University of Stuttgart

Institute of Geodesy

Estimation of inter-satellite and inter-track biases of satellite altimetry missions over lakes and reservoirs using surface area from satellite imagery

Masterthesis im Studiengang

GEOENGINE

University of Stuttgart

Shuhua Yu

Stuttgart, June 2022

Supervisor: Dr.-Ing. Mohammad Tourian
Universität 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

Including lakes, reservoirs, and rivers, inland water bodies cover only a small portion of the Earth's surface. However, they play an important role in maintaining life on Earth, the global water cycle, and climate change. Due to the declining number of gauge stations that provide the in-situ data, the muti-mission satellite altimetry has been applied to the monitoring of medium to large lakes and reservoirs, which enables computing a water-level time series with improved temporal and spatial resolution. However, inter-satellite and inter-track biases are still a problem for multi-mission. There have been studies conducted to determine absolute altimetry biases at calibration locations and global altimetry biases. But we still don't know everything about how satellites are biased over inland waters.

This thesis is dedicated to developing a method to resolve the biases between satellites and tracks over lakes and reservoirs. Our strategy for calculating the biases between overlapping and non-overlapping time series of water levels from various missions and tracks is based on satellite-derived time series of water area. With the help of the estimated area by the imagery, the relative biases can be estimated by modeling the area-height relationship. With water level observations and water area observations, the Gauss-Helmert model is chosen to adjust the area-height relationship. Due to the possible interpolating error and the gross error in both observations, two robust estimation methods have been used to deal with outliers. One is the expectation maximization method, which provides a robust estimate by iteratively downweighting the observation with large residuals, and another one is an outlier rejection method based on Baardas' data snooping, which detects outliers in the observations with statistical hypothesis tests, .

In the end, we also discuss the influence of the topography on the inter-track and inter-satellite biases. We calculate the standard deviation of the DEM of the intersaction area between the 2 km region along the track and the 5 km region along the lake to determine the relationship between topography and biases. The results show a high correlation between the inter-track biases and the topography.

We have employed the developed methodology on a number of lakes and reservoirs, and the findings are compared to in-situ water level data. The results reveal the existence of the inter-satellite and inter-track biases, which vary from the global bias estimates.

Key words: Satellite altimetry, Inter-satellite bias, Inland water, water extent

Contents

List of Figures

List of Tables

Chapter 1

Introduction

About 3% of the Earth's land surface is covered by lakes (Pekel et al., 2016). Although inland water bodies, including lakes, reservoirs, and rivers, cover only a small fraction of the Earth's surface, they play a significant role in sustaining life on Earth, in the global water cycle, and in climate change (Alsdorf et al., 2007). Therefore, monitoring lake water levels is important for a better understanding of their impact on the environment and for the wise management of freshwater resources (Shu et al., 2021).

Water level changes in lakes and reservoirs were traditionally derived from gauge data. The gauge stations provide highly precise observations. However, due to the high cost of the installation and the maintenance of gauge stations, only small portions of lakes are monitored by the gauge stations, especially for those in harsh environments and in remote developing countries (Hannah et al., 2011). In the last 20 years, satellite altimetry has shown that it's potential for monitoring the water-level time series of the world's major inland water bodies to close the data gap between measurements made on the ground and measurements made from space.

Since the 1980s, more than 13 satellite missions have been launched with different radar altimeters on board. Table 1.1 shows the planned and launched altimetry missions since the 1980s. Although satellite altimetry was first used to monitor sea level variation and ice sheet change at the beginning, monitoring the water level of large lakes and rivers has been the goal of research since the launch of the TOPEX/Poseidon and Envisat missions (Birkett, 1995, Crétaux et al., 2011). ICESat-1 with laser altimeter, which was launched by the National Aeronautics and Space Administration in January 2003 and functioned until February 2010, has benefits for measuring water levels in small bodies of water due to its small footprint size of about 70 m in diameter. However, compared with traditional radar satellites, ICESat-1 has short temporal coverage and sparse sampling. On the contrary, CryoSat-2, the radar altimetry mission, has relatively high spatial coverage and temporal resolution and operates in a 369-drfting orbit with a 30-dat subcycle. Due to its orbit configuration, ICESat-2 can visit a larger number of lakes and provide water levels for smaller lakes than any of the previous missions (Kleinherenbrink et al., 2014).

In the last few years, much progress has been made in altimeter capability to acquire quality measurements over inland waters. The accuracy has reached a few centimetres for lakes and a few dozens of centimetres for rivers (Schwatke et al., 2015). The use of satellite altimetry for inlandwater monitoring has been facilitated by the advent of two different developments (Tourian et al., 2021):

- Open-Loop Tracking Command (OLTC)
- Operation in Synthetic Aperture Radar (SAR) mode

The OLTC is an on-board feature used to set the altimeter waveform reception window, containing a priori elevation information built from a dedicated database of hydrological targets (Le Gac et al., 2021). With the assistance of OLTC, the SARAL/AltiKa, Jason-3, Jason-2, and Sentinel-3 are better able to follow inland water bodies, particularly those with coarse terrain.

The Sentinel-3 family follows CryoSat-2 as the first altimeter to offer SAR mode. The improved along-track resolution is accompanied by a naturally reduced level of land contamination thanks to the increased along-track resolution. Due to Delay-Doppler processing, SAR mode provides higher resolution along track (e.g. 300 km vs. a few km) than conventional (LRM) altimetry (Jensen and Raney, 1998, Wingham et al., 2006). Meanwhile, the capability of beam steering, which permits the measurement to be focused on a specific location, enables the estimation of the height from the processing of echoes returned by a limited band centered inside the river (taking advantage of the fact that water conceals a priori information). Accumulation of looks from the same footprint to raise the signal-to-noise ratio, hence improving the accuracy of the altimetric measurements and characterization of the surface reaction based on the look angles.

On the basis of the elevation measurements obtained by several satellite radar altimeters, databases as shown in Table 1.2 have been created to provide a time series of altimetry-derived water level estimations for the world's most significant inland lakes. In these datasets, the time series of water level estimates are made by combining elevation observations from many satellite radar altimeters and processing them in different ways.

Over lakes and reservoirs, it is often feasible to combine many altimeter missions and passes into a single multi-mission, multi-pass time series. Based on the premise that the water level of a lake follow an equipotential surface, this is achievable. By combining different altimetry passes and missions, the temporal resolution of the water-level time series is increased, depending on the number of passes over the water body (Boergens et al., 2017). However, the differences in lake size, geographic location, surrounding topography, and land cover type could significantly influence the accuracy of lake water levels retrieved by satellite radar altimeters (Maillard et al., 2015). Consequently, each radar altimetry mission covers distinct time periods and has varying degrees of measurement precision, and the multi-mission is hindered by inter-satellite bias. Several studies have been conducted to address the issue, with some focusing on lakes and reservoirs in particular. But there is still a lot we don't know about how satellites are biased over inland waterways.

The biases can be estimated relatively and absolutely. Compared with absolute estimation of the biases, which requires in situ data for calibration, estimation of relative biases is more applicable and has more advantages in coping with different situations on a global scale. Meanwhile, due to the different spatial and temporal resolution of different missions, only a small portion of lakes are monitored by sufficient missions with sufficient periods of overlap, which makes relative estimation more difficult. This lets geoid or altimetry correction errors within a single lake and over different tracks be taken into account.

In this study, we explore the feasibility of resolving satellite biases over lakes and reservoirs. This study focuses on the relative biases of each tracks. To avoid the limitation of sufficient periods of overlap, our strategy for calculating the biases between overlapping and non-overlapping time series of water levels from various missions and tracks relies on satellite-derived time series of surface area. The area estimated by the imaging connects the time series and the connection between area and height offers the foundation for determining the relative

biases. The relative biases are estimated by modeling the area-height relationship within a Gauss-Helmert model. The power law model is selected to present the relationship between the water area and water height. Even after retracking correction and geophysical correction, however, altimeter measurements continue to be noisy. In addition, the interpolation error and accuracy limitations of the interpolated water extent data will impact the quality of the results. The expectation maximization method and local test are finally used to get rid of the effect of outliers. This method creates a reliable estimate by changing the weights of the observations over and over again.

We examine our technique on a small sample of lakes and reservoirs and confirm the findings using in-situ water level data. Our findings demonstrate the occurrence of inter-satellite and inter-track biases at the decimeter level that vary from estimates of global bias. In the end, this study also discusses the influence of the topography on biases based on the DEM and finds a correlation between the biases and the topography.

mission	operated by	life time	height [km]	inclination [°]	rev./day	frequency
Geosat	NOAA	03.1985-09.1989	785	108.0	244/17	13.5
ERS-1	ESA	07.1991-03.1996	785	98.5	501/35	13.5
TOPEX/Poseidon	CNES,NASA	09.1992-01.2006	1336	66.0	127/10	13.6& 5.3
ERS-2	ESA	04.1995-09.2011	781	98.5	501/35	13.5
GFO	US-Navy	02.1998-11.2008	784	108.0	244/17	13.5
Jason-1	CNES,NASA	01.2002-06.2013	1336	66.0	127/10	13.6 & 5.3
ENVISAT	ESA	03.2002-04.2012	800	98.5	501/35	13.5 & 3.2
Jason-2	CNES,NASA,NOAA,EUMETSAT	06.2008-ongoing	1336	66.0	127/10	13.6 & 5.3
HY-2A	NSOAS	08.2011-ongoing	971	99.3	193/14	13.58 & 5.25
ICeSat	NASA	01.2003-10.2009	600	94.0	2723/183	Laser:1064&532 nm
SARAL/AltiKa	ISRO,CNES	02.2013-ongoing	800	98.5	501/35	35.75
Jason-3	CNES,NASA,NOAA,EUMETSAT	01.2016-ongoing	1336	66.0	127/10	13.6 & 5.3
Sentinel-3A	ESA,GMES	02.2016-ongoing	815	98.6	385/27	13.6 & 5.3
Sentinel-3B	ESA,GMES	04.2018-ongoing	815	98.6	385/27	13.6 & 5.3
ICeSat 2	NASA	09.2018-ongoing	480	92.0	1387/91	Laser:1064&532 nm
HY-2B	NSOAS	10.2018-ongoing	971	99.3	193/14	13.58 & 5.25
Sentinel-6 Michael Freilich	EUMETSAT,NASA	11.2020-ongoing	1336	66.0	127/10	13.5
SWOT	NASA,CNES	02.0222-02.2025	890	77.6	292/21	35.75

Table 1.1: *Satellite altimetry missions from 1985 and their characteristics, source:(Tourian et al., 2021)*

Product	oprated by	source	Remark
Hydroweb	CNES	hydroweb.theia-land.fr	NRT available for some lakes and rivers
River & Lake	ESA	altimetry.esa.int/riverlake/shared	no longer maintained
DAHITI	Deutsches Geodaetisches Forschungsinstitut(DGFI),TU Munich	dahiti.dgfi.tum.de	Kalman filter aprroach
HydroSat	Institude of Geodesy University of Stuttgart	hydrosat.gis.uni-stuttgart.de	High-Rate products are available
G-REALM	United States Department of Agriculture	ipad.fas.usda.gov/cropexplorer/global_reservoir	Lakes and reservoirs only
GRRATS	The Ohio State University	podaac.jpl.nasa.gov/dataset/-PRESWOT_HYDRO_GRRATS_L2_VIRTUAL_STATION_HEIGHTS_V1	Envisat- and Jason-series over 39 rivers
AltEx	USAID and NASA	altex.servirglobal.net	web application for exploring altimetry data Jason-2, Jason-3 and Saral/AltiKa
C3S LWL	CLS on behalf of Copernicus	cds.climate.copernicus.eu/cdsapp#!/dataset/10.2438	94 selected lakes are available
Water Level on VITO	Copernicus Global Land Operations CNES, LEGOS AND CLS	land.copernicus.eu/global/products/wl	NRT time series are available

Table 1.2: *Providers of water level time series from satellite altimetry, source:(Tourian et al., 2021)*

Chapter 2

Data and Case study

2.1 Data

2.1.1 Water level time series over lakes and reservoirs from HydroSat

Altbundle software package is used to produce the water level time series across lakes and reservoirs in this research. Figure 2.1 describes the method for creating the standard rate water level time series using satellite altimetry (Tourian et al., 2021).

A so-called "Virtual Station" (VS) may be calculated whenever the satellite's ground track encounters a hydrological item. Typically, the boundaries of a VS are determined by the kind of water body and the location of the altimetry track above the item. The altimetric water level is initially determined for each sample within the VS. First, distance measurements are adjusted for geophysical effects (solid earth tide and pole tide) and atmospheric route delays (wettropospheric, drytropospheric, and ionospheric). Subtracting the adjusted range from the satellite's height, the water level is then determined. The reference height is converted to geoid at the next step based on static gravity field models from XGM2019e (Pail et al., 2015), EGM2008 (Pavlis et al., 2012), or EIGEN6C3 (Förste et al., 2012). To guarantee a reliable estimate for each overpass, the median orthometric height within the VS is used as the representative height.

2.1.2 Surface water extent from satellite imagery from HydroSat

HydroSat provides water level and surface water extent time series of lakes, reservoirs, and river reaches using optical satellite photos. To estimate surface water extent, many spaceborne missions develop dynamic water masks with great temporal and geographical precision. Water masks are often created from optical pictures using pixel-based image segmentation techniques that set a threshold in the image pixel value histogram. Pixel-based algorithms cannot create realistic water masks, especially across rivers, despite their ease of development (Tourian et al., 2021).

To generate a river mask, a Posterior Markov Random Field (MAP-MRF) is estimated to account for spatial interactions between pixels and temporal pixel value fluctuations. Remove cloudy photos. Dynamic threshold water masks are created first. Joint conditional models, energy minimization, and graphs follow. Next, graph cuts discover the MAP solution. MAP changes the initial frequency coverage map and water masks to modify the graph. MAP solution river mask for revised graph. (Tourian et al., 2021). The algorithm is presented in Figure 2.2.

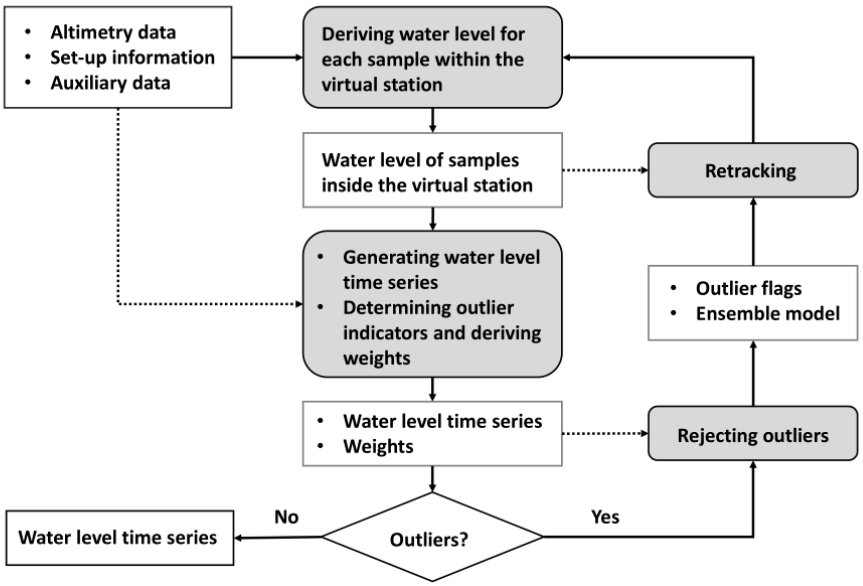


Figure 2.1: *Flowchart of obtaining water level time series for a single virtual station, source: (Tourian et al., 2021)*

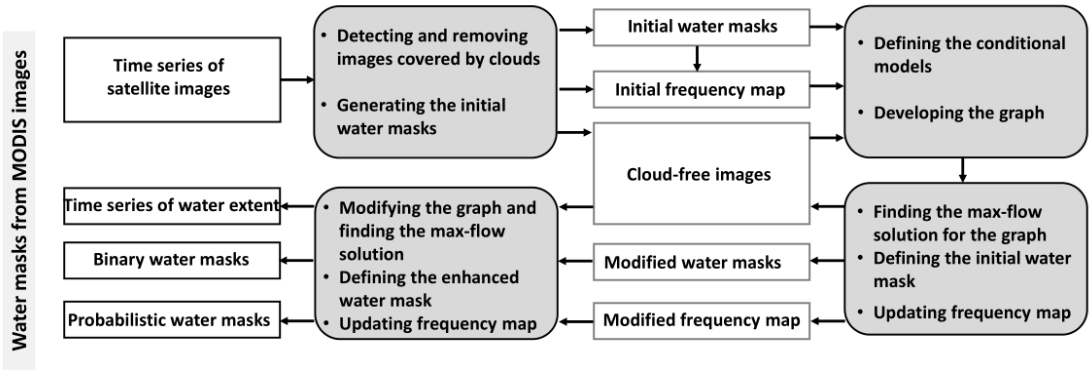


Figure 2.2: *Flow chart of generating time series of water extent, source: (Tourian et al., 2021)*

2.1.3 Digital elevation model (DEM)

NASADEM updates the Digital Elevation Model (DEM) and related products from SRTM data. Data products were created by reprocessing SRTM interferometric SAR data. Two-dimensional phase array and radar image interferograms improve elevation precision. Due to its nature, interferometric data must be wrapped and unwrapped to measure. NASADEM used the latest unwrapping methods and supplementary information that were unavailable during SRTM data processing. The improved method reduced data gaps and increased SRTM geographic coverage. The ASTER GDEM, PRISM on the Advanced Land Observing Satellite (ALOS), the USGS National Elevation Dataset (NED), and the Global DEM Specifications for Canada and Alaska filled more gaps. The Ice, Cloud, and Land Elevation Satellite (ICESat) mission adjusted data vertical and tilt using ground control points and laser profiles. This application enhanced mosaic vertical accuracy, swath uniformity, and homogeneity. The Land Processes Distributed Active Archive Center (LP DAAC) publishes NASADEM products with a one-arcsecond gap.

(OpenTopography URL: `https://portal.opentopography.org/datasetMetadata`).

2.1.4 In Situ data

To evaluate the accuracy of the bias estimation model, in situ data is needed as comparison data. In this study, in situ data is supplied by the USGS and ANA.

The USGS investigates the occurrence, quantity, quality, distribution, and movement of surface and underground waters and disseminates the data to the public, state and local governments, public and private utilities, and other federal agencies involved with managing water resources.

The in situ data of the lakes and reservoirs in Brazil is distributed by Natinal Water and Sanitation Agency, which is legally liable for implementing the National Water Resources Management System (SINGREH), created to ensure the sustainable use of rivers and lakes for the current and future generations.

2.2 Case Study

In this study, 8 lakes and reservoirs all over the world are selected to test the algorithm and the basic information of them is shown in Table 2.1, Figure 2.3 and Figure 2.4.

Maithon Dam is located 48 km from Dhanbad in the Indian state of Jharkhand. The lake is 65 km^2 in size. This dam was developed specifically for flood management and generates 60,000 kilowatts of electricity. (Ghosh et al., 2018).

Falls Lake is a 50 km^2 reservoir in Durham, Wake, and Granville counties in North Carolina, United States. Falls Lake spans 45 km up the Neuse River to the junction of the Eno, Little, and Flat rivers, its source. The lake serves as a source of drinking water for numerous local municipalities, including the city of Raleigh, as well as a recreational area and animal habitat (Golembesky et al., 2009).

The TresMarias reservoir, is an embankment dam on the Sao Francisco River at TresMarias in Minas Gerais, Brazil. It was built for the purposes of hydroelectric power generation and flood control. The dam is 2700 m long and 75 m tall, with a spillway on its east side and a power plant at its east base. The reservoir of the dam has a surface area of 1040 km^2 and a volume of 21 km^3 (Torres and Eterovick, 2010).

The Tiga Dam is in Kano State in the Northwest of Nigeria, constructed in 1971 to 1974. It is a major reservoir on the Kano River, the main tributary of the Hadejia River. The dam covers an area of 178 km^2 with maximum capacity of nearly 2 000 000 km^3 (Barau, 2007).

The Katta-Kurgan water reservoir is located in a natural depression to the south of Kattakurgan. The reservoir receives its water from the Kara-darya, a right tributary of the Zaravshan. The reservoir is replenished in the fall, winter, and spring. From May to June, water is used for irrigation purposes. The reservoir freezes briefly throughout the winter (Bakhtiyarovna et al., 2022).

The largest lake in Ethiopia and the source of the Blue Nile is Tana Lake. The lake is roughly 84 km long and 66 km wide, with a maximum depth of 15 m and an elevation of 1788 m (Vijverberg et al., 2009).

Lake Arapa is a lake in Peru, located in the Arapa and Chupa Districts in the Puno Region.

Itaparica Reservoir locate in Brazil, which has a regulated inflow of 2060 m^3/sec, a length of 149 km, and a sub-water basin of 93 040 km^2. Maximum depth is 101 m (mean depth - 13 m) (Gunkel et al., 2013).

