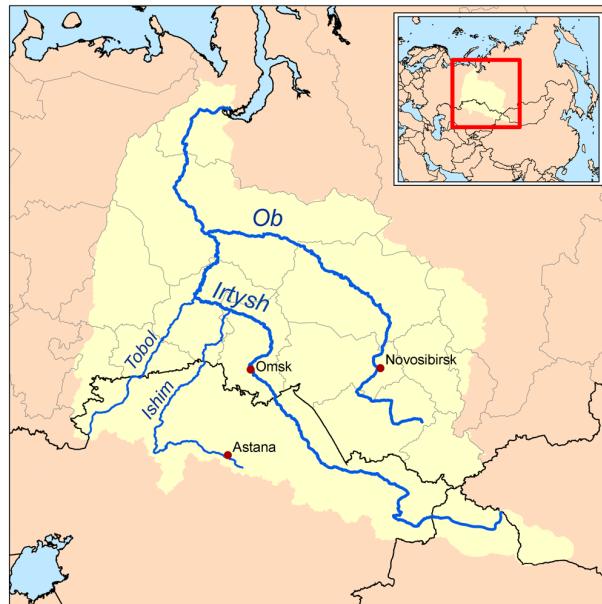




Understanding the positive trend in total water storage of Ob river basin using spaceborn observation



Bachelorarbeit im Studiengang
Geodäsie und Geoinformatik
an der Universität Stuttgart

Ziqing. Yu

Stuttgart, November 2020

Betreuer: Dr.-Ing. Mohammad Tourian
Universität Stuttgart

M.Sc. Peyman Saemian
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

In diesem kurzen Text erfolgt die Zusammenfassung der Arbeit oder – auf Englisch – das Abstract.

Contents

1	Introduction	1
1.1	Water Cycle	1
1.1.1	Precipitation	1
1.1.2	evapotranspiration	2
1.1.3	Runoff	3
1.2	Observation from satellite gravimetry	3
1.3	Motivation	5
1.4	Objective	6
2	Study Area	7
2.1	Physiography	8
2.2	Climate	8
2.3	Hydrology	9
2.4	Plants and animals	9
2.5	Human use	10
3	Data	11
3.1	GRACE and GRACE-FO	11
3.1.1	Spherical harmonics	11
3.1.2	Mass concentration	14
3.2	Precipitation	15
3.3	Evapotranspiration	17
3.4	Runoff	18
3.4.1	Global datasets	18
3.4.2	Water level height	20
4	Method	23
4.1	Estimating hydrological component combining different datasets	23
4.1.1	Adjustment using Gauss-Markov model	24
4.1.2	Validation of model-based estimations	25
4.1.3	Finding changing point	25
4.2	Estimating the quality of runoff datasets	26
4.3	Estimating the runoff using quantile function and water level height	26
5	Result	29
5.1	Equivalent water height	29
5.2	Precipitation	32
5.3	Evapotranspiration	33
5.4	Runoff	34
5.4.1	Runoff from global datasets	34
5.4.2	Estimating runoff using quantile function and satellite altimetry	35

5.5	Discussion about the quality	38
6	Conclusion and outlook	39
6.1	Conclusion	39
6.2	Outlook	39
A	Anhang	XV

List of Figures

1.1	hydrologic cycle	2
1.2	GRACE Component	4
1.3	GRACE-FO twin satellites	5
1.4	Trends in TWS (in centimetres per year) obtained on the basis of GRACE observations from April 2002 to March 2016. The cause of the trend in each outlined study region is briefly explained and colour-coded by category. The trend map was smoothed with a 150 km radius Gaussian filter for the purpose of visualization; however, all calculations were performed at the native 3° resolution of the data product.	6
1.5	Total water storage change from Apr.2002 to May.2020 in Ob river basin	6
2.1	River Basins Ob http://www.geologypage.com/2014/03/ob-river.html	7
2.2	The distribution of 40,962 geodesic grid tiles over the Earth used as a basis function for estimation of mass anomalies from GRACE for CSR mascon solutions. (top) Global view, (bottom left) South Pole view, and (bottom right) North Pole view [13]	9
3.1	The distribution of 40,962 geodesic grid tiles over the Earth used as a basis function for estimation of mass anomalies from GRACE for CSR mascon solutions. (top) Global view, (bottom left) South Pole view, and (bottom right) North Pole view [13]	15
3.2	Station Salekhard	19
4.1	EstimatEd water level from Envisat and available runoff for Ob river (top) and quantile function of water level and runoff (middle). A smoothed rating curve is obtained from the corresponding probabilities (bottom)	28
5.1	Total water storage anomaly from CSR, GFZ, ITSG and JPL between Apr.2002 and May.2020	29
5.2	Total water storage anomaly between Apr.2002 and May.2020	30
5.3	dS/dt between 2003 and 2020	30
5.4	Sudden change of dS/dt between 2003 and 2020 after using movmean	30
5.5	The mean value of dS/dt in three periods	31
5.6	Precipitation datasets	32
5.7	precipitation with uncertainties between 2003 and 2020 and the mean value in 3 periods	32
5.8	evapotranspiration datasets	33
5.9	evapotranspiration with uncertainties between 2003 and 2020 and the mean value in 3 periods	33
5.10	Runoff datasets	34
5.11	CDF of differences	35

5.12 water level time series and the location of virtual station	36
5.13 runoff timeseries calculated from water level (up) and the mean value in 3 periods(bottom)	37

List of Tables

3.1	precipitation datasets	15
3.2	evapotranspiration datasets	17
3.3	runoff datasets	19
5.1	RMSE for runoff datacenters	34
5.2	mean value of all water cycle component in 3 periods	38

Chapter 1

Introduction

1.1 Water Cycle

Water is one of the most important ingredient of life; it regulates climate and many industries need water as coolant, solvent, rea material, etc. Water covers about 71% of the earth's surface. However, 96.5 % of the water on the earth is in the oceans, 1.7% is in ice caps, which means only 1.4 % is the fresh water on land [5]. The water on earth is variable depending on a wide range of climatic variables. A very little variability of the hydrology cycle can have big effects on water resources. [3]

The hydrology cycle (Figure 1.1) includes 3 major compartments: evaporation, precipitation and runoff. The water evaporates from the oceans and the land surface as vapor to become part of the atmosphere along with water from evapotranspiration

transpiration, which is water transpired from plants and evaporated from the soil and the cooler temperature causes the vapor into clouds. The clouds fall out of the sky as precipitation, which includes rain, snow and ice. Most precipitation falls back into the oceans or onto land. Precipitated water may be intercepted by vegetation, become overland flow over the ground surface, flow through the soil as subsurface flow and discharge into streams as surface runoff. The process can be simplified as:

$$Pre - ET - R = \frac{dS}{dt} \quad (1.1)$$

where

Pre	Precipitation
ET	Evapotranspiration
R	Surface Runoff
dS/dt	total water storage change

1.1.1 Precipitation

Precipitation is any form of water particle, solid or liquid, that falls from the atmosphere and reaches the ground. Precipitation can include drizzle, rain, snow, sleet, and hail. Precipitation forms in the clouds when water vapor condenses into bigger and bigger droplets of water. When the drops are heavy enough, they fall to the Earth. If a cloud is colder, like it would be

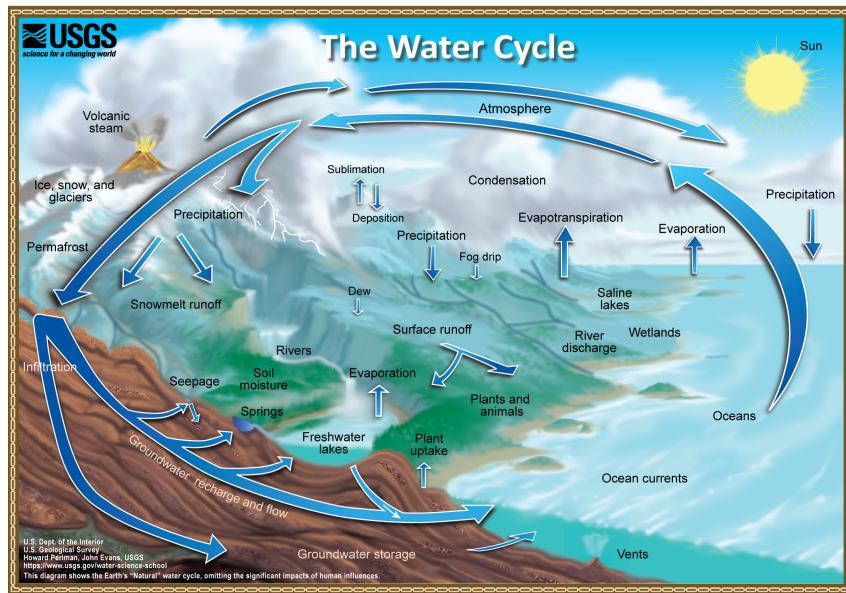


Figure 1.1: hydrologic cycle

at higher altitudes, the water droplets may freeze to form ice. These ice crystals then fall to the Earth as snow, hail, or rain, depending on the temperature within the cloud and at the Earth's surface. Most rain actually begins as snow high in the clouds. As the snowflakes fall through warmer air, they become raindrops.

Precipitation is responsible for depositing the fresh water on the planet. Approximately 505 000 cubic kilometres of water falls as precipitation each year; 398 000 cubic kilometres of it over the oceans and 107 000 cubic kilometres over land.[4] Given the Earth's surface area, that means the globally averaged annual precipitation is 990 millimetres , but over land it is only 715 millimetres

1.1.2 evapotranspiration

Evaporation is the process whereby liquid water is converted to water vapour (vaporization) and removed from the evaporating surface (vapour removal). Water evaporates from a variety of surfaces, such as lakes, rivers, pavements, soils and wet vegetation.

Transpiration consists of the vaporization of liquid water contained in plant tissues and the vapour removal to the atmosphere. Crops predominately lose their water through stomata. These are small openings on the plant leaf through which gases and water vapour pass. The water, together with some nutrients, is taken up by the roots and transported through the plant. The vaporization occurs within the leaf, namely in the intercellular spaces, and the vapour exchange with the atmosphere is controlled by the stomatal aperture. Nearly all water taken up is lost by transpiration and only a tiny fraction is used within the plant.

Evaporation and transpiration occur simultaneously and there is no easy way of distinguishing between the two processes. Apart from the water availability in the topsoil, the

evaporation from a cropped soil is mainly determined by the fraction of the solar radiation reaching the soil surface. This fraction decreases over the growing period as the crop develops and the crop canopy shades more and more of the ground area. When the crop is small, water is predominately lost by soil evaporation, but once the crop is well developed and completely covers the soil, transpiration becomes the main process. In Figure 2 the partitioning of evapotranspiration into evaporation and transpiration is plotted in correspondence to leaf area per unit surface of soil below it. At sowing nearly 100% of ET comes from evaporation, while at full crop cover more than 90% of ET comes from transpiration.[1]

1.1.3 Runoff

Runoff is quantity of water discharged in surface streams. Runoff includes not only the waters that travel over the land surface and through channels to reach a stream but also interflow, the water that infiltrates the soil surface and travels by means of gravity toward a stream channel (always above the main groundwater level) and eventually empties into the channel. Runoff also includes groundwater that is discharged into a stream; streamflow that is composed entirely of groundwater is termed base flow, or fair-weather runoff, and it occurs where a stream channel intersects the water table.

1.2 Observation from satellite gravimetry

It was extremely difficult to measure the global water storage change globally: the precipitation is neither constant in time nor in space. Observing phenomena like groundwater discharge is difficult due to the manifold number of possible flow paths. The total amount of water remains though constant, the contribution to the different subsystem changes over time and depends on the concerning regional and local characteristics. In some way, remote sensing with satellite is the perfect tool for hydrology research, which has the ability to provide the data globally in a long term.

The GRACE mission is a joint partnership between the National Aeronautics and Space Administration (NASA) in the United States, the Deutsche Forshungsanstalt fuer Luft und Raumfahrt (DLR) in Germany. The Grace Satellites launched on 17 March 2002, are making detailed measurements of Earth's gravity field, which are caused by monthly changes in mass. The mass changes can be thought of as concentrated in a very thin layer of water thickness changes near the Earth's surface by moving ocean, atmospheric and land ice masses and by mass exchanges between these Earth system compartments.

The two identical satellites orbit one behind the other in the same orbital plane at an approximate distance of 220 km (137 miles). As the pair circles the Earth, areas of slightly stronger gravity (greater mass concentration) will affect the lead satellite first, pulling it away from the trailing satellite, then as the satellites continue along their orbital path, the trailing satellite is pulled toward the lead satellite as it passes over the gravity anomaly. The change in distance would certainly be imperceptible to our eyes, but an extremely precise microwave ranging system on GRACE is able to detect these minuscule changes in the distance between the satellites. A highly accurate measuring device known as an accelerometer, located at each

satellite mass center, will be used to measure the non-gravitational accelerations (such as those due to atmospheric drag) so that only accelerations caused by gravity are considered. Satellite Global Positioning System (GPS) receivers will be used to determine the exact position of the satellite over the Earth to within a centimeter or less. Members of the GRACE science team can download all this information from the satellites, and use it to construct monthly maps of the Earth's average gravity field.

