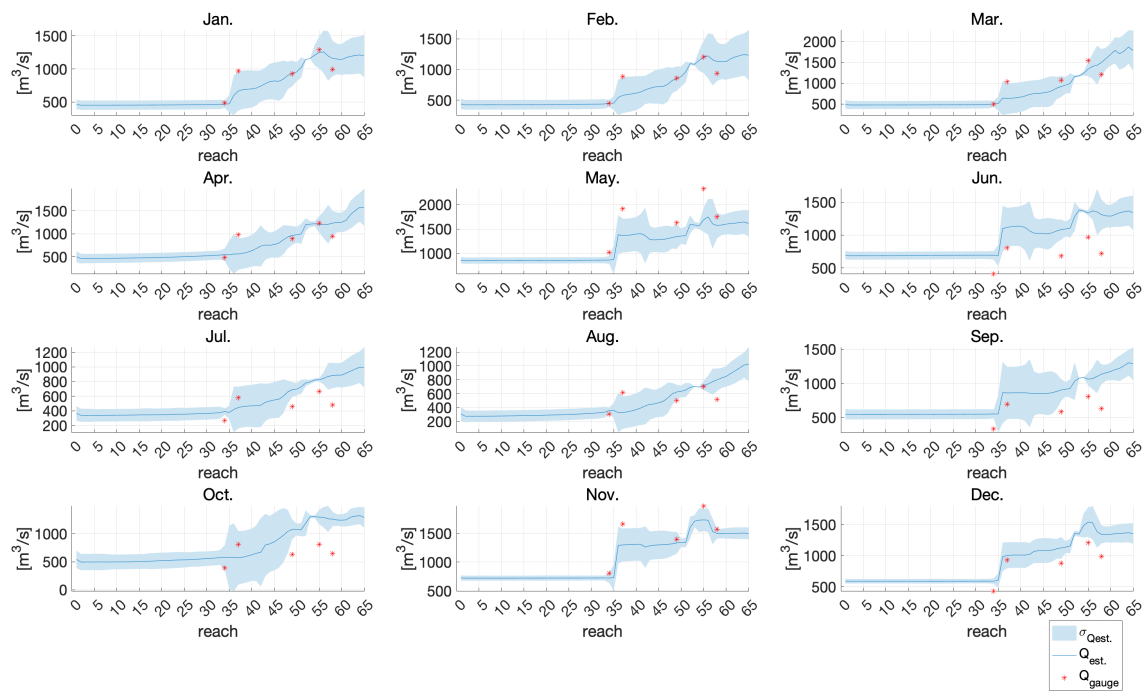


Figure A.1: Q_{est} . in 2002

Figure A.2: $Q_{est.}$ in 2003Figure A.3: $Q_{est.}$ in 2004

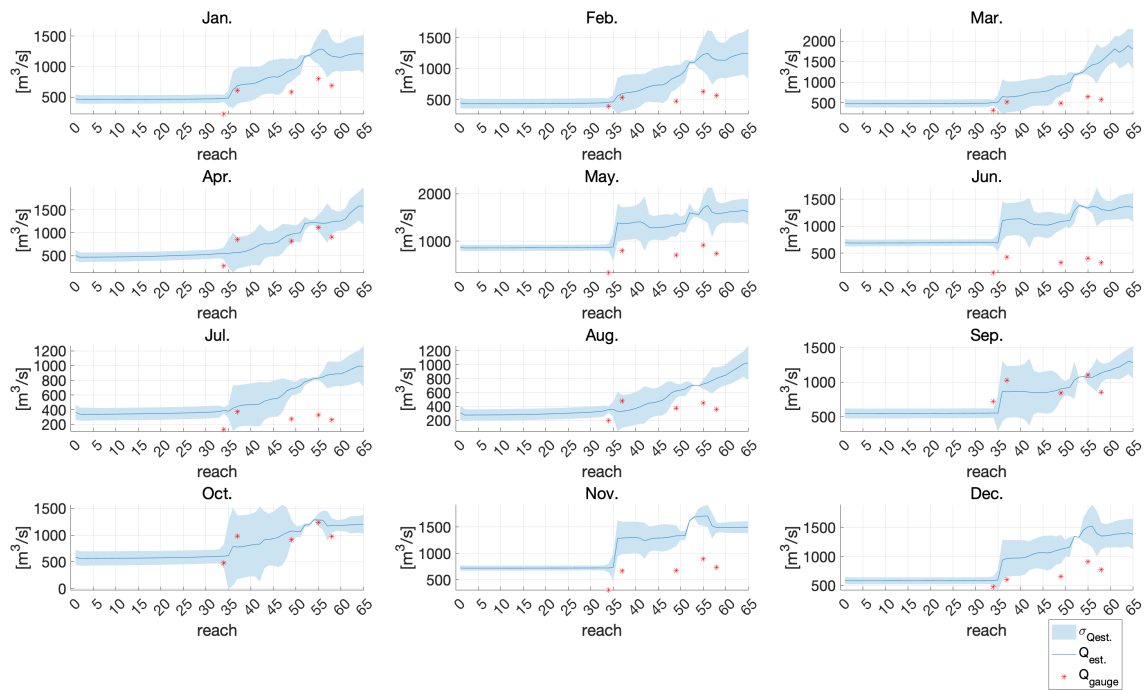


Figure A.4: $Q_{est.}$ in 2005

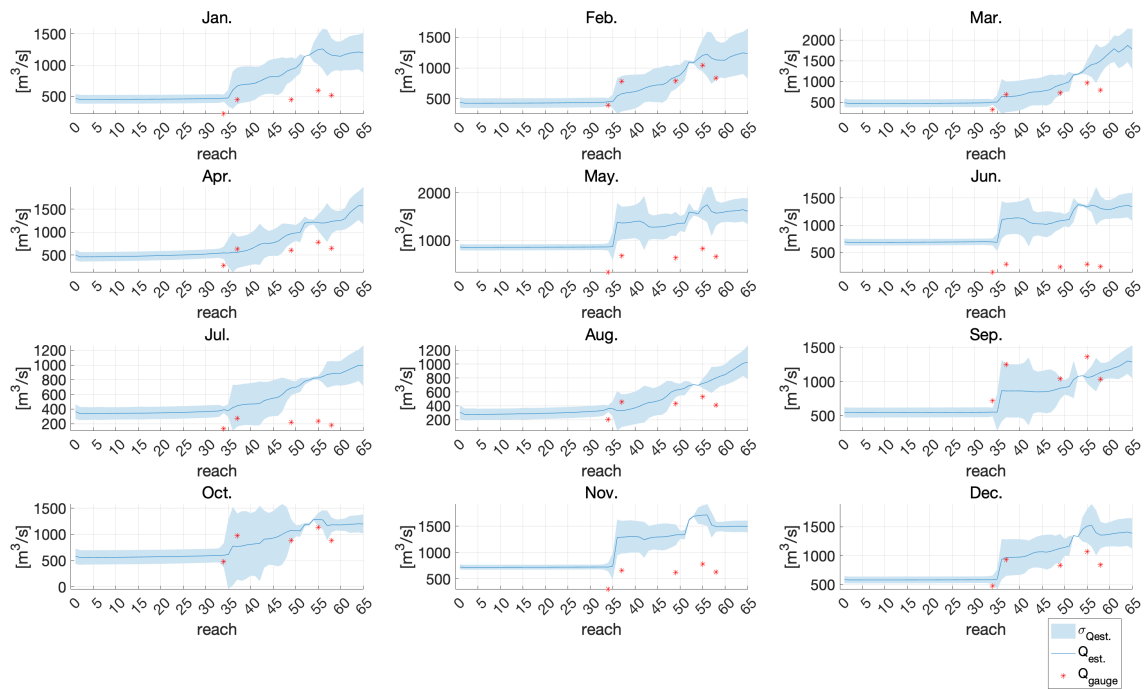
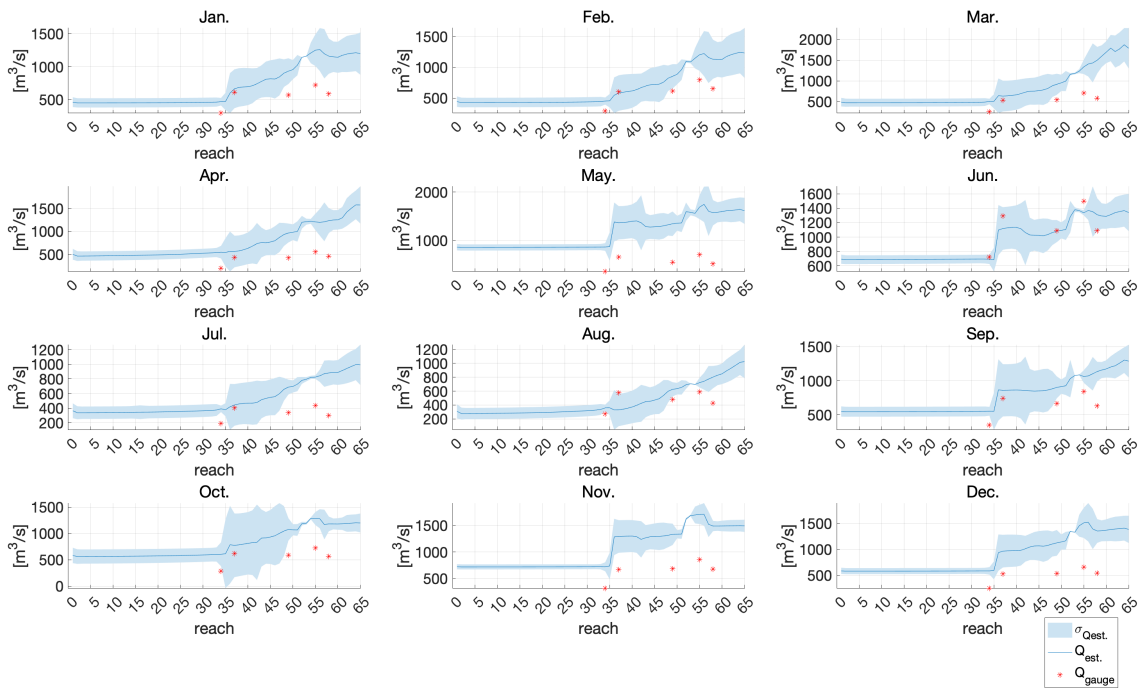
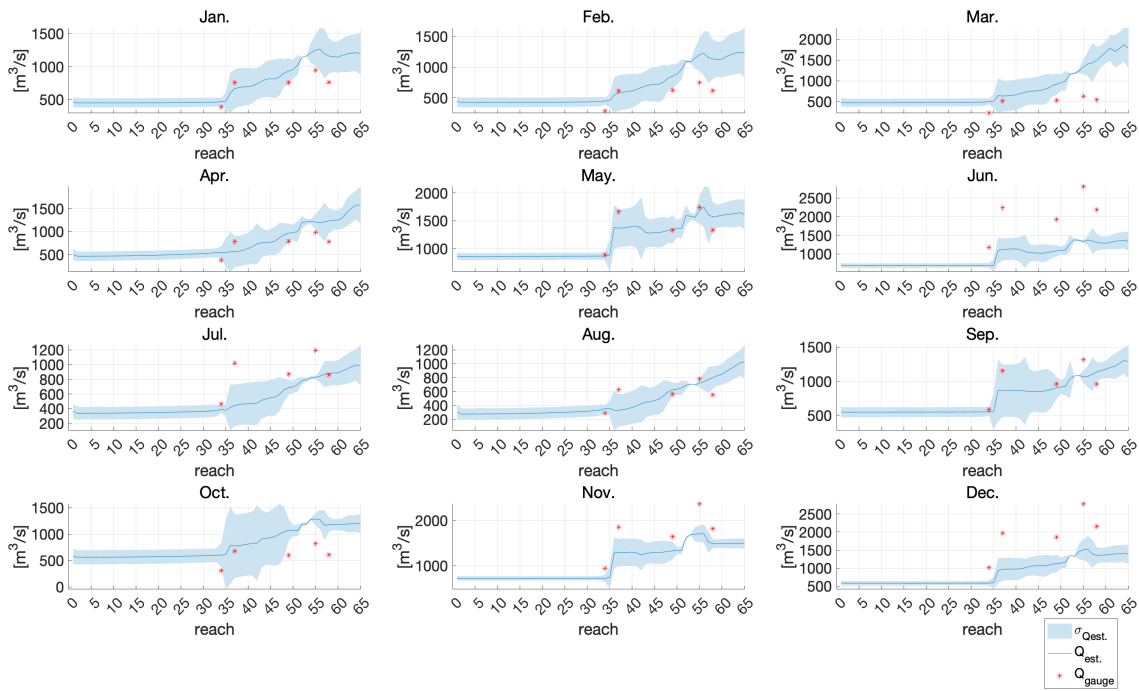