Poly ID	Lake Name	Lat (°)	Lon (°)	In situ data	DEM	Area (km^3)
376	MaithonReservoir	23.78	86.79	-	yes	65.67
133	Falls Lake	36.01	-78.74	yes	yes	48.06
205	TresMarias reservoir	-18.24	-45.34	yes	yes	1039.39
233	Itaparica Lake	-10.08	-42.32	yes	yes	1828.13
267	Tiga Lake	11.47	8.37	-	yes	178.20
357	Kattakurgan reservoir	39.79	66.20	-	yes	59.22
565	Tana Lake	11.81	37.26	-	yes	3055.10
1515	Arapa Lake	-15.19	-70.07	-	yes	126.17

Table 2.1: *The list of research lakes and reservoirs*

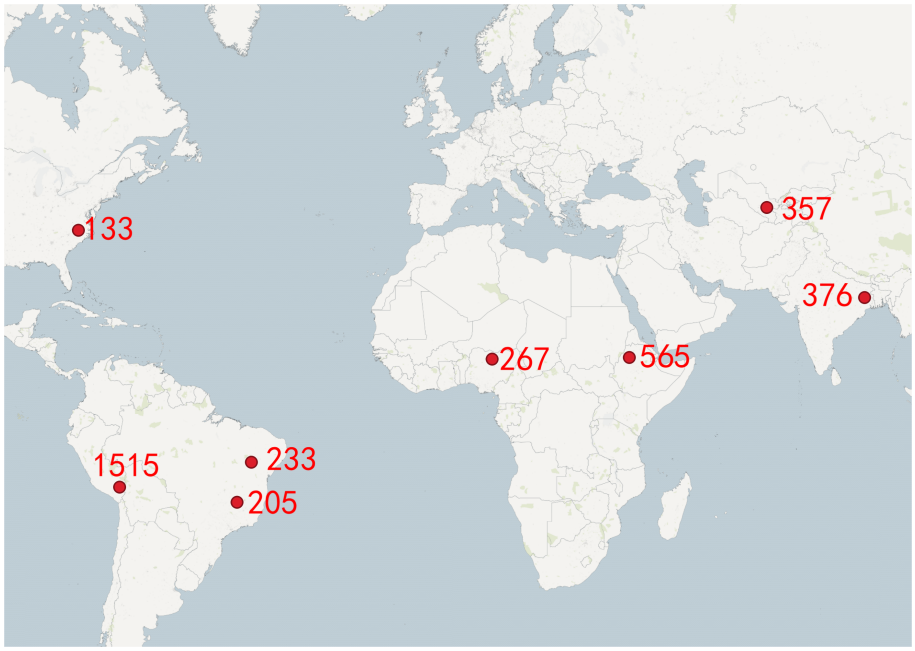
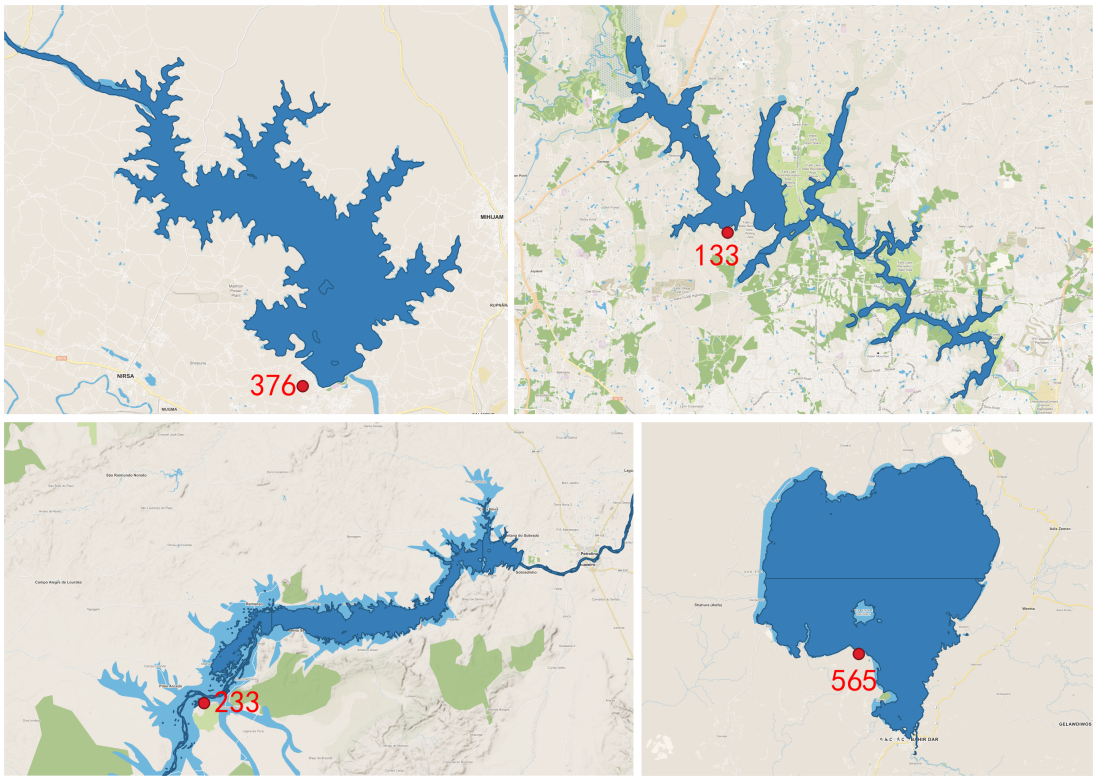
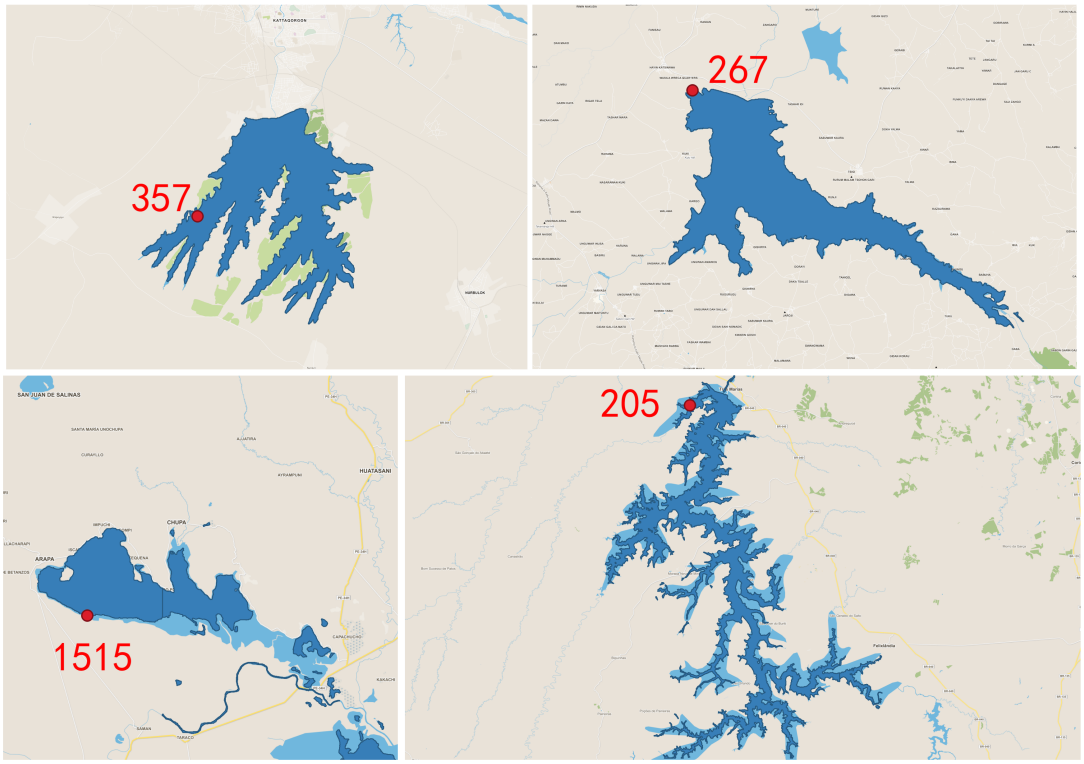


Figure 2.3: *Investigated 8 lakes and reservoirs (Overall)*



(a)



(b)

Figure 2.4: *Zoom in: 8 lakes and reservoirs*

Chapter 3

Methodology

3.1 Basic idea

Monitoring lake levels using data from many satellite altimeters necessitates bias correction due to systematic inaccuracies among satellite altimeters. Given that absolute assessment of biases needs along-track in situ data, relative calibration between the altimeters is frequently employed to correct biases.

Typically, the relative biases are estimated by minimizing the difference of the water level time series in overlapping period. However, a significant proportion of lakes and reservoirs are monitored by a small number of altimetry missions without sufficient overlap intervals. Therefore, this thesis attempts to connect the non-overlapping or overlapping time series by constructing the relationship between the water level and the long-term water area time series derived from satellite image. Utilizing remotely sensed surface area time series as an anchor for biased time series enables assessment of the relevant biases.

In this thesis, the power law model is chosen to represent the area height relationship and the functional expression of the relationship is adjusted using the Gauss-Helmert model. However, due to the gross error existing in the water area and water level observations, the resulting model will deviate from the real model, which will directly affect the accuracy of the bias estimates and even introduce the unexpected extra bias in results. The ordinary least squares estimation, on the other hand, can not withstand the large error. Therefore, two kinds of outlier rejection algorithms (outliers rejection based on the Baardas' data snooping and expectation maximization algorithm) have been sequentially introduced to reduce the influence of the error.

3.2 Relationship between the water area and water level

The relationship between the water area and water height is determined by the shape of the shore line and riverbed, which is complicated and hard to model simply. We can assume a linear, the second degree polynomial or power law function to fit the scatters to obtain the expression of the relationship between the water area and water height. An example of fitting results of the three regression model is shown in Figure 3.1 :
The fitted line using the linear function with a mean R^2 of 0.792:

$$S = aH + H0 \tag{3.2.1}$$

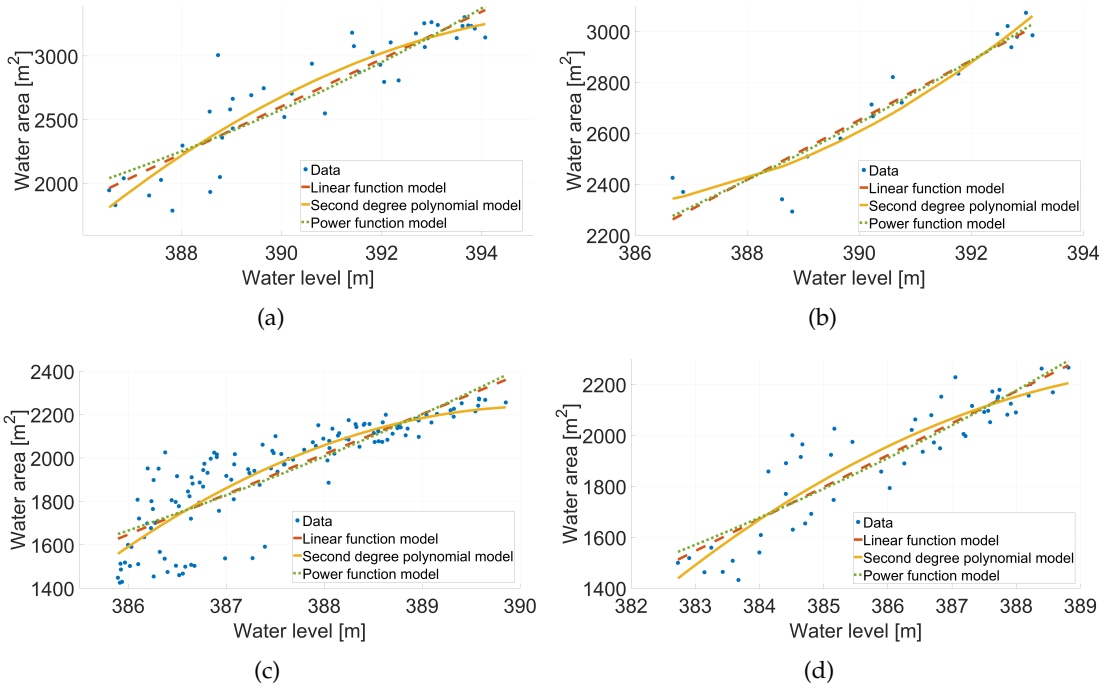


Figure 3.1: *An example of fitting results of the four regression model*

The fitted line using the second degree polynomial function with a mean R^2 of 0.831:

$$S = a(H)^b + c \tag{3.2.2}$$

The fitted line using the power function with a mean R^2 of 0.833:

$$S = a(H - H0)^c \tag{3.2.3}$$

As shown in the Figure 3.1, although the area-height relationship varies in different lakes and reservoirs, the scatters are better distributed near the power law function curve. By comparing the mean R^2 of different fitting function model, it can be found that power law function is a better model for the relationship between the area and height.

After selected the based mission, the area observations H for based mission is

$$H = H_{based} \tag{3.2.4}$$

for remaining missions the the area H can be presented as

$$H = H_i - b_i \qquad i = 1, 2, 3...n - 1(n \geq 2) \tag{3.2.5}$$

Where H_i is the observations of the remaining missions, b_i is the biases relative to water level observation of based mission and n is the number of the missions.

Because parameter b_i shows high correlation with parameter H0 in power law function, the area-height relationship will be defined with power law function without parameter H0 as

$$S = a(H)^c \tag{3.2.6}$$

Finally, the biases estimation model is defined as following:

$$\begin{aligned} S_{based} &= a(H_{based})c \\ S_1 &= a(H_1 - b_1)c \\ &\vdots \\ S_i &= a(H_i - b_i)c \end{aligned} \tag{3.2.7}$$

To facilitate the acquisition of the initial values for GHM, we take the logarithm on both sides of the equation so that the power law function model becomes a linear model. Thus the model eventually becomes

$$\begin{aligned} \log(S_i) &= \log(a) + c * \log(H_i - b_i) \\ &\vdots \\ \log(S_1) &= \log(a) + c * \log(H_1 - b_1) \\ \log(S_{\text{base}}) &= \log(a) + c * \log(H_{\text{based}}) \end{aligned} \tag{3.2.8}$$

3.3 Adjustment using Gauss-Helmert model

3.3.1 Principle

The Gauss Helmert model is defined as follows (Koch, 2014):

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y} + \boldsymbol{e}) = 0 \tag{3.3.1}$$

Where the vector $\boldsymbol{h}$ is the nonlinear, differentiable functions, $\boldsymbol{x}$ is the vector of the the unknowns, $\boldsymbol{y}$ is the vector of the observations and $\boldsymbol{e}$ is the vector of measurement errors (Koch, 2014).

Expected values $\hat{\boldsymbol{y}}$ of is defined with

$$\hat{y} = \boldsymbol{y} + \boldsymbol{e} \tag{3.3.2}$$

and the covariance matrix of y and e by

$$D(\boldsymbol{y}) = D(\boldsymbol{e}) = \sigma^2 \boldsymbol{\Sigma} = \operatorname{diag} \sigma^2 \left(\sigma_1^2, \cdots, \sigma_n^2\right) \tag{3.3.3}$$

Where σ^2 is the unknown variance factor and $\boldsymbol{\Sigma}$ is a known positive definite matrix.

A least squares adjustment will be used to estimate the unknowns x

$$\frac{1}{\sigma^2} \boldsymbol{e}' \boldsymbol{\Sigma}^{-1} \boldsymbol{e} \rightarrow \min \tag{3.3.4}$$

As condition equations Equation 3.3.1 is nonlinear, Taylor series expansion has to be taken using the approximate values $\boldsymbol{x}^{(0)}$ and $\boldsymbol{y}^{(0)}$ for the parameters and the observation.

$$\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{x^{(0)}, y^{(0)}} \left(\boldsymbol{x} - \boldsymbol{x}^{(0)}\right) + \left.\frac{\partial \boldsymbol{h}}{\partial y_E}\right|_{x^{(0)}, y^{(0)}} \left(\hat{y} - \boldsymbol{y}^{(0)}\right) + \boldsymbol{h}\left(\boldsymbol{x}^{(0)}, y^{(0)}\right) = 0 \tag{3.3.5}$$

After introducing the matrix $\boldsymbol{A}$ of derivatives with respect to $\boldsymbol{x}$ and the matrix of $\boldsymbol{B}$ of derivatives of derivatives with respect to $\hat{y}$, The condition equation is instead by

$$\boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{B}_e + \boldsymbol{w}_m = 0, \tag{3.3.6}$$

where $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{(0)}$ and $\boldsymbol{w}_m$ is the vector of misclosure

$$\boldsymbol{w}_m = \boldsymbol{h}(\boldsymbol{x}^{(0)}, \boldsymbol{y}^{(0)}) + B(\boldsymbol{y} - \boldsymbol{y}^{(0)}). \tag{3.3.7}$$

The Lagrange function is defined

$$w(\Delta\boldsymbol{x}, \boldsymbol{e}) = \frac{1}{\sigma^2}\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e} - \frac{2}{\sigma^2}\boldsymbol{k}'\left(\boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{B}\boldsymbol{e} + \boldsymbol{w}_m\right) \tag{3.3.8}$$

Where $-\frac{2}{\sigma^2}\boldsymbol{k}$ is the Lagrange multipliers.

To minimize the $\frac{1}{\sigma^2}e'\Sigma^{-1}e$, the following derivatives should be zero

$$\frac{\partial w(\Delta\boldsymbol{x}, \boldsymbol{e})}{\partial\Delta\boldsymbol{x}} = -\frac{2}{\sigma^2}\,\boldsymbol{A}'\boldsymbol{k} \tag{3.3.9}$$

$$\frac{\partial w(\Delta\boldsymbol{x}, \boldsymbol{e})}{\partial\boldsymbol{e}} = \frac{2}{\sigma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{e} - \frac{2}{\sigma^2}\,\boldsymbol{B}'\boldsymbol{k} = 0 \tag{3.3.10}$$

residuals $\hat{\boldsymbol{e}}$ is thus obtained by

$$\hat{\boldsymbol{e}} = \Sigma\boldsymbol{B}^{'}\boldsymbol{k}. \tag{3.3.11}$$

Together with Equation 3.3.10 ,the normal equations are derived

$$\begin{vmatrix} \boldsymbol{B}\Sigma\boldsymbol{B}' & \boldsymbol{A} \\ \boldsymbol{A}' & \boldsymbol{0} \end{vmatrix} \begin{vmatrix} \boldsymbol{k} \\ \Delta\boldsymbol{x} \end{vmatrix} = \begin{vmatrix} -\boldsymbol{W}_m \\ \boldsymbol{0} \end{vmatrix} \tag{3.3.12}$$

After solving the normal equations, following equations can be derived:

$$\hat{\Delta\boldsymbol{x}} = -\left(\boldsymbol{A}'\left(\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}'\right)^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\left(\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}'\right)^{-1}\boldsymbol{w}_m \tag{3.3.13}$$

$$\hat{\boldsymbol{k}} = (\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B})^{-1}\left(-\boldsymbol{w}_m - \boldsymbol{A}\hat{\Delta\boldsymbol{x}}\right) \tag{3.3.14}$$

$$\hat{\boldsymbol{e}} = \Sigma\boldsymbol{B}'\left(\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{Z}'\right)^{-1}\left(-\boldsymbol{w}_m - \boldsymbol{A}\hat{\Delta\boldsymbol{x}}\right) \tag{3.3.15}$$

Therefore, we can get

$$\hat{\boldsymbol{e}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{e}} = -\hat{\boldsymbol{k}}'\boldsymbol{w}_m \tag{3.3.16}$$

The unbiased estimate $\hat{\sigma}^2$ of σ^2 the variance factor is calculated by

$$\hat{\boldsymbol{\sigma}}^2 = \frac{1}{r-u}\hat{\boldsymbol{e}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{e}} \tag{3.3.17}$$

For Gauss Newton iterations, the initial approximate values $x^{(0)}$ and $y^{(0)}$ is needed for the first iteration, where $y^{(0)}$ is set to y.

The solution $\hat{\Delta\boldsymbol{x}}(0)$ and $\hat{\boldsymbol{e}}^{(0)}$ is given by $\hat{\boldsymbol{x}}^{(0)} = \boldsymbol{x}^{(0)} + \Delta\hat{x}^{(0)}, \hat{\boldsymbol{y}}_E^{(0)} = \boldsymbol{y} + \hat{\boldsymbol{e}}^{(0)}$.

At the k th iteration, the approximate values for Taylor expansion

$$\boldsymbol{x}^{(k)} = \hat{\boldsymbol{x}}^{(k-1)}, \boldsymbol{y}^{(k)} = \hat{\boldsymbol{y}}_E^{(k-1)} \tag{3.3.18}$$

and estimates is given

$$\hat{\boldsymbol{x}}^{(k)} = \boldsymbol{x}^{(k)} + \hat{\Delta \boldsymbol{x}}^{(k)}, \hat{\boldsymbol{y}}_E^{(k)} = \boldsymbol{y} + \hat{\boldsymbol{e}}^{(k)} \tag{3.3.19}$$

The convergence is checked by $\left|\hat{x}^{(k)} - x^{(k)}\right|$. Initial values have a significant impact. If the variances between the initial and actual values are too large, the function may converge to the local minimum rather than the global minimum, or it may not converge at all.

To facilitate the acquisition of the initial values for GHM, the logarithm has be taken on both sides of the equation so that the power law function becomes a linear model. Thus, the unknown parameters a and c can derived in linear model using Gauss-Markov model with holding the water level H_{based} fixed as

$$\log\left(\boldsymbol{S}_{\text{base}}\right) = \log(a) + c \cdot \log\left(\boldsymbol{H}_{\text{based}}\right) \tag{3.3.20}$$

The initial unknown parameter bias b_i^0 are calculated robustly by mean water level difference relative to based mission shown as

$$b_i^0 = \text{mean}\left(\boldsymbol{H}_{\text{based}}\right) - \text{mean}\left(\boldsymbol{H}_i\right) \tag{3.3.21}$$

Iteration of the Gauss-Helmert model is shown in Figure 3.2

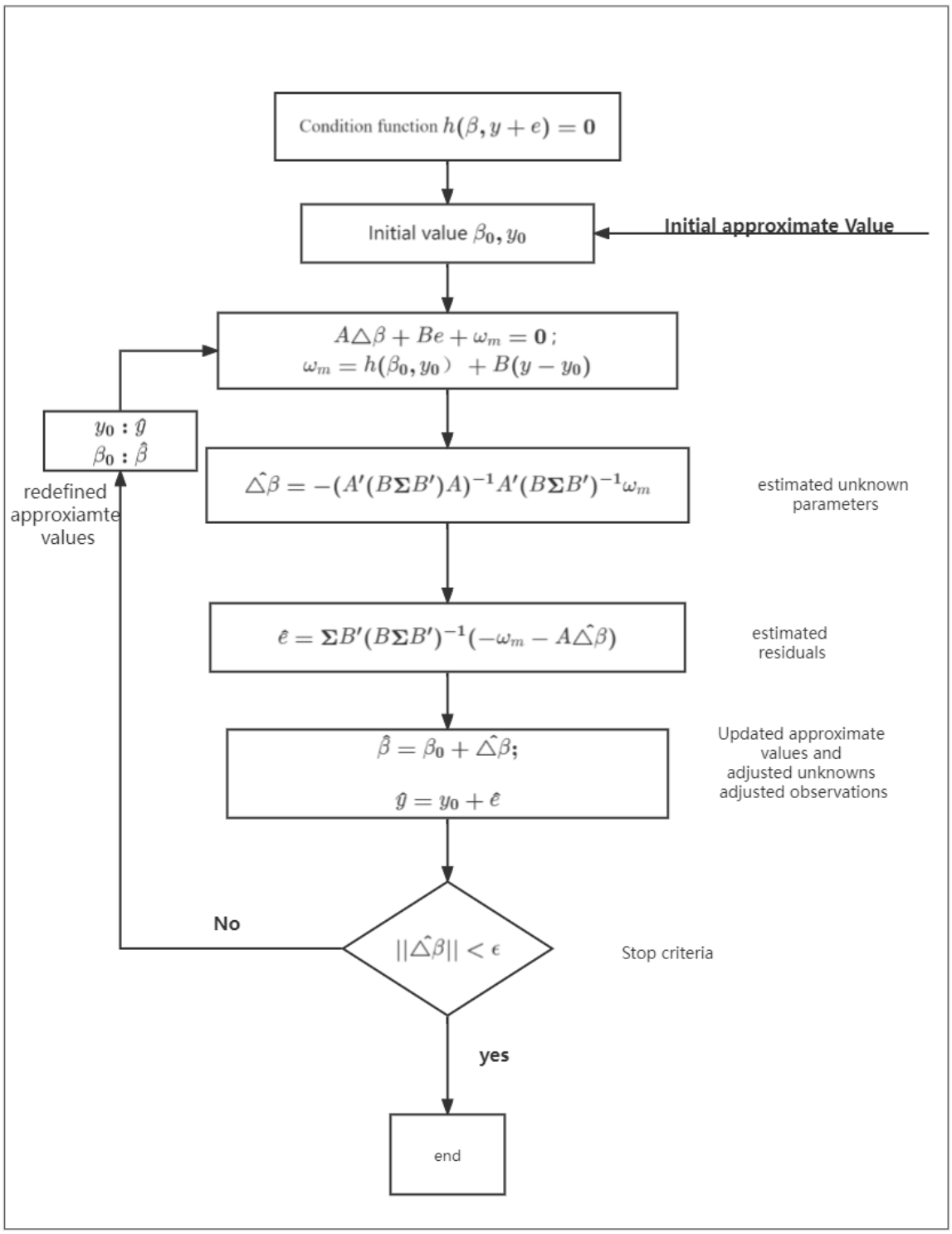


Figure 3.2: *Flow chart of Gauss Helmert Model*

3.4 Outlier Detection

Even after applying re-tracking adjustments and geophysical corrections, altimeter measurements are still fairly noisy. However the power model adjusting by Gauss-Helmert model is very sensitive to the outliers. Once the area-height points are dispersed due to the error in area and height observations, the fitting results becomes poor even iteration can not converge. In consequence, the fitting relationship between water area and water surface height will differ from reality and the extra biases will be introduced to final biases estimation. Therefore a robust estimation method is needed to improve the accuracy of the biases adjustment.

Since it is very difficult to remove all outliers by a certain method, a combination of various outlier criteria is often used. Two steps for outlier detection were implemented in this study. First, the outlier rejection based on the Baarda's data snooping is utilized to detect and eliminate the outliers with large residuals, employing statistical hypotheses to find outliers in the observations. The expectation Maximization (EM) algorithm, which decreases the effect of the outliers in GHM by repeatedly down-weighting the related observation, is presented to lessen the influence of outliers that cannot be discovered by the outlier rejection approach based on Baarda's data snooping.