The component parts of GRACE: (Figure 1.2) [10]

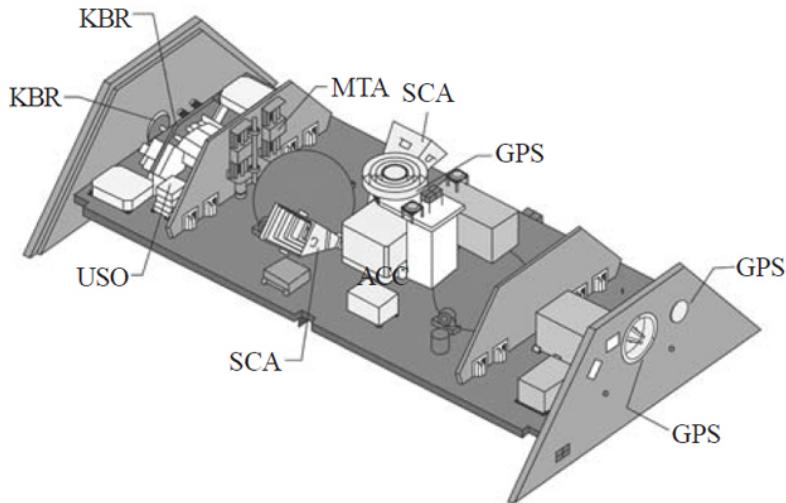


Figure 1.2: GRACE Component

- K-band Ranging System (KBR): Provides precise (within 10 micrometre) measurements of the distance change between the two satellites needed to measure fluctuations in gravity.
- Ultra Stable Oscillator (USO): Provides frequency generation for the K-band ranging system.
- SuperSTAR Accelerometers (ACC): Precisely measures the non-gravitational accelerations acting on the satellites.
- Star Camera Assembly (SCA): Precisely determines the two satellite's orientation by tracking them relative to the position of the stars.
- Coarse Earth Sun and Sensor (CES): Provides omnidirectional, reliable, and robust, but fairly coarse, Earth and Sun tracking. Used during initial acquisition and whenever GRACE operates in safe mode.
- Center of Mass Trim Assembly (MTA): Precisely measures the offset between the satellite's center of mass and the "acceleration-proof" mass and adjusts center of mass as needed during the flight.
- BlackJack GPS Receiver and Instrument Processing Unit (GPS): Provides digital signal processing; measures the distance change relative to the GPS satellite constellation.
- Globalstar Silicon Solar Cell Arrays (GSA): Covers the outer shell of the spacecraft and generates power.

It is shown, that GRACE delivers the highest temporal resolution and is thus able to observe monthly mass variation with a spatial resolution of less than 1000 km. In [15] it was predicted that GRACE would be able to measure these effects with an accuracy of about 2 mm of water equivalent heights. Though this accuracy has not yet been achieved because of the errors in spherical harmonic coefficients of short-wavelength, it was shown in many publications that the Stokes coefficients from GRACE indeed contain hydrological signals as the monthly solutions from GRACE showed a good agreement with mass variations from hydrological models.

The GRACE mission ended in the October 2017. GRACE-FO (Figure 1.3), which launched May 22, 2018, will continue the work of tracking Earth's water movement to monitor changes in underground water storage, the amount of water in large lakes and rivers, soil moisture, ice sheets and glaciers, and sea level caused by the addition of water to the ocean. This leads to a data gap in time series.

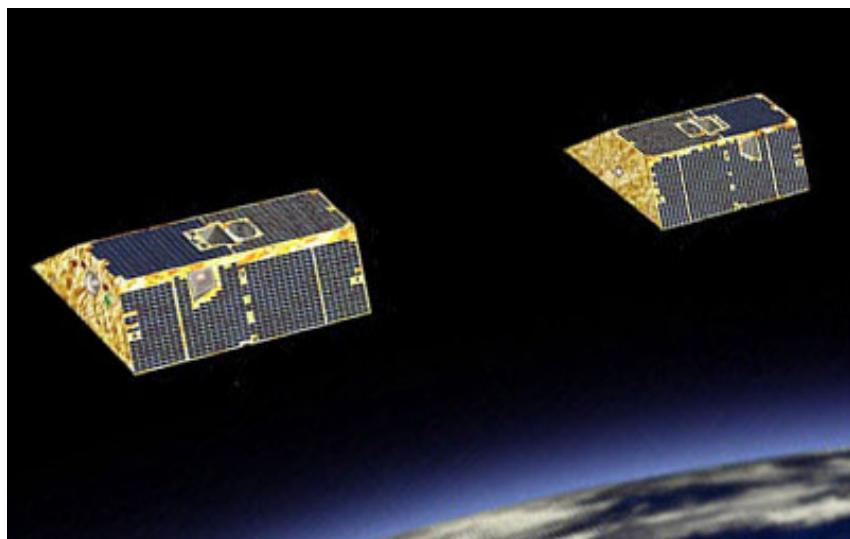


Figure 1.3: GRACE-FO twin satellites

1.3 Motivation

In hydrology, most variables are observed in time series, including Total Water Storage Anomaly (TWSA). In the hydrological cycle, this should reflect seasonal behavior and is in long term relatively stable. However, it was shown that since 2002 the TWSA of many big basins has increased (Figure 1.4 [12]). One important basin of them is Ob basin in west Siberia (Figure 1.5). How did this trend happen is a very interesting topic. Through the analysis of the trend of the time series, it is possible to further understand the changes that have taken place before and future changes can also be predicted based on the stationary analysis.

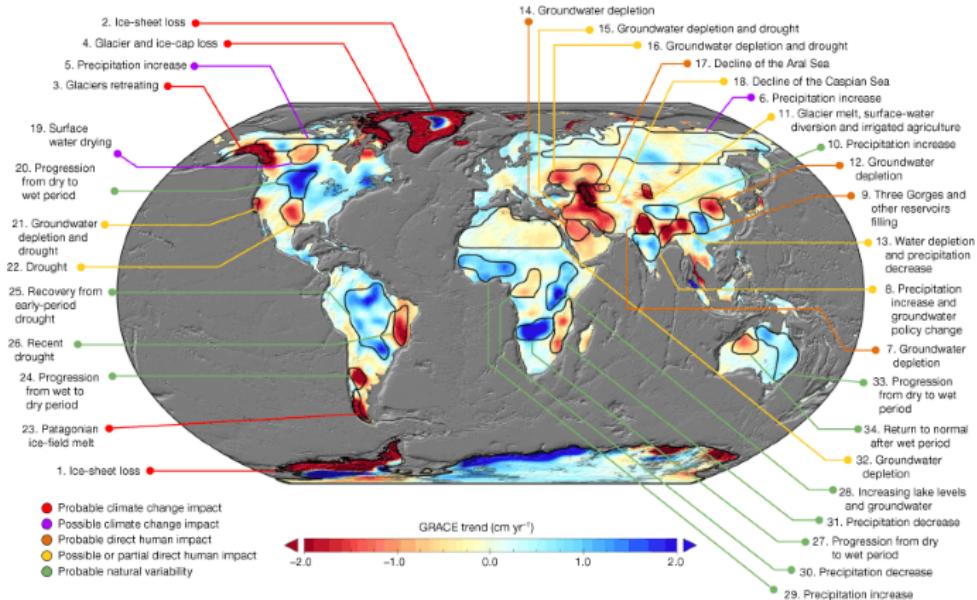


Figure 1.4: Trends in TWS (in centimetres per year) obtained on the basis of GRACE observations from April 2002 to March 2016. The cause of the trend in each outlined study region is briefly explained and colour-coded by category. The trend map was smoothed with a 150 km radius Gaussian filter for the purpose of visualization; however, all calculations were performed at the native 3° resolution of the data product.

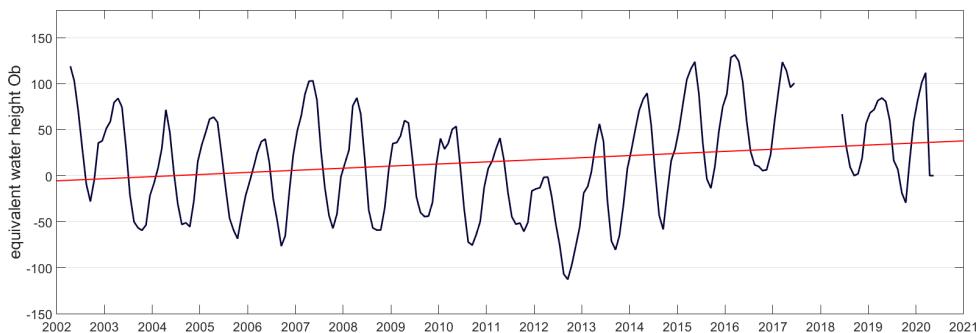


Figure 1.5: Total water storage change from Apr.2002 to May.2020 in Ob river basin

1.4 Objective

In this work this positive trend in total water storage of Ob river basin will be discussed. The time series of total water storage change will be generated with data from GRACE and GRACE-FO mission. Then, the changing points of the trend will be determined and thus the time interval of this change can be found. At last, by analyzing the precipitation, evapotranspiration and runoff from different data centers in this period the reason for this change can hopefully be discovered.

Chapter 2

Study Area

Ob River (Figure 2.1), river of central Russia. One of the greatest rivers of Asia, the Ob flows north and west across western Siberia in a twisting diagonal from its sources in the Altai Mountains to its outlet through the Gulf of Ob into the Kara Sea of the Arctic Ocean. It is a major transportation artery, crossing territory at the heart of Russia that is extraordinarily varied in its physical environment and population. Even allowing for the barrenness of much of the region surrounding the lower course of the river and the ice-clogged waters into which it discharges, the Ob drains a region of great economic potential.

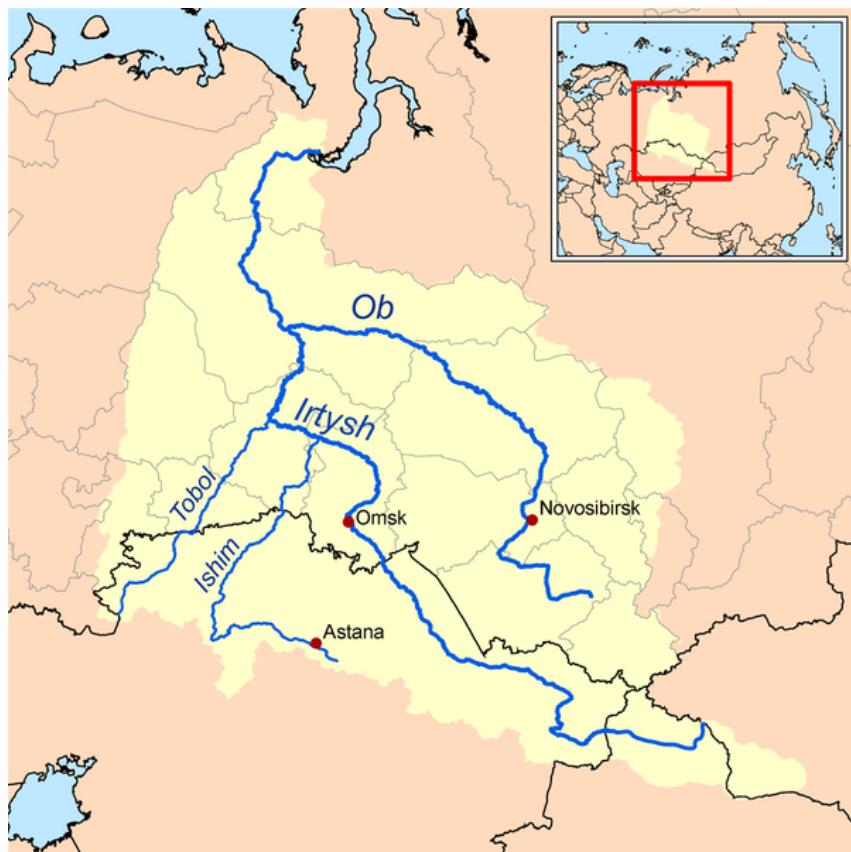


Figure 2.1: River Basins Ob <http://www.geologypage.com/2014/03/ob-river.html>

2.1 Physiography

The Ob proper is formed by the junction of the Biya and Katun rivers, in the foothills of the Siberian sector of the Altai, from which it has a course of 3 650 km. If, however, the Irtysh River is regarded as part of the main course rather than as the Ob's major tributary, the maximum length, from the source of the Black (Chorny) Irtysh in China's sector of the Altai, is 5 410 km, making the Ob the seventh longest river in the world. The catchment area is approximately 2 975 000 square km. Constituting about half of the drainage basin of the Kara Sea, the Ob's catchment area is the sixth largest in the world. The drainage basin is classified as cropland (36%), forest (30%), wetland (11%), grassland (10%), shrub (5%) , developed (5%) and irrigated cropland (3%).[11]

The West Siberian Plain covers about 85 percent of the Ob basin.[9] The rest of the basin comprises the terraced plains of Turgay (Kazakhstan) and the small hills of northernmost Kazakhstan in the south and the Kuznetsk Alatau range, the Salair Ridge, the Altai Mountains and their foothills and outliers in the southeast.

