Frequentist and Bayesian Approaches to State-Space Modelling of Ice Dynamics in Antarctic Ice Basins

Author
Mats Bergeman
4909496

Thesis Committee
prof. dr.-ing. habil. R. (Roland) Klees
dr.ir. T. (Taco) Broerse
dr. M. (Marc) Schleiss

Master Thesis Applied Earth Sciences
Earth Observation

Geoscience & Remote Sensing
Delft University of Technology
The Netherlands

July 2024

Abstract

The melting of the Antarctic ice sheet is anticipated to play a significant role in sea level rise over the upcoming decades. Long-term mass and volume changes of the Antarctic ice sheet are predominantly caused by changes in the movement of the ice layer, referred to as ice dynamics. Mass and volume changes of the Antarctic ice sheet are monitored by gravimetry and altimetry satellites. Their data are corrected for glacial isostatic adjustment and changes in the surface climate, namely cumulated surface mass balance anomalies and firn thickness changes, to obtain ice-dynamical mass and volume time series. These time series are modelled using dynamic state-space models in this thesis.

State-space models decompose the data into several components. These components consist of parameters that are constant with time and states that vary with time. This thesis considers two modelling approaches to estimate the parameters and states, referred to as the the Frequentist and Bayesian approaches. The Frequentist approach entails estimating model parameters using maximum likelihood estimation and subsequently determining the states at each epoch using the Kalman Filter and Smoother. The Frequentist parameter estimates are deterministic. As a result, potential stochasticity of parameters is not accounted for when determining the states. This may lead to overconfident small uncertainties in Frequentist models. The Bayesian approach remedies this by considering model parameters to be stochastic. Following the Bayesian approach, the marginal posterior distributions of the parameters are estimated using Markov Chain Monte Carlo methods. Samples from these distributions are used to estimate the conditional distributions of the states at each epoch using the simulation smoother. Because parameter samples are used to estimate the states, stochasticity of the parameters is accounted for when sampling the states following the Bayesian approach.

Three distinct Antarctic ice drainage basins are investigated in this thesis. For their ice-dynamical ice mass data, significant differences between the uncertainties of Frequentist and Bayesian models are found. The Bayesian uncertainties are consistently larger than the Frequentist uncertainties. The largest differences between Frequentist and Bayesian uncertainties are found for the slope components of state-space models. Depending on the complexity of the data that are being modelled, the Bayesian uncertainty of a slope component can be up to 4 times larger than its Frequentist uncertainty.

Frequentist and Bayesian methods to combine the trends of gravimetry-based and altimetry-based state-space models of ice-dynamical ice mass in Antarctic ice basins are also investigated in this thesis. Epoch-wise weighted averaging, with weights based on the uncertainties in the gravimetry-based and altimetry-based models, is done to combine the trends. It is found that the gravimetry data is of significantly higher quality than the altimetry data, having much smaller uncertainties. As a result, the weighted average of the trends aligns closely to the gravimetry-based trend. Finally, several choices that have to be made when working with Bayesian Markov Chain Monte Carlo methods for state-space modelling and their impact on the results are discussed.

Preface

Before you lies my thesis for the degree of Master of Science in Applied Earth Sciences with a specialization in Earth Observation, at the Delft University of Technology. Thanks to prof. dr.-ing. habil. R. Klees, dr.ir. T. Broerse and dr. M.A. Schleiss, my supervisors while working on this thesis, for their continuous support and feedback. Further thanks go out to Dr. M.O. Willen, for providing the data used in this thesis.

List of symbols

α	Metropolis-Hastings acceptance ratio
$\bar{l}_n$	Weighted average of the GRACE-based and altimetry-based level components at epoch n
Δt	Step size in Hamiltonian Markov Chain Monte Carlo
ϵ_n	Irregular vector
$\hat{\theta}$	Maximum likelihood parameter estimates
$\hat{x}_n$	smoothed state vector at epoch n
$\hat{y_n}$	Observed states at epoch n
ω_n	Disturbance vector
$\omega_{AR,n}$	White noise process driving the AR(3) process at epoch n
$\omega_{c*,n}$	Disturbance term of a cycle component's auxiliary term at epoch n
$\omega_{c,n}$	Disturbance term of a cycle component's coefficient at epoch n
$\omega_{r,n}$	Slope disturbance term at epoch n
ψ	Argument of the cosine and sine terms of cycle components
σ^2_ϵ	Irregular variance
σ^2_{AR}	Variance of the white noise process which drives the AR(3) process
$\sigma^2_{cycle,1}$	Disturbance variance of the annual cycle
$\sigma^2_{cycle,2}$	Disturbance variance of the SA tidal alias cycle
σ^2_r	Slope disturbance variance
Θ	Parameter space
θ	Vector of parameters
$\tilde{x}_{n+1}$	Filtered states at epoch n+1
φ_{n-1}	AR(3) process at epoch n-1
φ_{n-2}	AR(3) process at epoch n-2
φ_n	AR(3) process at epoch n
a_1	First autoregressive coefficient of the AR(3) process
a_2	Second autoregressive coefficient of the AR(3) process
a_3	Third autoregressive coefficient of the AR(3) process
A_n	Transition matrix at epoch n
A_{n-1}	Transition matrix at epoch n-1
AIC	Akaike information criterion
BIC	Bayesian information criterion
c_n	Cycle component at epoch n (volume model)
c^*_n	Auxiliary term of a cycle component at epoch n (volume model)
$c_{1,n}*$	Auxiliary term for the annual cycle component at epoch n (mass model)
$c_{1,n}$	Annual cycle component at epoch n (mass model)
$c_{2,n}*$	Auxiliary term for the SA tidal alias cycle component at epoch n (mass model)
$c_{2,n}$	SA tidal alias cycle component at epoch n (mass model)
dt	Time between consecutive epochs
$f(\theta_{new})$	A function proportional to the probability density function of the posterior distribution that is sampled, using a new set of parameters
$f(\theta_{old})$	A function proportional to the probability density function of the posterior distribution that is sampled, using the previous set of parameters