3.4.1 Outliers rejection based on the Baardas'data snooping

3.4.1.1 Principle

Baardas' data snooping, which uses statistical hypothesis testing in the GH model to find outliers in the observations, is used in this part to implement the outliers rejection. Assuming that there is just one outlier, $\nabla_i, i = 1, \cdots, n$, which will be characterized as bias and used as an estimable deterministic parameter, the observation model becomes (Ettlinger and Neuner, 2020)

$$\begin{aligned} &\boldsymbol{y} = E\boldsymbol{y} - \boldsymbol{c}_i\nabla_i - \boldsymbol{e} \\ &\text{with} \quad \boldsymbol{y} \sim N(E\boldsymbol{y}, \sigma^2\Sigma_{yy}) \text{ and } \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2\Sigma_{yy}) \end{aligned} \tag{3.4.1}$$

If $\nabla_i = 0$, one arrives at the observation model used in the GHM. Using the linearized functional model which is same to that in GHM, two following equivalent null hypothesis can be formulated (Ettlinger and Neuner, 2020)

$$H_0 : \nabla_i = 0 \equiv \boldsymbol{w} + \boldsymbol{A}d\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{e}} = \boldsymbol{0} \tag{3.4.2}$$

$$H_{A,i} : \nabla_i \neq 0 \equiv \boldsymbol{w} + \boldsymbol{A}d\hat{\boldsymbol{x}}' + \boldsymbol{B}\hat{\boldsymbol{e}}' + \boldsymbol{b}_i\hat{\nabla}_i = \boldsymbol{0} \tag{3.4.3}$$

With the new introducing parameter $\hat{\nabla}_i$, $d\hat{\boldsymbol{x}}'$ and $\hat{\boldsymbol{e}}'$ will be numerically different from those in $\boldsymbol{H_0}$. However, $\boldsymbol{w}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ keep the same where misclosure vector $w \in R^{b\times 1}$, design matrix $\boldsymbol{A} \in R^{b\times u}$ and observation matrix $\boldsymbol{B} \in R^{b\times n}$ (Ettlinger and Neuner, 2020).

Under the premise that the error from linearization is minor compared to the systematic deviation induced by $\hat{\nabla}_i$, all further deductions will be made relative to the initial iteration.

Following test value is applied, using a priori variance of unit weight σ_0^2.

$$T_F(\nabla_i) = \frac{\hat{\nabla}_i^2}{\sigma_0^2 \cdot q_{\hat{\nabla}_i}} \sim F_{1,\text{inf}} \mid H_0 \tag{3.4.4}$$

$$T_F(\nabla_i) = \frac{\hat{\nabla}_i^2}{\sigma_0^2 \cdot q_{\hat{\nabla}_i}} \sim F_{1,\text{inf},\lambda_{F_i}} \mid H_{A_i'} \tag{3.4.5}$$

$$T_N(\nabla_i) = \frac{\hat{\nabla}_i}{\sigma_0 \sqrt{q_{\hat{\nabla}_i}}} \sim N(0,1) \mid H_0 \tag{3.4.6}$$

$$T_N(\nabla_i) = \frac{\hat{\nabla}_i}{\sigma_0 \sqrt{q_{\hat{\nabla}_i}}} \sim N(\lambda N, i, 1) \mid H_{A_i} \tag{3.4.7}$$

The square root of $\sqrt{T_F(\nabla_i)}$ follows the central standardized normal distribution $N(0,1)$ in case of H_0 is true and otherwise it follows the non-central standardized normal distribution $N(\lambda_{N,i}, 1)$ with non-centrality parameter $\lambda_{N,i}, 1$ (Ettlinger and Neuner, 2020).

Test values $T_F(\nabla_i)$ and $T_N(\nabla_i)$ for the i-th observation is computed as:

$$T_F(\nabla_i) = \frac{\hat{\boldsymbol{k}}^T \boldsymbol{b}_i \boldsymbol{b}_i^T \hat{\boldsymbol{k}}}{\sigma_0^2 \cdot \boldsymbol{b}_i^T \boldsymbol{Q}_{\hat{k}\hat{k}}} = \frac{\hat{\boldsymbol{e}}^T \boldsymbol{B}^T N^{-1} \boldsymbol{b}_i \boldsymbol{b}_i^T \boldsymbol{N}^{-1} \boldsymbol{B} \hat{\boldsymbol{e}}}{\sigma_0^2 \boldsymbol{b}_i^T \boldsymbol{N}^{-1} \boldsymbol{B} \boldsymbol{Q}_{\hat{e}} \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{b}_i} \tag{3.4.8}$$

$$T_N(\nabla_i) = \frac{\boldsymbol{b}_i^T \hat{\boldsymbol{k}}}{\sigma_0 \sqrt{\boldsymbol{b}_i^T \boldsymbol{Q}_{\hat{k}\hat{k}} \boldsymbol{b}_i}} = \frac{\boldsymbol{b}_i^T \boldsymbol{N}^{-1} \boldsymbol{B} \hat{\boldsymbol{e}}}{\sigma_0 \sqrt{\boldsymbol{b}_i^T \boldsymbol{N}^{-1} \boldsymbol{B} \boldsymbol{Q}_{\hat{e}} \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{b}_i}} \tag{3.4.9}$$

$\hat{\nabla}_{0,i}, T_N(\nabla_i)$ will be used for the derivation of the minimal detectable bias (MDB). It is an essential measure of inner reliability and specifies how large an outlier in the i-th observation must be for it to be identified by a statistical test with a predefined significance level α_0 and power β_0. Since these three numbers are not independent of one another, setting α_0 and β_0 implies a fixed $\lambda_{0,N}$. $\lambda_{0,N}$ can be calculated with

$$\lambda_{0,N} \approx \phi^{-1}(1 - \frac{\alpha_0}{2}) + \phi^{-1}(\beta_0) \tag{3.4.10}$$

where ϕ_{-1} is the inverse of the density function of the standard normal distribution. An outlier may be identified at levels α_0 and β_0 if it results in a non-centrality larger than $\lambda_{0,N}$. This limit is equivalent to the MDB.

$$\hat{\nabla}_{0,i} = \lambda_{0,N} \cdot \sigma_0 \cdot \sqrt{q_{\hat{\nabla}_i}} = \frac{\lambda_{0,N} \cdot \sigma_0}{\sqrt{\boldsymbol{b}_i^T \boldsymbol{Q}_{\hat{k}\hat{k}} \boldsymbol{b}_i}} = \frac{\lambda_{0,N} \cdot \sigma_0}{\sqrt{\boldsymbol{b}_i^T \boldsymbol{N}^{-1} \boldsymbol{B} \boldsymbol{Q}_{\hat{e}} \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{b}_i}} \tag{3.4.11}$$

Since $\boldsymbol{Q}_{ll}$ is a diagonal matrix, $\hat{\nabla}_i$, q_{∇_i}, $T_N(\hat{\nabla}_{0,i})$ and $\hat{\nabla}_{0,i}$ can be calculated more simply.

$$q_{\hat{\nabla}_i} = \frac{q_{l_i}}{r_{ii}} \tag{3.4.12}$$

$$\hat{\nabla}_i = \frac{e_i}{r_{ii}} \tag{3.4.13}$$

$$T_F(\nabla_i) = \frac{e_i^2}{\sigma_0^2 \cdot q_{l_i} \cdot r_{ii}} \tag{3.4.14}$$

$$T_N(\nabla_i) = \frac{v_i}{\sigma_0 \cdot \sqrt{q_{l_i} r_{ii}}} \tag{3.4.15}$$

$$\hat{\nabla}_{0,i} = \lambda_{0,N} \cdot \sigma_0 \sqrt{\frac{q_{l_i}}{r_{ii}}} \tag{3.4.16}$$

Here, r_{ii} are the diagonal elements of the redundancy matrix.

$$\boldsymbol{R} = \boldsymbol{Q}_{\hat{e}\hat{e}} \boldsymbol{Q}_{yy}^{-1} = \boldsymbol{Q}_{ll} \boldsymbol{B}^T \boldsymbol{N}^{-1} \left(\boldsymbol{I} - \boldsymbol{A} \left(\boldsymbol{A}^T \boldsymbol{N}^{-1} \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T \boldsymbol{N}^{-1} \right) \boldsymbol{B} \tag{3.4.17}$$

Which is also called redundancy numbers.

Under the premise that there is only one unknown outlier, each observation must be checked, resulting in n local tests or alternative Hypothesis HA, i. To translate the preceding ideas into the context of DIA, it is necessary to specify the detection stage. After each adjustment task, a global exam should be performed (Ettlinger and Neuner, 2020).

$$T_F(\sigma_0^2) = \frac{\hat{\sigma}_0^2}{\sigma_0^2} = \frac{\hat{\boldsymbol{v}}^T \boldsymbol{Q}_{ll}^{-1} \hat{\boldsymbol{v}}}{(b-u)\,\sigma_0^2} \sim F_{b-u,\text{inf}} \mid H_0 \tag{3.4.18}$$

$$T_F(\sigma_0^2) = \frac{\hat{\sigma}_0^2}{\sigma_0^2} \sim F_{b-u,\text{inf},\lambda_{F,i}} \mid H_{A,i} \tag{3.4.19}$$

The global test may also serve as the detection stage, revealing whether or not observations include outliers. The significance number α of the global test $T_F(\sigma_0^2)$ as well as the significance numbers α_0 of the n local tests $T_N(\nabla_i)$ (or $T_F(\nabla_i)$) have to be matched by using the Bonferroni equation

$$\alpha_0 \approx \frac{\alpha}{n} \tag{3.4.20}$$

such that all the tests have similar sensitivity to an outlier (Ettlinger and Neuner, 2020). The flowchart of the local test is shown in Figure 3.3.

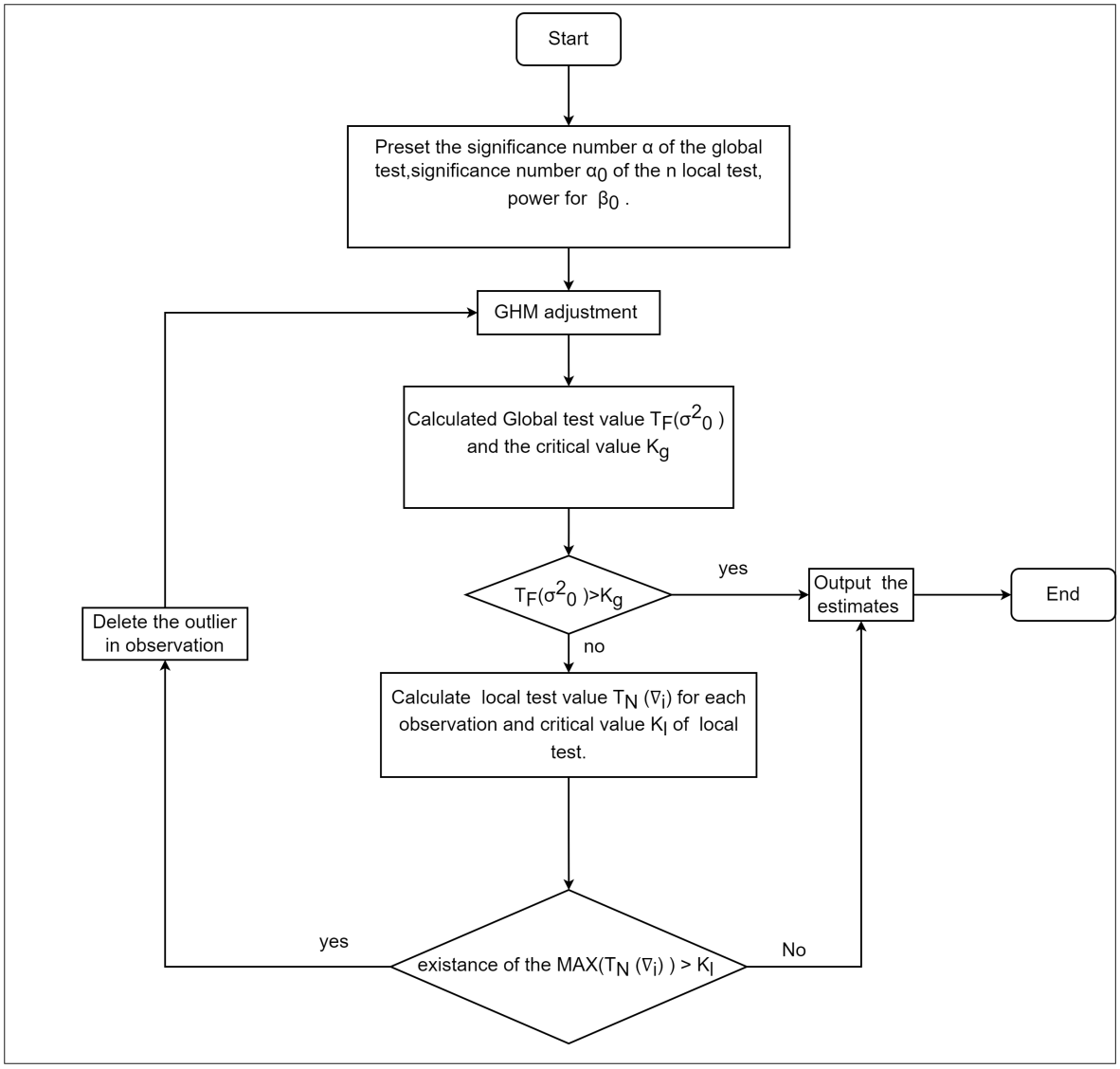


Figure 3.3: *Flow chart of local test based on Baarda'snooping*

3.4.2 EM algorithm based on the variance-inflation model

3.4.2.1 Principle

To In this section, the EM algorithm is introduced for robust estimation (Koch, 2014).

In the EM algorithm, the unknown independent weight $w_i i \in \{1, \ldots, n\}$ for observation y_i will be introduced as missing data which gets the gamma distribution.

According to Equation 3.3.2 and Equation 3.3.3, the following normal distribution is given:

$$e_i \mid w_i \sim N\left(0, \sigma^2 \sigma_i^2 / w_i\right) \tag{3.4.21}$$

In the EM algorithm for the GH model, The unknown parameters vector x', unkonwn variance σ^2 and unknown degree of freedom v are collected together in the vector Θ as

$$\Theta = \left|\boldsymbol{x}', \sigma^2, \boldsymbol{e}\right|' \tag{3.4.22}$$

the EM algorithm have two steps which are applied iteratively. In the expectation step the log-likelihood function $\log p(\boldsymbol{e}, \boldsymbol{w} \mid \boldsymbol{\Theta})$ is defined which is the function of Θ with given e. The conditional expectation $Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}\right)$ of $\log p(\boldsymbol{e}, \boldsymbol{w} \mid \boldsymbol{\Theta})$ with respect to the conditional density $p\left(w \mid e, \boldsymbol{\Theta}^{(t)}\right)$ for $\boldsymbol{w}$ given $\boldsymbol{e}$ and the current estimate $\boldsymbol{\Theta}^{(t)}$ of the unknown parameters follows from citep13

$$\begin{aligned} Q\left(\Theta, \Theta^{(t)}\right) &= E\left(\log p(\boldsymbol{e}, \boldsymbol{w} \mid \boldsymbol{\Theta}) \mid \boldsymbol{e}, \boldsymbol{\Theta}^{(t)}\right) \\ &= \int_{\mathcal{W}} \log p(\boldsymbol{e}, \boldsymbol{w} \mid \boldsymbol{\Theta}) p\left(\boldsymbol{w} \mid \boldsymbol{e}, \boldsymbol{\Theta}^{(t)}\right) d\boldsymbol{w} \end{aligned} \tag{3.4.23}$$

where $\mathcal{W}$ is the domain of w. Thus, the weights w_i are integrated out and replaced by their conditional expectations which are denoted by $w_i^{(t)}$ and result from citep13

$$W_i^{(t)} = \frac{v^{(t)} + 1}{v^{(t)} + e_i^{2(t)} / \left(\sigma^{2(t)} \sigma_i^2\right)} \tag{3.4.24}$$

in which large residuals $e_i^{(t)}$ implies small weights $w_i^{(t)}$.

After Substituting $w_i^{(t)}$ in $Q\left(\Theta, \Theta^{(t)}\right), W^{(t)} = \operatorname{diag}\left(w_1^{(t)} / \sigma_1^2, \ldots, w_n^{(t)} / \sigma_n^2\right)$ is derived.

$$\begin{aligned} Q\left(\Theta, \Theta^{(t)}\right) = &- \frac{1}{2\sigma^2} \boldsymbol{e}' \boldsymbol{W}^{(t)} \boldsymbol{e} \\ &- \frac{n}{2} \log \sigma^2 + \frac{nv}{2} \log v - n \log \Gamma\left(\frac{v}{2}\right) \\ &+ \frac{nv}{2} \left[\Psi\left(\frac{v^{(t)} + 1}{2}\right) - \log\left(v^{(t)} + 1\right)\right. \\ &\left. + \frac{1}{n} \sum_{i=1}^{n} \left(\log w_i^{(t)} - w_i^{(t)}\right)\right] \end{aligned} \tag{3.4.25}$$

where constants have been neglected and $\psi(\ldots)$ denotes the psi function.

In the maximization step, new estimate $\boldsymbol{\Theta}^{(t+1)}$ is derived through maximizing the

$$\boldsymbol{\Theta}^{(t+1)} = \arg\max_{\Theta} Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}\right) \tag{3.4.26}$$

Maximizing the expectation $Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}\right)$ can be achieved by minimizing the $\boldsymbol{e}'\boldsymbol{W}^{(t)}\boldsymbol{e}$. Therefore the unknown parameter estimates $x^{(t)}$ and residuals $\hat{e}^{(t)}$ can be derived by GH model with $\Sigma = \boldsymbol{W}^{-1(t)}$ (Koch, 2014).

Unknown parameter σ^2 results from

$$\frac{\partial}{\partial\sigma^2} Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}\right) = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\boldsymbol{e}^{(t)'}\boldsymbol{W}^{(t)}\boldsymbol{e}^{(t)} = 0 \tag{3.4.27}$$

Therefore, $\sigma^{2^{(t)}} = \frac{1}{n}\boldsymbol{e}^{(t)'}\boldsymbol{W}^{(t)}\boldsymbol{e}^{(t)}$.

The estimate $V^{(t+1)}$ of v is obtained from

$$\begin{aligned} &\log v^{(t+1)} - \psi\left(\frac{v^{(t+1)}}{2}\right) \\ &+\left[\psi\left(\frac{v^{(t)}+1}{2}\right) - \log\left(v^{(t)}+1\right)\right. \\ &\left.+\frac{1}{n}\sum_{i=1}^{n}\left(\log w_i^{(t)} - w_i^{(t)}\right) + 1\right] = 0 \end{aligned} \tag{3.4.28}$$

$V^{(t+1)}$ can only be solved by searching methodically within a certain interval for a zero solution.

Iteration begins using the biased estimates $\boldsymbol{x}^{(1)}$ and $\sigma^{2(1)}$ from the first GHM iteration. For the first approximation $v^{(1)}$, if there are many outliers in observations, $v > 2$ and when no outliers exist inobservaions $v = 120$. In the end, $\boldsymbol{w_i^{(1)}}$ results from $\boldsymbol{x}^{(1)}$ and $\sigma^{2(1)}$. Outliers with Large residuals are iteratively down-weighted by changing the weight. The convergence is checked by the differences of $Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}\right)$ of subsequent iterations or the difference of unknown $\boldsymbol{\Theta}^{(t)}$. The iterations have to be performed repeatedly with $\boldsymbol{v}^{(t+1)}$ as approximation $v^{(1)}$ until there is no dependence on $v^{(1)}$ (Koch, 2014).

The flow chart of the algorithm is shown in Figure 3.4.

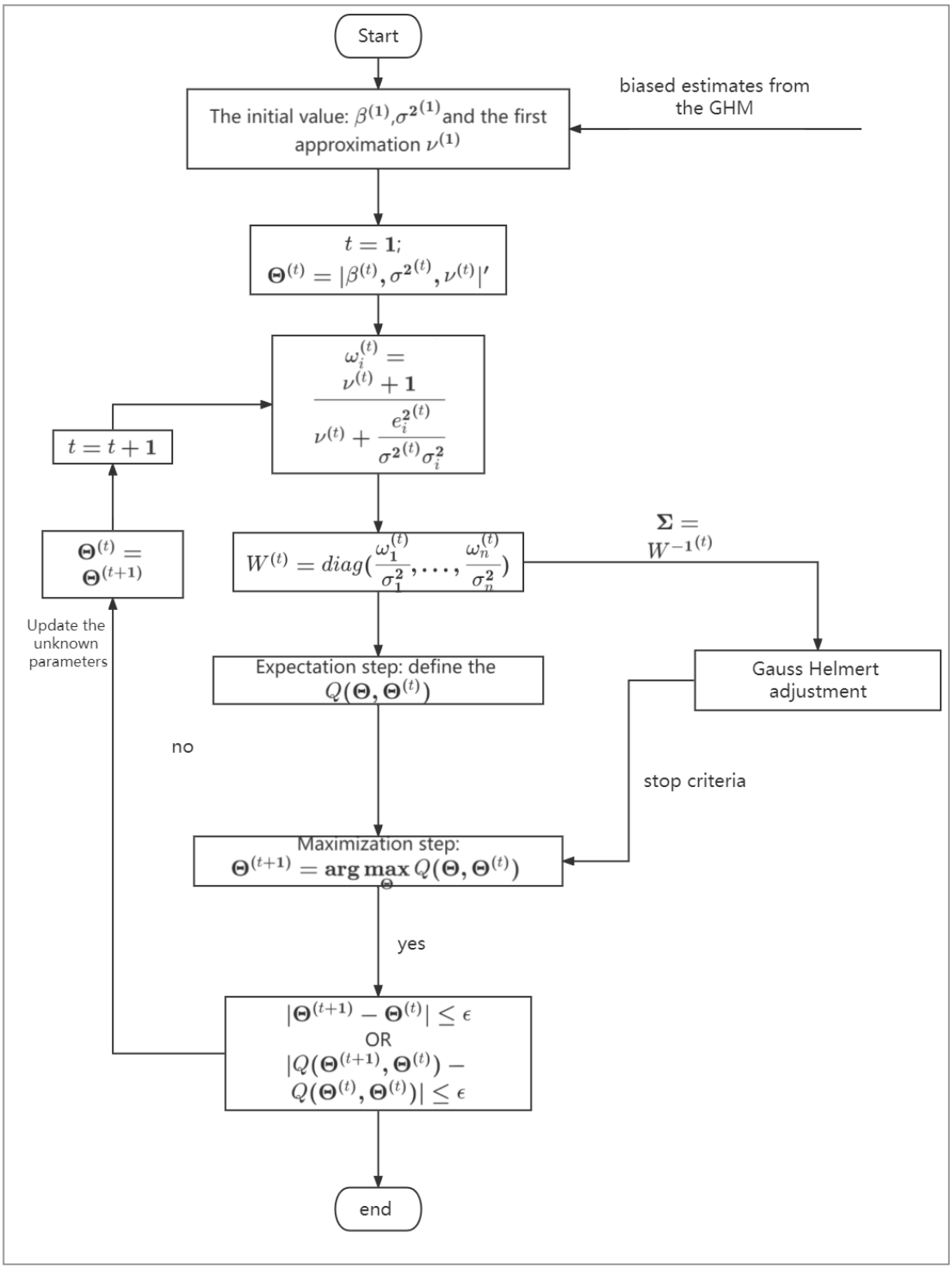


Figure 3.4: *the flowchart of the EM algorithm*

Chapter 4

Results and analysis

Now we apply our algorithm to determine the inter-track biases. As mentioned before, 8 lakes and reservoirs have been investigated to test the algorithm. With the time series and water extent data, the inter-track biases are estimated based on the height-are relationship in the Gauss-helmet model. To decrease the influence of the outliers, both two outlier rejection methods are used to provide the robust estimation. In this chapter, firstly, the results of the bias estimation will be presented. Secondly, we will analyze the relationship between the inter-track biases and topography.