The huge basin of the Ob stretches across a number of natural zones. Semidesert prevails in the far south around Lake Zaysan (recipient of the Black Irtysh and source of the Irtysh proper), bordered on the north by steppe grassland. The central regions of the West Siberian Plain i.e., more than half of the basin-consist of taiga (swampy coniferous forest), with great expanses of marshland. In the north there are vast stretches of tundra (low-lying, cold-tolerant vegetation).

2.2 Climate

The Ob basin has short, warm summers and long, cold winters. Average January temperatures range from -28°C on the shores of the Kara Sea to -16°C in the upper reaches of the Irtysh. July temperatures for the same locations, respectively, range from 4°C to above 20°C . The absolute maximum temperature, in the arid south, is 40°C ,[9] and the minimum, in the Altai Mountains, is -60° . Rainfall, which occurs mainly in the summer, averages less than 400 mm per year in the north, 500 to 600 mm in the taiga zone, and 300 to 400 mm on the steppes. The western slopes of the Altai receive as much as 1 575 mm per year. Snow cover lasts for 240 to 270 days in the north and for 160 to 170 days in the south. It is deepest in the forest zone, where it ranges from 60 to 90 cm, and in the mountains, where it averages 200 cm per year. It is much shallower on the tundra, ranging from 30 to 50 cm, and very thin on the steppe, where 20 to 40 cm fall.[9]

On the upper Ob the spring floods begin early in April, when the snow on the plains is melting; and they have a second phase, ensuing from the melting of snow on the Altai Mountains. The middle Ob, scarcely affected by the upper Ob's phases, has one continuous spring-summer period of high water, which begins in mid April. For the lower Ob, high water begins in late April or early May. Levels, in fact, begin to rise when the watercourse is still obstructed by ice; and maximum levels, which occur by May on the upper Ob, may not be reached until June, July, or even August on the lower reaches. For the upper Ob, the spring floods end by July, but autumn rains bring high water again in September and October; in the

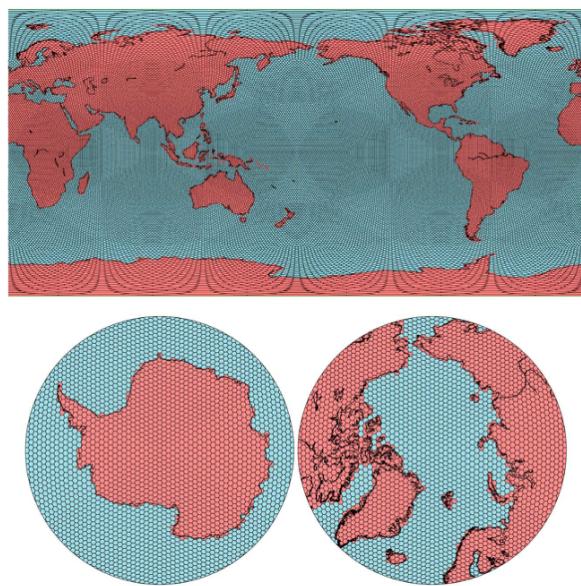


Figure 2.2: The distribution of 40,962 geodesic grid tiles over the Earth used as a basis function for estimation of mass anomalies from GRACE for CSR mascon solutions. (top) Global view, (bottom left) South Pole view, and (bottom right) North Pole view [13]

middle and lower Ob, the spring and summer floodwaters gradually recede until freezing sets in. On the lower reaches, flooding may last four months. Flooding of the Ob proper and of the Irtysh obstructs the minor tributaries' drainage.

2.3 Hydrology

The Ob has the third greatest discharge of Siberia's rivers, after the Yenisey and the Lena. On average, it pours some 400 cubic km of water annually into the Arctic Ocean about 12% of that ocean's total intake from drainage.

The volume of flow at Salekhard, just above the delta, is about 42 000 cubic metres per second at its maximum and 2 000 cubic metres per second at its minimum, while for Barnaul, on the upper Ob, the corresponding figures are 9 600 and 200 cubic metres per second. The average annual discharge rate at the river's mouth is about 12 700 cubic metres per second. Most of the water comes from the melting of seasonal snow and from rainfall; much less of it comes from groundwater, mountain snow, and glaciers.[9]

2.4 Plants and animals

Pine, cedar, silver fir, aspen, and birch grow on the banks and occasionally constitute isolated forests on the higher ground of the floodplain. Large areas near the river are covered with willow, snowball trees, bird cherry, buckthorn, currant bushes, and wild roses.

Fur-bearing mammals of the Ob valley include European and Siberian mole, Siberian and American mink, ermine, fox, wolf (in the taiga), elk, white hare, water rat, muskrat, otter, and beaver. Among more than 170 species of birds breeding in the floodplain are grouse, partridge, goose, and duck.

2.5 Human use

Basin total population is about 27 million, with 39 cities having a population of more than 100 000. The Ob's immense hydroelectric potential is estimated at some 250 billion kilowatts. Three main stations have been built: one on the Ob proper, at Novosibirsk, and the other two on the mountainous reaches of the Irtysh, at Bukhtarma and Oeskemen. Both industry and agriculture have been intensively developed in the Ob basin. Cities such as Omsk, Novosibirsk, and Barnaul are major industrial and manufacturing centres. The steppe zone, in the southern Ob basin, is the major producer of spring wheat in Russia. The west Siberian oil and gas fields, located in the taiga and tundra zones of the middle and lower Ob, are the most important in Russia, contributing about two-thirds of the country's crude oil and natural gas output.

Chapter 3

Data

3.1 GRACE and GRACE-FO

3.1.1 Spherical harmonics

The variations in the gravity field impacts on both GRACE satellites at different times, such deviations cause a change in the inter-satellite range, which is measured with very high accuracy from the K-band measurement unit. The measured inter-satellite range can be transformed into changes in the Earth's gravity field, which is described with the Stokes coefficients \tilde{C}_{lm} and \tilde{S}_{lm} .

There is only one Earth gravity field and all centers start off with identical GRACE Level-1 observations, but deriving month-to-month gravity field variations from GRACE observations requires a complex inversion of relative ranging observations between the two formation-flying GRACE spacecraft, in combination with precise orbit determination via GPS and various corrections for spacecraft accelerations not related to gravity changes. Many parameter choices and solution strategies are possible, and have been explored by different data centers. In this thesis the solutions from JPL, CSR, GFZ and ITSG are used, which allows the calculation of gravity field and equivalent water height anomaly.

The shape of the geoid, i.e. the distance between the reference ellipsoid and the geoid surface N , can be expanded in a sum of spherical harmonics.

$$N(\theta, \lambda) = R \sum_{l=0}^{\infty} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) (\tilde{C}_{lm} \cos m\lambda + \tilde{S}_{lm} \sin m\lambda) \quad (3.1)$$

where The time-dependent change in the geoid heights ΔN is reflected by the difference be-

$N(\theta, \lambda)$ geoid height at a point with the spherical coordinates θ, λ

R radius of the Earth

\tilde{P}_{lm} normalized associated Legendre functions of degree l and order m

$\tilde{C}_{lm}, \tilde{S}_{lm}$ normalized Stokes coefficients

tween the spherical harmonic coefficients $\tilde{C}_{lm}, \tilde{S}_{lm}$. In this case, the equation can be written as:

$$\Delta N(\theta, \lambda) = R \sum_{l=0}^{\infty} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) (\Delta \tilde{C}_{lm} \cos m\lambda + \Delta \tilde{S}_{lm} \sin m\lambda) \quad (3.2)$$

By assuming that $\Delta N(\theta, \lambda; t) \neq 0$, it is clear that there had to be a change in the Earth's gravity field caused by mass fluctuations in, on and above the Earth's surface. This change is denoted as a change in the Earth's density distribution. In[15], it was found that there is a connection between this quantity and its representation in spherical harmonic coefficients.

$$\begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix} = \frac{3}{4\pi R \rho_{ave} (2l+1)} \int \int \Delta \rho(r, \theta, \lambda; t) \tilde{P}_{lm}(\cos \theta) \times \left(\frac{r}{R} \right)^{l+2} \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \sin \theta d\theta d\lambda \quad (3.3)$$

where r is the distance of the computation point from the center of the Earth and ρ_{ave} is the average density of the Earth. However, an accurate determination of $\Delta \rho(r, \theta, \lambda; t)$ is nearly impossible, because it requires prior knowledge about the inner density distribution of the Earth. But all short periodic mass variations can be assumed to happen only in a thin layer on the Earth's surface, which can be detected by GRACE satellites. The thickness is mostly determined by the thickness of the atmosphere and is of the order of 10 to 15 km [15].

The change in this thick layer is called surface density $\Delta \rho_S$, which can be defined as the radial integral of $\Delta \rho$ through this layer and since the layer is thick enough, it can be assumed that $r \approx R$, so the equation can be simplified as

$$\begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix}_{\text{surf mass}} = \frac{3}{4\pi R \rho_{ave} (2l+1)} \int \int \Delta \rho(\theta, \lambda; t) \tilde{P}_{lm}(\cos \theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \sin \theta d\theta d\lambda \quad (3.4)$$

This equation now connects the density redistribution in this thin layer with the spherical harmonic coefficients. Thus, it describes the contribution to the geoid from the direct gravitational attraction of the surface mass [15]. The mass fluctuations on the surface also deform the underlying Earth, which implicates a change in the gravitational potential, and thus a change in the geoid shape, as well. This effect is considered by the so called *Love number* k_l , which were derived from[6]. The contribution from the deformed solid earth may then be written as

$$\begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix}_{\text{solid Earth}} = \frac{3k_l}{4\pi R \rho_{ave} (2l+1)} \int \int \Delta \rho(\theta, \lambda; t) \tilde{P}_{lm}(\cos \theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \sin \theta d\theta d\lambda \quad (3.5)$$

The total geoid change is obtained by adding (3.4) and (3.5)

$$\begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix} = \begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix}_{\text{surf Earth}} + \begin{Bmatrix} \Delta \tilde{C}_{lm}(t) \\ \Delta \tilde{S}_{lm}(t) \end{Bmatrix}_{\text{solid Earth}} \quad (3.6)$$

Inserting (3.6) into (3.2) leads to the so called *isotropic transfer coefficients*, which define the quantity of a spherical harmonic series expansion. In the case of a surface mass density, they are defined as

$$\Lambda_l = \frac{R \rho_{ave}}{3} \frac{2l+1}{1+k_l} \quad (3.7)$$

Then an expression for the surface mass density in terms of the spherical harmonic coefficients can be written as

$$\Delta \rho_S(\theta, \lambda) = \frac{R \rho_{ave}}{3} \sum_{l=0}^{\infty} \frac{2l+1}{1+k_l} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) (\Delta \tilde{C} \cos m\lambda + \Delta \tilde{S} \sin m\lambda) \quad (3.8)$$

The gravity field change can be assumed as the change of the thin layer of water on the Earth's surface. The relation between the water equivalent heights and the surface mass density is

$$h_W(\theta, \lambda) = \frac{\Delta \rho_S(\theta, \lambda)}{\rho_W} \quad (3.9)$$

where ρ_W is the average density of water and thus

$$h_W(\theta, \lambda; t) = \frac{R\rho_{ave}}{3\rho_W} \sum_{l=0}^{\infty} \frac{2l+1}{1+k_l} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) (\Delta \tilde{C} \cos m\lambda + \Delta \tilde{S} \sin m\lambda) \quad (3.10)$$

For simplicity, this formula can be written as

$$h_W(\theta, \lambda; t) = \sum_{l=0}^{\infty} \Lambda_l \sum_{m=0}^l \tilde{Y}_{lm}(\theta, \lambda) \Delta \tilde{K}_{lm}(t) \quad (3.11)$$

where

- $\Lambda_l = \frac{R\rho_{ave}}{3\rho_W} \frac{2l+1}{1+k_l}$: isotropic spectral transfer coefficients
- $\tilde{Y}_{lm}(\theta, \lambda) = \tilde{P}_{lm}(\cos \theta) (\cos m\lambda \quad \sin m\lambda)^T$: normalized surface spherical harmonics
- $\Delta \tilde{K}_{lm}(t) = (\Delta \tilde{C}_{lm} \quad \Delta \tilde{S}_{lm})^T$: normalized Stokes coefficients

The associated Legendre functions are given by

$$\tilde{P}_{n,m}(t) = \sqrt{(2 - \delta_{m0})(2n+1) \frac{(n-m)!}{(n+m)!}} \sqrt{1-t^2}^m \frac{d^{n+m}}{dt^{n+m}} \frac{1}{2^n n!} (t^2 - 1)^n \quad (3.12)$$

where n is degree, m is order and $t = \cos \theta$ is a substitution. The Legendre functions can be calculated by the recursion.

$$\tilde{P}_{0,0}(t) = 1 \quad (3.13)$$

$$\tilde{P}_{m,m}(t) = W_{m,m} \sin \theta \tilde{P}_{m-1,n-1}(t-1) \quad \text{for } m > 0 \text{ and } m = n \quad (3.14)$$

$$\tilde{P}_{n,m} = W_{n,m} [\cos \theta \tilde{P}_{n-1,m}(t) - \frac{1}{W_{n-1,m}} \tilde{P}_{n-2,m}(t)] \quad \text{for } m \neq n \quad (3.15)$$

with the factors

$$W_{n,m} = \begin{cases} \sqrt{3} & \text{for } n = 1 \text{ and } m = 0, 1 \\ \sqrt{\frac{2n+1}{2n}} & \text{if } n = m \text{ and } n > 1 \\ \sqrt{\frac{(2n+1)(2n-1)}{(n+m)(n-m)}} & n > 1 \text{ and } m \neq n \end{cases} \quad (3.16)$$

and the convention $\tilde{P}_{n,m}(t) = 0$ for $m > n$. This algorithm is shown to be stable until degree $n \approx 1800$. In this thesis they are up to 96.