F_n	Variance-covariance matrix of prediction errors
H_n	Design matrix
k	Number of model parameters (length of θ)
K_n	Kalman gain
L_n	Smoothing gain
l_n	Level at epoch n
$l_{1,n}$	Grace-based level at epoch n
$l_{2,n}$	Altimetry-based level at epoch n
l_{n+1}	Level at epoch n+1
m	Number of steps in Hamiltonian Markov Chain Monte Carlo
m_1	Mean initial states
M_n	Variance-covariance matrix of the smoothed states at epoch n
N	Total number of epochs
n	Epoch
$P(\theta\|y)$	Marginal posterior distribution for parameters
$P(\theta)$	Prior distribution of parameters
$P(x_{1:N}\|\theta, y)$	Conditional distribution of the states at each epoch given parameters and data
$P(y_{1:N}\|\hat{\theta})$	Likelihood given the maximum likelihood estimates
P_1	Mean initial state variance-covariance matrix
P_{n+1}	Error variance-covariance matrix of the filtered states at epoch n+1
Q_n	Disturbance variance-covariance matrix
r_n	Slope at epoch n
r_{n+1}	Slope at epoch n+1
r_{n-1}	Weighted sum of prediction errors at epoch n-1
T	Period of a cycle component
U_{n-1}	Weighted sum of inverse variances
V_n	Irregular variance-covariance matrix
v_n	Prediction errors (innovations)
$w_{1,n}$	Weight given to the GRACE-based level at epoch n
$w_{2,n}$	Weight given to the altimetry-based level at epoch n
x_1	Initial state vector
x_n	State vector at epoch n
x_{n+1}	State vector at epoch n+1
y_n	Vector of observations at epoch n
$y_{1:N}$	Full data record
$y_{1:n}$	Observations up to epoch n

Contents

1 Introduction

The melting of the Antarctic ice sheet is anticipated to play a significant role in sea level rise over the upcoming decades (Alley et al., 2005; Bronselaer et al., 2018; Bulthuis et al., 2019). Mass changes in the Antarctic ice sheet are composed of two main constituents: surface mass balance and ice dynamics (Willen et al., 2021). The surface mass balance encompasses mass changes caused by processes in the surface climate such as meltwater runoff, precipitation and snow sublimation. These processes are the primary source of short-term mass changes in the Antarctic ice sheet and primarily take place in its upper layers, comprised by snow and firn. Ice dynamics, the primary driver of long-term mass changes in the Antarctic ice sheet, refers to mass changes caused by motion of the ice layer, ultimately outputting ice mass into the oceans.

To gain insight into mass changes of the Antarctic ice sheet and the underlying processes involved, the Antarctic ice sheet is monitored. This is primarily done using satellite observations. The Gravity Recovery and Climate Experiment (GRACE) satellite mission consisted of two satellites that measured changes in the Earth's gravitational field from 2002 until 2017. A follow-up mission, named GRACE-FO, was launched in June 2018 and is active to this day. GRACE data for the Antarctic ice sheet consist of mass changes caused by both surface mass balance changes and ice dynamics (van Wessem et al., 2018). Several prior studies have used GRACE data to investigate time series of Antarctic ice mass (Chen et al., 2006; Groh et al., 2019; Velicogna, 2009).

To investigate how the Antarctic ice sheet evolves, time series of both ice mass and ice volume are investigated. Ice volume data can be inferred from satellite altimetry missions such as the ESA's CryoSat or NASA's ICESat. Altimetry satellites measure changes in surface elevation, including the ones caused by changes in the surface climate, such as firn compaction, and those caused by ice dynamics. When height changes are multiplied by a surface area, volume change is obtained. Various studies have used satellite altimetry data for time series analysis of Antarctic ice volume (Helm et al., 2014; Zhang et al., 2020). Note that the studies cited here are illustrative examples, but there are many other studies of similar research using satellite gravimetry and altimetry to obtain time series of the mass and volume of the Antarctic ice sheet.