4.1 inter-track biases

4.1.1 Estimate the bias in Gauss-helmet model without outlier rejection

To investigate the influence of the outliers and validate the necessity of the outlier rejection, inter-track biases are estimated in the Gauss-helmet model without outlier rejection at first. To judge the quality of the results, Sobradino Lake and TresMarias Lake, as shown in Figure 4.1 and Figure 4.2, were chosen to estimate the biases. The area-height scatters with fit cueve and multi-track time series after biases correction have been shown in Figure 4.1 through Figure 4.4, and the RMSE of the model and RMSE referring to in-situ data are displayed in Table 4.1

Poly ID	Name	Standard deviation of residuals (m)	RMSE (m)
233	Itaparica reservoir	1.3118	1.2506
205	Repressa TresMatrias reservoir Lake	0.6522	0.9506

Table 4.1: *The results of the estimation without rejection*

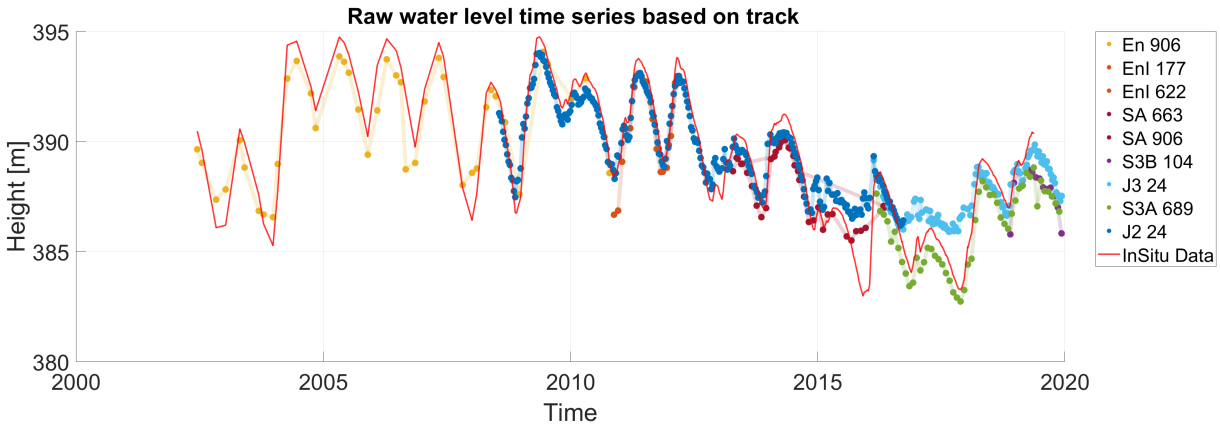


Figure 4.1: *Raw water level time series of the case Itaparica reservoir*

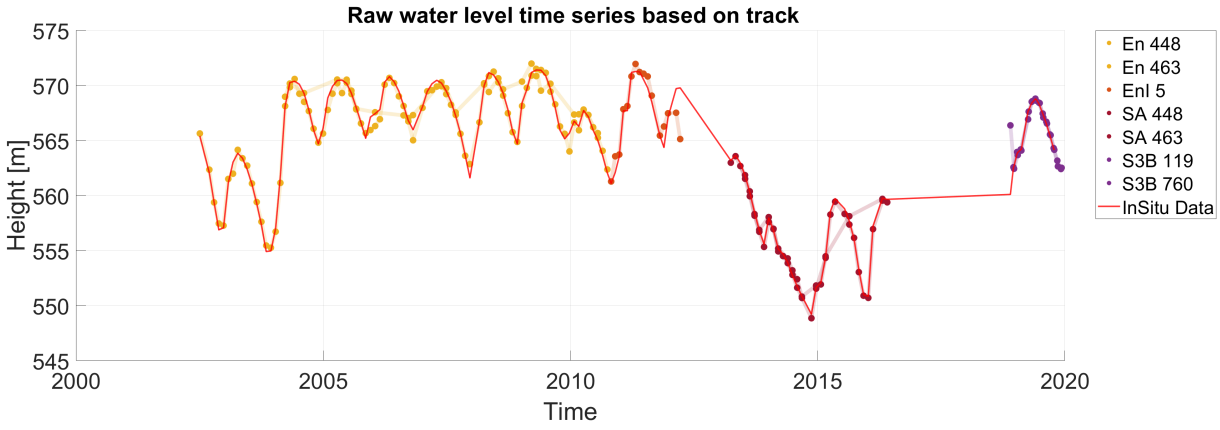


Figure 4.2: *Raw water level time series of the case Represa TresMatrias reservoir*

Due to the effect of the outliers in both observations, the findings of the Gauss-Helmet model's performance are not always desirable. As shown in Figure 4.3, there are distinct and obvious outliers distributed away from the fitted curve, which distort the fitting results relative to the true value. The distortion will result in wrong biases even extra unexpected biases in time series. In Figure 4.4, We can see that there was a very small deviation between raw water level time series from the S3A track 689 and J2 track 24, but after the correction of the biases, an extra significant deviation was introduced.

One possible source of outliers is the interpolation error. The sampling time of the water area observation varies from lake to lake, and the sampling time of the water level observation also varies from mission to mission due to the different temporal resolution of the satellites. Thus, the interpolation of the water area is necessary to acquire the water area data and the water level observation at the same epoch. However, interpolation errors will occur during the interpolation, especially when the sampling interval of the water area observation is large and the water level changes rapidly in a short time.

In the meanwhile, the influence of the uncertainty of the area and height observations can not be ignored. For the water level observations, although the water level time series used in this study has been processed by re-tracking and outlier rejection, there are still a small number of outliers in the final results. The accuracy of the observation is also limited by the water extent, which is derived from the optical satellite imagery.

In addition, it is noticeable that the water level observations from satellite Jason 3 track 24 are much higher than those from sential sentinel 3A track 689 in the case of Itaparica reservoir. when water level lower than about 387 m, the pattern of area-height scatters from Jason2 and Jason3 shows an odd curve that differs from other missions. However, when the water level increase higher than 387 m, the trend is identical to that of the other missions. The reason can be determined from water occurrance frequnency map in the area plot. As seen in Figure 4.33, the portion of the river where the Topex missions pass is narrow, shallow, and far from the river's main channel. Once the water level declines, this section is likely to dry out and separate from the main body of the lake, resulting in a river with a different height than the main body. This why the water level observations from the track crossing this district is lower than that from the track crossing the main body of the lake.

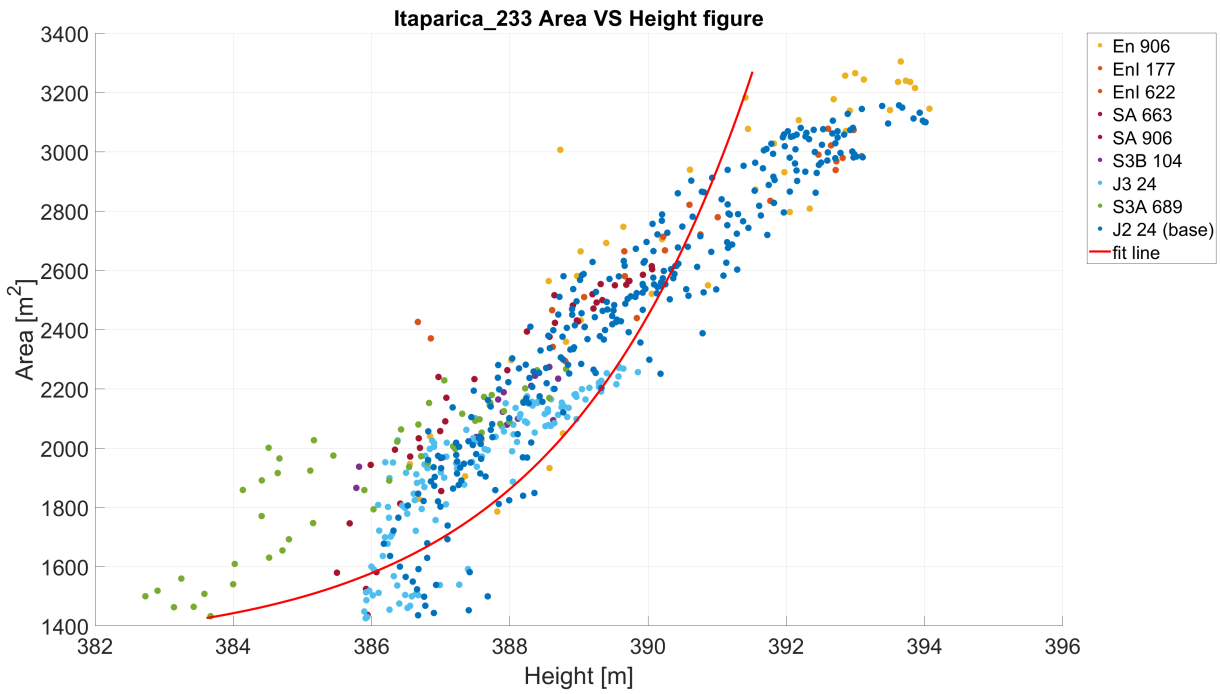


Figure 4.3: *Area-height figure of the case Itaparica reservoir without outlier rejection*

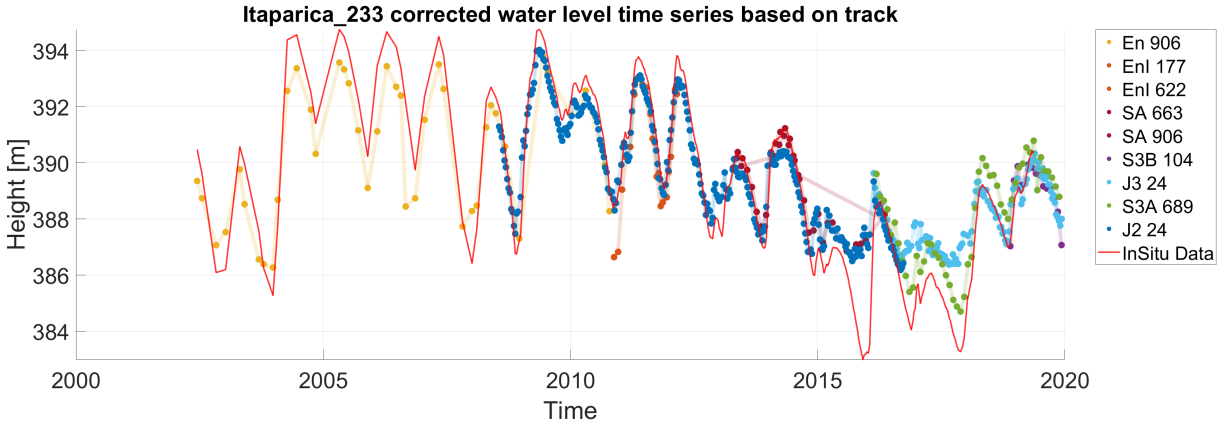


Figure 4.4: *Biases corrected water level of the case Itaparica reservoir without outlier rejection*

Sat Track	J2 24 (base)	En 906	EnI 177	EnI 622	SA 663	SA 906	S3B 104	J3 24	S3A 689
Bias (m)	0	0.296	0.161	0.037	-0.904	-1.165	-1.243	-0.470	-1.967
Uncertainty (m)	0.000	0.034	0.201	0.076	0.140	0.048	0.0.092	0.016	0.031

Table 4.2: *The final inter-track biases of Itaparica reservoir*

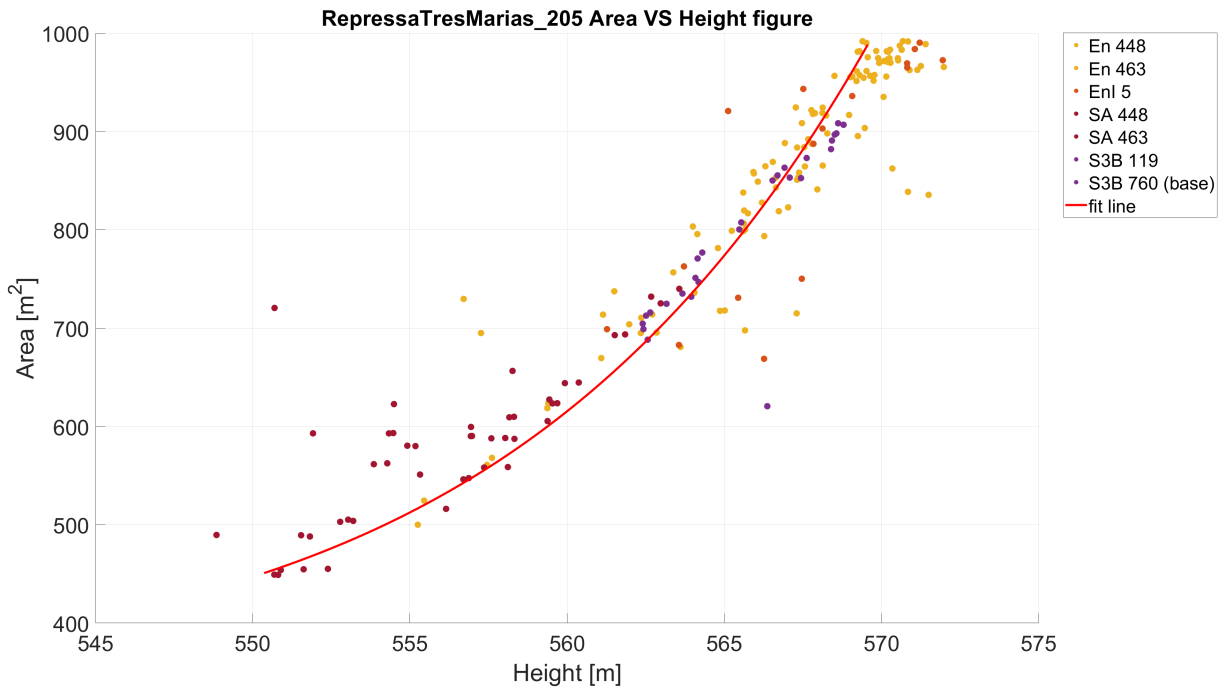


Figure 4.5: *Area-height figure of the case Repressa TresMatrias reservoir without outlier rejection*

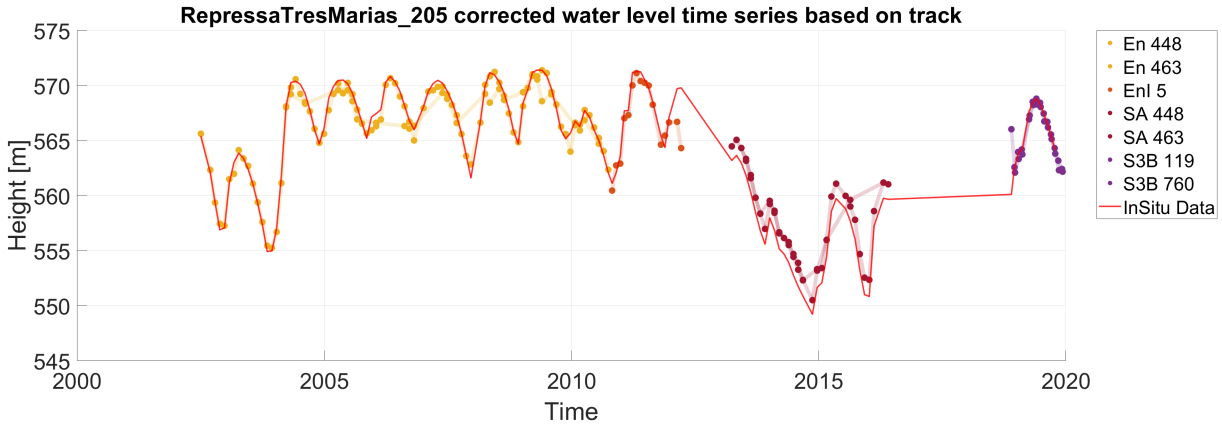


Figure 4.6: *Biases corrected water level time series of the case Repressa TresMatrias reservoir without outlier rejection*

Sat Track	S3B 760 (base)	En 448	En 463	EnI 5	SA 448	SA 463	S3B 119
Bias (m)	0.000	0.965	0.020	0.823	-1.484	-1.638	0.352
Uncertainty (m)	0.000	0.290	0.220	0.347	0.405	0.370	0.367

Table 4.3: *The final inter-track biases of Repressa TresMatrias*

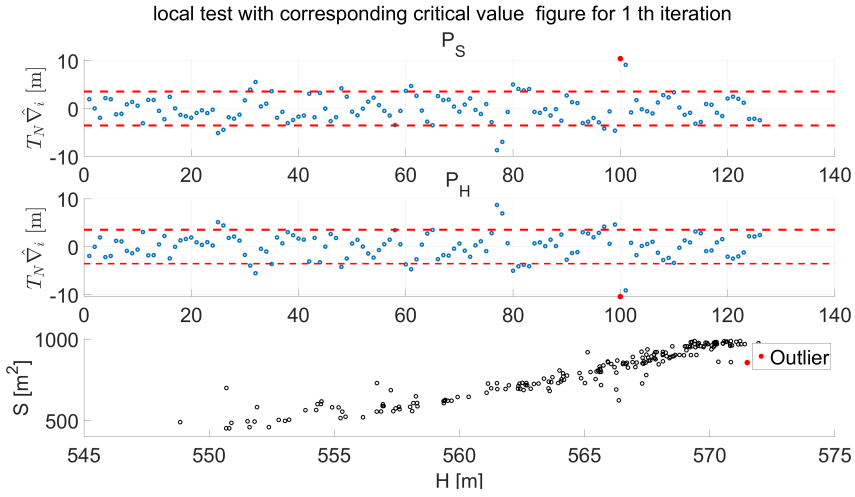
4.1.2 Outlier rejection using local test based on Baardas' data snooping

As mentioned in chapter 3, the algorithm starts with local test based on Baardas'snooping, which aims to detect and eliminate outliers with large residuals in the observations with statistical hypothesis tests (Ettlinger and Neuner, 2020). Therefore the preset significant number of the global tests was set relatively large $\alpha = 0.1$ and the power was set to $\beta = 0.8$. Meanwhile, the priori variance of the unit weight σ_2 is needed, which is set to 1 in this study for all cases. After applying the local test, the quality of the results are shown in Table 4.4.

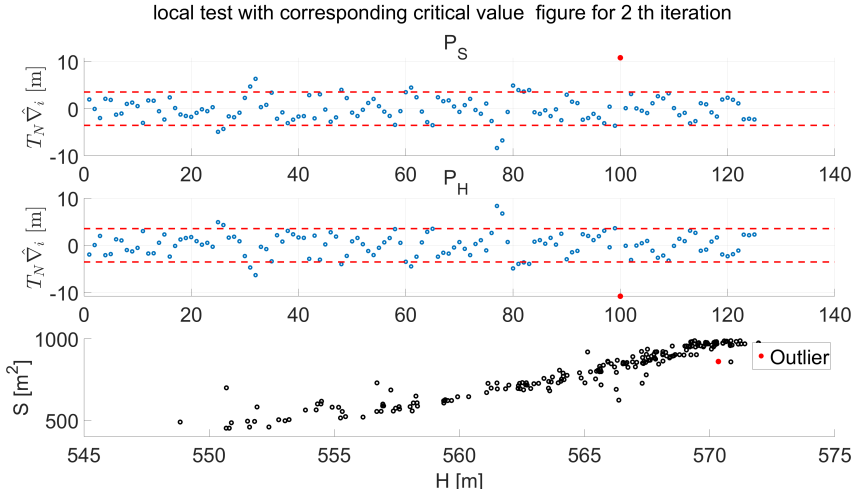
Poly ID	Name	Standard deviation of residuals (m)	RMSE (m)
233	Itaparica reservoir	0.3370	0.6433
205	Repressa TresMatrias reservoir Lake	0.1962	0.2304

Table 4.4: *The results of the estimation with EM algorithm*

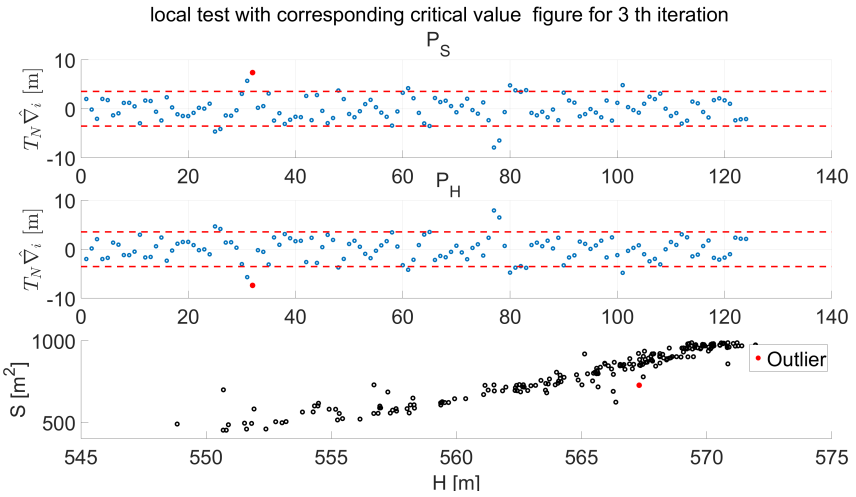
Figure 4.7 shows the test values $T_N(\nabla_N)$ of the n local tests for identifying the erroneous observations. It can be seen, that the erroneous point is correctly identified but not specifically single type of observations. Both water area and water level observations are identified as being subject to an outlier. As the observations with large residuals was successfully detected and removed, RMSE and RMSE refer to in-situ data for both case have decreased significantly comparing the Table 4.1 and Table 4.4. From Figure 4.8 through Figure 4.11, it is also clear that while the majority of outliers with latge residuals have been successfully identified by the algorithm, some observations with relatively small error still exist, due to relatively large preseted significant number of the global tests. Consequently, the additional outlier rejection procedure to reduce the impact of the undetected outliers is required.