It is obvious that only $\Delta \tilde{K}_{lm}$ is time dependent while Λ_l and $\tilde{Y}_{lm}(\theta, \lambda)$ are constant in time, by using the methods of forwards and backward-differences a rate of mass variations in terms of water equivalent heights can be obtained.

$$\dot{h}_W(\theta, \lambda; t) = \sum_{l=0}^{\infty} \Lambda_l \sum_{m=0}^l \tilde{Y}_{lm}(\theta, \lambda) \Delta \dot{\tilde{K}}_{lm}(t) \quad (3.17)$$

This computation of the area weighted rate of change of water equivalent heights of one particular region χ , defined by a set of k grid cell centers $(\theta_i, \lambda_i), j = 1, 2, 3 \dots, k$, can be done according to

$$\dot{h}_W(\chi; t) = \sum_{j=1}^k \frac{a(\theta_i, \lambda_i)}{a(\chi)} \sum_{l=0}^{\infty} \Lambda_l \sum_{m=0}^l \tilde{Y}_{lm}(\theta_i, \lambda_i) \Delta \dot{\tilde{K}}_{lm}(t) \quad (3.18)$$

$\dot{h}_W(\chi; t)$	rate of mass change in catchment χ
k	number of date points in the catchment
$a(\theta_i, \lambda_i)$	area of the grid cell j
$a(\chi)$	total area of the catchment χ

In this thesis, the size of the cell is $0.5^\circ \times 0.5^\circ$, which means there are 360×720 cells. With the help of the *shbundle*, *EWHbundle* and the basin mask from the Institute of Geodesy (GIS), University of Stuttgart, this process can easily be done and the equivalent water heights of Ob area between 2002 and 2020 are acquirable.

3.1.2 Mass concentration

Spherical harmonics have been well studied and widely used in satellite geodesy for several decades, based on the computational efficiency of the parameterization, and because the satellite sensitivity is dependent on the spatial wavelength of the mass variations which is implicit in the harmonic basis function. However, unconstrained harmonic solutions from GRACE have typically suffered from poor observability of east-west gradients, resulting in the so-called "stripes" that are conventionally removed via empirical smoothing and-or "destriping" algorithms. Although quite effective, especially for larger spatial scales, the destriping also removes some real geophysical signal along with the stripes, and the size shape, and orientation of the signals strongly affect the effectiveness of destriping.[16]

Thus, to confirm the reliability of spherical harmonic, another common function would be taken into consideration to estimate mass flux from GRACE, which is called mass concentration(mascon). Each mass tile are defined as a finite truncated spherical harmonic representation up to degree and order 120, which are in turn related to the range-rate observation via their partial derivatives. The size of each tile is approximately 1° equatorial longitudinal distance. The mass anomaly for each of the mass tile is estimated using the KBR range-rate observations and the associated spherical harmonic partial derivatives and the singular estimation process is stabilized using Tikhonov regularization solutions with time-variable regularization matrix. [13].

This mascon solutions have no stripe errors and capture all the signals observed by GRACE within the measurement noise level. The solutions are not tailored for specific applications and are global in nature.[13] In this thesis the mascon solution from CSR are compared with the results from spherical harmonic solutions.

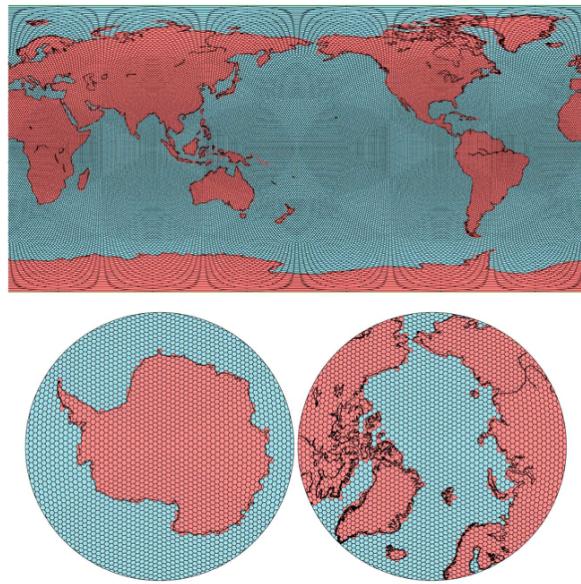


Figure 3.1: The distribution of 40,962 geodesic grid tiles over the Earth used as a basis function for estimation of mass anomalies from GRACE for CSR mascon solutions. (top) Global view, (bottom left) South Pole view, and (bottom right) North Pole view [13]

3.2 Precipitation

In this thesis the precipitation data from 9 data resources are gathered and processed (Table 3.1). The methods dealing with these data sets will be discussed later in this chapter:

Dataset	Timespan	Temporal Resolution	Spatial Resolution	Spatial Coverage
PREC/L	1948 - present	Monthly	$0.5^\circ \times 0.5^\circ$	Global
CPC	1979 - present	Monthly & Daily	$0.5^\circ \times 0.5^\circ$	Global
GPCP	1979 - present	Monthly	$2.5^\circ \times 2.5^\circ$	Global
CMAP	1979 - present	Monthly	$2.5^\circ \times 2.5^\circ$	Global
PERSIANN-CDR	1983 - present	Monthly & Daily	$0.25^\circ \times 0.25^\circ$	$60S^\circ - 60N^\circ$
NCEP1	1948 - present	Monthly & Daily	$2.5^\circ \times 2.5^\circ$	Global
NCEP2	1979 - present	Monthly & Daily	$1.875^\circ \times 1.875^\circ$	Global
ERA5	1979 - present	Monthly	$31 \text{ km} \times 31 \text{ km}$	Global
MERRA-2	1980- present	Monthly & Daily	$0.5^\circ \times 0.625^\circ$	Global

Table 3.1: precipitation datasets

Precipitation reconstruction over land (PREC/L) PREC/L if provided from National Oceanic and Atmospheric Administration (NOAA).It is derived from gauge observations from over 17 000 stations collected in the Global Historical Climatology Network(GNCN), and the Climate Anomaly Monitoring System(CAMS) datasets. By using OI analysis procedure, the monthly gridded analyses of precipitation over the global land area since 1948 are presented. The mean

distribution and annual cycle of precipitation observed in the PREC/L showed good agreement with those in several published gauge-based datasets. [2]

CPC unified gauge-based analysis of global daily precipitation This dataset is provided from Climate Prediction Center(CPC). A gauge-based analysis of daily precipitation has been constructed over the global land areas. Gauge reports from over 30 000 stations are collected from multiple sources including GTS, COOP, and other national and international agencies. Quality control is performed through comparisons with historical records and independent information from measurements at nearby stations, concurrent radar / satellite observations, as well as numerical model forecasts. Quality controlled station reports are then interpolated to create analyzed fields of daily precipitation with consideration of orographic effects [18]. The daily analysis is constructed on a 0.125 degree lat/lon grid over the entire global land areas, and released on a 0.5 degree lat/lon grid over the global domain for a period from 1979 to the present.[17] This dataset has two components: (a) the "retrospective version" which uses 30K stations and spans 1979-2005 and (b) the "real-time version" which uses 17K stations and spans 2006-present.

Global Precipitation Climatology Project (GPCP) In this dataset, precipitation data from rain gauge stations, satellites and sounding obsevations have been merged to estimate monthly rainfall on a 2.5 degree global grid since 1979. It provides a consistent analysis of global precipitation from the integration of various satellite data sets of lands and oceans and a gauge analysis over land.

CPC Merged Analysis of Precipitation (CMAP) This data set is also constructed from an analysis of gauge data and satellite-derived precipitation estimates. The overlapping satellite and reanalysis-based estimates are weighted according to their fit with the gauge-based analysis. The quality of this is in general the best in the tropics and weakens towards the polar regions.

Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Climate Data Record(PERSIANN-CDR) This data set provides daily rainfall estimates at a spatial resolution of 0.25 degrees in the latitude band 60S - 60N from 1983 to the near-present. The precipitation estimate is produced using the PERSIANN algorithm on GridSat-B1 infrared satellite data, and the training of the artificial neural network is done using the National Centers for Environmental Prediction (NCEP) stage IV hourly precipitation data. The PERSIANN-CDR is adjusted using the Global Precipitation Climatology Project (GPCP) monthly product.

NCEP/NCAR Reanalysis 1&2 The NCEP/NCAR Reanalysis 1 project is using a state-of-the-art analysis/forecast system to perform data assimilation using past data from 1948 to the present. The system has been designed with advanced quality control and monitoring components, and can produce 1 mon of reanalysis per day on a Cray YMP/8 supercomputer. Different types of output archives are being created to satisfy different user needs[8]. The NCEP Reanalysis 2 is the improvement of the NECP Reanalysis 1.

ERA5 ERA5 is produced using 4D-Var data assimilation in CY41R2 of ECMWF's Integrated Forecast System (IFS), with 137 hybrid sigma/pressure (model) levels in the vertical, with the top level at 0.01 hPa. Atmospheric data are available on these levels and they are also interpolated to 37 pressure, 16 potential temperature and 1 potential vorticity level(s). [7]"Surface or single level" data are also available, containing 2D parameters such as precipitation, 2m temperature, top of atmosphere radiation and vertical integrals over the entire atmosphere. The IFS is coupled to a soil model, the parameters of which are also designated as surface parameters, and an ocean wave model.

The ERA5 dataset contains one (hourly, 31 km) high resolution realisation (referred to as "reanalysis" or "HRES") and a reduced resolution ten member ensemble (referred to as "ensemble" or "EDA"). Generally, the data are available at a sub-daily and monthly frequency and consist of analyses and short (18 hour) forecasts, initialised twice daily from analyses at 06 and 18 UTC.[7] Most analysed parameters are also available from the forecasts. There are a number of forecast parameters, e.g. mean rates and accumulations, that are not available from the analyses. The daily total precipitation can then be calculated from ERA5 data using python with the help of CDS API.

Modern-Era Retrospective analysis for research and Applications, version 2 (MERRA-2)
This is a global atmospheric reanalysis produced by the NASA Global Modeling and Assimilation Office(GMAO). It spans the satellite observing era from 1980 to the present. The goals of MERRA-2 are to provide a regularly-gridded, homogeneous record of the global atmosphere, and to incorporate additional aspects of the climate system including trace gas constituents (stratospheric ozone), and improved land surface representation, and cryospheric processes.

3.3 Evapotranspiration

Evaporation is the process whereby liquid water is converted to water vapour (vaporization) and removed from the evaporating surface (vapour removal). Water evaporates from a variety of surfaces, such as lakes, rivers, pavements, soils and wet vegetation. 6 datasets are used in this thesis (Table 3.2).

Dataset	Timespan	Temporal Resolution	Spatial Resolution	Spatial Coverage
GLDAS NOAH	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
GLDAS CLSM	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
GLDAS VIC	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
FLDAS	2002 - present	Monthly	$0.1^\circ \times 0.1^\circ$	$60^\circ S - 90^\circ N$
SSEBop	2003- present	Monthly	$0.0097^\circ \times 0.0097^\circ$	$60^\circ S - 60^\circ N$
ERA5	1979 - present	Monthly	31 km \times 31 km	Global

Table 3.2: evapotranspiration datasets

The Global Land Data Assimilation System (GLDAS) Direct measurements and data acquisition of ET are very difficult and expensive, especially at the global level. Therefore, modeling is one common alternative for estimating ET. GLDAS has been generating quality-controlled, spatially and temporally consistent, terrestrial hydrologic data, including ET and other variants. The goal of GLDAS is to ingest satellite- and ground-based observational data products, using advanced land surface modeling and data assimilation techniques, in order to generate optimal fields of land surface states and fluxes.

The high-quality, global land surface fields provided by GLDAS support several current and proposed weather and climate prediction, water resources applications, and water cycle investigations. The project has resulted in a massive archive of modeled and observed, global, surface meteorological data, parameter maps, and output which includes 1-degree and 0.25-degree resolution 1948-present simulations of the NOAH, Common Land Model (CLM), Variable Infiltration Capacity Model (VIC), Mosaic, and Catchment Land Surface Models (CLSM). NOAH, CLSM and VIC are used in this thesis.

Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) The goal of the FLDAS project is to achieve more effective use of limited available hydroclimatic observations and is designed to be adopted for routine use for FEWS NET decision support. It is a custom instance of the NASA Land Information System (LIS) that has been adapted to work with domains, data streams, and monitoring and forecast requirements associated with food security assessment in data-sparse, developing country settings. Adopting LIS allows FEWS NET to leverage existing land surface models and generate ensembles of soil moisture, ET, and other variables based on multiple meteorological inputs or land surface models.

optional Simplified Surface Energy Balance (SSEBop) The SSEBop seTup is based on the Simplified Surface Energy Balance (SSEB) approach with unique parameterization for operational application combines ET fractions generated from remotely sensed MODIS thermal imagery, acquired every 8 days, with reference ET using a thermal index approach. The unique feature of the SSEBop parameterization is that it uses pre-defined, seasonally dynamic, boundary conditions that are unique to each pixel for the hot/dry and cold/wet reference points.

3.4 Runoff

3.4.1 Global datasets

Runoff is quantity of water discharged in surface streams. Runoff includes not only the waters that travel over the land surface and through channels to reach a stream but also interflow, the water that infiltrates the soil surface and travels by means of gravity toward a stream channel (always above the main groundwater level) and eventually empties into the channel. Runoff also includes groundwater that is discharged into a stream; streamflow that is composed entirely of groundwater is termed base flow, or fair-weather runoff, and it occurs where a stream channel intersects the water table.

The in-situ run off data are from Global Runoff Data Center (GRDC). The GRDC is an international archive of data up to 200 years old, and fosters multinational and global long-term hydrological studies. Originally established three decades ago, the aim of the GRDC is to help earth scientists analyse global climate trends and assess environmental impacts and risks. Operating under the auspices of WMO the database of quality controlled "historical" mean daily and monthly discharge data grows steadily and currently comprises river discharge data of more than 9,900 stations from 159 countries.

However, the discharge data from GRDC of Ob river are only up to 2010, which doesn't fit the purpose totally. Thus, like precipitation and evapotranspiration, models from different data centers play a role. The datasets (Table 3.3), which fit the in-situ discharge the best, are chosen while others are denied. The chosen data will then be processed. But it is shown later, that not all these datasets are good enough. In addition to those hydrological models, the altimetry measurement using satellite is also taken into consideration since the station for in-situ discharge(Salekhard)(Figure 3.2) is known.

Dataset	Timespan	Temporal Resolution	Spatial Resolution	Spatial Coverage
GLDAS NOAH	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
GLDAS CLSM	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
GLDAS VIC	1948 - present	Monthly	$0.25^\circ \times 0.25^\circ$	$60^\circ S - 90^\circ N$
ERA5	1979 - present	Monthly	31 km \times 31 km	Global

Table 3.3: runoff datasets



Figure 3.2: Station Salekhard

3.4.2 Water level height

Among the space-borne sensors, satellite altimetry can provide surface water height successively with repeat periods of 10 and 35 days. Although satellite altimetry was initially designed for oceanography, four decades of altimetry missions have provided an opportunity to study the continental hydrological cycle as well. By using the methods mentioned later, the runoff can be estimated from the satellite water level.

Altimetry satellites basically determine the distance from the satellite to a target surface by measuring the satellite-to-surface round-trip time of a radar pulse. However, this is not the only measurement made in the process, and a lot of other information can be extracted from altimetry.

The magnitude and shape of the echoes (or waveforms) also contain information about the characteristics of the surface which caused the reflection. The best results are obtained over the ocean, which is spatially homogeneous, and has a surface which conforms with known statistics.

Envisat Envisat (Environmental Satellite) is a large inactive Earth-observing satellite which is still in orbit. Operated by the European Space Agency (ESA), it was the world's largest civilian Earth observation satellite. It was launched on 1 March 2002 aboard an Ariane 5 from the Guyana Space Centre in Kourou, French Guiana, into a Sun synchronous polar orbit at an altitude of 790 ± 10 km. It orbits the Earth in about 101 minutes, with a repeat cycle of 35 days. After losing contact with the satellite on 8 April 2012, ESA formally announced the end of Envisat's mission on 9 May 2012.

In working towards the global and regional objectives of the mission, numerous scientific disciplines currently use the data acquired from the different sensors on the satellite to study such things as atmospheric chemistry, ozone depletion, biological oceanography, ocean temperature and colour, wind waves, hydrology (humidity, floods), agriculture and arboriculture, natural hazards, digital elevation modelling (using interferometry), monitoring of maritime traffic, atmospheric dispersion modelling (pollution), cartography and study of snow and ice.

Saral SARAL (Satellite with ARgos and ALtiKa) is a cooperative altimetry technology mission of Indian Space Research Organisation (ISRO) and CNES (Space Agency of France). SARAL performs altimetric measurements designed to study ocean circulation and sea surface elevation. The payloads of SARAL are The ISRO built satellite with payloads modules (ALTIKA altimeter), DORIS, Laser Retro-reflector Array (LRA) and ARGOS-3 (Advanced Research and Global Observation Satellite) data collection system provided by CNES was launched by Indian Polar Satellite Launch Vehicle rocket into the Sun-synchronous orbit (SSO). SARAL was successfully launched on 25 February 2013. It will fill the gap between Envisat and the Sentinel 3 mission of the European GMES program.

Sentinel-3 Sentinel-3 is an Earth observation satellite constellation developed by the European Space Agency as part of the Copernicus Programme. The Sentinel-3 mission's main objective is to measure sea-surface topography, sea- and land-surface temperature and ocean- and land-surface colour with accuracy in support of ocean forecasting systems, and for environmental and climate monitoring. Sentinel-3 builds directly on the heritage pioneered by ERS-2 and Envisat satellites. Near-real time data will be provided for ocean forecasting, sea-ice charting, and maritime safety services on the state of the ocean surface, including surface temperature, marine ecosystems, water quality and pollution monitoring. The satellite orbit provides a 27-day repeat for the topography package, with a 4-day sub-cycle.

Chapter 4

Method

4.1 Estimating hydrological component combining different datasets

As mentioned before, the terrestrial water balance can be written as:

$$Pre - ET - R = \frac{dS}{dt} \quad (4.1)$$

where

Pre	Precipitation
ET	evapotranspiration
R	Surface Runoff
dS/dt	total water storage change

For each, time series from several datasets are processed

There are several centers that post process Level-1 data of GRACE and provide solutions in terms of spherical harmonics. It should be noted that there is a one-year gap between the GRACE and GRACE-FO and several monthly gaps. This one-year gap would be ignored since we are interested in the long term trend and these monthly gaps are dealt using interpolation. For the equivalent water height we use spline interpolation while for uncertainty we use linear interpolation.

If two known points are given by the coordinates (x_0, y_0) and (x_1, y_1) , the linear interpolant is the straight line between these points. For a value x in the interval (x_0, x_1) , the value y along the straight line is given from the equation of slopes

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0} \quad (4.2)$$

Spline was a term of elastic rulers that were bent to pass through a number of predefined points (knots). The approach to mathematically modelling the shape of such elastic rulers fixed by $n + 1$ knots is to interpolate between all the pairs of knots and new point with polynomials. In MatLab, both of these interpolation methods can be finished using function *griddedInterpolant*.

We have 4 EWH time series from CSR, GFZ, ITSG and JPL with different uncertainties. To generate one time series of EWH for further analyse we use Gaus-Markov model.

4.1.1 Adjustment using Gauss-Markov model

Gauss-Markov model is known as the adjustment with observation equations. The model is as follows

$$\mathbf{y} = \mathbf{Ax} + \mathbf{e} \quad (4.3)$$

where \mathbf{y} is a vector of observations, \mathbf{A} is the design matrix, \mathbf{x} is a vector of unknowns and \mathbf{e} is a vector of measurement errors. Define the *Lagrangian* or *cost function*;

$$\mathcal{L}_a(\mathbf{x}) = \frac{1}{2} \mathbf{e}^T \mathbf{e} \quad (4.4)$$

Then, the adjusted observations can be estimated by using least square criterion, the best \mathbf{x} can be found with the minimum cost function. The equations can be solved as following:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (4.5)$$

$$\hat{\mathbf{y}} = \mathbf{Ax} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (4.6)$$

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T] \mathbf{y} \quad (4.7)$$

In many cases, the observations are not equal weighted, which means they have different quality. To solve this problem, we use a matrix \mathbf{P} to describe the weight. The cost function is formed as:

$$\mathcal{L}_a(\mathbf{x}) = \frac{1}{2} \mathbf{e}^T \mathbf{Pe} \quad (4.8)$$

the weighted least squares estimations are:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{PA})^{-1} \mathbf{A}^T \mathbf{Py} \quad (4.9)$$

$$\hat{\mathbf{y}} = \mathbf{Ax} = \mathbf{A}(\mathbf{A}^T \mathbf{PA})^{-1} \mathbf{A}^T \mathbf{Py} \quad (4.10)$$

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{PA})^{-1} \mathbf{A}^T \mathbf{P}] \mathbf{y} \quad (4.11)$$

For each month, there are 4 EWH values $S_{CSR}(t), S_{GFZ}(t), S_{ITSG}(t), S_{JPL}(t)$ along with their uncertainty $\sigma_{CSR}(t), \sigma_{GFZ}(t), \sigma_{ITSG}(t), \sigma_{JPL}(t)$. We use the uncertainty to build the weight matrix \mathbf{P} , then we have:

$$\mathbf{y} = \begin{pmatrix} S_{CSR}(t) \\ S_{GFZ}(t) \\ S_{ITSG}(t) \\ S_{JPL}(t) \end{pmatrix} \quad (4.12)$$

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sigma_{CSR}^2(t)} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_{GFZ}^2(t)} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_{ITSG}^2(t)} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_{JPL}^2(t)} \end{pmatrix} \quad (4.13)$$

$$\mathbf{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad (4.14)$$

By inserting Equation 4.12, Equation 4.13 and Equation 4.14 into Equation 4.9 we are able to get one EWH for one month

$$\hat{x}_{1 \times 1} = (\underset{1 \times 4}{A^T} \underset{4 \times 4}{P} \underset{4 \times 1}{A})^{-1} \underset{1 \times 4}{A^T} \underset{4 \times 4}{P} \underset{4 \times 1}{y} \quad (4.15)$$

where \hat{x} is the EWH for one month. By repeating Equation 4.15, we are able to get one whole time series for equivalent water height. The uncertainty for this new time series would be calculated using least squares error:

$$\sigma(t) = \frac{1}{4} \sqrt{\sigma_{CSR}^2(t) + \sigma_{GFZ}^2(t) + \sigma_{ITSG}^2(t) + \sigma_{JPL}^2(t)} \quad (4.16)$$

The derivation of the TWSA would be calculated using central difference:

$$\frac{dS(t)}{dt} = \frac{(dS(t + \Delta t) - dS(t)) + (dS(t) - dS(t - \Delta t))}{2\Delta t} \quad (4.17)$$

$$= \frac{dS(t + \Delta t) - dS(t - \Delta t)}{2\Delta t} \quad (4.18)$$

where Δt is one month.

4.1.2 Validation of model-based estimations

The method used in subsection 4.1.1 also works for precipitation, evapotranspiration and runoff. However, the uncertainties of these time series were unknown. Therefor, it's necessary to get the uncertainty before the adjustment.

We assume that the precipitation and evapotranspiration are normal distributed during 2002 to 2020 (though it is not the case). Under this assumption, the precipitation and evapotranspiration in the same month every year can be regarded as a constant with random errors. We then can use the standard deviation as the uncertainty.

$$\sigma_{Pre_{Jan}} = \sqrt{\frac{\sum_{i=1}^n (Pre(i)_{Jan} - \bar{Pre}_{Jan})^2}{n-1}} \quad (4.19)$$

where $\sigma_{Pre_{Jan}}$ is the standard deviation of the Precipitation in January, $Pre(i)_{Jan}$ is the precipitation in January in different years and \bar{Pre}_{Jan} is the mean of all the precipitation in January. With the same method we are able to obtain the uncertainty for precipitation and evapotranspiration in other eleven months.

4.1.3 Finding changing point

The changing points can be found using moving average (MA). In statistics, a moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set. The reason for calculating the moving average is to help smooth out the data by creating a constantly updated average data. In MatLab, this could be down with *movmean* function. The size of subsets is 12 in this thesis because of the number of months in one year.

After that, a moving average time series is generated. By computing the abrupt changes in the mean of the this new series, the sudden change of the time series can be founded. This can be achieved using MatLab function *ischange*.

4.2 Estimating the quality of runoff datasets

It was mentioned in section 3.4 that the in-situ runoff data is existing til 2010, which allows us to select the better model to do the analyze. A simple way to do that is to calculate the RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (R(t)_{insitu} - R(t)_{model})^2}{T}}$$

Meanwhile in probability theory and statistics ,the cumulative distribution function (CDF) of a real-valued random variable X , or just distribution function of X , evaluated at x , is the probability that X will take a value less than or equal to x . Quantiles are cut points dividing the range of a probability distribution into continuous intervals with equal probabilities, or dividing the observations in a sample in the same way. The quantile function, associated with a probability distribution of a random variable, specifies the value of the random variable such that the probability of the variable being less than or equal to that value equals the given probability. It is also called the percent-point function or invers cumulative distribution function.