Figure A.5: $Q_{est.}$ in 2006

Figure A.6: $Q_{est.}$ in 2007Figure A.7: $Q_{est.}$ in 2008

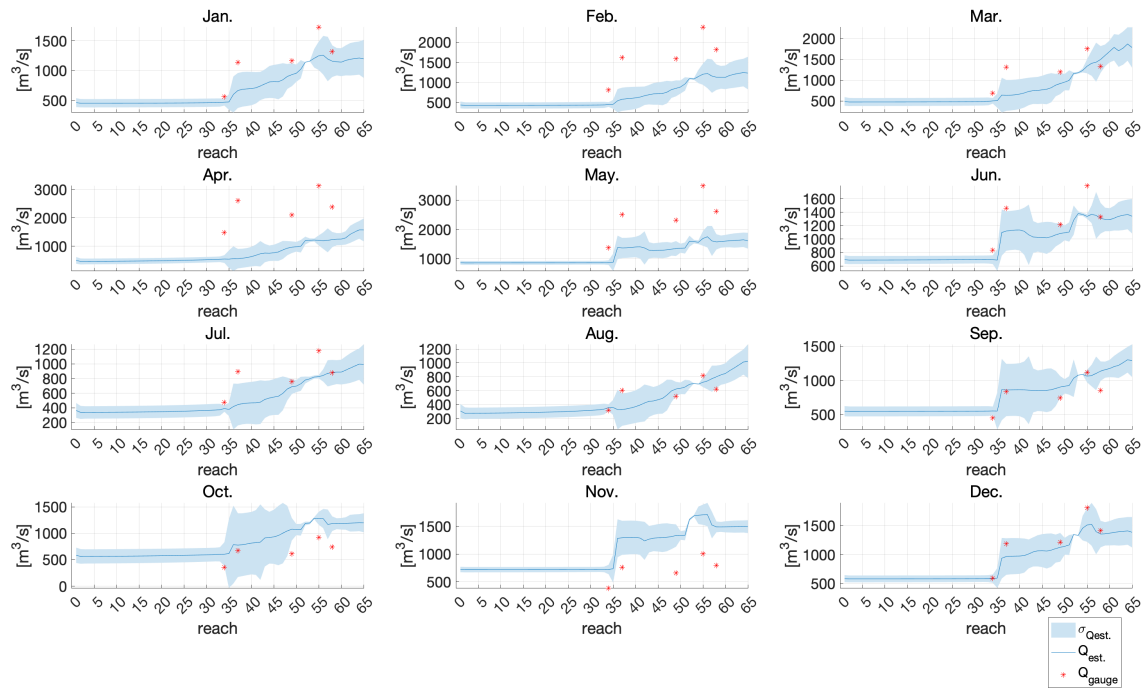


Figure A.8: $Q_{est.}$ in 2009

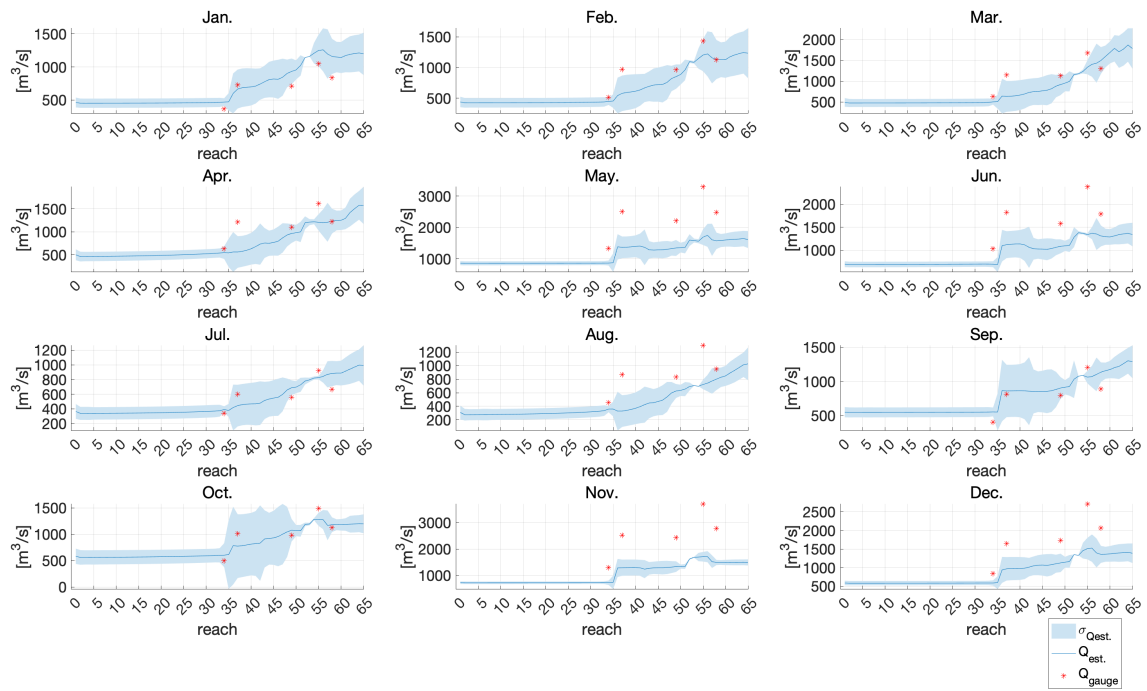
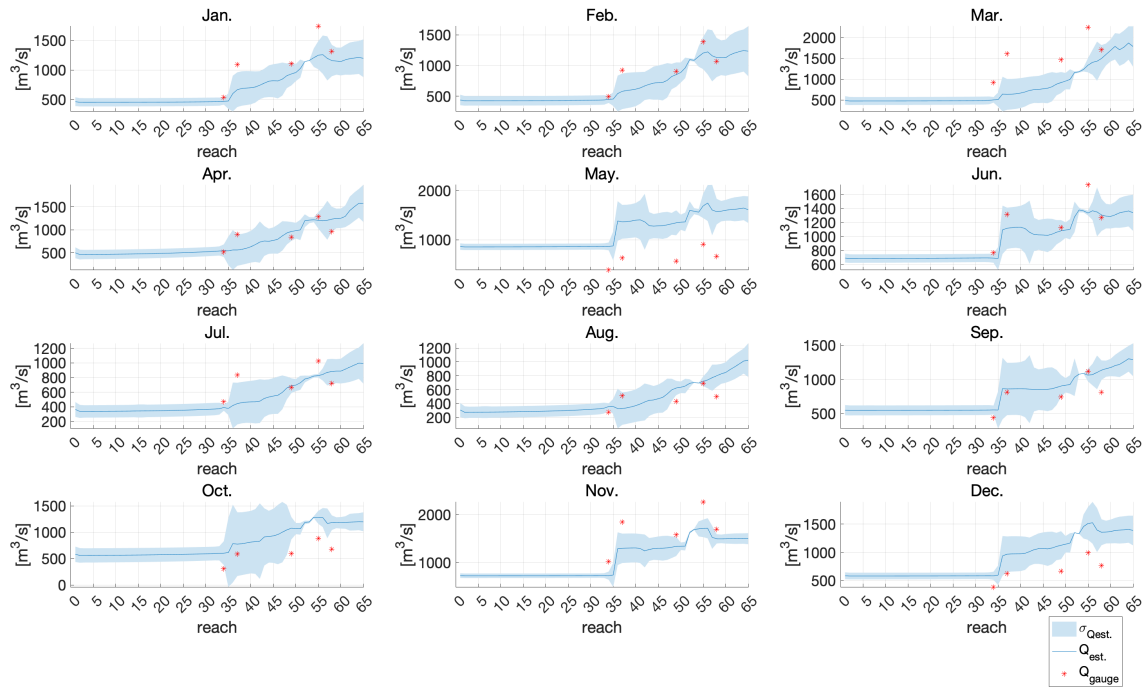
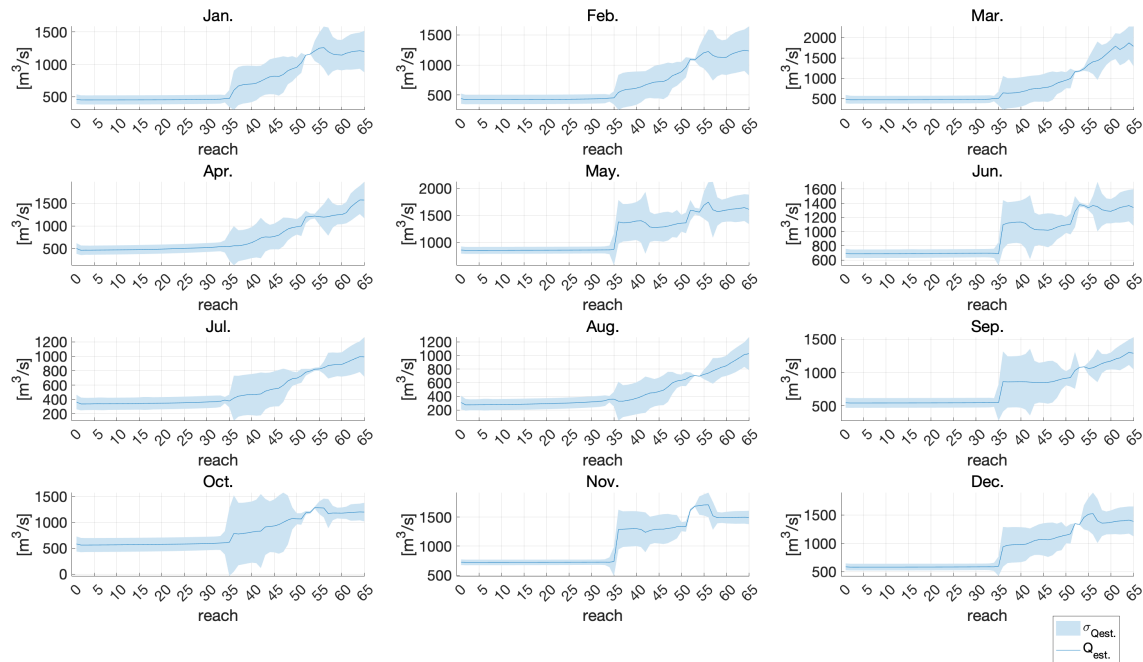