(a)



(b)



(c)

Figure 4.7: *local test with corresponding critical values (red horizontal lines). The values exceeding* $\alpha_0/2$ *and* $1-\alpha_0/2$ *are filled with red color.*

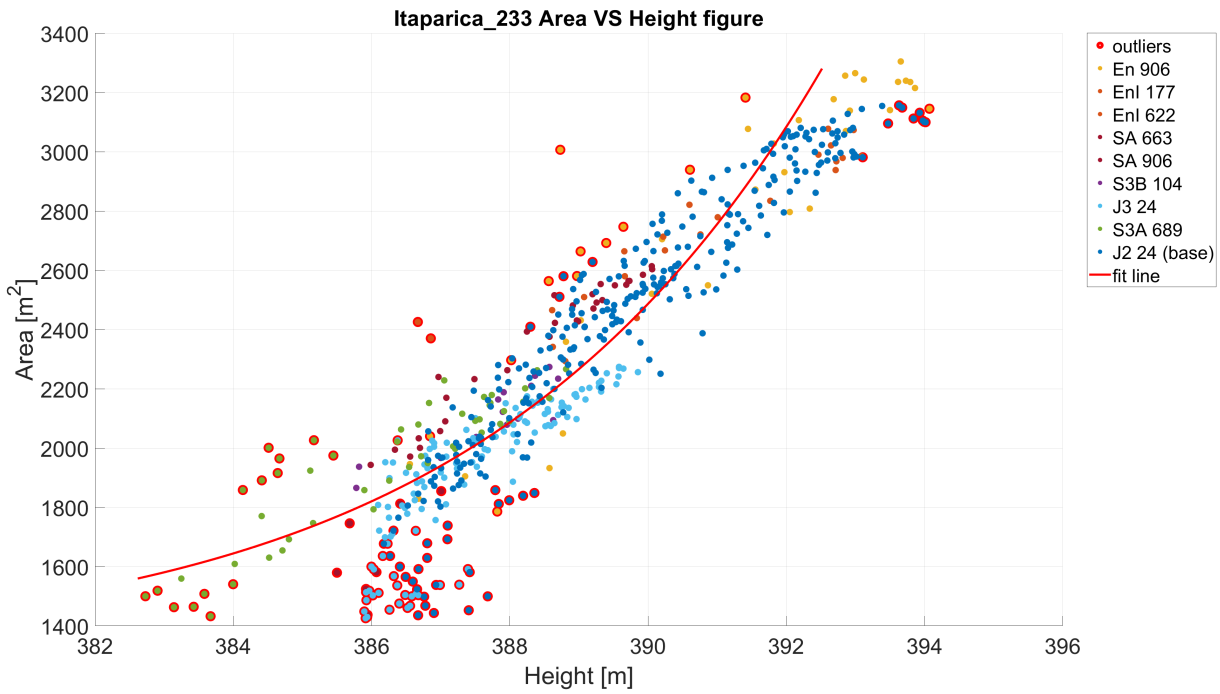


Figure 4.8: *Biases corrected water level time series of the case Itaparica reservoir with local test*

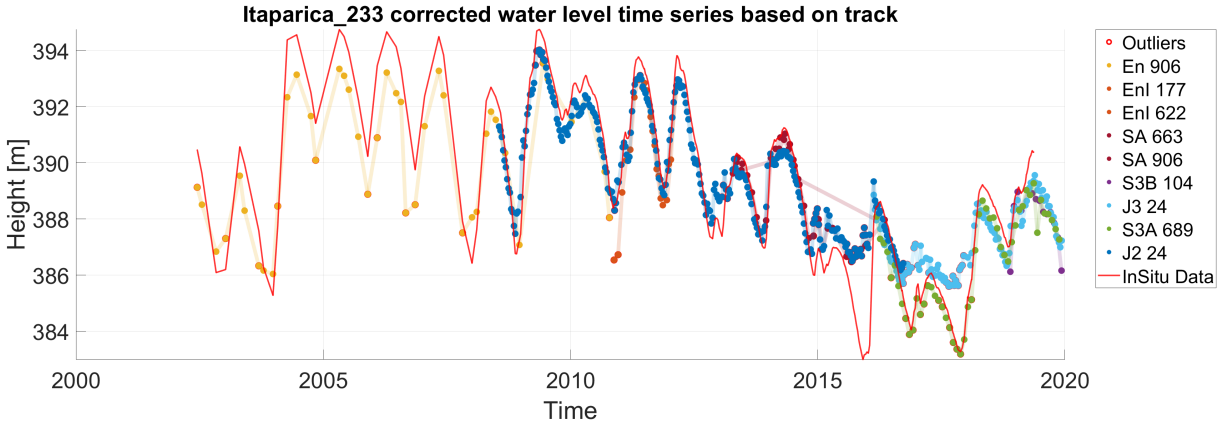


Figure 4.9: *Biases corrected water level time series of the case Itaparica reservoir with local test*

Sat Track	J2 24 (base)	En 906	EnI 177	EnI 622	SA 663	SA 906	S3B 104	J3 24	S3A 689
Bias (m)	0	0.522	-0.112	0.138	-0.762	-0.971	0.333	0.302	-0.450
Uncertainty (m)	0.000	0.013	0.062	0.026	0.042	0.019	0.028	0.006	0.013

Table 4.5: *The inter-track biases of Itaparica reservoir estimated by local test*

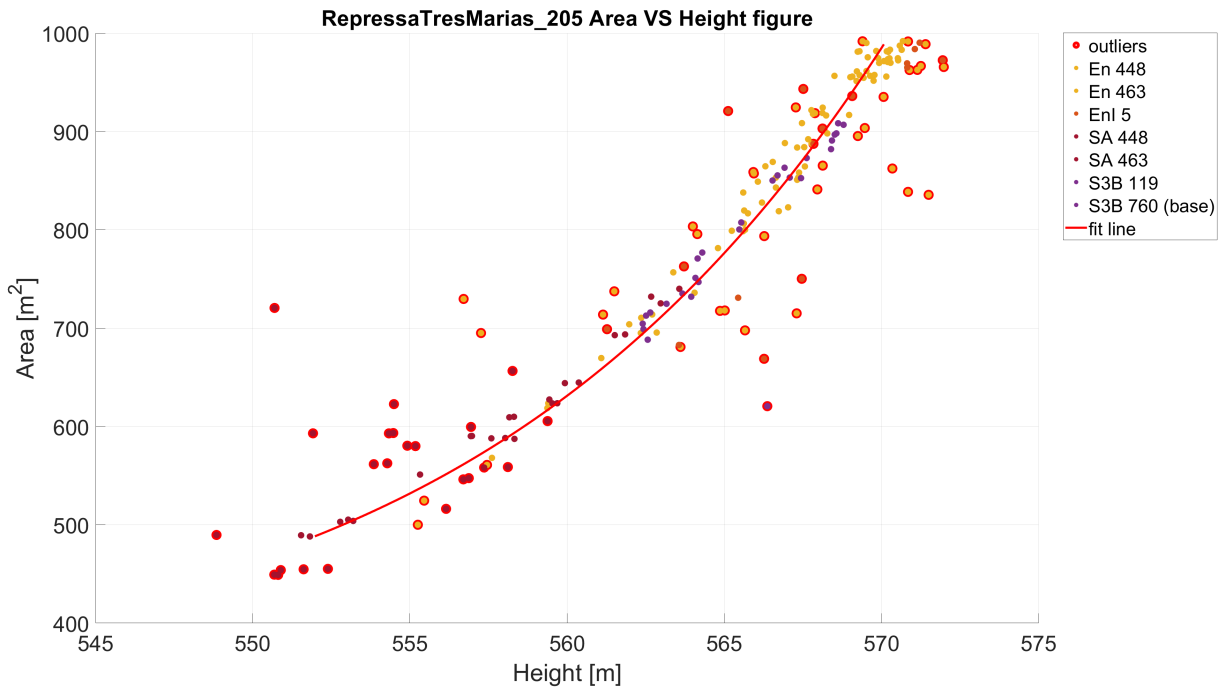


Figure 4.10: *Area-height figure of the case Repressa TresMatrias reservoir with local test*

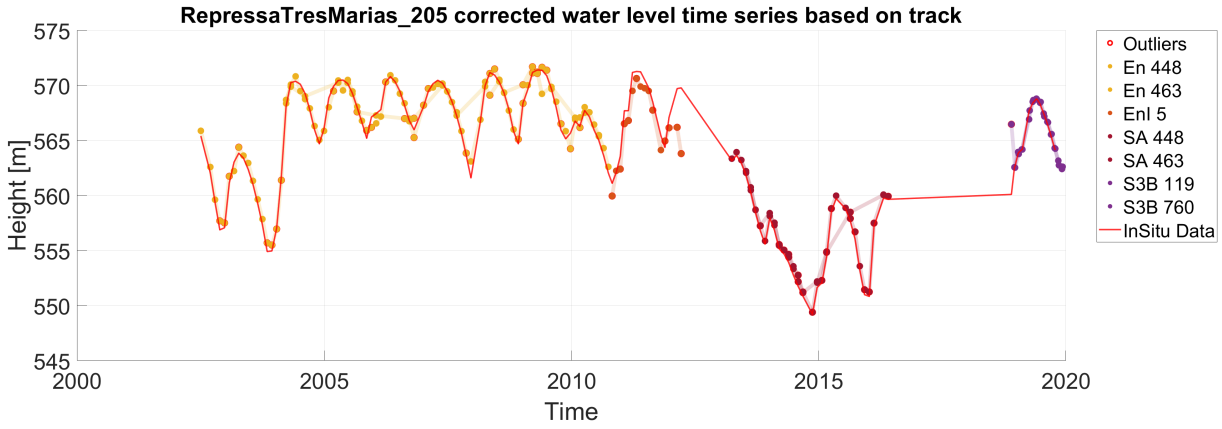


Figure 4.11: *Biases corrected water level time series of the case Repressa TresMatrias reservoir with local test*

Sat Track	S3B 760 (base)	En 448	En 463	EnI 5	SA 448	SA 463	S3B 119
Bias (m)	0	0.417	-0.091	1.100	-0.420	-0.398	-0.110
Uncertainty (m)	0.000	0.037	0.026	0.054	0.059	0.052	0.043

Table 4.6: *The inter-track biases of Repressa TresMatrias reservoir estimated by local test*

4.1.3 Outlier rejection EM algorithm

To further reduce the impact of the outliers, the Expectation Maximization algorithm (EM) applied. RMSE and RMSE refer to in-situ data are shown in Table 4.7

Poly ID	Name	Standard deviation of residuals (m)	RMSE (m)
233	Itaparica reservoir	0.3370	0.5833
205	Repressa TresMatrias reservoir Lake	0.1852	0.2152

Table 4.7: *The results of the estimation with EM algorithm*

As mentioned in chapter 3, the expectation maximization algorithm decreases the influence of the outliers by down-weighting the corresponding observations. 10 points of the observation in the case of Sobradino lake are selected to illustrate how the weight of the outliers with large residuals changes iteratively. Tables Table 4.8 to Table 4.11 show the details of points in 4 iterations by increasing weights. From the table, we can see the residuals of points 35, 37, 25.17 are significantly larger than the others. The weight of point number 35 has jumped from 0.729 to 0.592 after 4 iterations.

Point number	Residual e_i (m)	Weight w_i
35	1.9251	0.729
37	1.0284	0.802
25	0.5401	0.961
17	-0.4558	0.974
113	-0.0811	1.223
162	0.0171	1.410

Table 4.8: *iteration =1*

Point number	Residual e_i (m)	Weight w_i
35	2.0421	0.652
37	1.0293	0.795
25	0.548	0.912
17	-0.4372	0.988
113	-0.0805	1.245
162	0.0169	1.419

Table 4.9: *iteration = 2*

Point number	Residual e_i (m)	Weight w_i
35	2.0453	0.631
37	1.0284	0.802
25	0.5401	0.961
17	-0.4558	0.974
113	-0.081	1.223
162	0.0167	1.427

Table 4.10: *iteration = 3*

Point number	Residual e_i (m)	Weight w_i
35	2.0621	0.592
37	1.0317	0.713
25	0.5372	0.982
17	-0.4531	0.978
113	-0.0538	1.392
162	0.0161	1.824

Table 4.11: *iteration = 4*

With slightly changes in weight of the outliers during every iteration, the EM algorithm based on the variance -inflation model converge with the degree of the of the v after 734 iterations in the end. From Figure 4.12 to Figure 4.15, it is clearly that the the time series from the several tracks converge into a single, continuous time series after removing the biases and outliers. After taking two outliers rejection algorithm, we can observe from Table 4.12 through Table 4.13 that the estimated distances between the tracks have been computed successfully at the decimeter level.

The results of remaining 6 lakes and reservoirs have been presented in Figure 4.16 to Figure 4.21. Clearly, the algorithm relies on the quality of the fit line between height and area scatters. The area and height scatter plot indicates that, in addition to the inaccuracies in area and height observations, the bias estimate process is also sensitive to the riverbank slope. When the riverbank slope is very steep, the change in water area is insensitive to the change in water level, resulting in the dispersion of area-height plots. In this scenario, it is difficult for an algorithm to fit the relationship between lake height and area. When the riverbank is more sloping, a little change in water level will result in a significant change in area observation, resulting in more concentrated area-height plots. In this instance, relationship between water area and height is clear and easier to be be determined by fit algorithm.

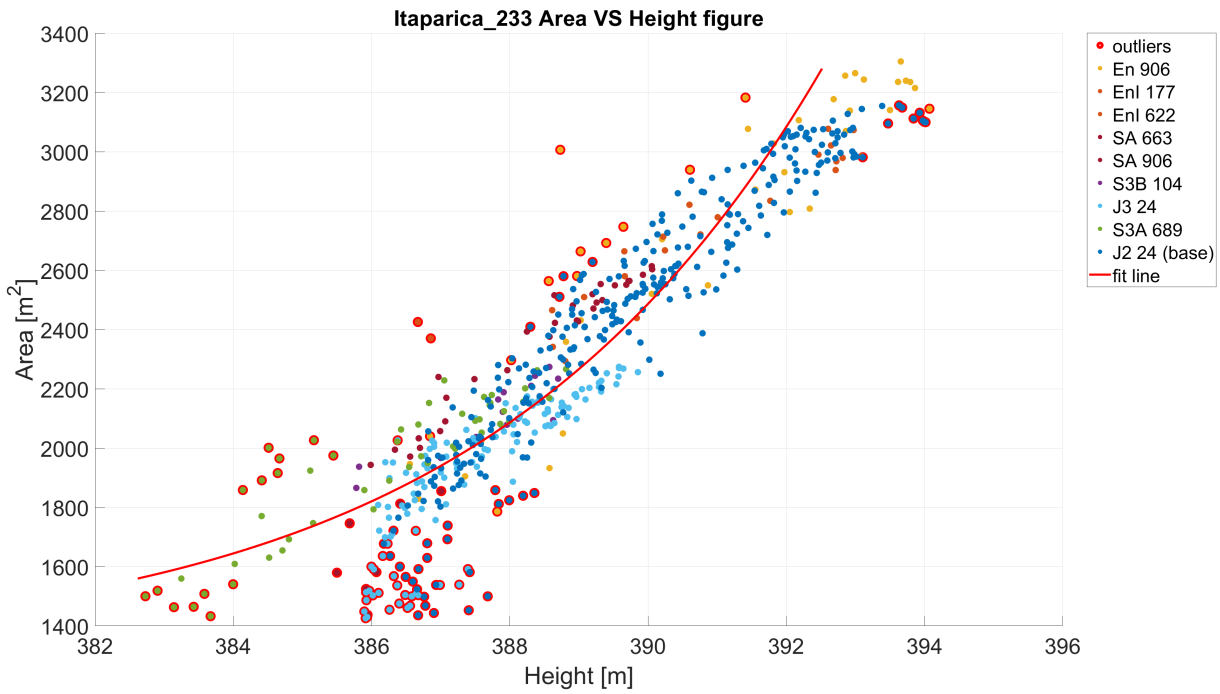


Figure 4.12: *Area-height figure of the case Itaparica reservoir with EM algorithm*

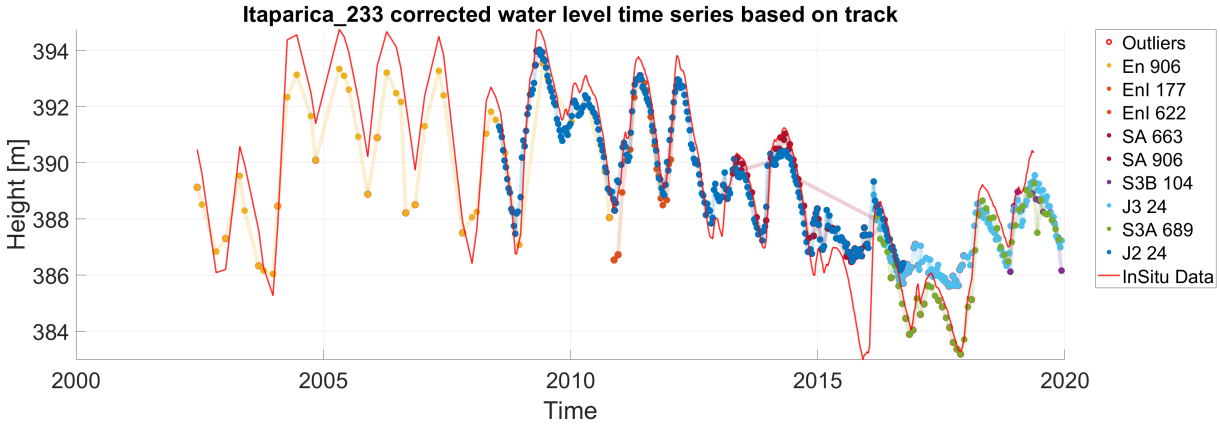


Figure 4.13: *Biases corrected water level time series of the case Itaparica reservoir with EM Algorithm*

Sat Track	J2 24 (base)	En 906	EnI 177	EnI 622	SA 663	SA 906	S3B 104	J3 24	S3A 689
Bias (m)	0	0.6365	-0.051	0.148	-0.784	-0.991	0.364	0.312	-0.446
Uncertainty (m)	0.000	0.013	0.063	0.027	0.033	0.016	0.024	0.006	0.013

Table 4.12: *The inter-track biases of Itaparica reservoir estimated by EM algorithm*

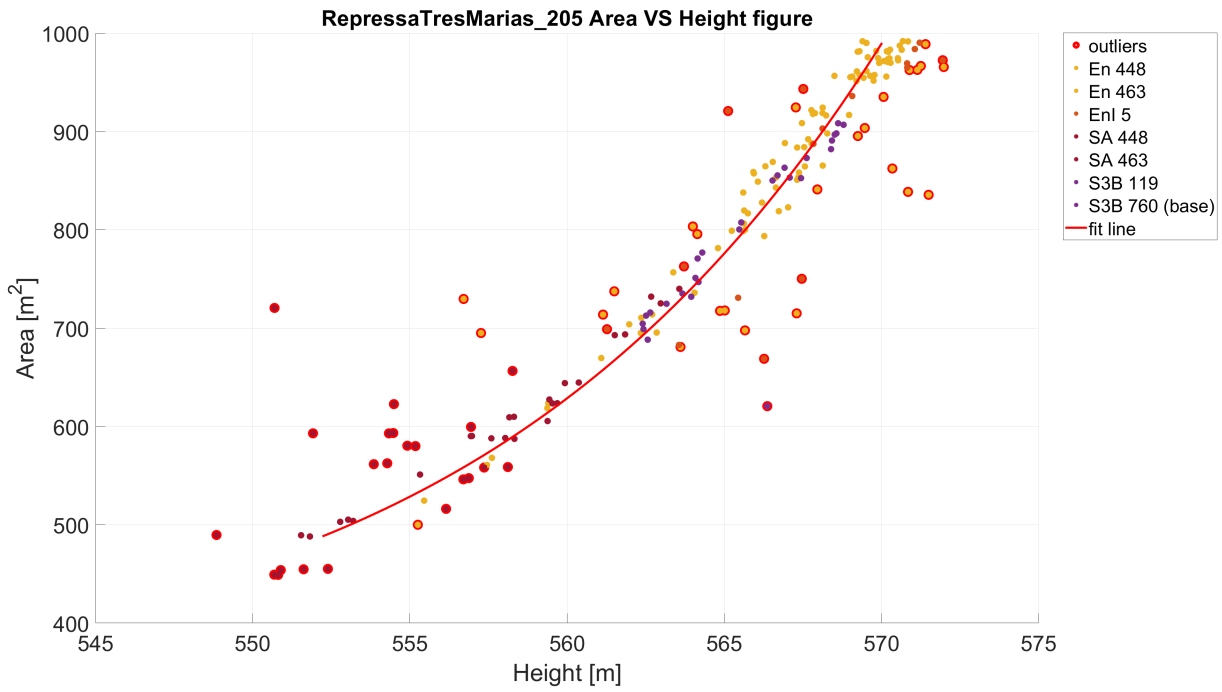


Figure 4.14: *Area-height figure of the case Repressa TresMatrias reservoir with EM algorithm*

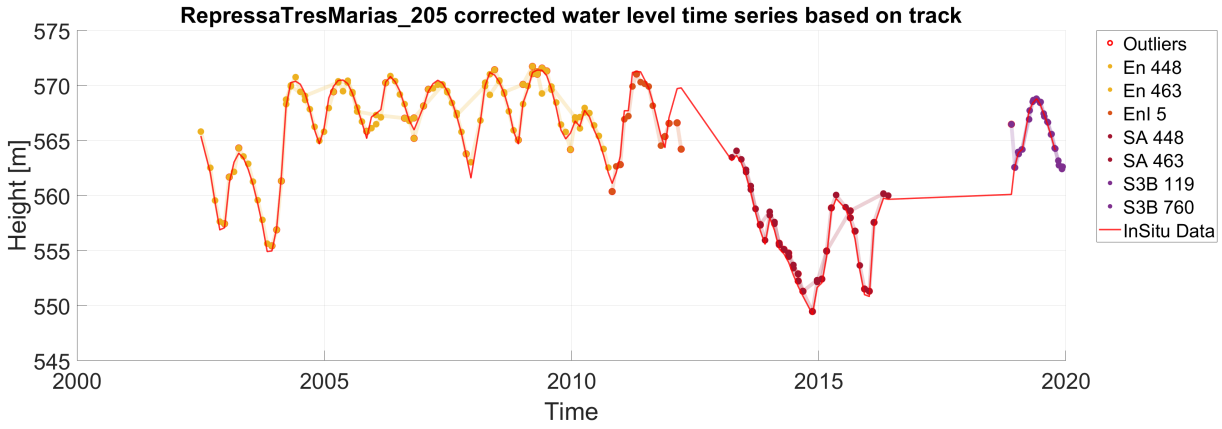


Figure 4.15: *Biases corrected water level time series of the case Repressa TresMatrias reservoir With EM algorithm*

Sat Track	S3B 760 (base)	En 448	En 463	EnI 5	SA 448	SA 463	S3B 119
Bias (m)	0	0.407	-0.095	1.093	-0.413	-0.523	-0.110
Uncertainty (m)	0.000	0.037	0.025	0.053	0.055	0.052	0.041

Table 4.13: *The inter-track biases of Repressa TresMatrias estimated by EM algorithm*