First of all, we calculate the difference between the model and the in-situ data:

$$d(t) = R(t)_{insitu} - R(t)_{model} \quad (4.20)$$

By presenting the CDF of d , the quality of the model can be estimate. By setting the appropriate quantile we are able to choose better time series for further analyse.

4.3 Estimating the runoff using quantile function and water level height

The river discharge at the selected gauges is typically determined from an empirical functional relation between water level estimated by satellite altimetry and measured discharges. This relation, referred to as a rating curve, is specific to each gauging station and location of altimetry. However, this technique has many limitations. First, first, this technique is limited by the availability of in situ discharge measurements simultaneous with altimetry data. Second, the location of the altimetry foot- print can also limit the usage of the technique.

Fortunately, [14] provides a methods, which can represent a direct connection between the quantile functions at the corresponding probability and the relationship between runoff and water level.

first, we get the quantile functions for altimetric water level, $Q_R(p)$ and discharge from in situ measurement, $Q_W(p)$:

$$\begin{aligned} Q_R(p) &= \inf(X_R \in R : p \leq F(X_R)) \\ Q_W(p) &= \inf(X_W \in R : p \leq F(X_W)) \end{aligned}$$

where X_R and X_W refer to the runoff and water level values and F represents the CDF. The quantile function specifies, for a given probability $0 < p < 1$, the maximum value that X_R or X_W can attain with that probability.

By achieving the function T between the quantile functions, we can get the function between the runoff and water level since T is a non decreasing function.

$$Q_R = T(Q_W) \implies X_R = T(X_W) \quad (4.21)$$

In principle, the obtained relationship can be used as look-up table implying the desired rating curve. However, as this study aims to compare the statistical and empirical rating curves, a similar way of approximation for both is needed. In this thesis, the simple quadratic estimation is used for modeling the rating curve. Thus, the statistical rating curve is achieved by fitting a quadratic curve over the obtained statistical relationship. Figure 4.1 showed an example of achieving the runoff using water level data from Envisat.

By repeating this process to Saral and Sentinel missions we can generate an whole runoff time series from 2002 to 2020. (note there is a gap between Envisat and Saral)

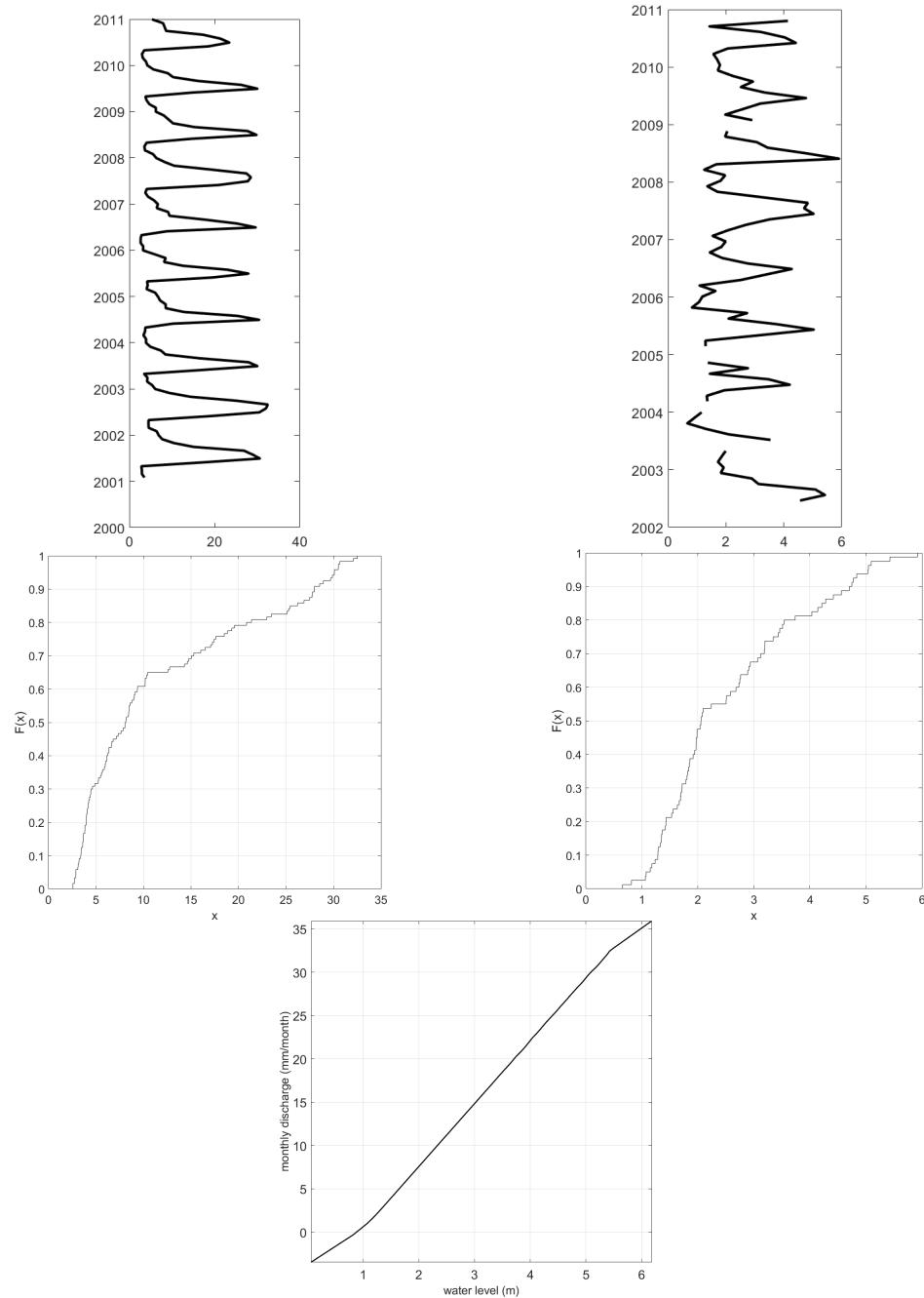


Figure 4.1: Estimated water level from Envisat and available runoff for Ob river (top) and quantile function of water level and runoff (middle). A smoothed rating curve is obtained from the corresponding probabilities (bottom)

Chapter 5

Result

Now we can use our method to determine dS/dt , Pre , ET and R . After that, we can combine these results and discuss the trend of water storage in Ob basin and the reason for it.

5.1 Equivalent water height

As mentioned in chapter 3, 4 time series of TWSA from Apr.2002 to May.2020 are handled in this work Figure 5.1. It is shown that before 2010 no obvious trend in TWSA can be seen, from 2012 to 2016 TWSA has gained consistently. By using the method in subsection 4.1.1, these four time series can be summarized into one time series. Figure 5.2. The summarized time series shows the same tendency as in Figure 5.1. To prove that, the deviation of TWSA can be reached by using centre difference, since data in 2002 and 2020 are not for whole year available, the time series from Jan.2013 to Dec.2019 are chosen. Figure 5.3. note that dS/dt is between -50 and 50 mm/month.

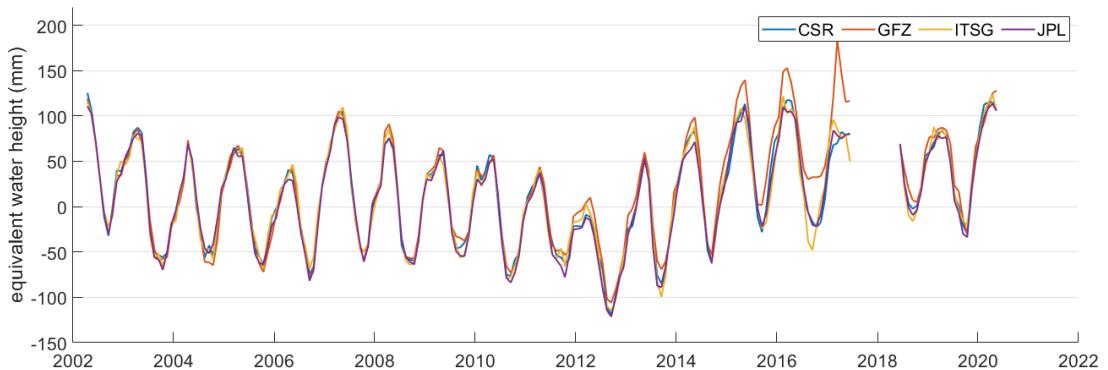


Figure 5.1: Total water storage anomaly from CSR, GFZ, ITSG and JPL between Apr.2002 and May.2020

2003 - 2012	2013 - 2015	2015 - 2019
-------------	-------------	-------------

From Figure 5.5 we can see, that the total water storage in the first and the third period are relative stable. However, between 2013 and 2015, the water in Ob river basin has gained from outside.

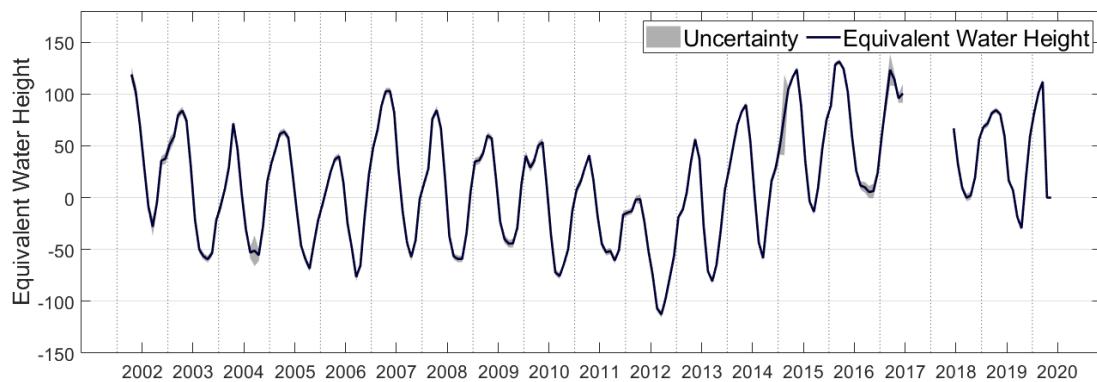


Figure 5.2: Total water storage anomaly between Apr.2002 and May.2020

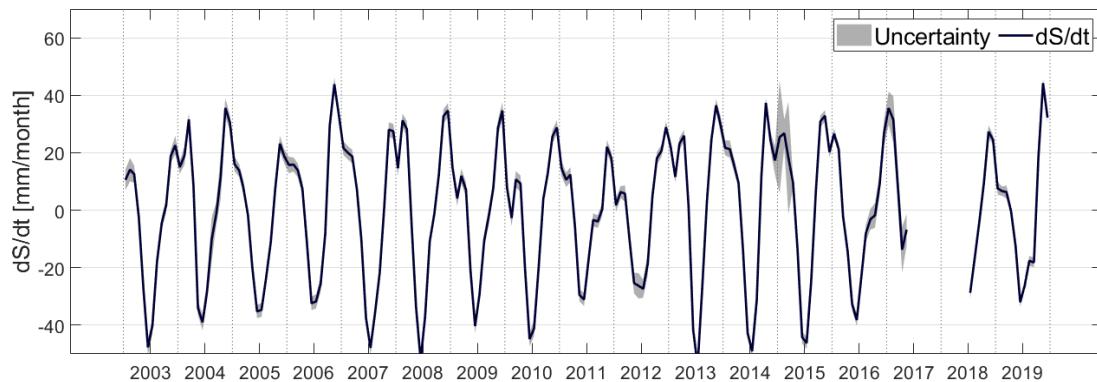


Figure 5.3: dS/dt between 2003 and 2020

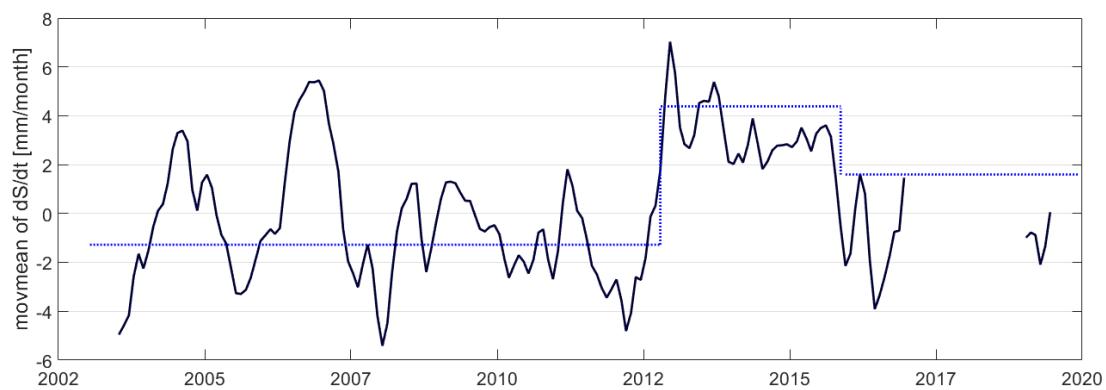


Figure 5.4: Sudden change of dS/dt between 2003 and 2020 after using movmean

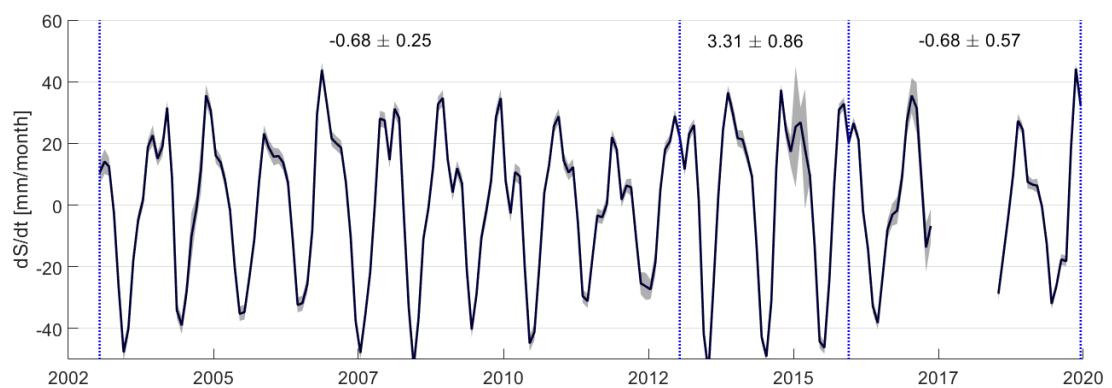


Figure 5.5: The mean value of dS/dt in three periods

5.2 Precipitation

It was mentioned in chapter 3 that we have precipitation data from 9 resources Figure 5.6, and one timeseries with uncertainty can be generated by combining them. The spatial behavior of precipitation in this area is also interesting to us.