Figure A.9: $Q_{est.}$ in 2010

Figure A.10: Q_{est} . in 2011Figure A.11: Q_{est} . in 2012

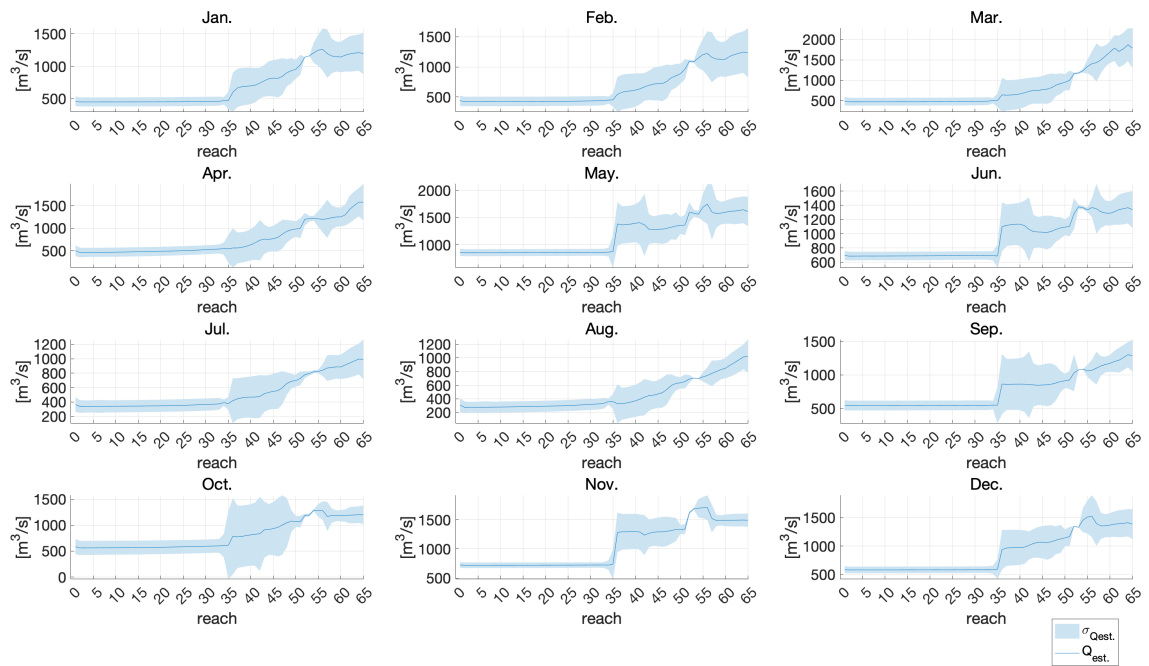


Figure A.12: Q_{est} . in 2013

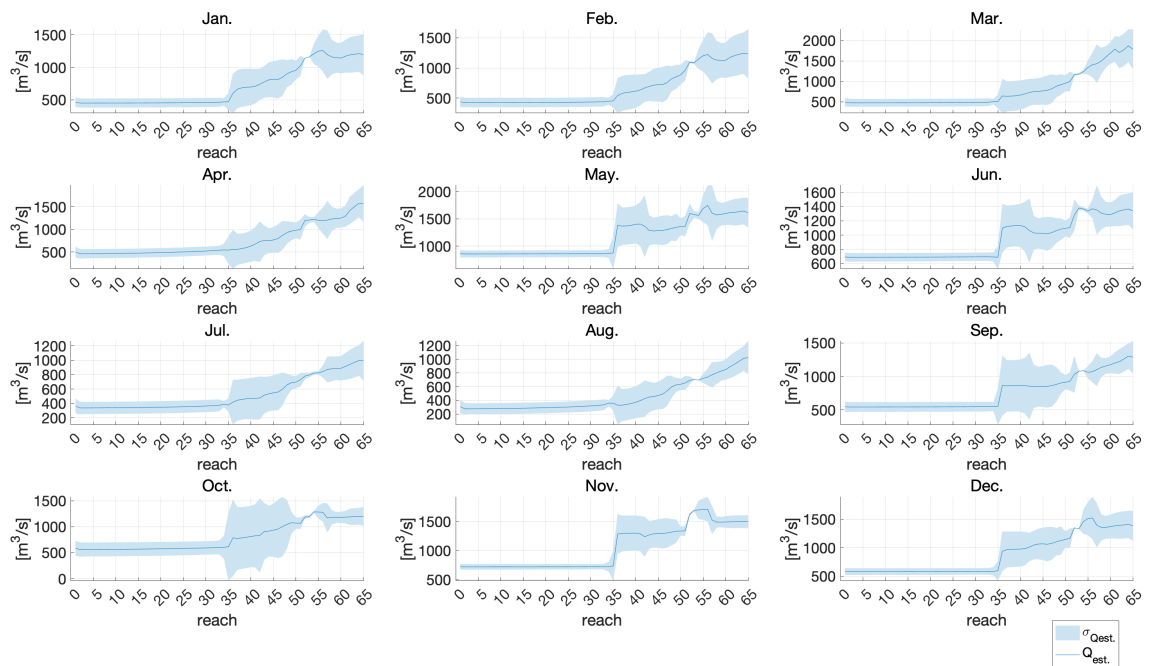
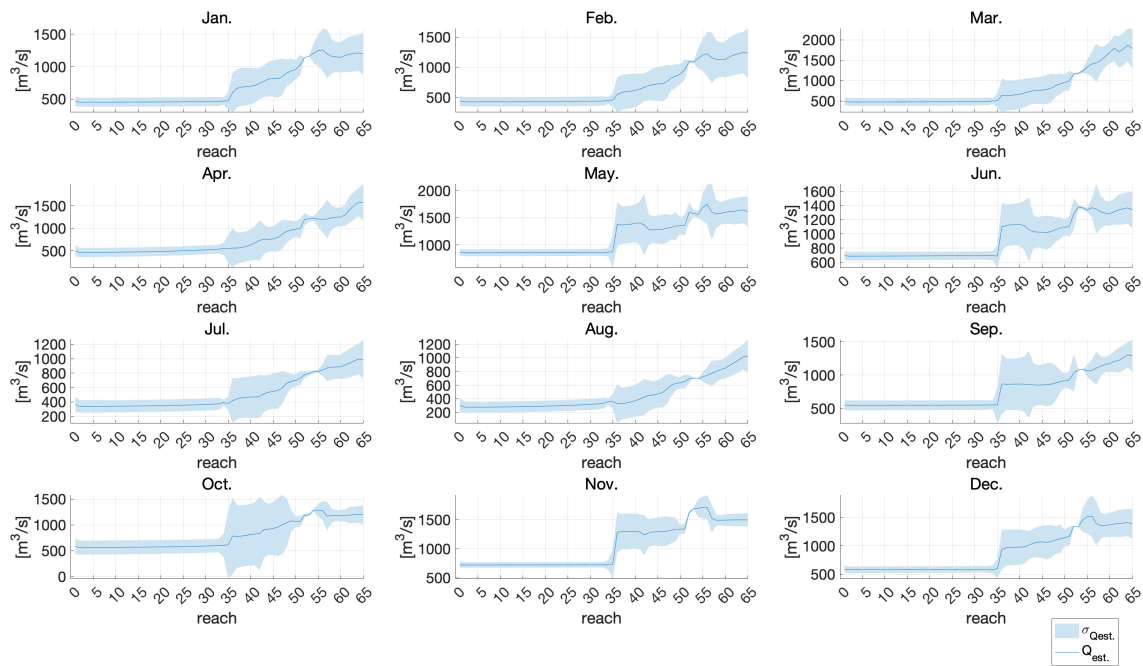
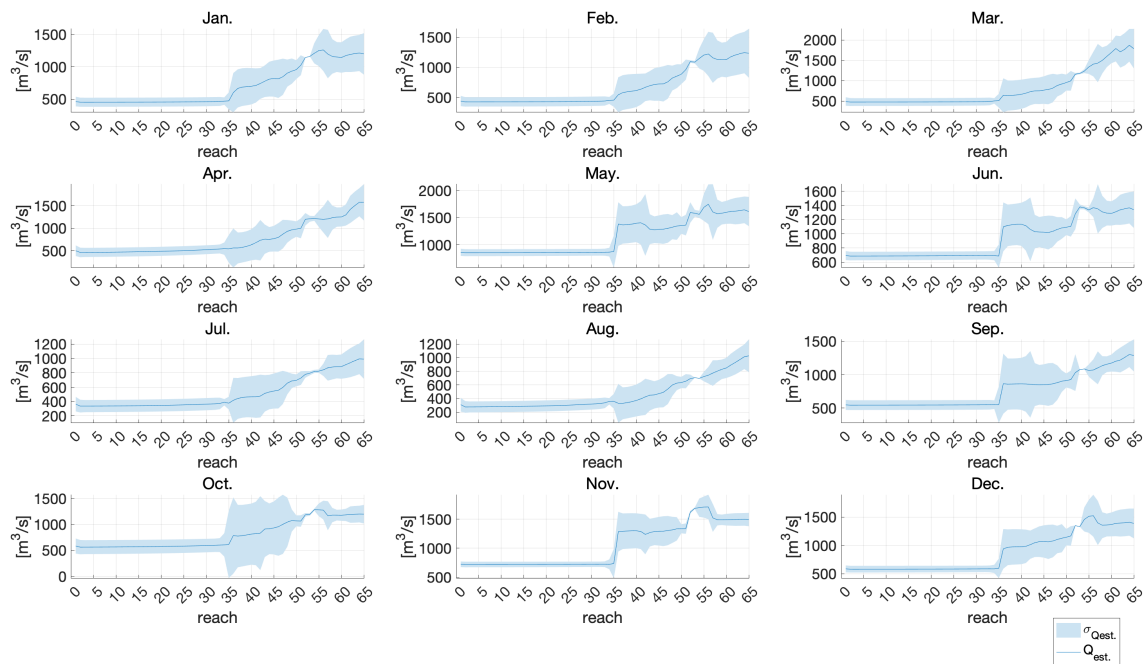