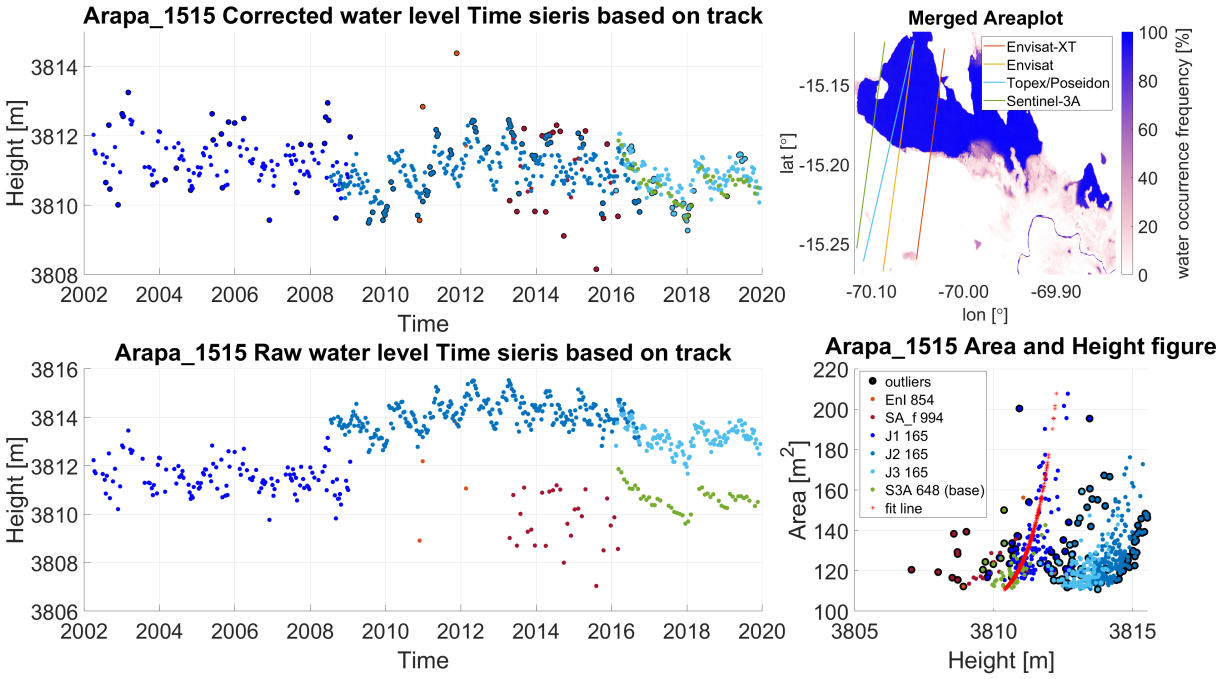


Figure 4.16: *The final result of Arapa Lake*

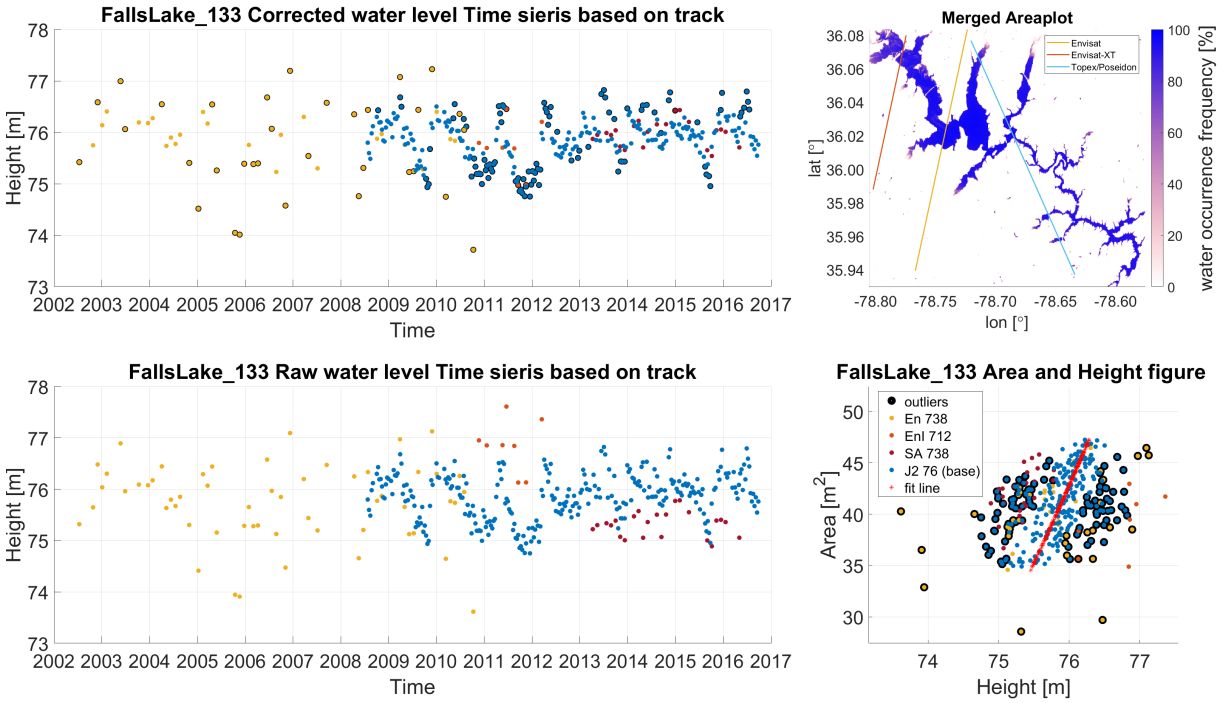


Figure 4.17: *The final result of FallsLake*

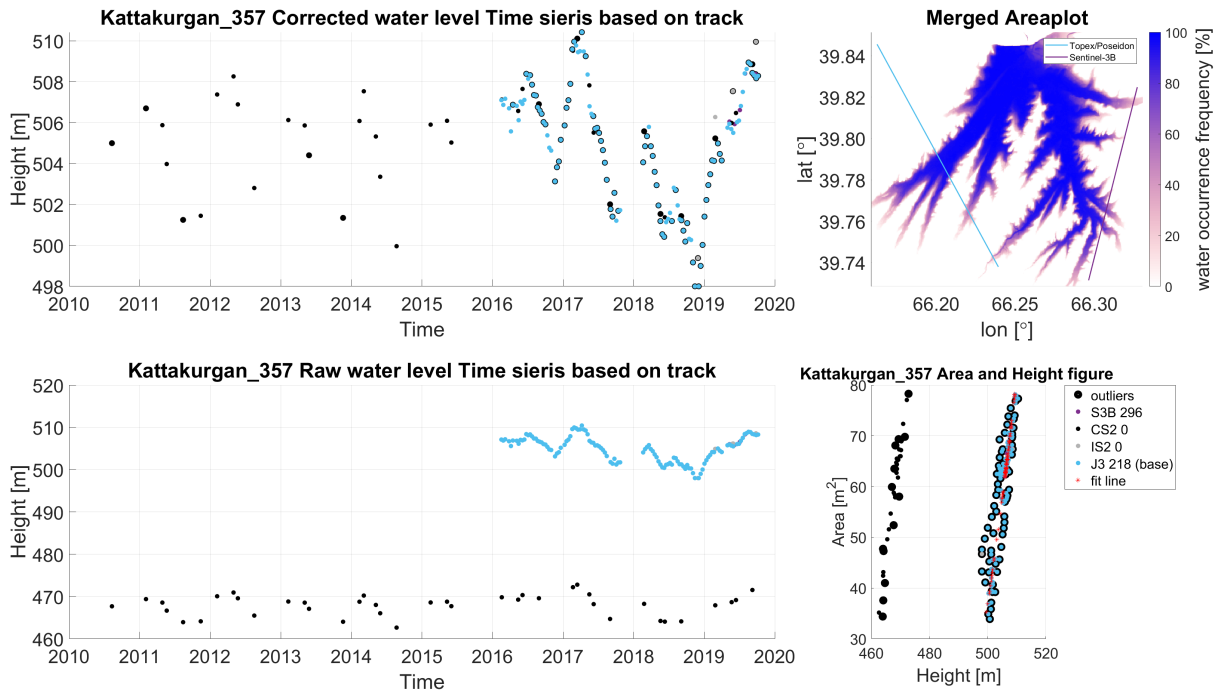


Figure 4.18: *The final result of Kattakurgan Lake*

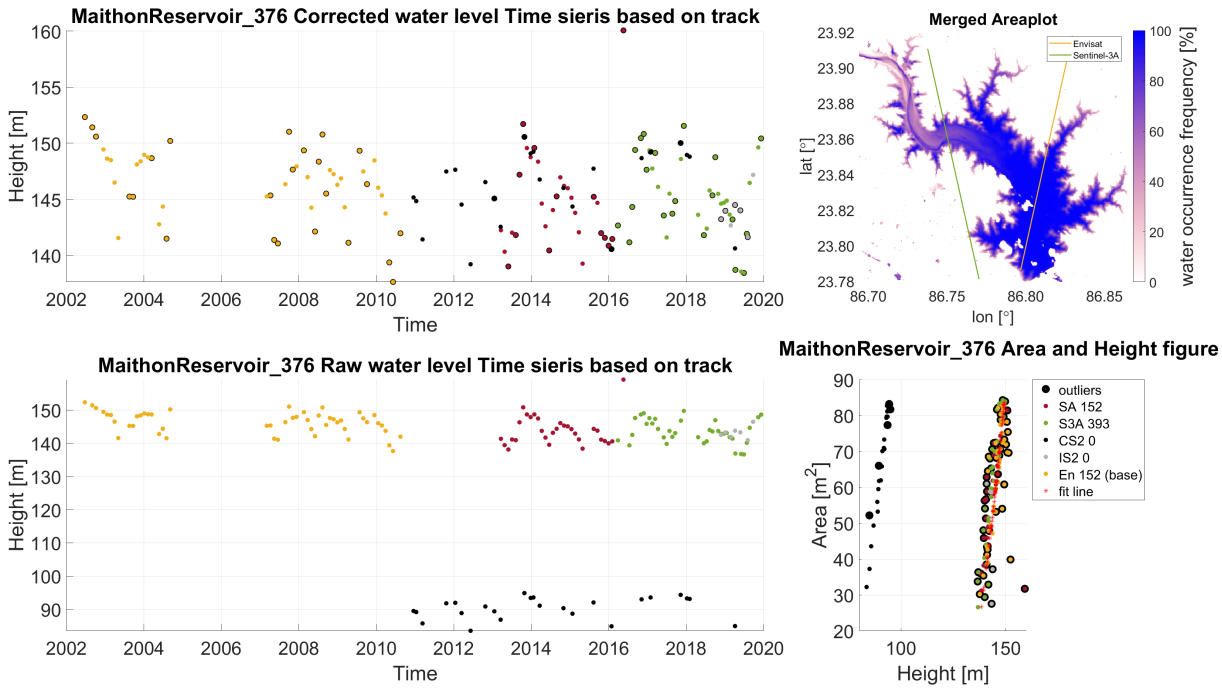


Figure 4.19: *The final result of Maithon Reservoir*

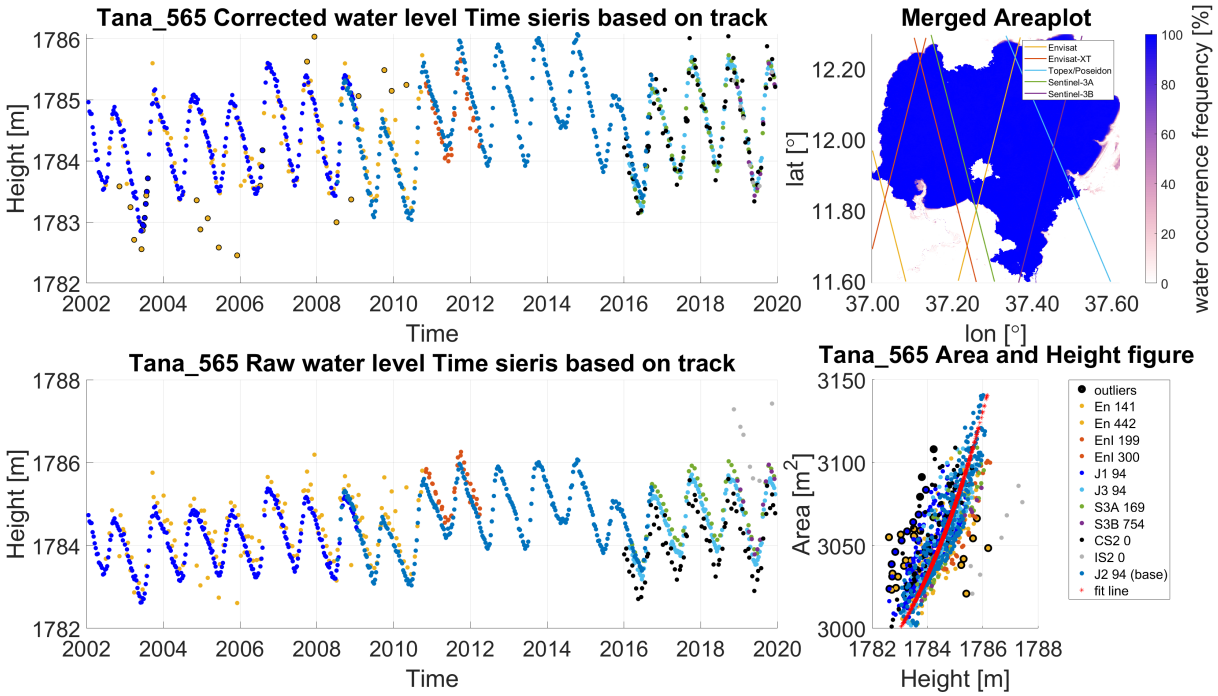


Figure 4.20: *The final result of Tana Lake*

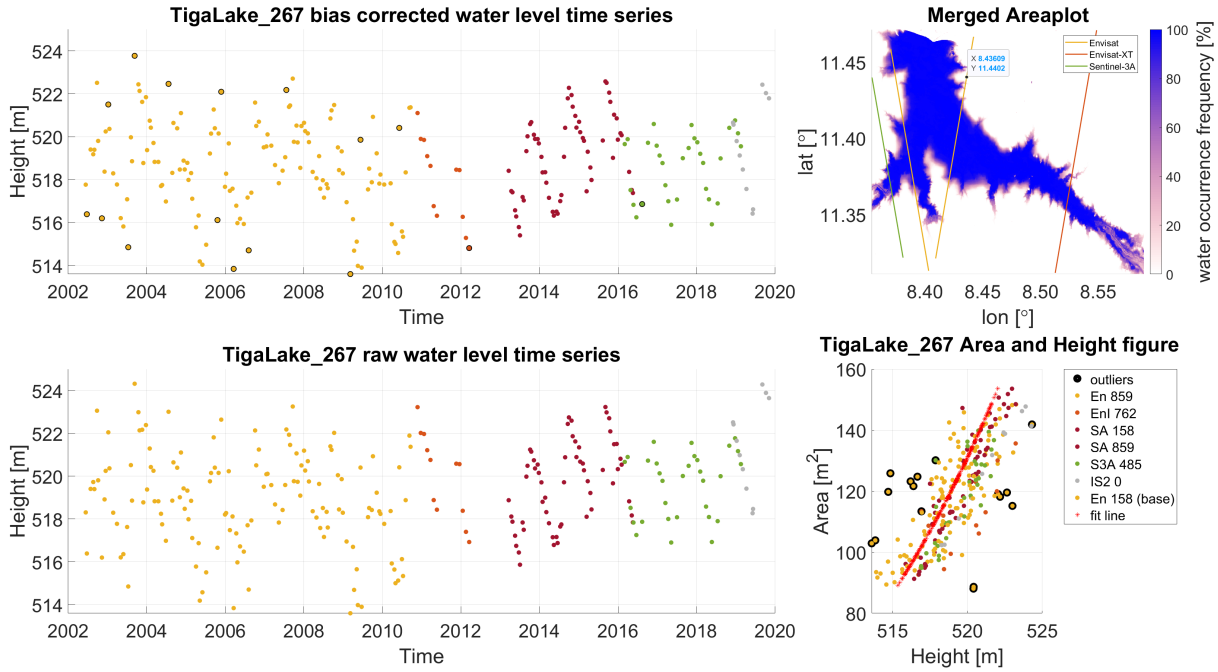


Figure 4.21: *The final result of Tiga Lake*

Sat Track	S3A 648 (base)	EnI 854	SA_f 994	J1 165	J2 165	J3 135
Bias (m)	0	-0.655	-1.112	0.199	3.073	2.404
Uncertainty (m)	0.000	0.032	0.042	0.035	0.037	0.048

Table 4.14: *The inter-track biases of Arapa Lake*

Sat Track	J2 76 (base)	En 738	EnI 712	SA 738
Bias (m)	0	-0.058	1.06	-0.584
Uncertainty (m)	0.000	0.021	0.031	0.027

Table 4.15: *The inter-track biases of FallsLake*

Sat Track	J3 218 (base)	S3B 296	CS2 0	IS2 0
Bias (m)	0	-0.055	-37.314	-1.376
Uncertainty (m)	0.000	0.153	0.231	0.159

Table 4.16: *The inter-track biases of Kattakurgan Lake*

Sat Track	En 152 (base)	SA 152	S3A 393	CS2 0	IS2 0
Bias (m)	0	-0.896	-1.8	-55.667	-0.71
Uncertainty (m)	0.000	0.109	0.117	0.116	0.115

Table 4.17: *The inter-track biases of Maithon Reservoir*

Sat Track	J2 94 (base)	En 141	En 442	EnI 199	EnI 300	J1 94
Bias (m)	0	0.157	0.075	0.509	0.625	-0.23
Uncertainty (m)	0.000	0.023	0.022	0.028	0.028	0.021
Sat Track	**S3A 169**	**S3B 754**	**CS2 0**	**IS2 0**	**IS2 0**	
Bias (m)	0.301	0.353	-0.436	1.971	1.971	
Uncertainty (m)	0.021	0.023	0.031	0.022	0.022	

Table 4.18: *The inter-track biases of Tana Lake*

Sat Track	En 158 (base)	En 859	EnI 762	SA 158	SA 859	S3A 485	IS2 0
Bias (m)	0	0.541	2.116	0.463	0.648	1.007	1.853
Uncertainty (m)	0.000	0.039	0.040	0.073	0.043	0.045	0.045

Table 4.19: *The inter-track biases of Tiga Lake*

4.2 Influence of the Topography

As mentioned in chapter 1, the inter-track biases could be influenced the differences in lake size, geographic location, surrounding topography, and land cover type. This section focuses on the effect of the topography along the track and surrounding lakes based on standard deviation of the elevation and mean slop of the area, with the help of the NASADEM. The target area is defined as the overlapping area of the buffer along the track and the buffer along the lake bank. Since Footprints are varied with the different missions and size of effective footprints decreases due to undulation of the terrain, the buffer along the track is designed to be a rectangular with 8 km width for all lakes. The buffer along the river is defined by the area along the river shape file with 8 km width. To accelerate the calculation, the buffer along the missions except the Saral/AltiKa and ICESat2 is calculated by the nominal orbit and the DEM have been been down sampled for large lakes.

4.2.1 Mean slope estimated

The mean slope indicates how rapidly the elevation changes around the river. The elevation difference is estimated by the mean slope calculated from the gradient in the buffer. Thus, we interpolate the elevation value in the buffer with a fixed step length. The along-track gradient and the elevation difference can then be computed by

$$\nabla E'_i = \frac{E'_i - E'_{i-1}}{d_i}$$

$$\text{Elevation difference} = \bar{\nabla E'}$$

where E'_i is the elevation value, which is interpolated in the buffer, n is the number of elevation values interpolated in the buffer.

The inter-track biases in this study are calculated relatively, which means the positive and negative signs of the biases in the topography analysis have no meaning. Therefore, the absolute values of inter-track biases are used to investigate the relationship between the biases and the mean slop of the topography. The results are shown from Figure 4.22 to Figure 4.29.

4.2.2 Elevation variance estimated for topography

The variance of elevation indicates how large the elevation changes around the river are. The lake with highly undulating surroundings has a larger standard deviation, especially for those near the mountains and hills. By selecting places in the buffer zone and calculating the standard deviation, root-mean-square of the difference between the real elevation and the mean height, elevation variation is determined as

$$\bar{E} = \frac{\sum_{i=1}^{n} E_i}{n}$$

$$\text{Elevation variance} = \sqrt{\frac{\sum_{i=1}^{n} \left(E_i - \bar{E}\right)^2}{n}}$$

Where n is the number of points within the buffer.

The results are shown from Figure 4.30 to Figure 4.37.