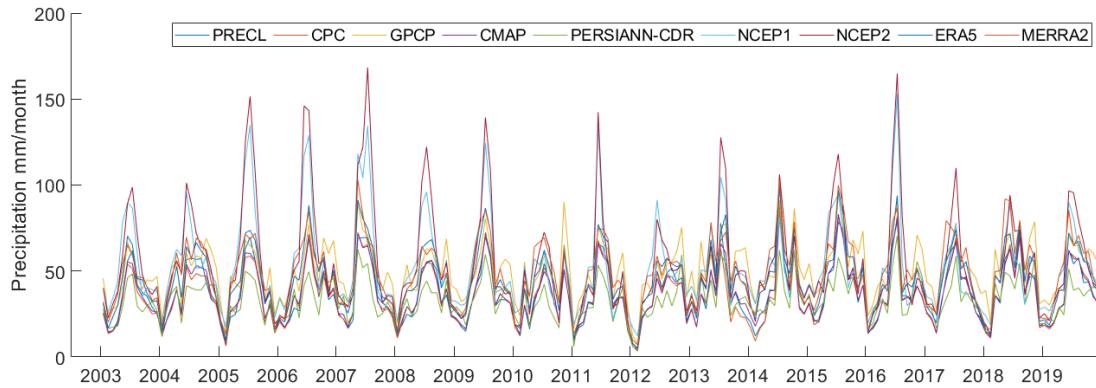


Figure 5.6: Precipitation datasets

Since we have already divide the whole period into 3 phases, we are interested what we see the behavior of precipitations in this 3 phases. Figure 5.7

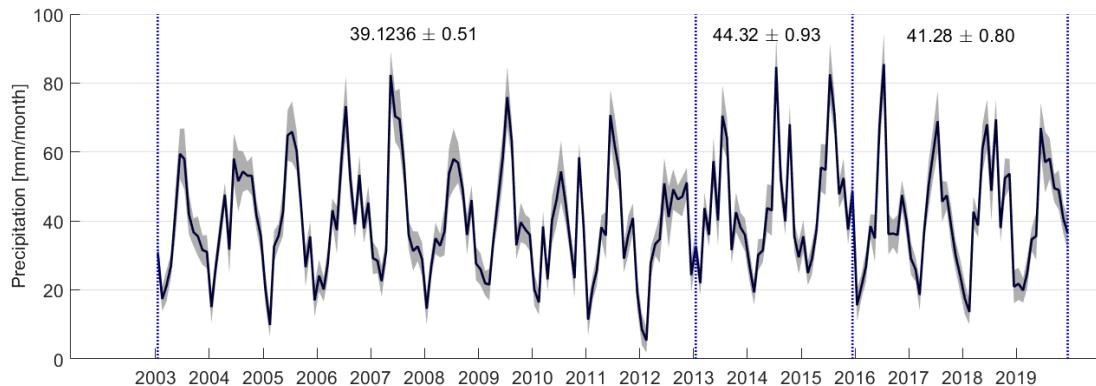


Figure 5.7: precipitation with uncertainties between 2003 and 2020 and the mean value in 3 periods

We can see the precipitation in the second period is obviously higher than the other 2 periods. After 2016, the amount of precipitation has decreased.

5.3 Evapotranspiration

Just as precipitation, we present the evapotranspiration temporally and spatially and cut the timeseries into 3 periods.

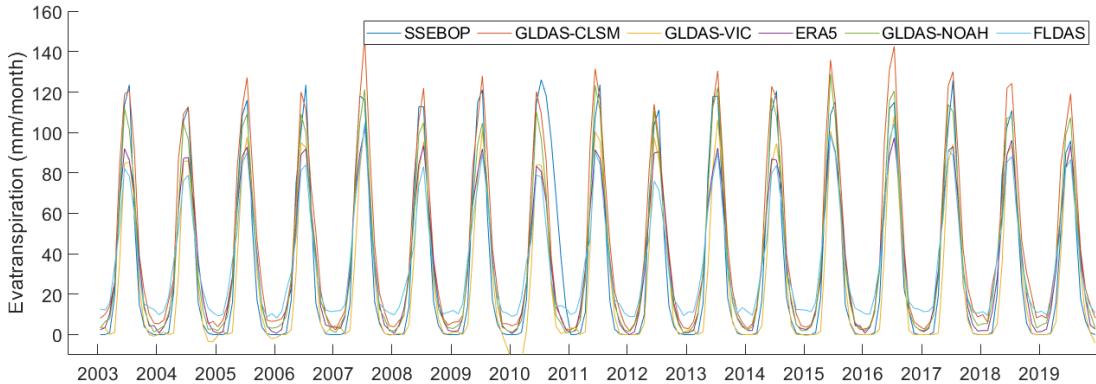


Figure 5.8: evapotranspiration datasets

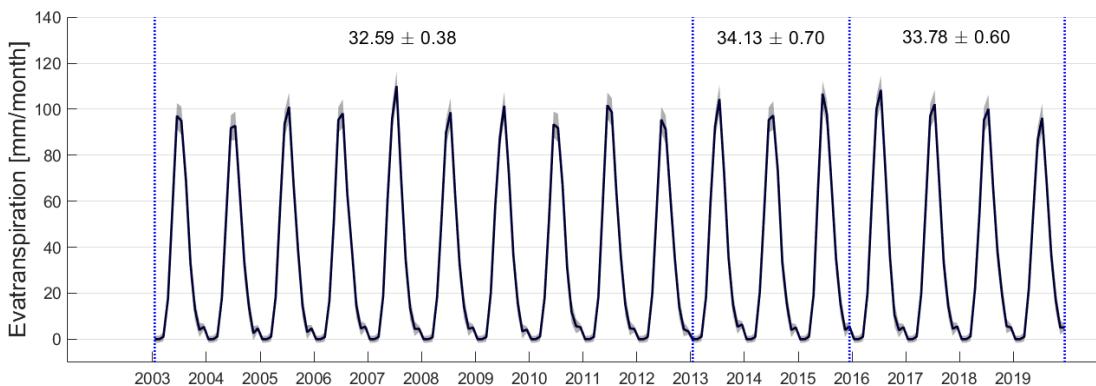


Figure 5.9: evapotranspiration with uncertainties between 2003 and 2020 and the mean value in 3 periods

The behavior of evapotranspiration is similar to precipitation, in second period the evapotranspiration reached the summit and then decreased, which suits theory of water cycle.

5.4 Runoff

5.4.1 Runoff from global datasets

Like precipitation and evapotranspiration, we have runoff data estimated from several models and meanwhile we have in-situ data till end of 2010. We can plot them in a way we did for precipitation and evapotranspiration. From the Figure 5.10 we can assume that there are large differences between these models and in-situ. By calculating RMSE for these models Figure 5.4.1 this assumption can be confirmed. The smallest RMSE for those datasets are bigger than $8.18 \frac{\text{mm}}{\text{month}}$, which is far from what we needed.

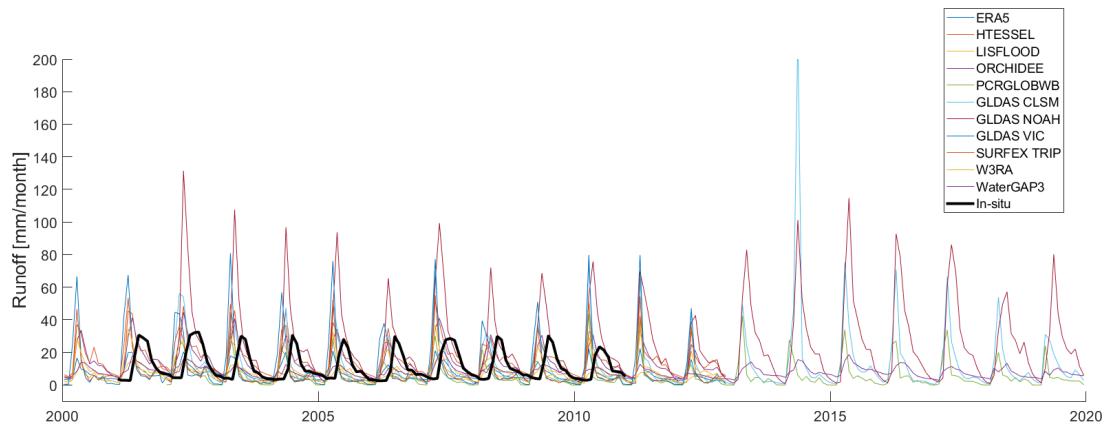


Figure 5.10: Runoff datasets

Datacenter	RMSE (mm/month)
ERA5	8.18
HTESSEL	10.40
LISFLOOD	11.92
ORCHIDEE	10.24
PCRGLOBWB	9.48
GLDAS CLSM	13.68
GLDAS NOAH	17.01
GLDAS VIC	25.95
SURFEX-TRIP	22.05
W3RA	16.47
WaterGAP3	11.03

Table 5.1: RMSE for runoff datacenters

Then, if we take the CDF of the difference from those models and in-situ data and set 10% of mean in-situ runoff as quantile Figure 5.11 , it is shown that none of these datasets has achieved probability of 90%

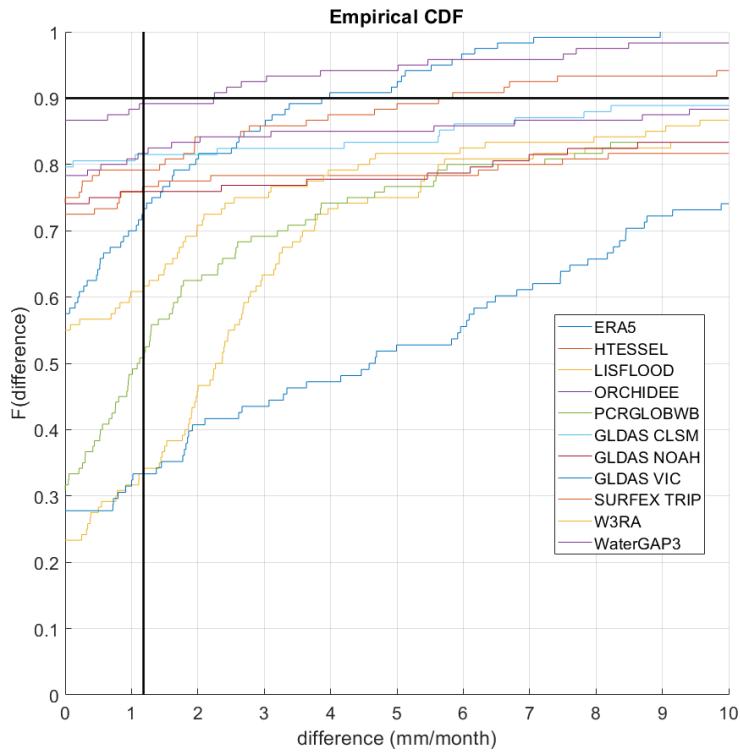


Figure 5.11: CDF of differences

5.4.2 Estimating runoff using quantile function and satellite altimetry

As mentioned in section 3.4, we can use the satellite to determine the water level height in Ob river basin, and we also have the in-situ data, using the methods in section 4.3, the runoff can be estimated. We have chosen several virtual station for each satellite mission. After denying the virtual stations in bad locations we get 2 water level timeseries from Envisat, 2 from Saral and 2 from Sentinel. In order to get the most accurate runoff, we choose the water level timeseries from the virtual stations, which are closer to Salekhard, where the in-situ runoff are measured.