Figure A.13: Q_{est} . in 2014

Figure A.14: Q_{est} . in 2015Figure A.15: Q_{est} . in 2016

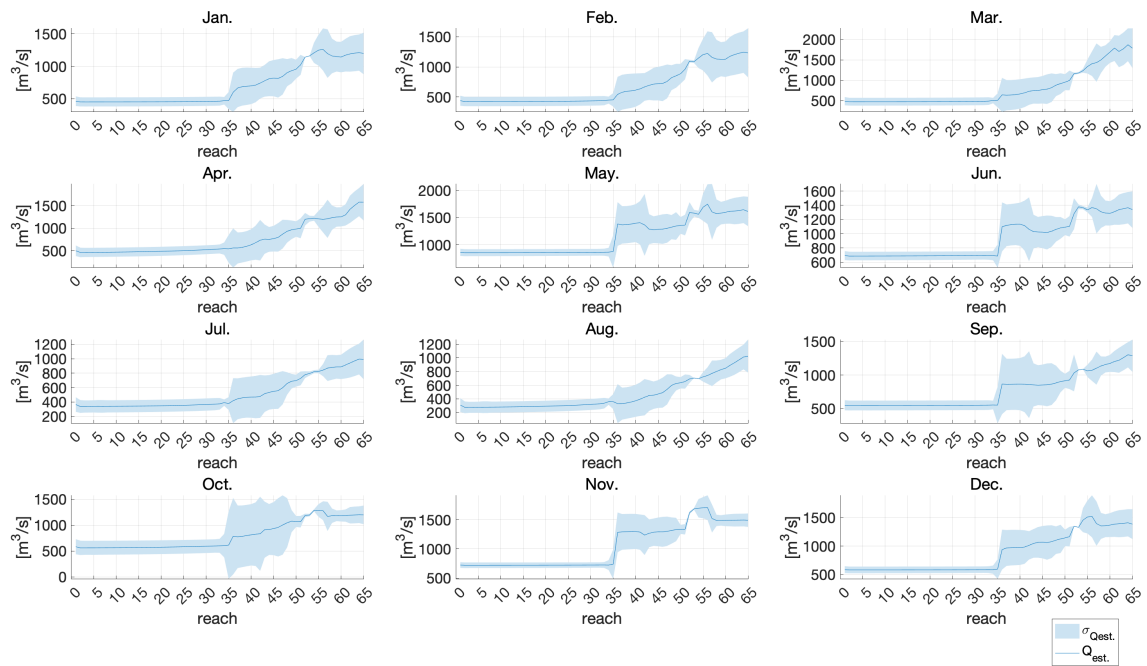


Figure A.16: $Q_{est.}$ in 2017

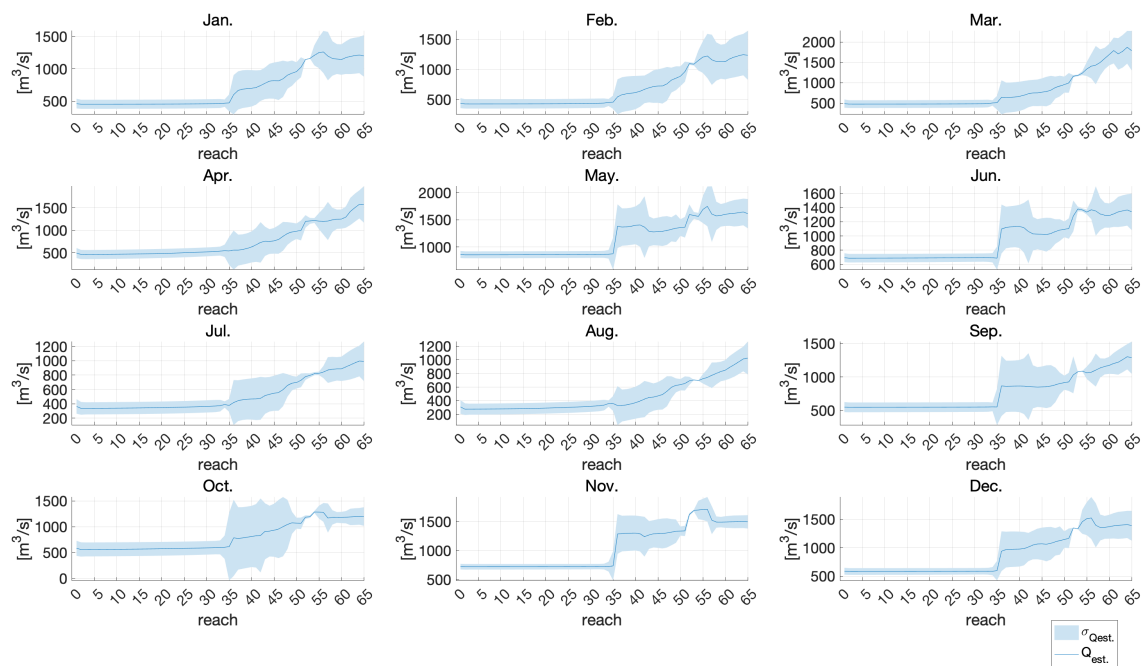
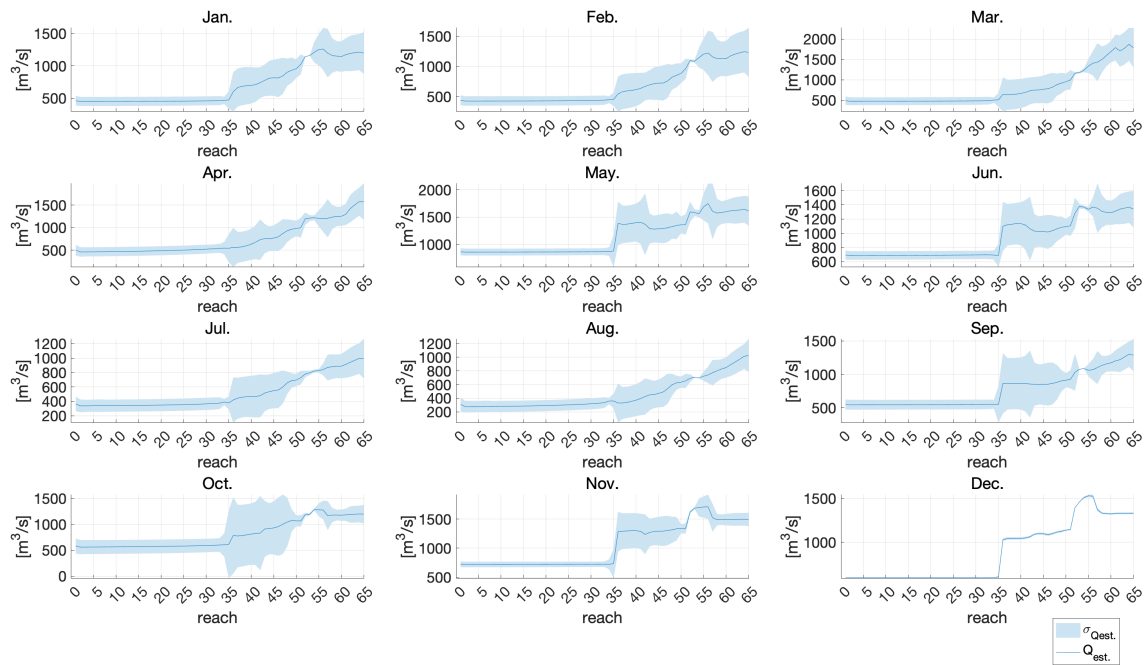