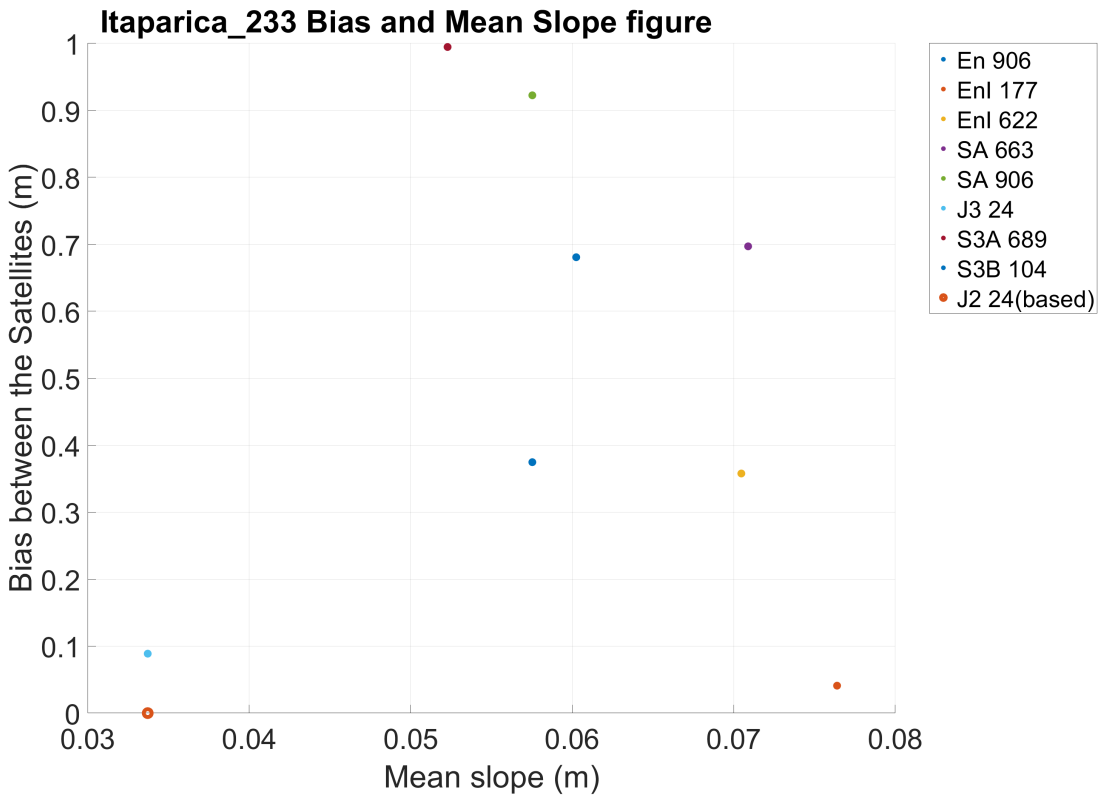


Figure 4.22: *The mean slope and bias figure of Itaparica reservoir*

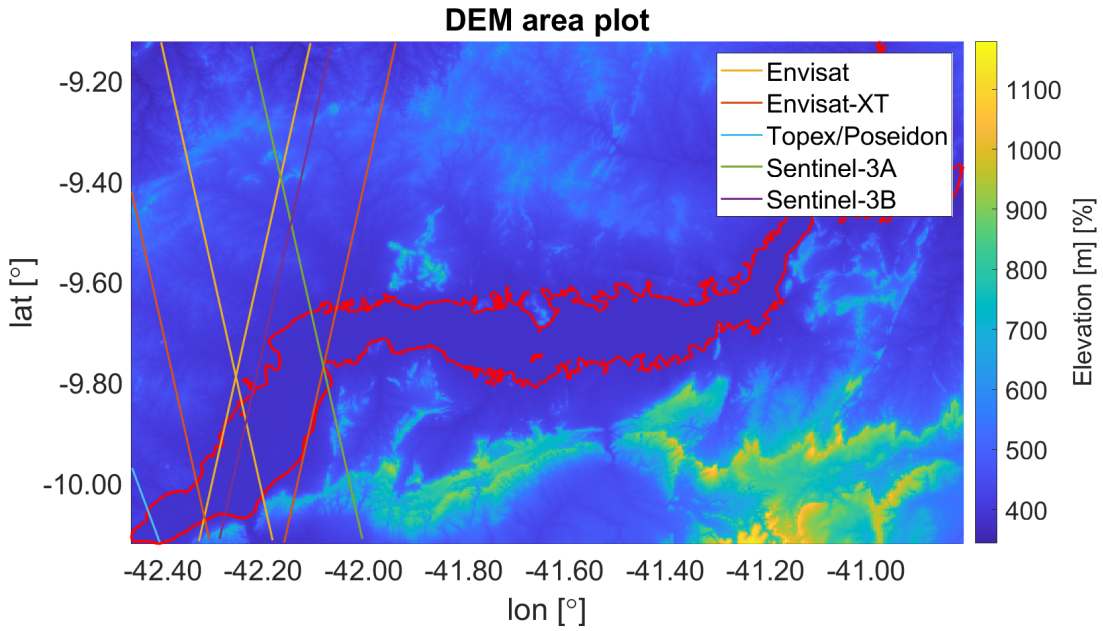


Figure 4.23: *The DEM area plot of Itaparica reservoir*

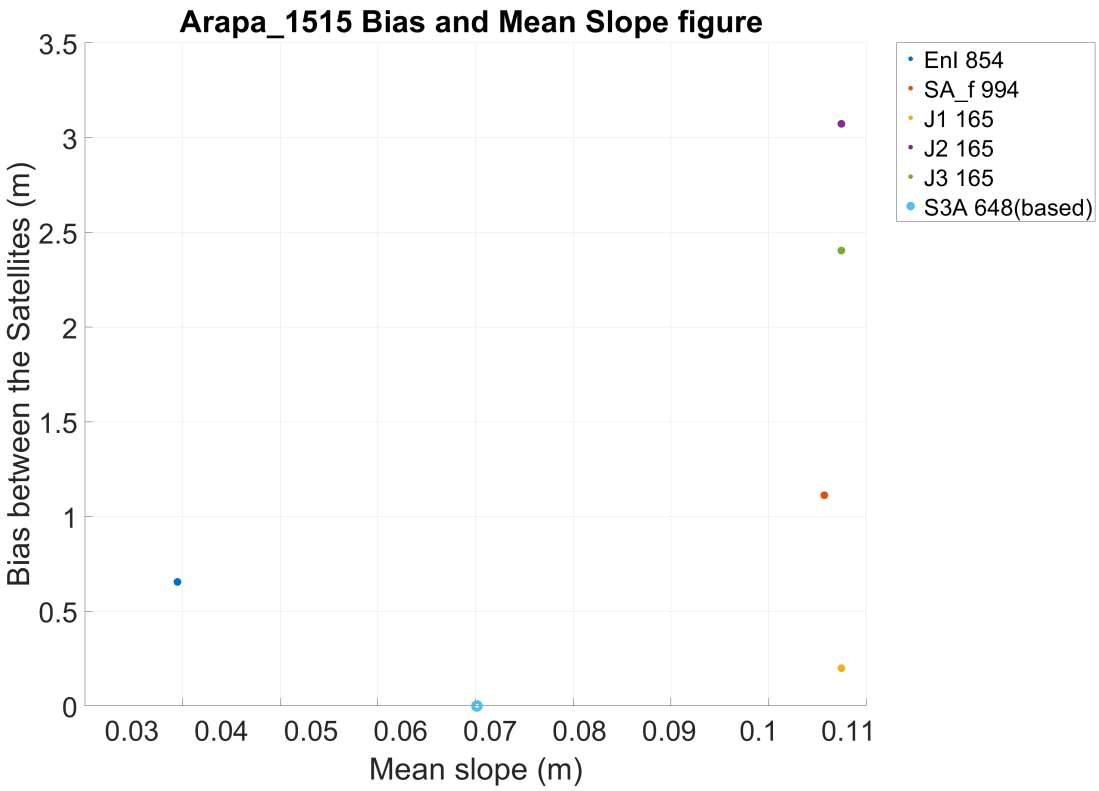


Figure 4.24: *The mean slope and bias figure of Arapa Lake*

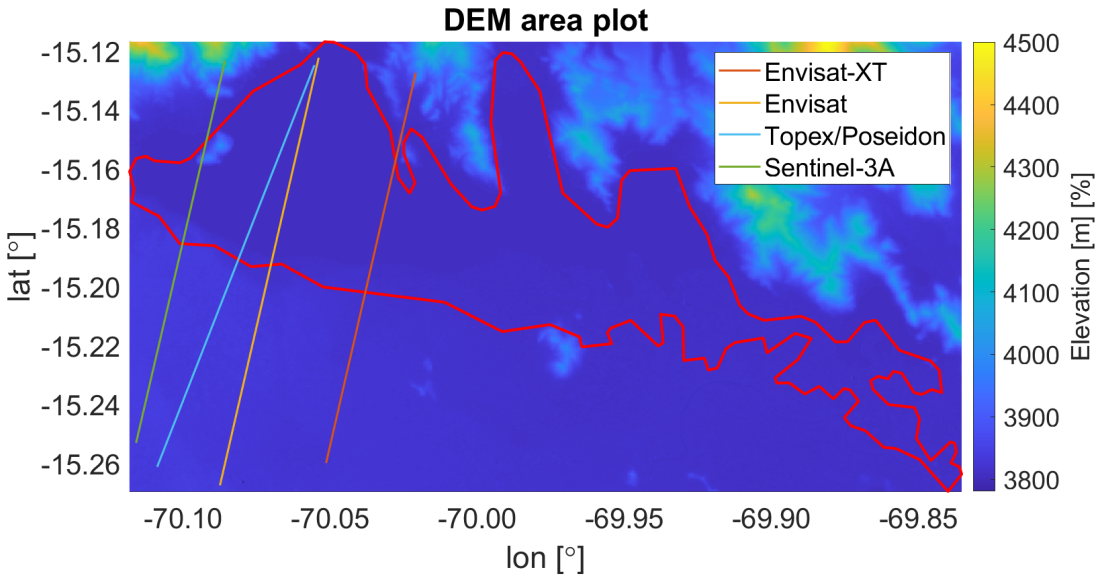


Figure 4.25: *The DEM area plot of Arapa Lake*

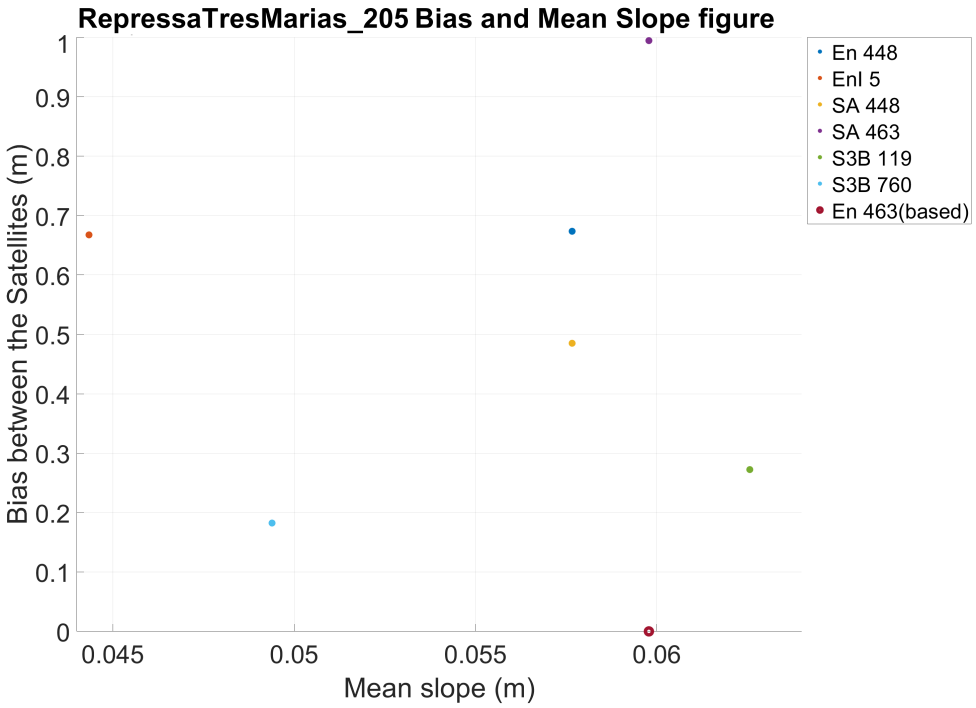


Figure 4.26: *The mean slope and bias figure of TresMarias Reservoir*

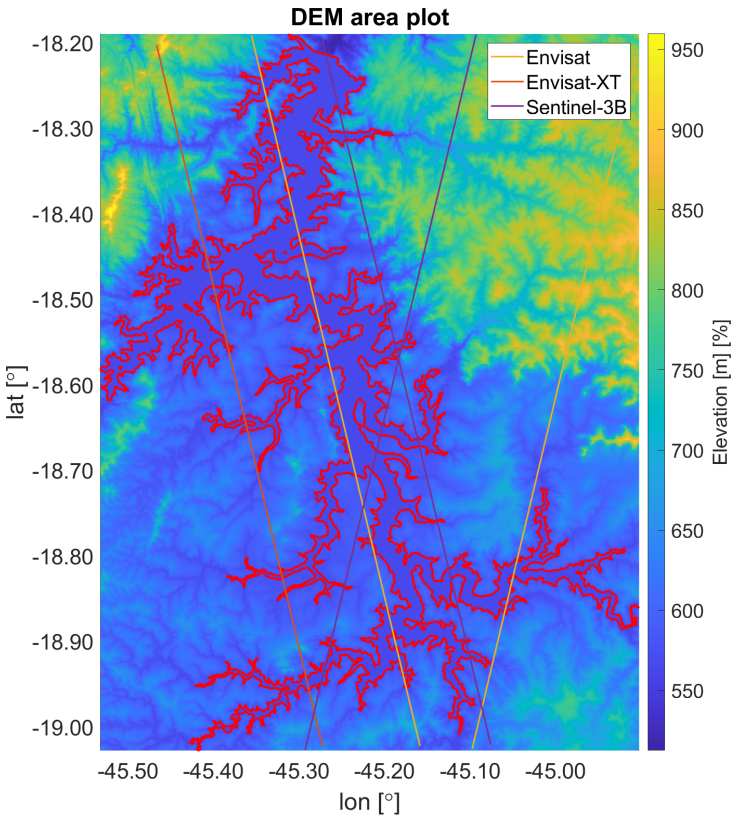


Figure 4.27: *The DEM area plot of TresMarias Reservoir*

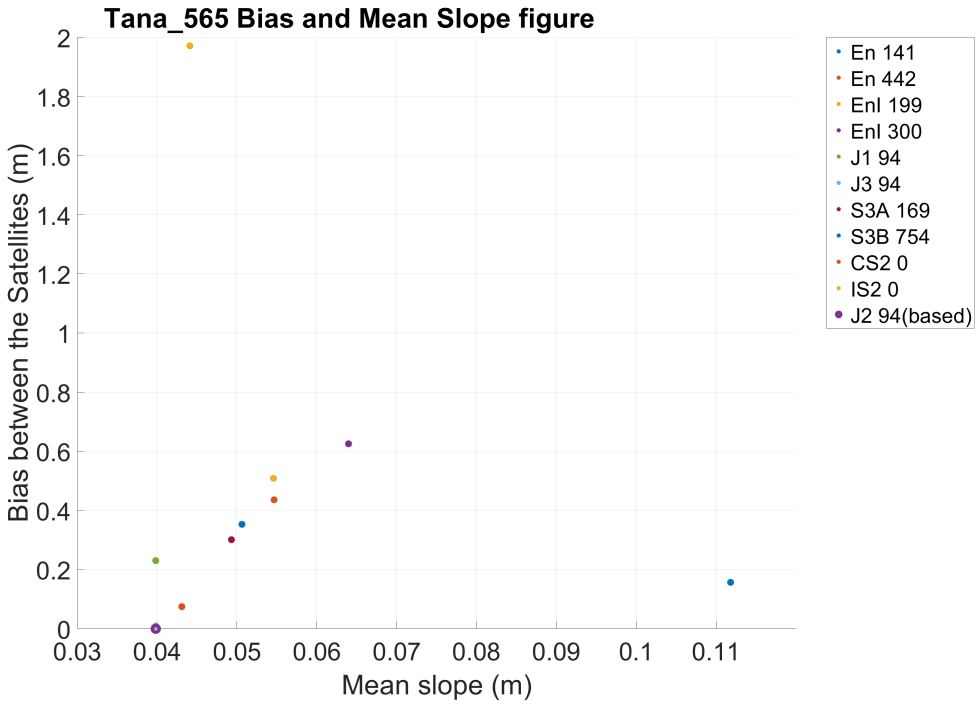


Figure 4.28: *The mean slope and bias figure of Tana reservoir*

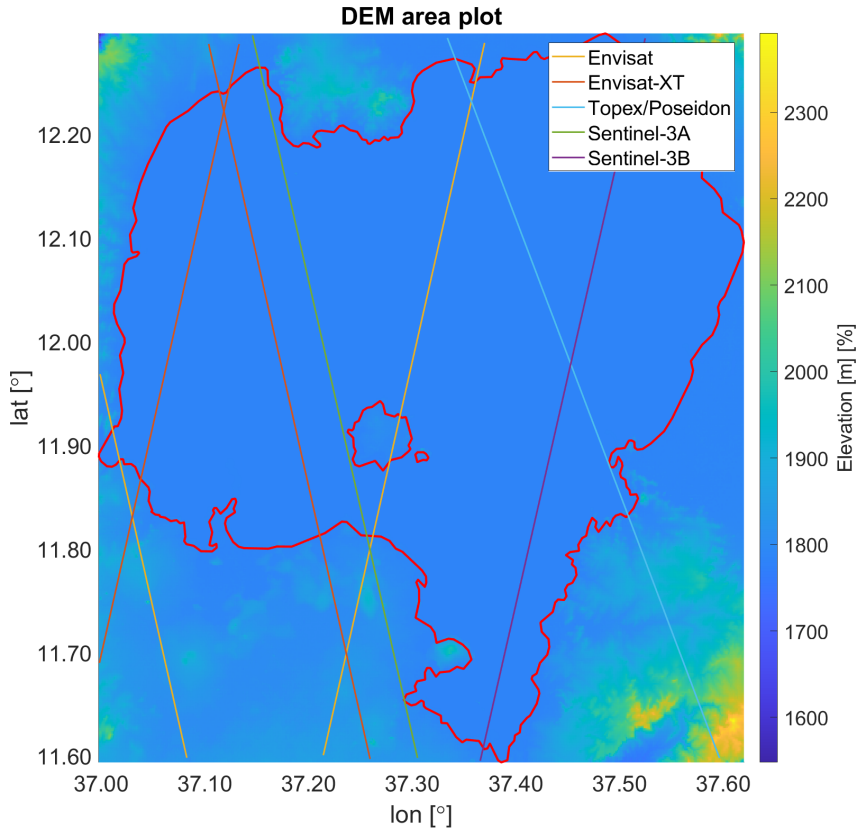


Figure 4.29: *The DEM area plot of Tana reservoir*

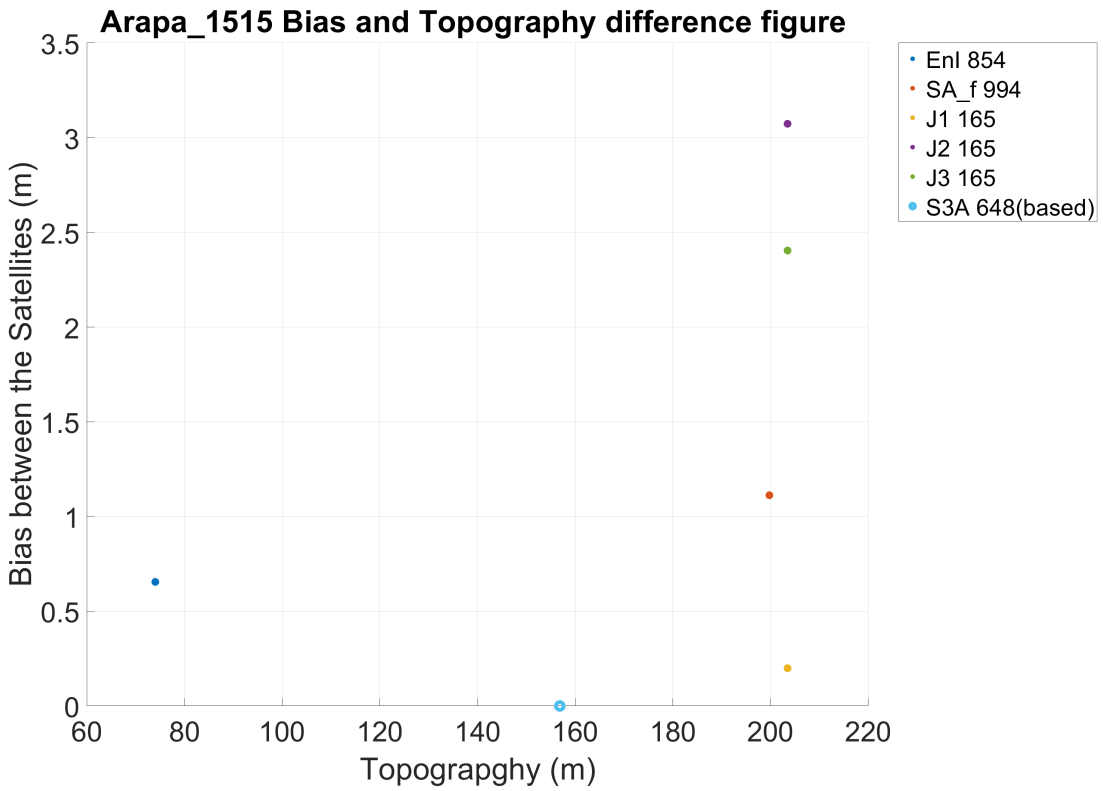


Figure 4.30: *The topography and bias figure of Arapa Lake*

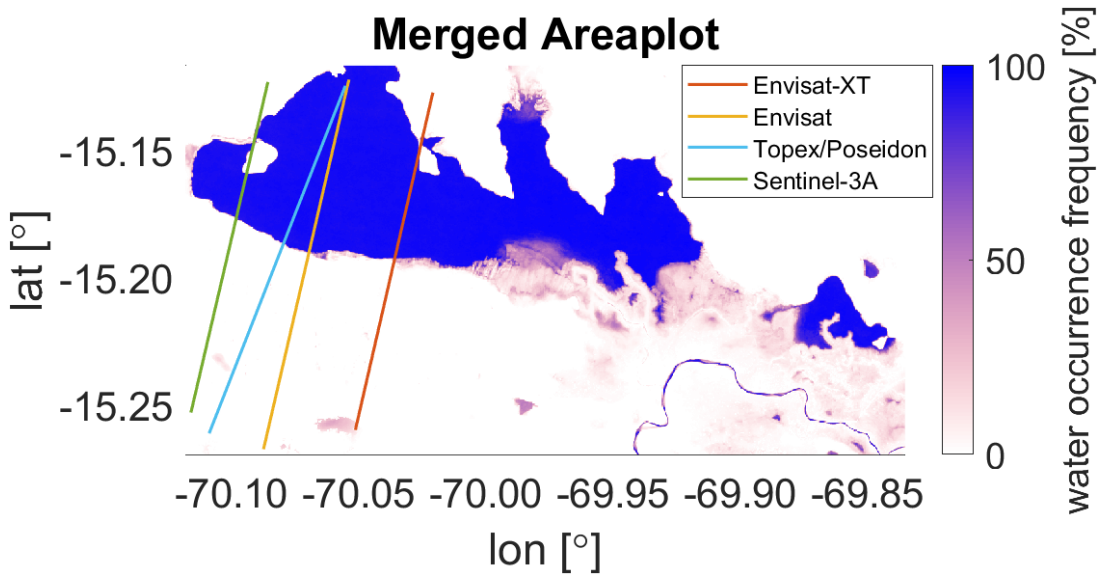


Figure 4.31: *The water occurrance area plot of Arapa Lake*

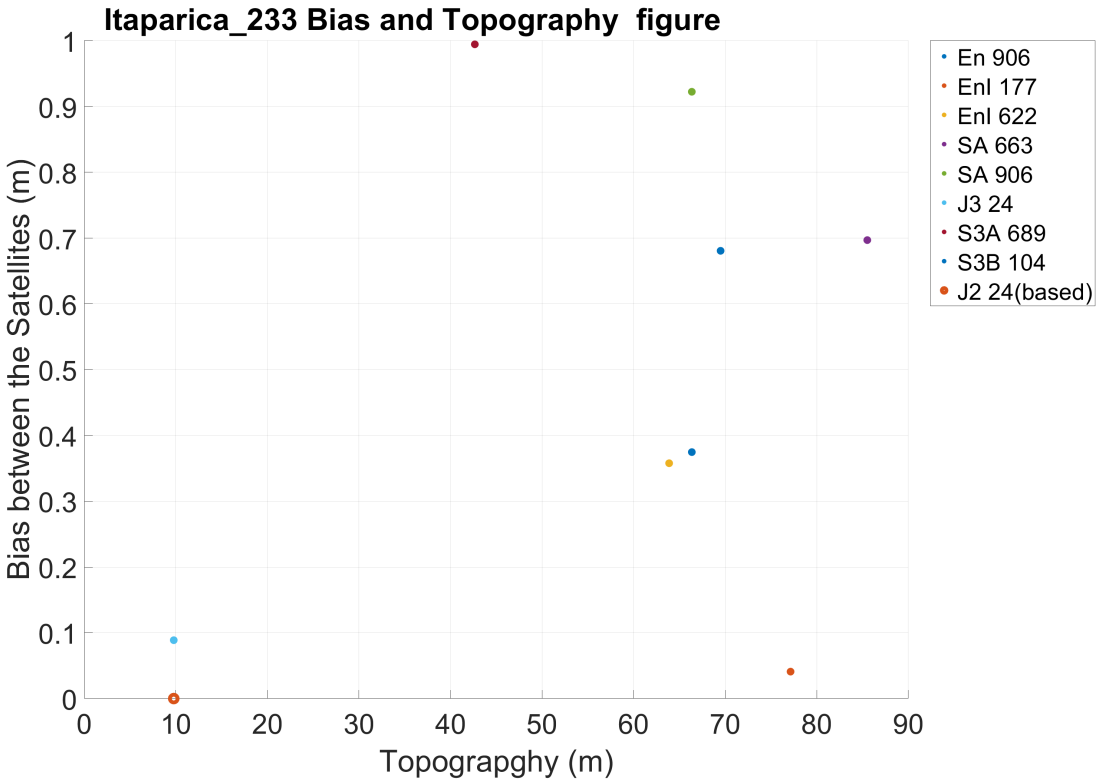


Figure 4.32: *The topography and bias figure of Itaparica Lake*

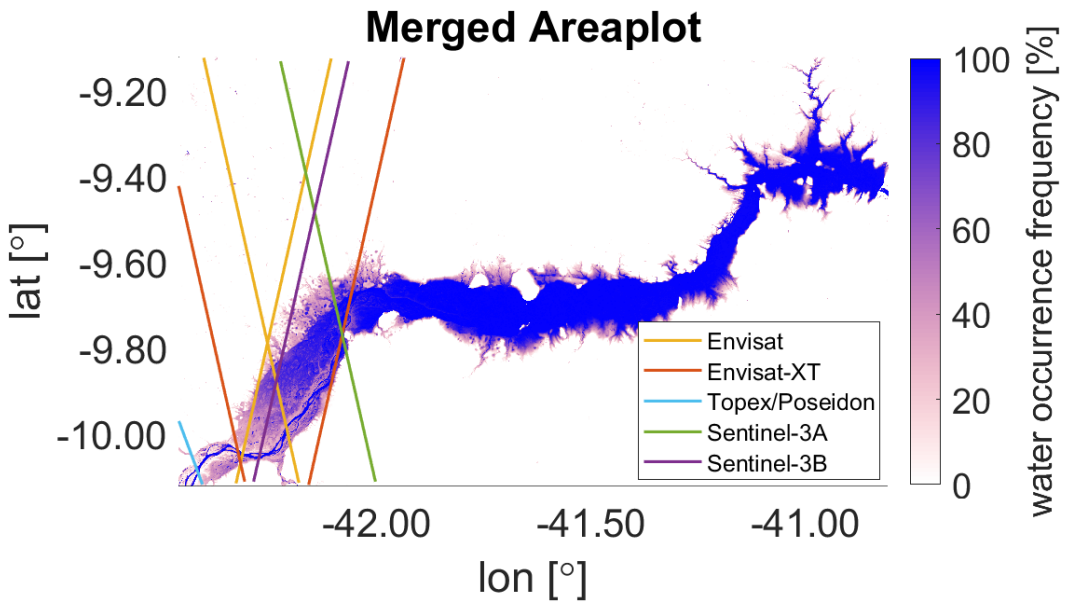


Figure 4.33: *The water occurrance area plot of Itaparica reservoir*

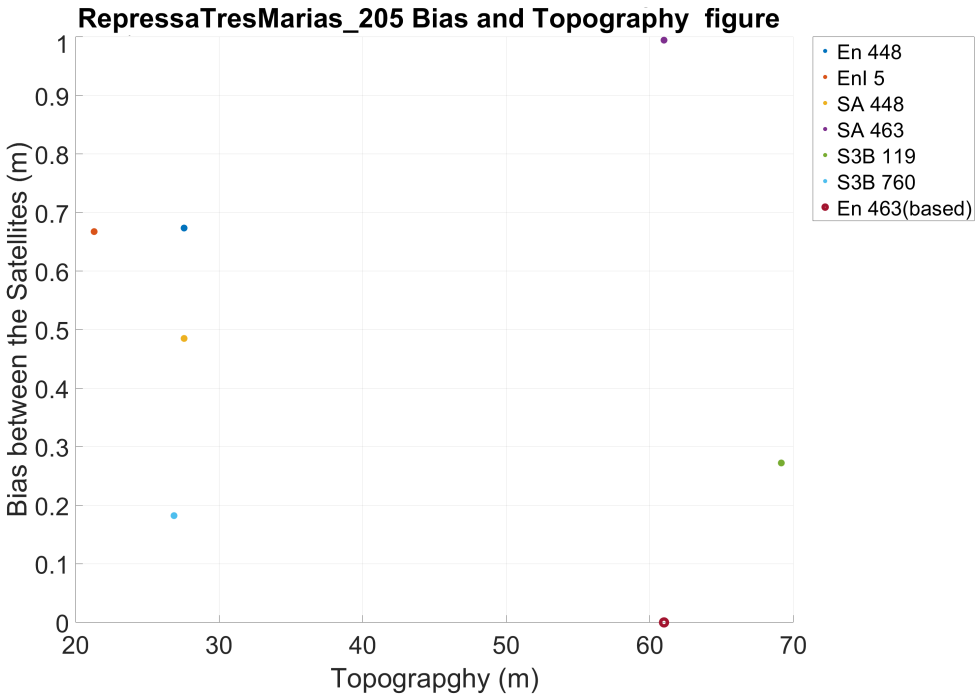


Figure 4.34: The topography and bias figure of TresMarias Reservoirs

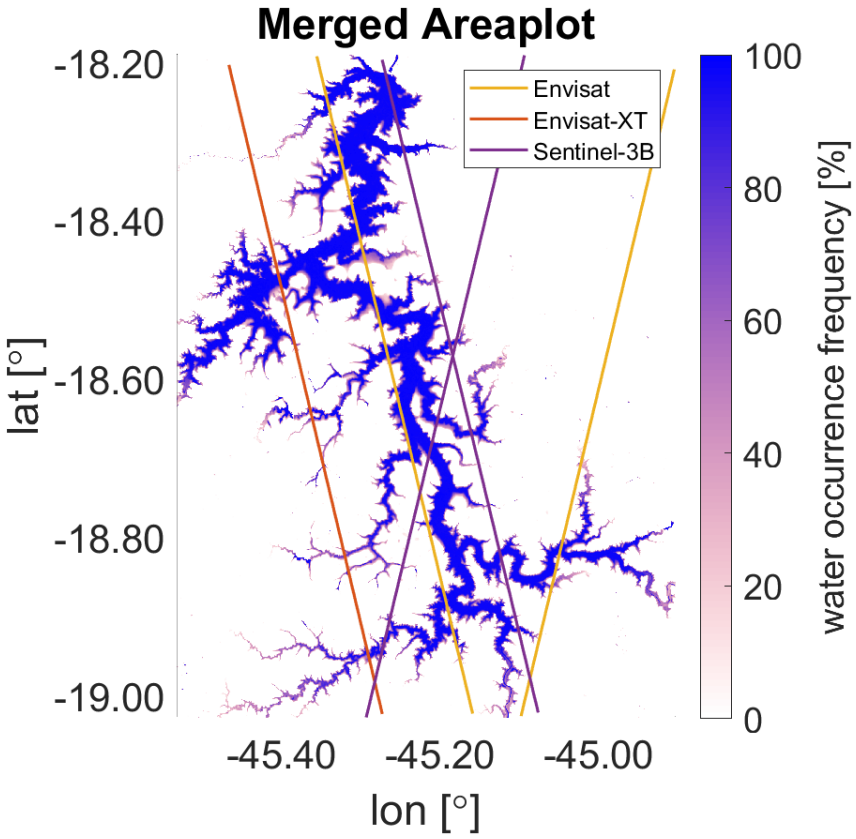


Figure 4.35: The water occurrance area plot of TresMarias Reservoirs

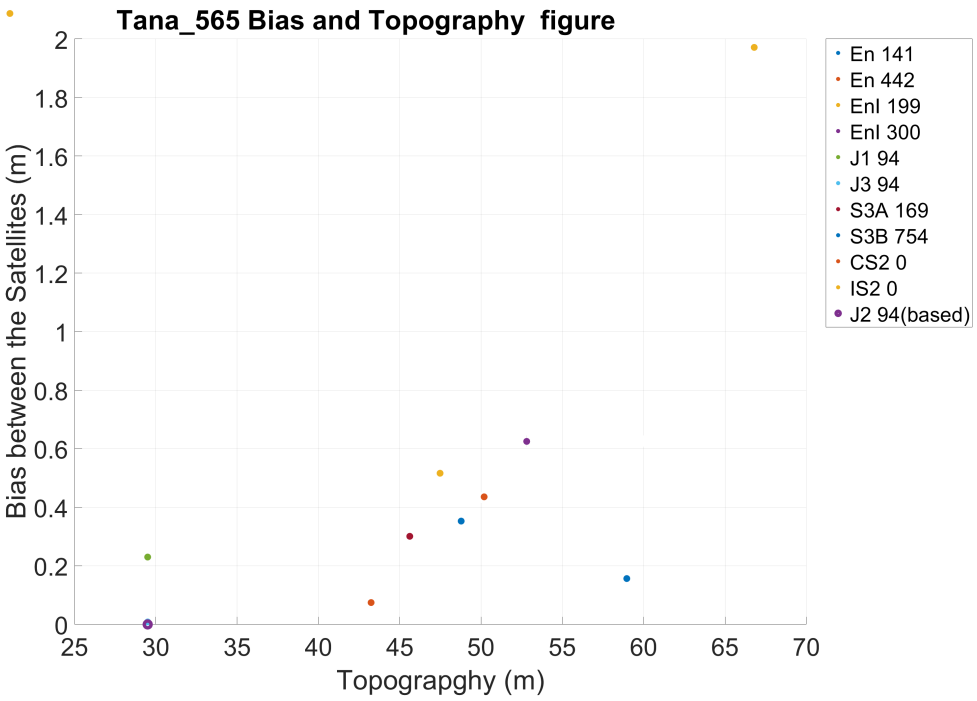


Figure 4.36: *The topography and bias figure of Tana Lake*

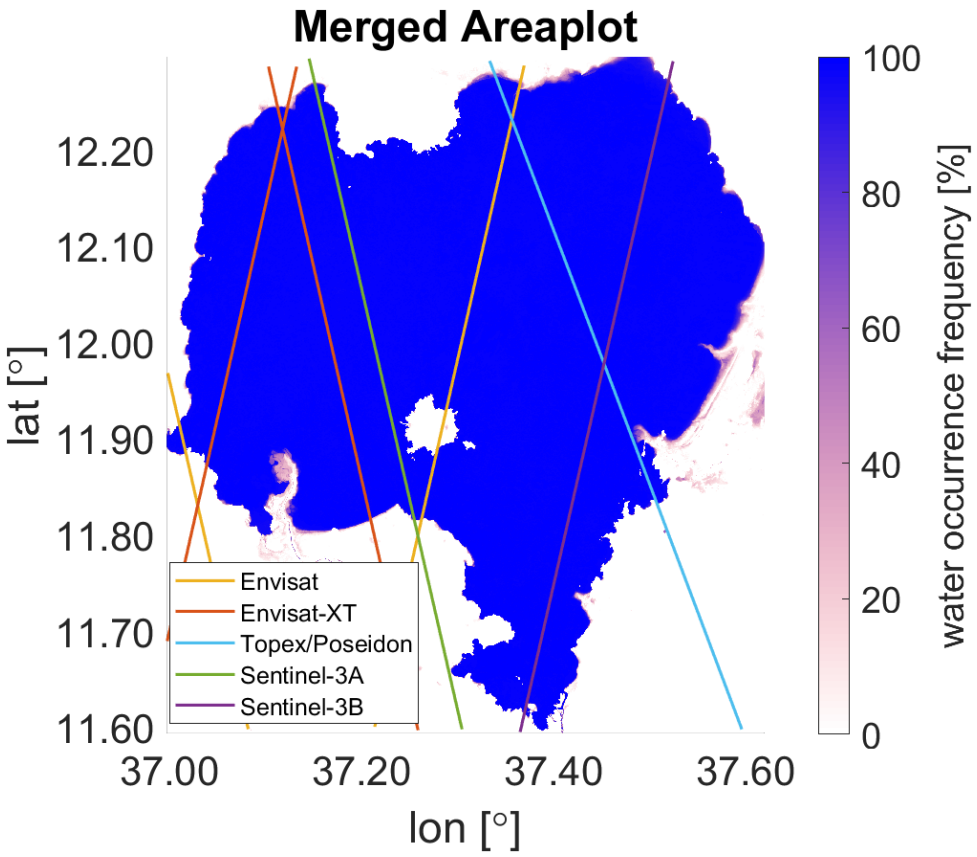


Figure 4.37: *The water occurrance area plot of Tana Lake*

4.2.3 Analysis of the influence of the topography

Both elevation variance and mean slop do not show significant correlation with inter-track biases and scatter plots are quiet noisy. The main reasons are as follows:

- The buffer for topography calculation is determined based on nominal orbits of missions except Saral/Altika and ICESat2, which differ from actual satellite orbit. Therefore, we can see from Figure 4.30 and Figure 4.34 that, although the observations come from the same track with same topography, relative biases differ significantly.
- As mentioned before, the shape of the water body and water occurrance also have impact on the bias estimation. From Figure 4.33, we can see that different parts of the river have different water occurance and river width, which will make the water level observation from different orbit inconsistent with each other in same period. However, comparing DEM area plot (Figure 4.23) and the water occurance area plot (Figure 4.33), we can see that this influence can not reflect from DEM data. Therefore, comparing with other case, the mean slope and bias figure of Itaparica reservoir (Figure 4.23) is noisier and more dispersed.

Ignoring the influence of the norminal orbit and observation from CryoSat-2, we can still find that those missions with topograph smaller than based missions show a negative correlation. However, those missions with topography larger than based missions show a positive correlation. The greater the elevation variance difference compared to the based track is, the greater the inter-track bias is.

4.2.4 Analysis water level biases between different satellites

The baises between different satellites have been shown in Figure 4.38 and Figure 4.39. Because the biases between CryoSat and other satellites are significantly larger, only biases based on the satellites except CryoSat have been plotted. Although, due to the limited number of study cases, we are unable to determine the fixed biases between various satellites using mean inter-satellite biases, which are easily influenced by special situations, we can still find that:

- As it shown from Table 4.14 to Table 4.19, CryoSat has bigger mean biases than other satellites, particularly with EnviSat, Saral\Altika, Jason 3, Sentinel 3A, Sentinel 3B and ICESat2. It is also notable that the mean biases of cryoSat, Envisat, Jason 3, and Sentinel 3B are similar whcih are approximately 20 m.
- For the remaining missions, mean baises are relatively small, ranging from 0.01 m to 4.19 m, and mostly distributed in the interval of 0.5 m to 1.5 m.
- Furthermore, it can be observed that the inter-satellite biases between some pairs of satellites have a relatively limited variance compared to others, indicating the existence of probable fixed inter-satellite biases. Taking biases relative to Jason 2 as an example, the variance of biases between Jason 2 and Envisat, Saral\Altika is relatively minor when compared to those of other individuals.

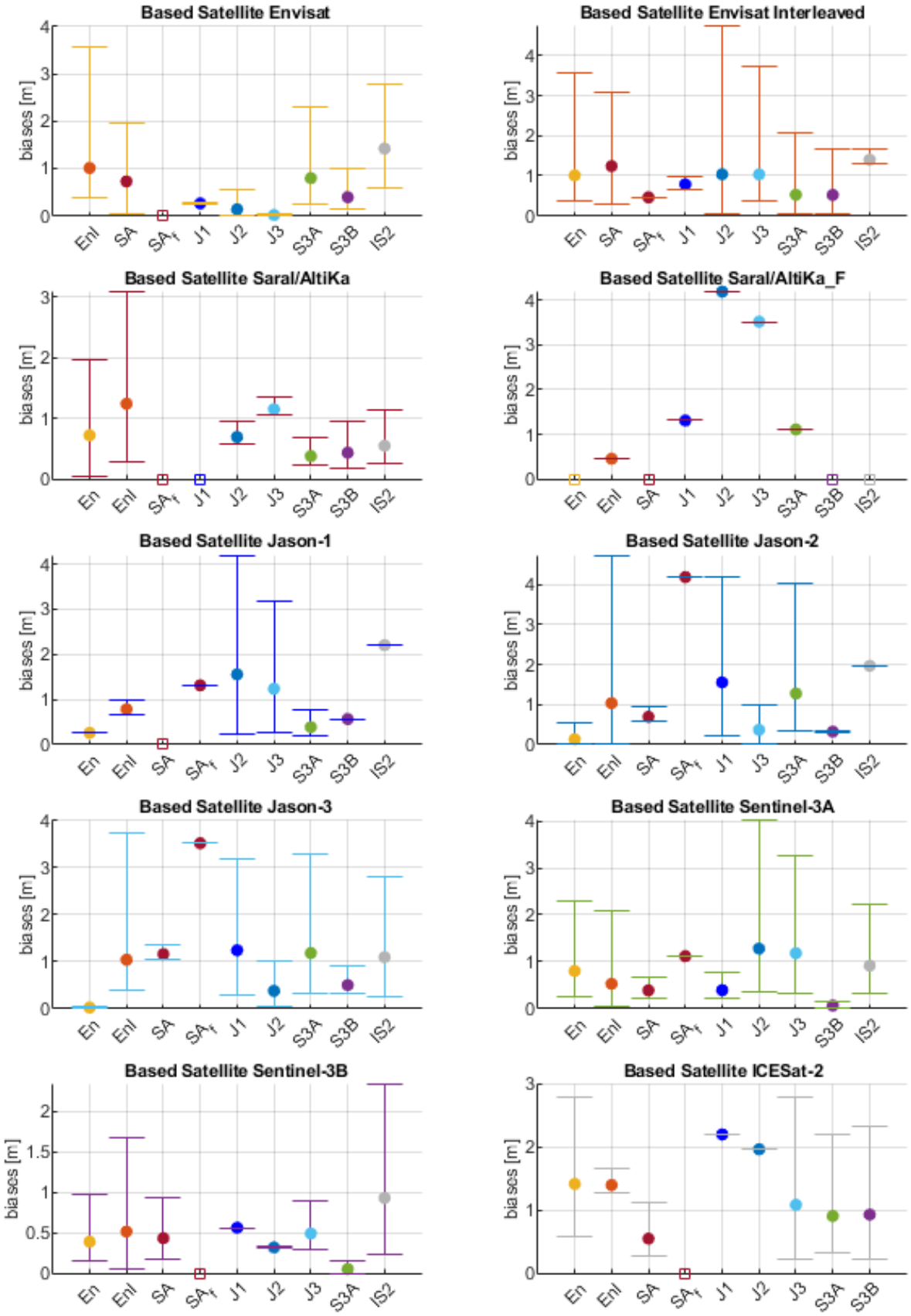


Figure 4.38: *The plot of the mean inter-mission biases*

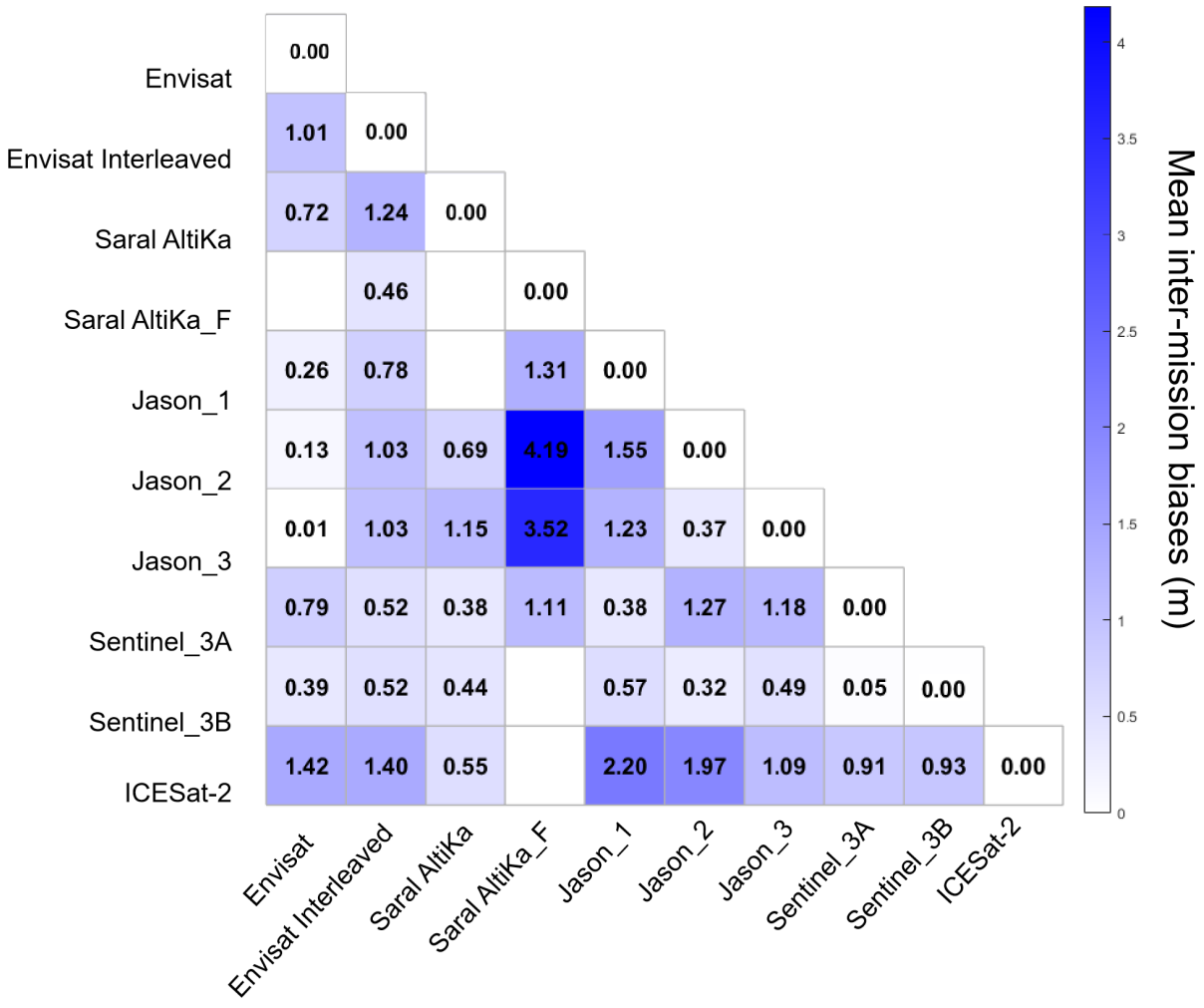


Figure 4.39: *The heatmap of the mean inter-mission biases*

Chapter 5

Conclusion and outlook

5.1 Summary and conclusion

In the last few decades, satellite altimetry has been extensively used to determine the surface level of inland water bodies. However, inter-satellite bias is still a problem for the multi-mission. In this thesis, bias is estimated based on the track rather than satellites. To estimate the biases between overlapping and non-overlapping water level time series, the time series of surface area derived from imagery is used to construct the area-height relationship, which allows the relative biases to be estimated.

The primary contributions of this thesis can be summed up as follows:

- We compared four fitting models and chose the power model to construct the relationship between the water extent and water level.
- Choose the Gauss-Helmet model for bias adjustment and address the source of the outliers.
- Due to the outliers existing in observations, the Expectation Maximization (EM) algorithm and local test are applied for robust estimation.
- Investigate the influence of the topography on the biases by calculating the DEM standard deviation and mean slope of each track.

We have applied our bias estimation model over 8 lakes and calculated the elevation difference and mean slope to investigate the influence of the topography. Finally, the validation shows that the proposed method can determine the inter-biases in diameter.

5.2 Outlook