In Figure 5.12 the location of those virtual stations and the water level timeseries are represented. we finally choose the blue line from Envisat, the blue line from Saral and timeseries from Sentinel-3A to generate discharge.

Now we have in-situ runoff data from 2001 to 2010 and water level, Using the methods in section 4.3, we are able to find the probability connections between them and after all obtain a smoothed rating curve. We use this method for Envisat, Saral and Sentinel separately and combine them in one time series. There are still several gaps in the series, we use interpolation to fill the gaps. Since we don't have uncertainty information for each month, we take the RMSE as the uncertainty for 3 periods.

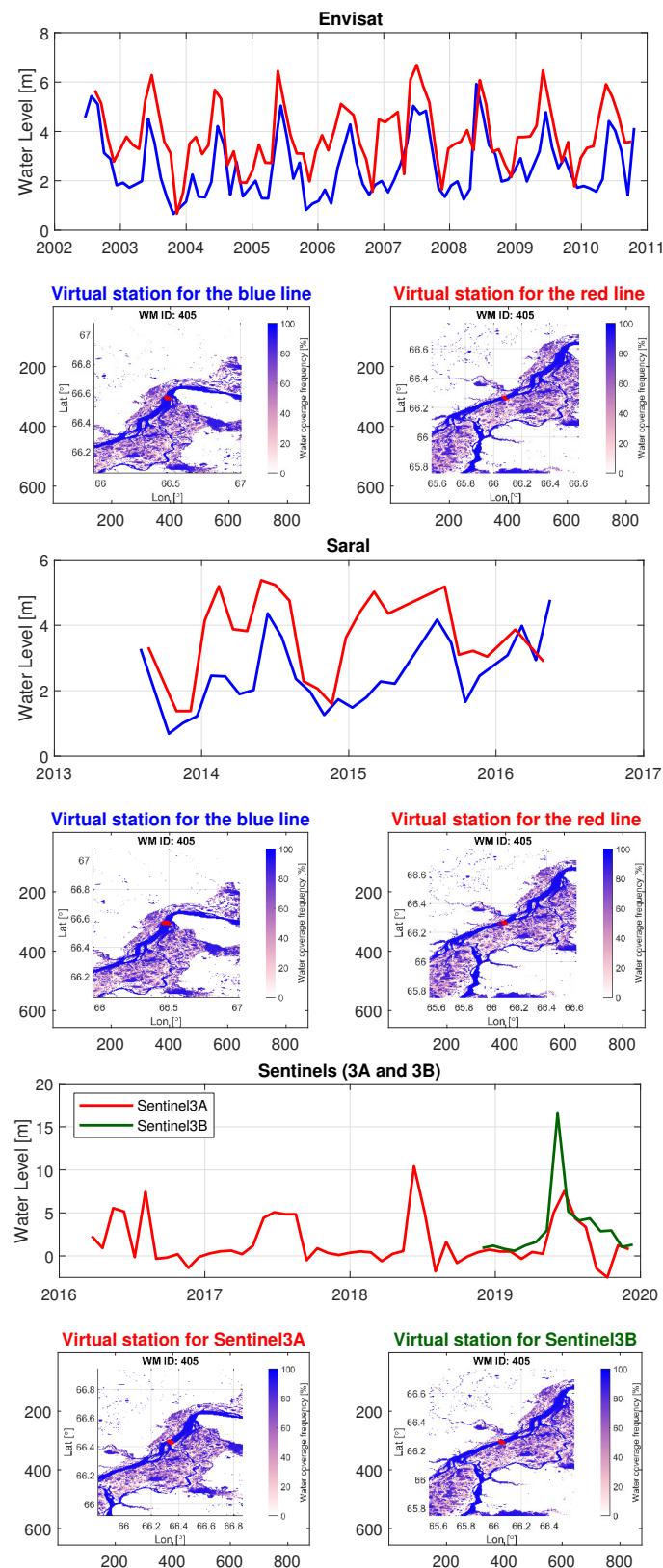


Figure 5.12: water level time series and the location of virtual station

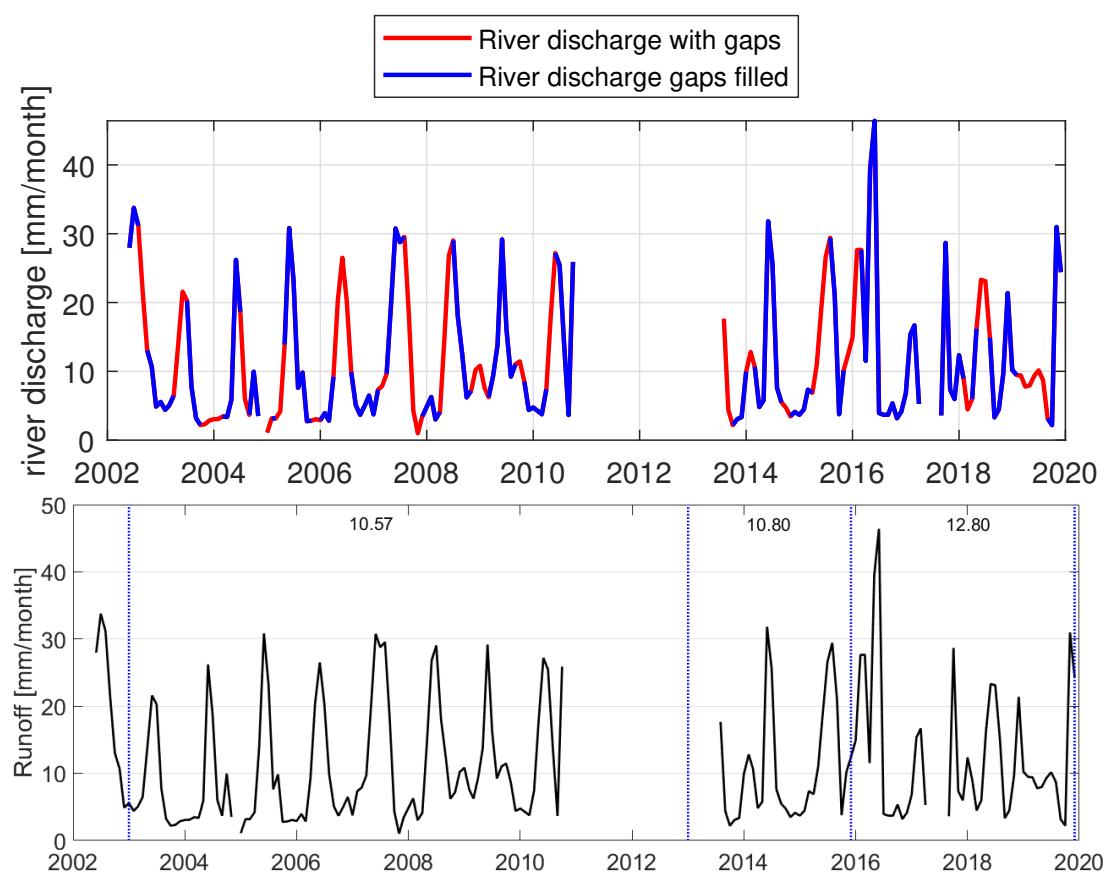


Figure 5.13: runoff timeseries calculated from water level (up) and the mean value in 3 periods(bottom)

5.5 Discussion about the quality

Since the total water storage in the first period can be seen as stable, we take this period as the baseline for further analyze. Relative dS/dt , precipitaion, evapotranspiration and runoff numbers of other 2 period are obtained based on that.

In any case, the equation of the water cycle has to be a true statement, which means the result of $Pre - \frac{dS}{dt} - ET - R$ should be 0. By setting results of both periods into this statement, it is shown that the results for the second period are proved while the results for the third periods are not ideal. This can be caused by the gap between GRACE and GRACE-FO. This gap is between Jul.2017 and May.2018, therefore, we only have full year data of 2016 and 2019 in this period. Though we tried linear regression to make the best use of what we have, this gap can still increase the instability of the final results because the period is not long enough.

Period	2003 - 2012	2013 - 2015	2016 - 2019
dS/dt	-0.68 ± 0.25	3.31 ± 0.86	-0.80 ± 0.57
Precipitaion	39.12 ± 0.51	44.32 ± 0.93	41.28 ± 0.80
evapotranspiration	32.59 ± 0.38	34.13 ± 0.70	33.78 ± 0.6
Runoff	10.57 ± 4.16	10.80 ± 4.16	12.80 ± 4.16
Based on the first period			
dS/dt		4.00 ± 0.45	-0.11 ± 0.52
Precipitaion		5.20 ± 0.53	2.15 ± 0.61
evapotranspiration		1.53 ± 0.40	1.19 ± 0.46
Runoff		0.23 ± 2.94	2.23 ± 2.94
$Pre - dS/dt - ET - R$		-0.57 ± 0.76	-1.14 ± 0.76

Table 5.2: mean value of all water cycle component in 3 periods

Chapter 6

Conclusion and outlook

6.1 Conclusion

In this work the behavior of total water storage along with other water cycle component in Ob river basin since 2003 are discussed. The equivalent water height is through GRACE and GRACE-FO mission determined, precipitation and evapotranspiration are obtained from various data models and Runoff is measured using satellite altimetry indirectly. For each of these component, one final time series is generated using mathematic and statistic methods. The whole time series divide into 3 periods and the change of each component in these periods are analyzed.

The GRACE data showed that before 2013, the total water storage in Ob river basin is quite stable but from 2013 to 2015 it has a large increase. This is due to the increase of the precipitation in this area. Meanwhile, runoff changed not much, evapotranspiration has also increased but not as much as precipitation.

After 2015, the runoff of basin was shown to have a obvious growth according to space measurement but both of precipitation and evapotranspiration were lower than in the second period (still higher than the first period). However, because of a one-year gap between GRACE and GRACE-FO, the average change of total water storage during this period may be not very reliable. However, one thing is clear, the water cycle in this area is no more stable since 2013.

6.2 Outlook

As mentioned, the trend of total water storage after 2016 is still not clear because of the gap between GRACE and GRACE-FO. This issue can be easily solved in the future when more data is acquirable.

Bibliography

- [1] Richard G Allen, Luis S Pereira, Dirk Raes, Martin Smith, et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109, 1998.
- [2] Mingyue Chen, Pingping Xie, John E Janowiak, and Phillip A Arkin. Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3(3):249–266, 2002.
- [3] Terry E Evans. The effects of changes in the world hydrological cycle on availability of water resources. *Global Climate Change and Agricultural Production: Direct and Indirect Effects of Changing Hydrological, Pedological and Plant Physiological Processes*, pages 15–48, 1996.
- [4] F Ghassemi, AJ Jakeman, and HA Nix. Salinization of land and water resources. *CAB International, Wallingford, UK*, 1995.
- [5] Peter H Gleick. Water in crisis. *Pacific Institute for Studies in Dev., Environment & Security. Stockholm Env. Institute, Oxford Univ. Press*. 473p, 9, 1993.
- [6] Dazhong Han and John Wahr. The viscoelastic relaxation of a realistically stratified earth, and a further analysis of postglacial rebound. *Geophysical Journal International*, 120(2):287–311, 1995.
- [7] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.
- [8] Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The ncep/n-car 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–472, 1996.
- [9] Philip P. Micklin Liliya Konstantinovna Malik, Lewis Owen. Ob river.
- [10] NASA. Grace mission. https://www.nasa.gov/mission_pages/Grace/spacecraft/index.html.
- [11] C. Revenga, World Resources Institute, S. Murray, A. Hammond, and J.N. Abramovitz. *Watersheds of the World: Ecological Value and Vulnerability*. World Resources Institute report. World Resources Institute, 1998.
- [12] Matthew Rodell, JS Famiglietti, DN Wiese, JT Reager, HK Beaudoin, Felix W Landerer, and M-H Lo. Emerging trends in global freshwater availability. *Nature*, 557(7707):651–659, 2018.

- [13] Himanshu Save, Srinivas Bettadpur, and Byron D Tapley. High-resolution csr grace rl05 mascons. *Journal of Geophysical Research: Solid Earth*, 121(10):7547–7569, 2016.
- [14] MJ Tourian, N Sneeuw, and A Bárdossy. A quantile function approach to discharge estimation from satellite altimetry (envisat). *Water Resources Research*, 49(7):4174–4186, 2013.
- [15] John Wahr, Mery Molenaar, and Frank Bryan. Time variability of the earth’s gravity field: Hydrological and oceanic effects and their possible detection using grace. *Journal of Geophysical Research: Solid Earth*, 103(B12):30205–30229, 1998.
- [16] Michael M Watkins, David N Wiese, Dah-Ning Yuan, Carmen Boening, and Felix W Landerer. Improved methods for observing earth’s time variable mass distribution with grace using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, 120(4):2648–2671, 2015.
- [17] Pingping Xie, M Chen, and W Shi. Cpc unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc*, volume 2, 2010.
- [18] Pingping Xie, Mingyue Chen, Song Yang, Akiyo Yatagai, Tadahiro Hayasaka, Yoshihiro Fukushima, and Changming Liu. A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007.

Appendix A

Und das hier ist noch der Anhang...