Figure A.17: $Q_{est.}$ in 2018

Figure A.18: Q_{est} . in 2019

Appendix B

Estimated posterior discharges

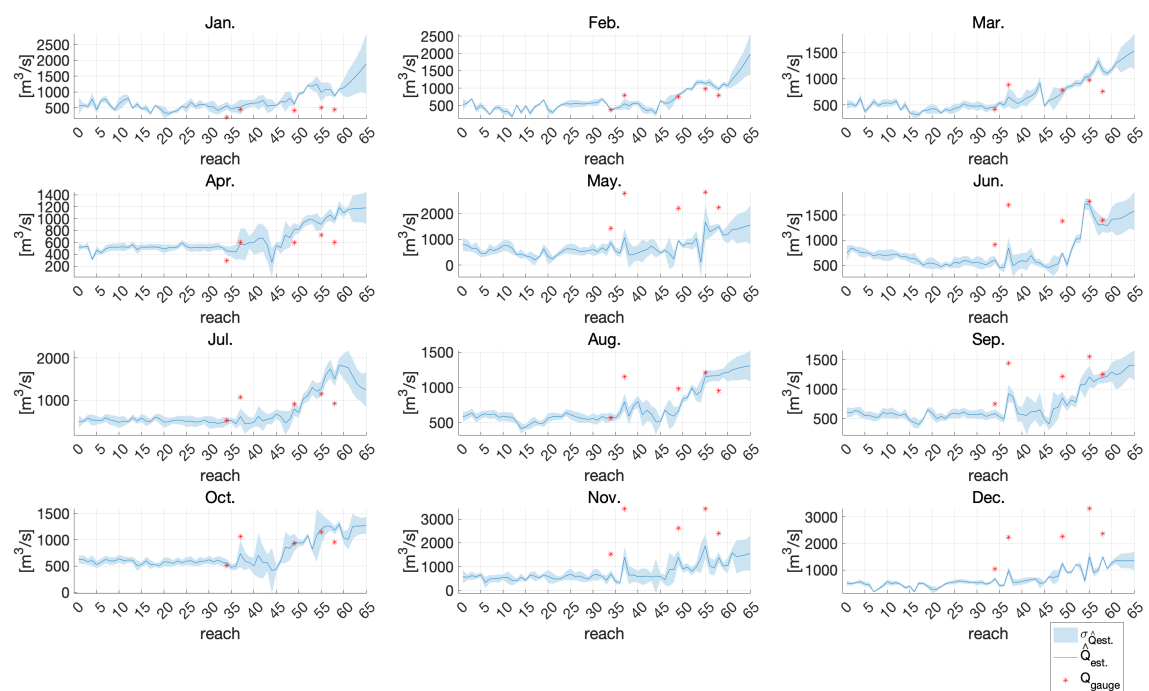


Figure B.1: \hat{Q}_{est} in 2002

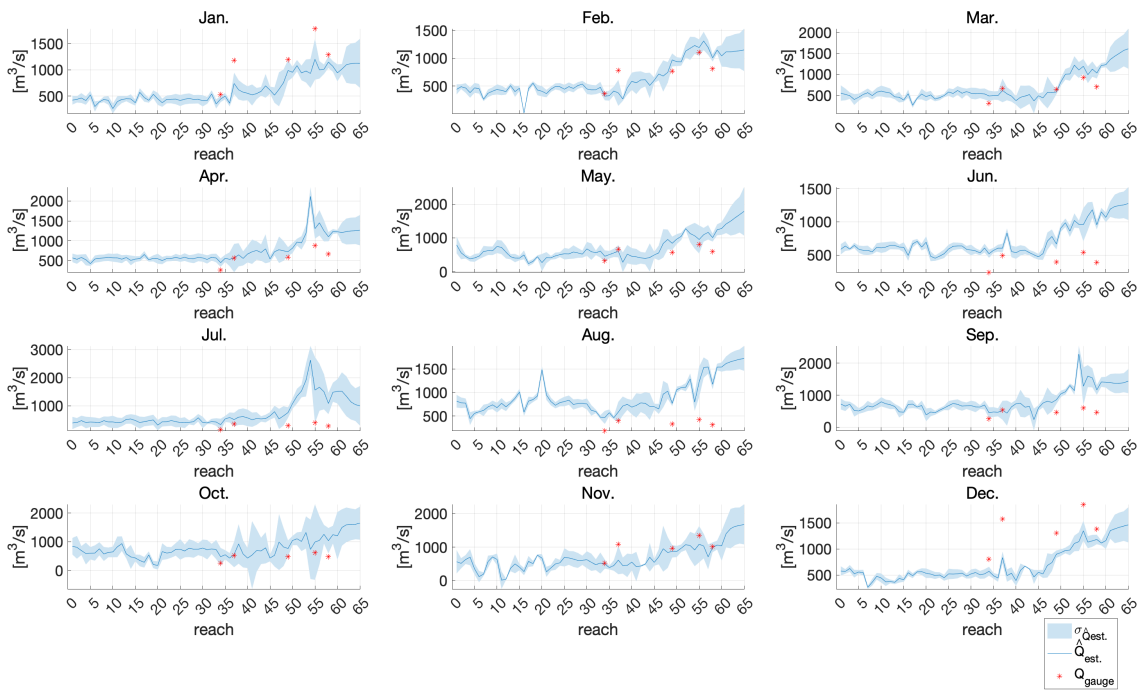


Figure B.2: $\hat{Q}_{est.}$ in 2003