Although inter-track biases may be calculated in diameter by establishing the link between water area and water height, the accuracy of the entire technique is highly dependent on the precision of water area observations. Despite the application of outlier rejection techniques such as the expectation maximization approach and the local test, the error of the water area cannot always be eliminated. Thus in future research, one might offer a novel strategy for detecting and eliminating outliers. In addition, the algorithm in this study does not account for overlapping time series, which are more trustworthy and make it easier to identify outliers. Thus, we may enhance the density and weight of the data in time periods that overlap, which

can improve the accuracy of the results. In the meanwhile, the method is inapplicable without area observation. Consequently, we anticipate a method that may estimate biases based solely on time series. In addition, the power model was applied to all examples in this study, which may not be appropriate for all lakes. Using the DEM and optical satellite photos, it is possible to establish a more exact link between the area and water level.

Bibliography

D. E. Alsdorf, E. Rodríguez, and D. P. Lettenmaier. Measuring surface water from space. *Reviews of Geophysics*, 45(2), 2007.

I. Z. Bakhtiyarovna, G. L. Akramovna, and M. R. Masharipovich. Current condition and usage of soils around kattakurgan reservoir. *Web of Scientist: International Scientific Research Journal*, 3(5):1927–1932, 2022.

A. S. Barau. The great attractions of kano. *Research and Documentation Directorate, Government House Kano*, 2007.

C. Birkett. The contribution of topex/poseidon to the global monitoring of climatically sensitive lakes. *Journal of Geophysical Research: Oceans*, 100(C12):25179–25204, 1995.

E. Boergens, S. Buhl, D. Dettmering, C. Klüppelberg, and F. Seitz. Combination of multi-mission altimetry data along the mekong river with spatio-temporal kriging. *Journal of Geodesy*, 91(5):519–534, 2017.

J.-F. Crétaux, A. Arsen, S. Calmant, A. Kouraev, V. Vuglinski, M. Bergé-Nguyen, M.-C. Gennero, F. Nino, R. A. Del Rio, A. Cazenave, et al. Sols: A lake database to monitor in the near real time water level and storage variations from remote sensing data. *Advances in space research*, 47(9):1497–1507, 2011.

A. Ettlinger and H. Neuner. Assessment of inner reliability in the Gauss-Helmert model. *Journal of Applied Geodesy*, 14(1):13–28, 2020. doi:doi:10.1515/jag-2019-0013. URL `https://doi.org/10.1515/jag-2019-0013`.

C. Förste, S. L. Bruinsma, F. Flechtner, J. Marty, J.-M. Lemoine, C. Dahle, O. Abrikosov, H. Neumayer, R. Biancale, F. Barthelmes, et al. A preliminary update of the direct approach goce processing and a new release of eigen-6c. In *AGU Fall meeting abstracts*, volume 2012, pages G31B–0923, 2012.

S. Ghosh, S. Bid, and C. Mondal. Impact of maithon dam-a micro level study on salanpur block of paschim bardhaman district, west bengal. *International Journal of Computational Engineering Research (IJCER)*, 8(08):51–64, 2018.

K. Golembesky, A. Sankarasubramanian, and N. Devineni. Improved drought management of falls lake reservoir: Role of multimodel streamflow forecasts in setting up restrictions. *Journal of Water Resources Planning and Management*, 135(3):188–197, 2009.

G. Gunkel, F. Selge, and M. Sobral. Re-oligotrophication of tropical water reservoirs to minimize environmental impact. *WIT Transactions on Ecology and the Environment*, 172:313–326, 05 2013. doi:10.2495/RBM130261.

D. M. Hannah, S. Demuth, H. A. van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl, L. M. Tallaksen, et al. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, 25(7):1191–1200, 2011.

J. R. Jensen and R. K. Raney. Delay/doppler radar altimeter: Better measurement precision. In *IGARSS'98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings.(Cat. No. 98CH36174)*, volume 4, pages 2011–2013. IEEE, 1998.

M. Kleinherenbrink, P. Ditmar, and R. Lindenbergh. Retracking cryosat data in the sarin mode and robust lake level extraction. *Remote sensing of environment*, 152:38–50, 2014.

K.-R. Koch. Robust estimations for the nonlinear gauss helmert model by the expectation maximization algorithm. *Journal of geodesy*, 88(3):263–271, 2014.

S. Le Gac, F. Boy, D. Blumstein, L. Lasson, and N. Picot. Benefits of the open-loop tracking command (oltc): Extending conventional nadir altimetry to inland waters monitoring. *Advances in Space Research*, 68(2):843–852, 2021.

P. Maillard, N. Bercher, and S. Calmant. New processing approaches on the retrieval of water levels in envisat and saral radar altimetry over rivers: A case study of the são francisco river, brazil. *Remote Sensing of Environment*, 156:226–241, 2015.

R. Pail, R. Bingham, C. Braitenberg, A. Eicker, M. Horwath, E. Ivins, L. Longuevergne, I. Panet, B. Wouters, G. Balsamo, et al. Observing mass transport to understand global change and and to benefit society: Science and user needs-an international multi-disciplinary initiative for iugg. 2015.

N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor. The development and evaluation of the earth gravitational model 2008 (egm2008). *Journal of geophysical research: solid earth*, 117(B4), 2012.

J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, 2016.

C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz. Dahiti–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, 19(10):4345–4364, 2015.

S. Shu, H. Liu, R. A. Beck, F. Frappart, J. Korhonen, M. Lan, M. Xu, B. Yang, and Y. Huang. Evaluation of historic and operational satellite radar altimetry missions for constructing consistent long-term lake water level records. *Hydrology and Earth System Sciences*, 25(3): 1643–1670, 2021.

P. F. Torres and P. C. Eterovick. Anuran assemblage composition and distribution at a modified environment in tres marias reservoir, south-eastern brazil. *Journal of Natural History*, 44 (43-44):2649–2667, 2010.

M. J. Tourian, O. Elmi, Y. Shafaghi, S. Behnia, P. Saemian, R. Schlesinger, and N. Sneeuw. Hydrosat: a repository of global water cycle products from spaceborne geodetic sensors. *Earth System Science Data Discussions*, pages 1–42, 2021.

J. Vijverberg, F. A. Sibbing, and E. Dejen. Lake tana: source of the blue nile. pages 163–192, 2009.

D. Wingham, C. Francis, S. Baker, C. Bouzinac, D. Brockley, R. Cullen, P. de Chateau-Thierry, S. Laxon, U. Mallow, C. Mavrocordatos, et al. Cryosat: A mission to determine the fluctuations in earth,Äôs land and marine ice fields. *Advances in Space Research*, 37(4):841–871, 2006.