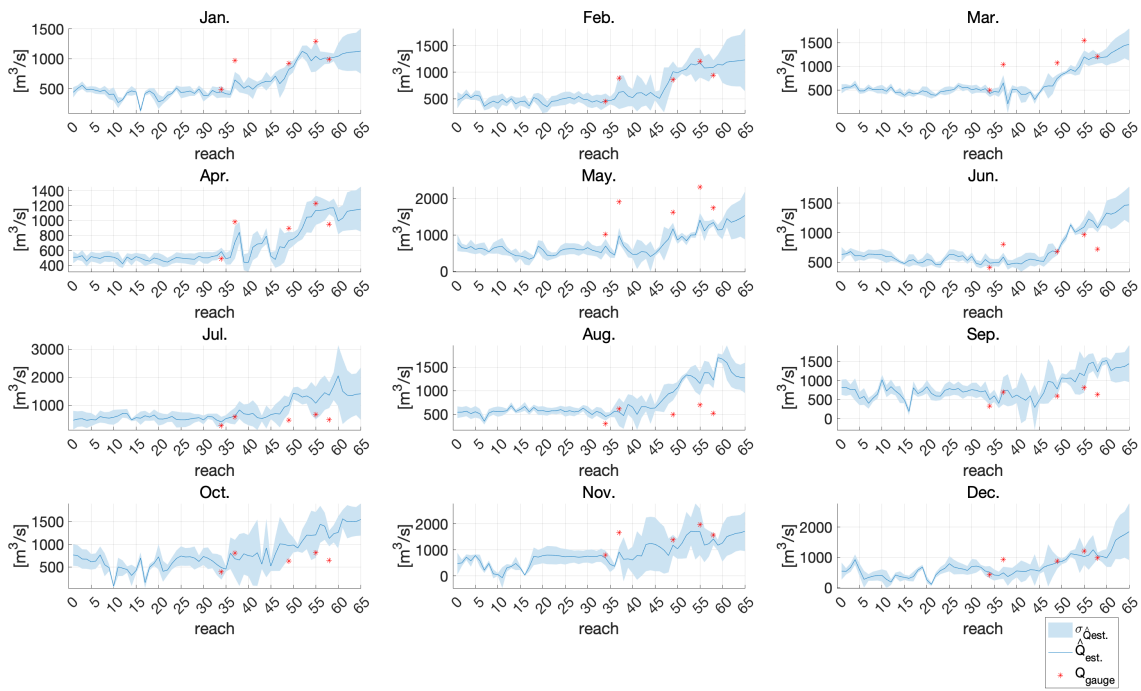


Figure B.3: $\hat{Q}_{est.}$ in 2004

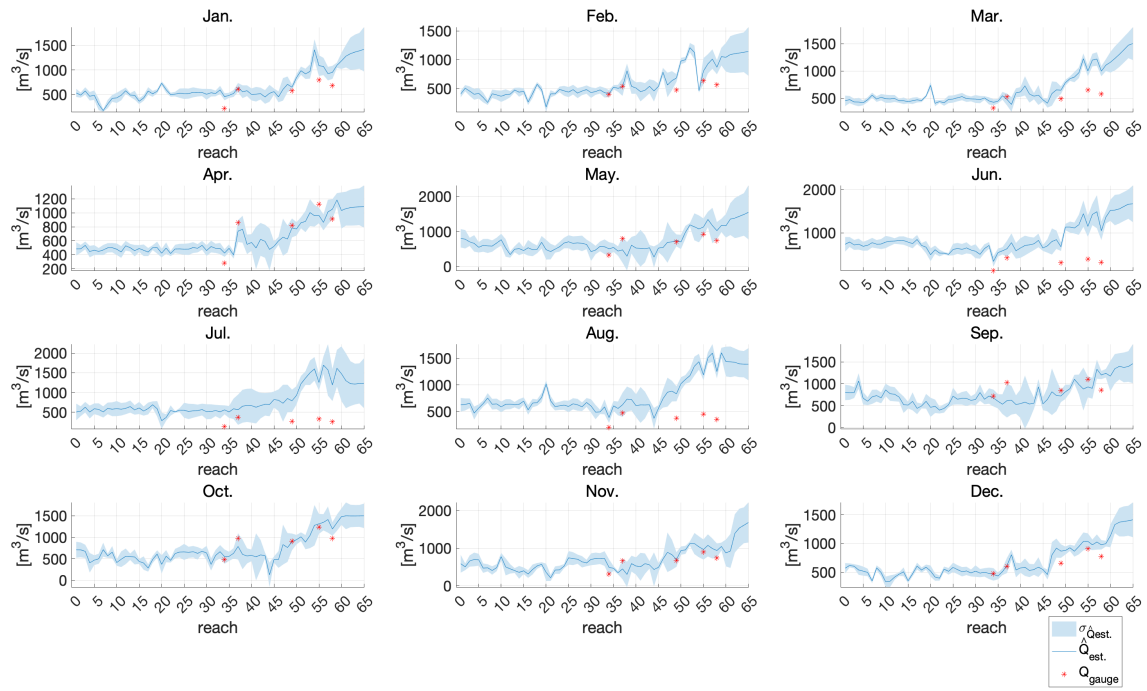


Figure B.4: $\hat{Q}_{est.}$ in 2005

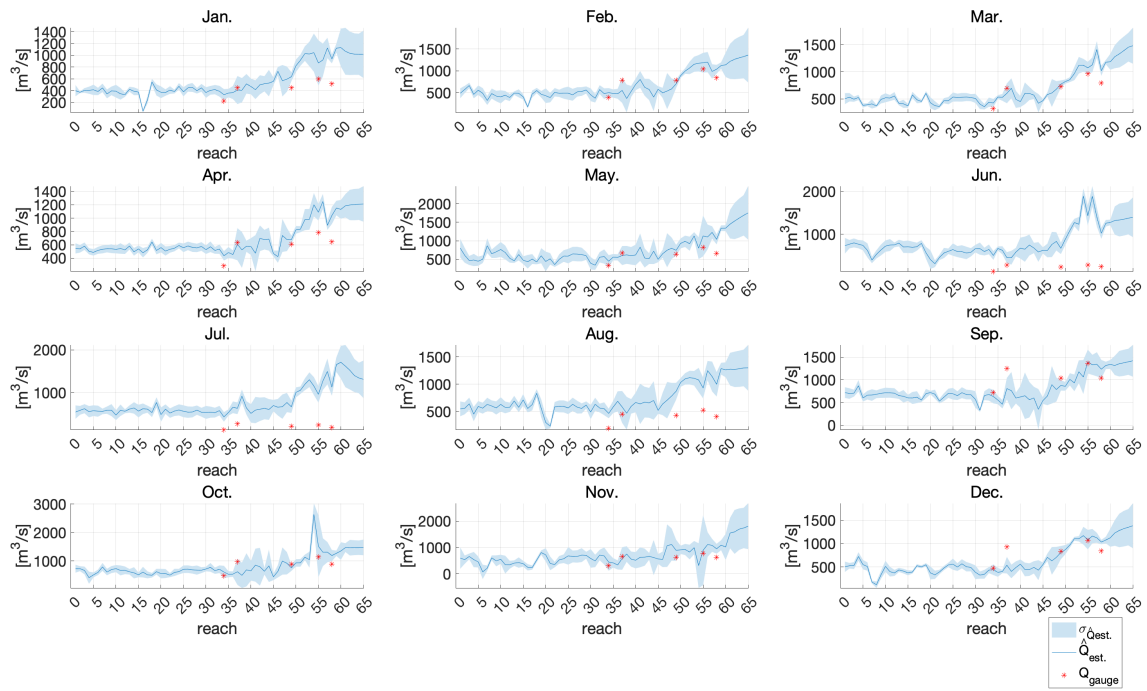
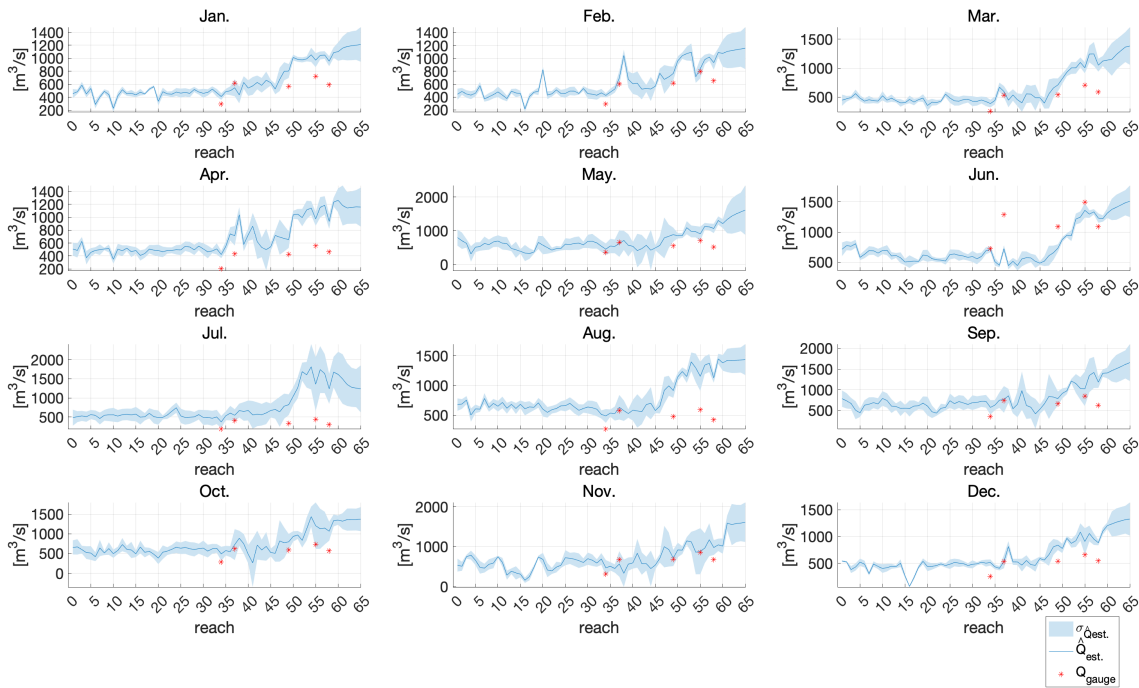
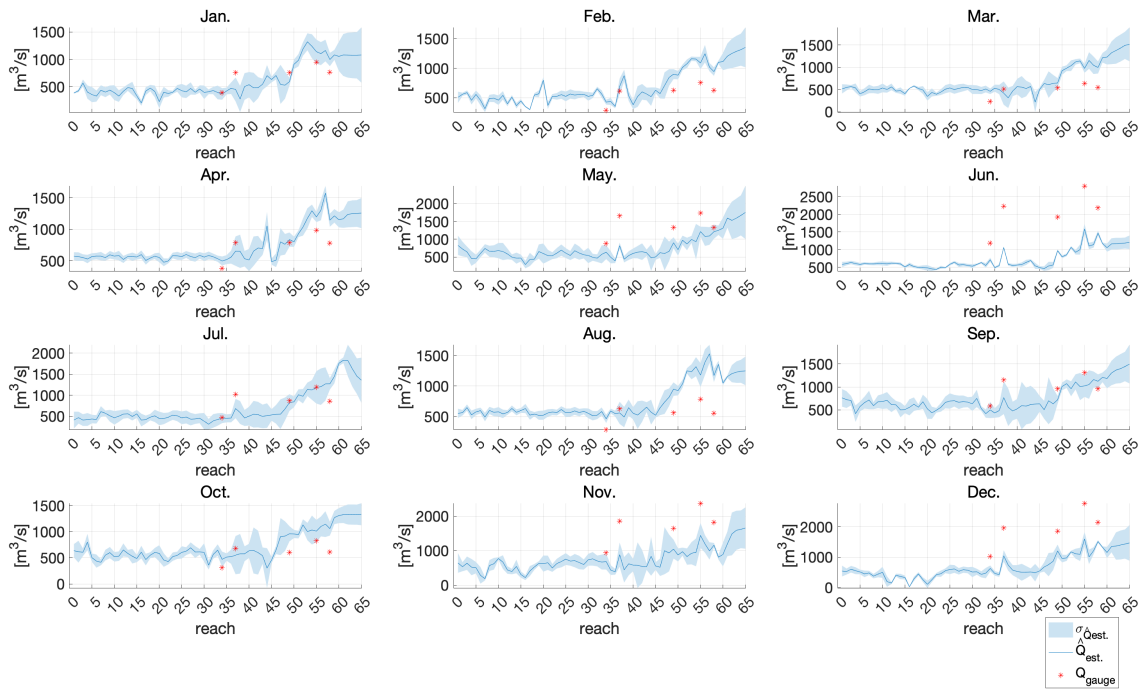


Figure B.5: $\hat{Q}_{est.}$ in 2006

Figure B.6: \hat{Q}_{est} . in 2007Figure B.7: \hat{Q}_{est} . in 2008

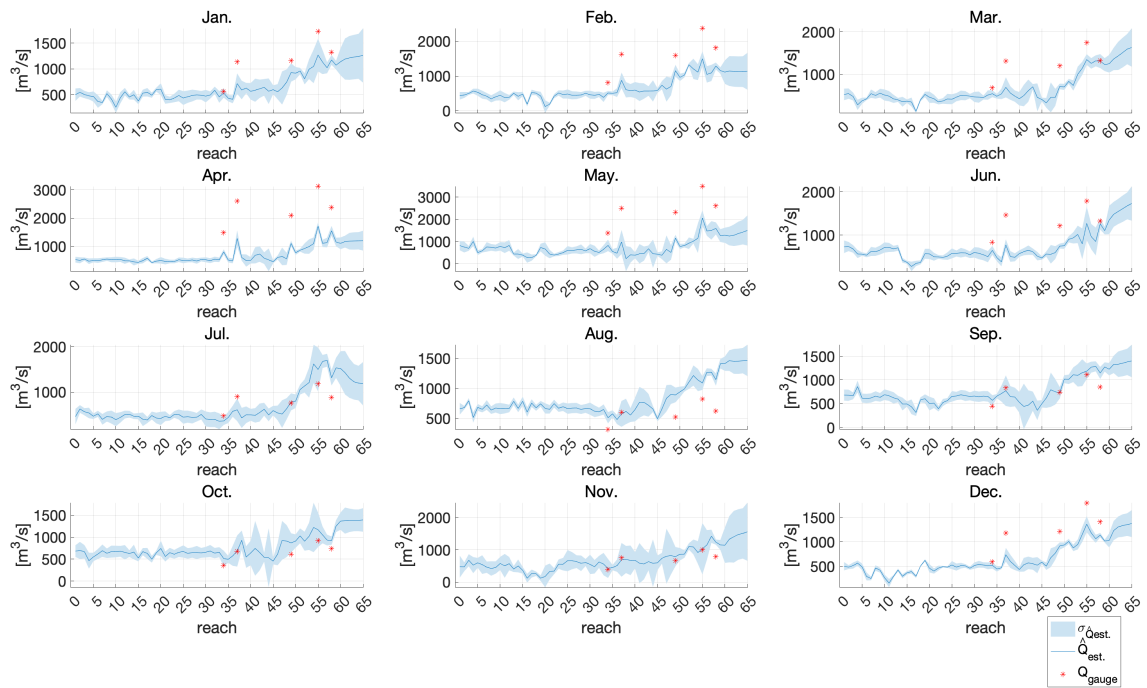


Figure B.8: $\hat{Q}_{est.}$ in 2009



Figure B.9: $\hat{Q}_{est.}$ in 2010



Figure B.10: \hat{Q}_{est} . in 2011

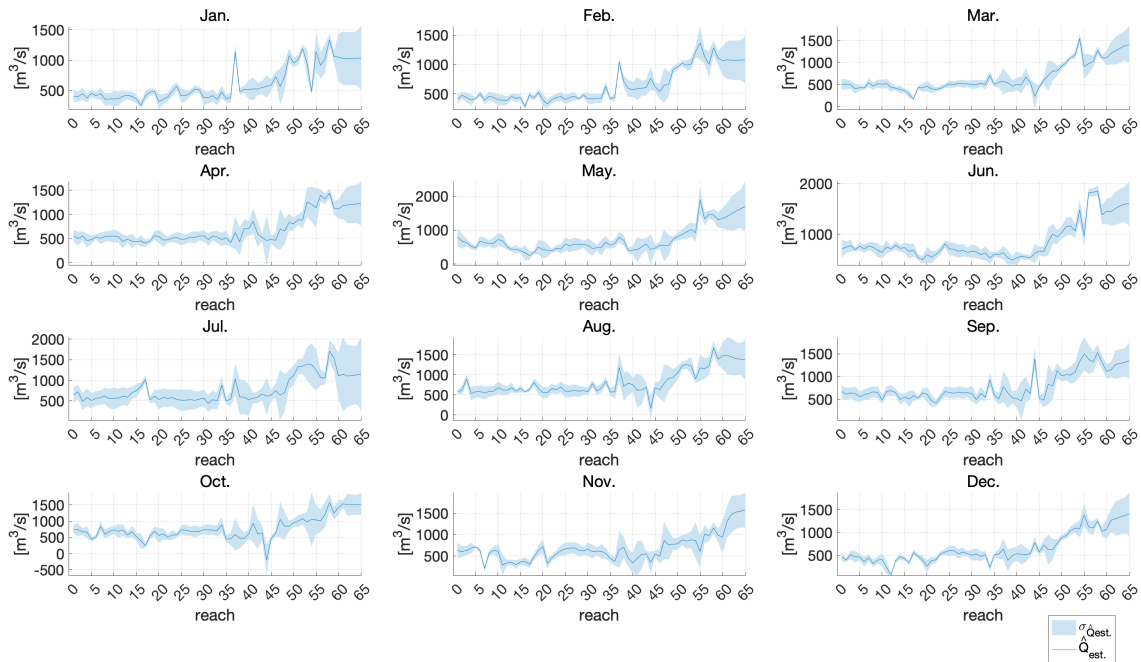


Figure B.11: \hat{Q}_{est} . in 2012

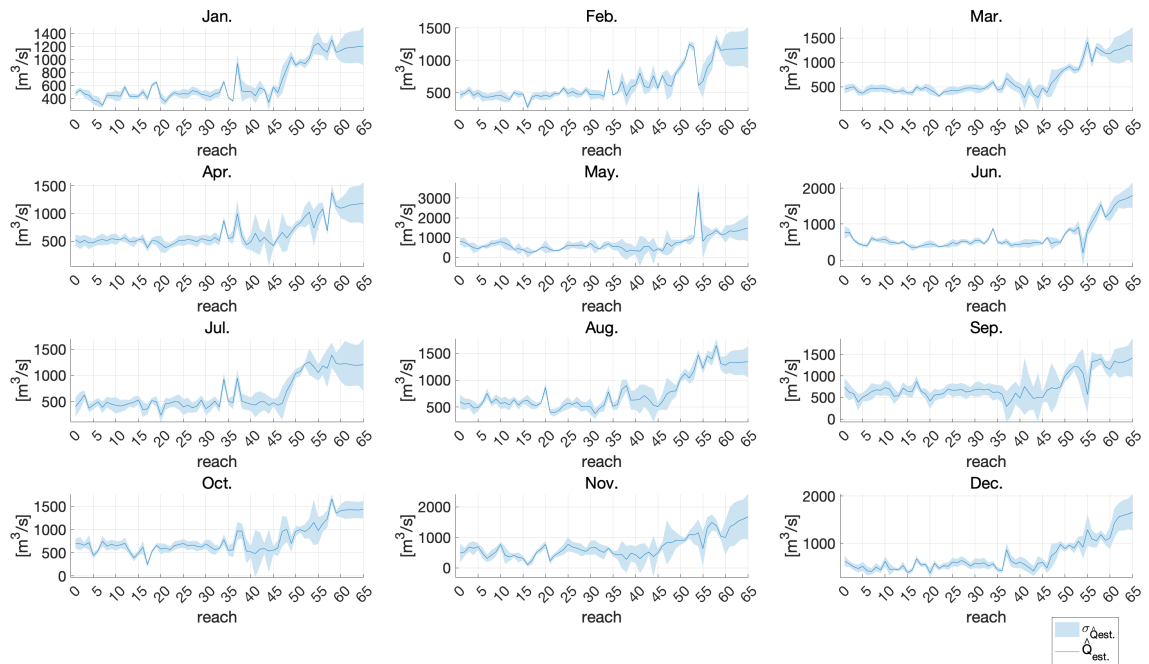


Figure B.12: $\hat{Q}_{est.}$ in 2013

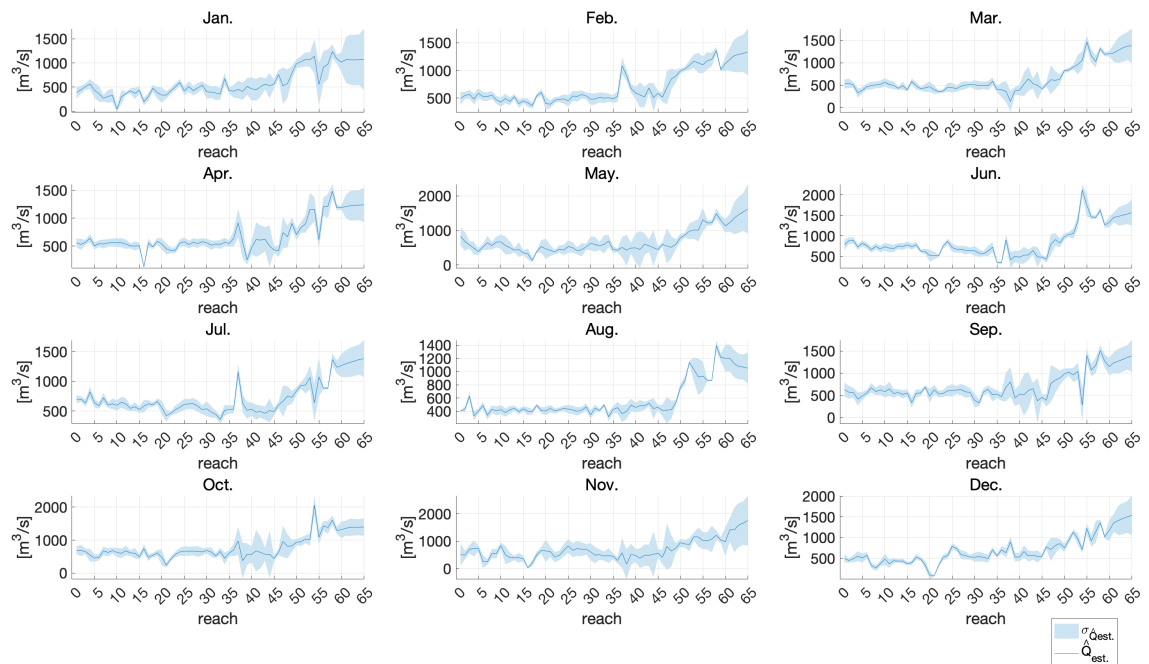


Figure B.13: $\hat{Q}_{est.}$ in 2014

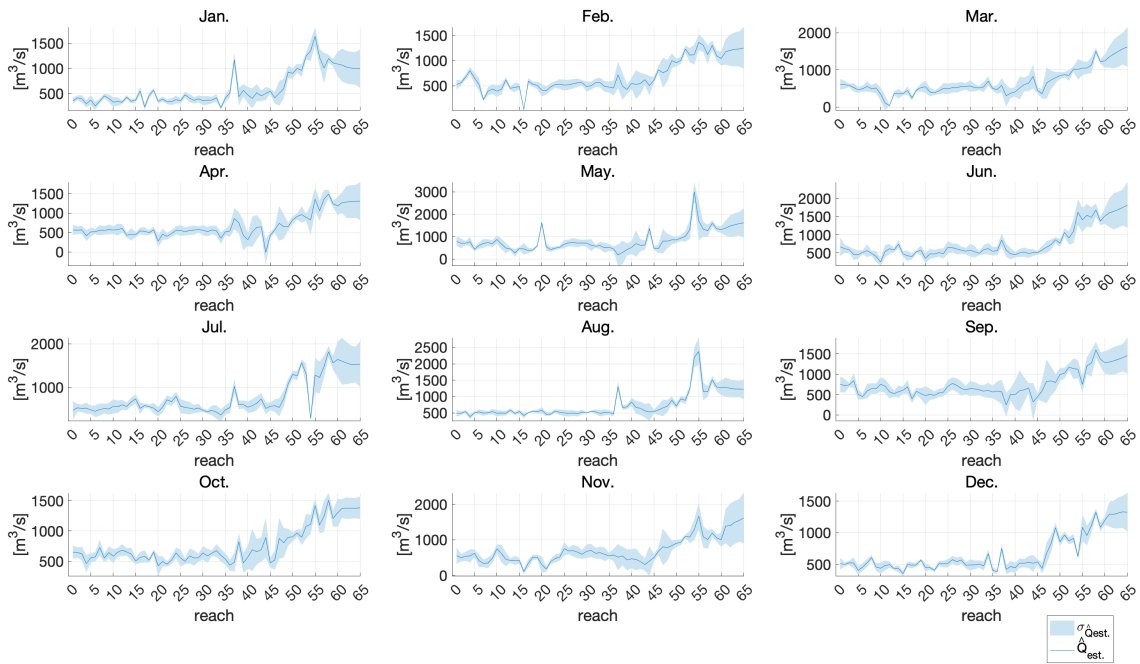


Figure B.14: $\hat{Q}_{est.}$ in 2015

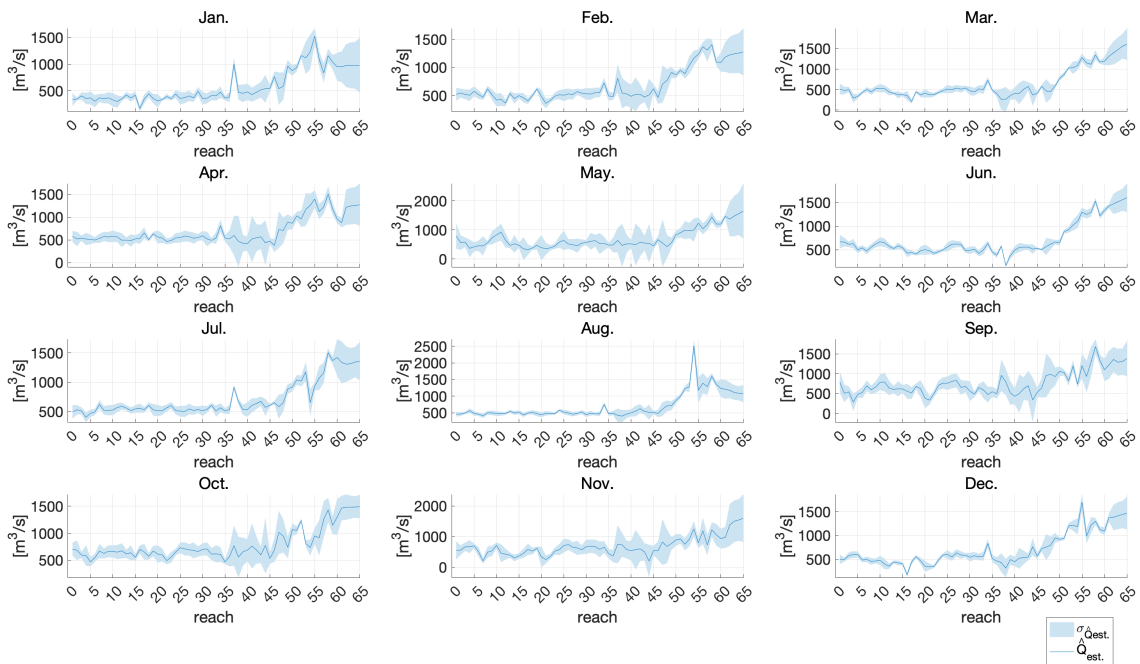


Figure B.15: $\hat{Q}_{est.}$ in 2016

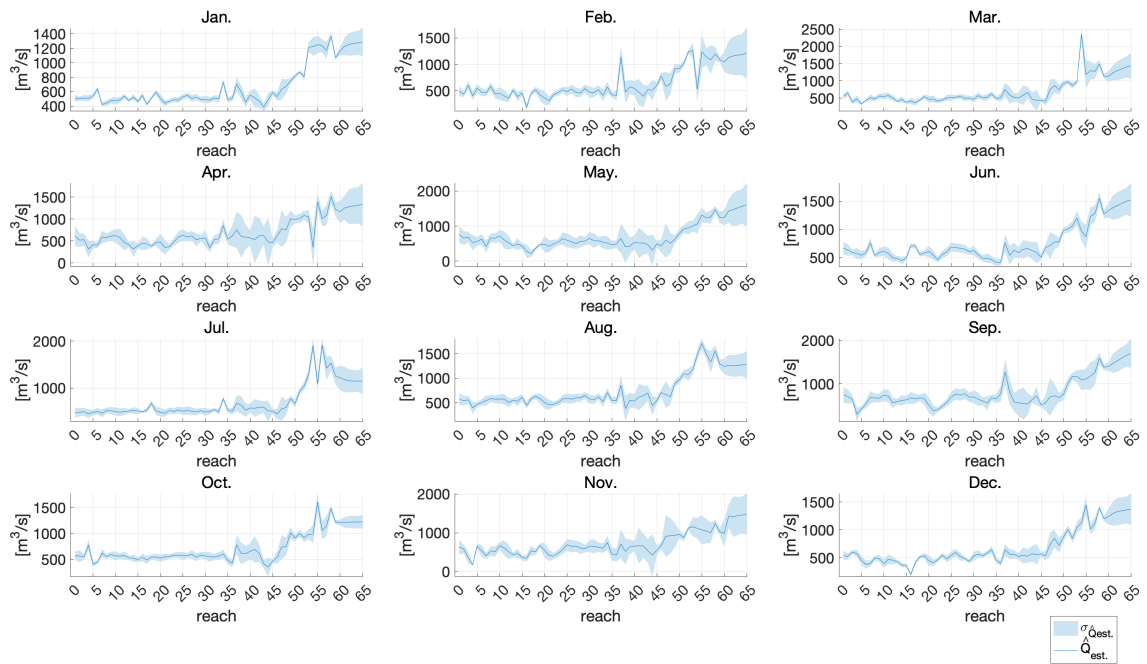


Figure B.16: $\hat{Q}_{est.}$ in 2017

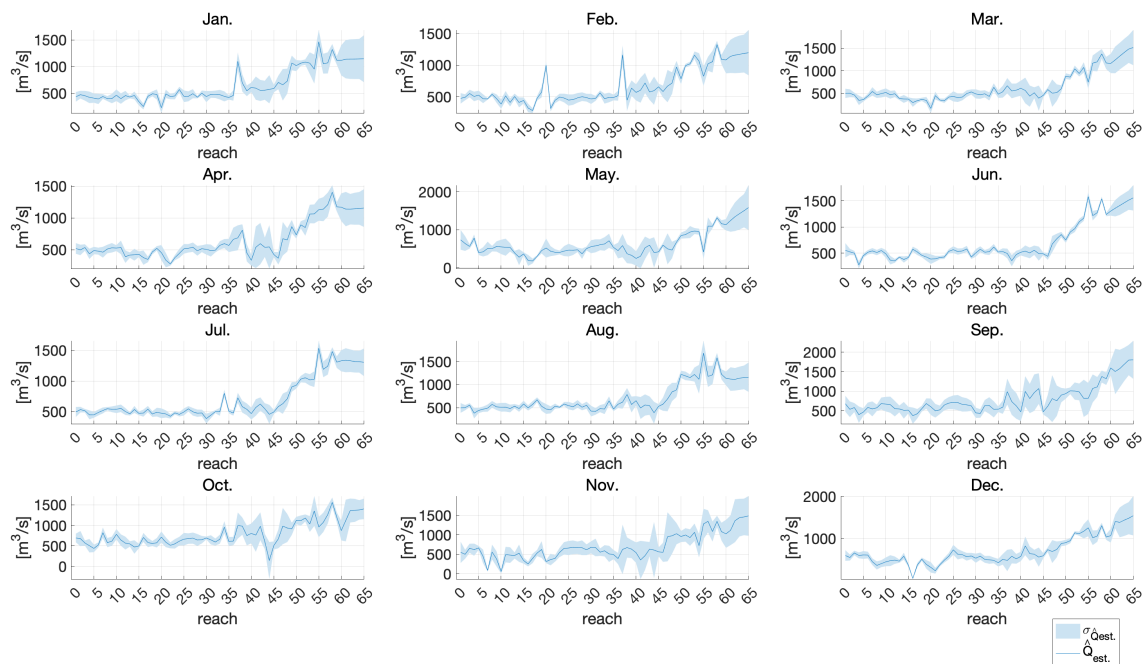


Figure B.17: $\hat{Q}_{est.}$ in 2018

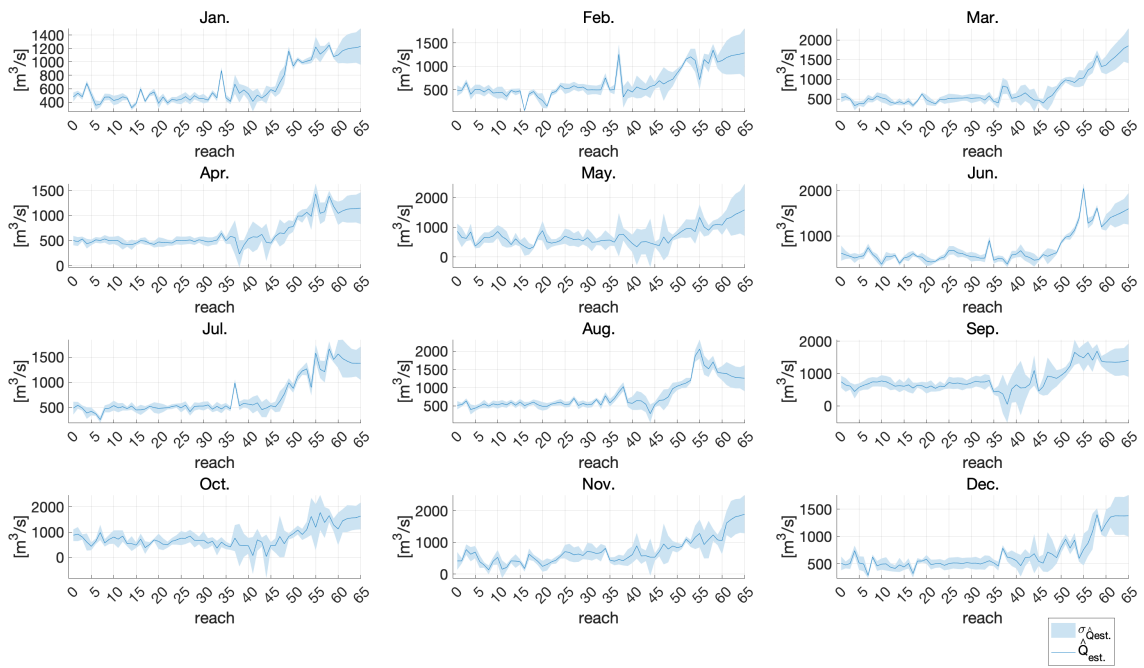


Figure B.18: \hat{Q}_{est} . in 2019

References

- Altenau, E. H., Pavelsky, T. M., Durand, M. T. & Frasson, R. P. d. M. (2020), 'Sword: Product description document'.
URL: <http://gaia.geosci.unc.edu/SWORD>
- Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M. & Bendezu, L. (2021), 'The surface water and ocean topography (swot) mission river database (sword): A global river network for satellite data products', *Water Resources Research* **57**(7), e2021WR030054. e2021WR030054 2021WR030054.
URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021WR030054>
- Britannica (2019), 'Po river', *Encyclopedia Britannica* .
URL: <https://www.britannica.com/place/Po-River>
- Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006), 'Measurement error in nonlinear models: A modern perspective, second edition (2nd ed.)', *Chapman and Hall/CRC* .
- Durand, M. (2018), 'Computing cross-sectional area changes from height and width observations'.
URL: <https://github.com/mikedurand/SWOTAprimeCalcs/blob/master>
- Durand, M., Fu, L.-L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E. & Esteban-Fernandez, D. (2010), 'The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies', *Proceedings of the IEEE* **98**(5), 766–779.
- Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., Rodriguez, E., Bates, P. D., Pavelsky, T. M., Monnier, J., Chen, X., Di Baldassarre, G., Fiset, J.-M., Flipo, N., Frasson, R. P. d. M., Fulton, J., Goutal, N., Hossain, F., Humphries, E., Minear, J. T., Mukolwe, M. M., Neal, J. C., Ricci, S., Sanders, B. F., Schumann, G., Schubert, J. E. & Vilmin, L. (2016), 'An inter-comparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope', *Water Resources Research* **52**(6), 4527–4549.

URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR018434>

Durand, M., Gleason, C. J., Pavelsky, T. M. & de Moraes Frasson, R. P. (2021), 'A framework for estimating global river discharge from the surface water and ocean topography satellite mission', *Water Resources Research* .

Durand, M., Neal, J., Rodríguez, E., Andreadis, K. M., Smith, L. C. & Yoon, Y. (2014), 'Estimating reach-averaged discharge for the river severn from measurements of river water surface elevation and slope', *Journal of Hydrology* **511**, 92–104.

URL: <https://www.sciencedirect.com/science/article/pii/S002216941400002X>

Durand, M. T., Gleason, C. J., Frasson, R. P. M., Pavelsky, T., Allen, G. H., Bates, P. D., Dudley, R. W., Emery, C. M., Fenoglio-Marc, L., Garambois, P. A., Hossain, F., Larnier, K., Lin, P., Matte, P., Monnier, J., Oubanas, H., Pan, M., Rodriguez, E., Schaperow, J., Tarpanelli, A., Tourian, M. J. & Wang, J. (2020), A global framework for SWOT discharge, in 'AGU Fall Meeting Abstracts', Vol. 2020, pp. H040–0012.

Elmer, N. J. (2020), 'Cnes large scale swot simulator'.

URL: <https://swot.jpl.nasa.gov/system/documents/files>

Elmer, N. J., Hain, C., Hossain, F., Desroches, D. & Pottier, C. (2020), 'Generating proxy swot water surface elevations using wrf-hydro and the cnes swot hydrology simulator', *Water Resources Research* **56**(8), e2020WR027464.

URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027464>

Elmi, O., Tourian, M. J. & Sneeuw, N. (2016), 'Dynamic river masks from multi-temporal satellite imagery: An automatic algorithm using graph cuts optimization', *Remote Sensing* **8**(12).

URL: <https://www.mdpi.com/2072-4292/8/12/1005>

Fekete, B. & Vörösmarty, C. (2002), 'The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements', *Proceedings of the PUB Kick-off Meeting* **309**, 20–22.

Fugazza, M., Magri, P., Natale, L. & Savi, F. (2008), 'The october 2000 flood at the po river and sesia river junction', *WIT Transactions on Ecology and the Environment* **118**, 37–45.

JPL (2021), 'Swot: Overview'.

URL: <https://swot.jpl.nasa.gov/mission/overview/>

- Manning, R., Griffith, J. P., Pigot, T. F. & F., V.-H. L. (1891), *On the flow of water in open channels and pipess*, Transactions of the Institution of Civil Engineers of Ireland, Ireland.
- Oki, T. & Kanae, S. (2006), 'Global hydrological cycles and world water resources', *Science (New York, N.Y.)* **313**, 1068–72.
- Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C., Allen, G. H. & Miller, Z. F. (2014), 'Assessing the potential global extent of swot river discharge observations', *Journal of Hydrology* **519**, 1516–1525.
URL: <https://www.sciencedirect.com/science/article/pii/S0022169414006544>
- Tourian, M. J., Schwatke, C. & Sneeuw, N. (2017), 'River discharge estimation at daily resolution from satellite altimetry over an entire river basin', *Journal of Hydrology* **546**, 230–247.
URL: <https://www.sciencedirect.com/science/article/pii/S0022169417300100>
- Tourian, M. J., Tarpanelli, A., Elmi, O., Qin, T., Brocca, L., Moramarco, T. & Sneeuw, N. (2016), 'Spatiotemporal densification of river water level time series by multimission satellite altimetry', *Water Resources Research* **52**(2), 1140–1159.
URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017654>