To gain a better understanding of how the ice mass and ice volume of the Antarctic ice sheet evolve over time, time series models are made. From such models, patterns, trends, and other behaviors in the data that vary over time can be inferred. Many time series models exist. A distinction can be made between two general types: static and dynamic models (Kantz and Schreiber, 2004). Static models assume that model components, such as the trend or the phase offset and amplitude of a periodic signal, remain constant over time while dynamic models allow model components to change with time.

In this thesis, dynamic time series models are made for the ice-dynamical ice mass and ice volume of Antarctic ice basins. Two approaches towards fitting these models to data are compared: the Frequentist and Bayesian approaches. Additionally, GRACE-based ice mass models are combined with altimetry-based ice volume models.

1.1 State-space models

The dynamic time series models used in this thesis are state-space models. State-space models are structural time series models, which decompose data into several components. Examples of components are a level, or cycle. These components consist of parameters and states. Model parameters stay constant with time, whereas states vary with time. The time-varying nature of the states makes state-space models dynamic.

State-space models have historically often been used in the field of econometrics (Hamilton, 1994). This is because economic processes, such as stock prices, are often modelled as random walks (Menkveld et al., 2007). The use cases of dynamic state-space models are not limited to just economic models, however. Natural processes, such as ice loss in the Antarctic, are impacted by many factors, both known and unknown. Therefore, variations in such processes can be considered random. Because of this, dynamic state-space modelling are suitable to model time series of such natural processes as well.

State-space models have been used to model climatic time series in several research projects. For instance, state-space models have been used to model sea level change (Frederikse, 2018). More closely related to the topic of this thesis, time series of Antarctic ice mass and volume have previously been modelled using state-space models (Willen et al., 2021). In the study by Willen et al. (2021), long-term mass and volume changes in ice basins that are primarily caused by ice dynamics were separated from short-term mass changes that mainly resulted from variations in the surface climate. To achieve this, GRACE-based mass time series and altimetry-based volume time series were first corrected for glacial isostatic adjustment by a glacial isostatic adjustment model. Mass changes due to surface mass balance anomalies, retrieved from a regional climate model, were subtracted from the GRACE data. Firn thickness changes obtained by a firn densification model were subtracted from satellite altimetry data. By correcting the GRACE and altimetry data for short-term mass changes occurring on the surface, long-term ice mass and volume changes caused by ice dynamics were obtained.

Updated versions of the ice-dynamical ice mass and volume data from the study by Willen et al. (2021) are modelled with state-space models in this thesis.

1.2 Frequentist and Bayesian approaches

State-space models may contain unknown parameters, which are often estimated using maximum likelihood estimation. The model states are subsequently estimated using the Kalman Filter and Smoother. This way of modelling is known as the Frequentist approach. Following the Frequentist approach, the model parameters, once estimated, are considered deterministic. These deterministic estimates are then used to estimate the model states. As a result, stochasticity that may be present in the model parameters is not taken into account when estimating the states. This may lead to overconfident small uncertainties in Frequentist time series models.

An alternative to the Frequentist approach, is the Bayesian approach. In the Bayesian approach, model parameters are stochastic variables, described by probability density functions. Samples from model parameters' posterior distributions are used to estimate the states. Consequently, the stochastic nature of model parameters is propagated to the estimated states. As a result, the Bayesian approach leads to larger, but more realistic uncertainties in the fitted model, than the Frequentist approach. The significance of the difference in the uncertainty found by both methods, depends on how much stochasticity is contained in the model parameters. The efficacy of the Frequentist approach compared to the Bayesian approach is a divisive topic in time series modelling (Albers et al., 2018; Bayarri and Berger, 2004; Pek and van Zandt, 2020).

Generally, the probability distribution of a parameter is not (fully) known. However, if data are available, the posterior distribution of a parameter may be estimated. Bayesian estimation of model parameters' posterior distributions is often done using Markov Chain Monte Carlo (MCMC) methods. MCMC methods sample from the marginal posterior distribution for the parameters $P(\theta|y)$, where θ is the vector of unknown model parameters and y is the vector of observations. The term 'marginal' indicates that $P(\theta|y)$ is irrespective of the states in x_n, meaning the parameters are sampled no matter what the states are.

Intrinsic to working with Bayesian methods such as MCMC, is the use of priors. Priors encode known information about the parameters before data have become available. An example of a prior is constraining a certain parameter to only take on non-negative values when it is known that negative values for that parameter are impossible. This is the case for a variance parameter, for example.

After estimating the marginal distributions of the model parameters given the data $P(\theta|y)$ using MCMC, samples of the conditional distribution of the state vectors given the data and parameters $P(x_{1:N}|\theta, y)$ can be obtained using the simulation smoother (De Jong and Shephard, 1995). Statistics or samples from $P(\theta|y)$ and $P(x_{1:N}|\theta, y)$ are used to obtain the observed states, approximating the data.

Many different MCMC samplers exist. A well-known and relatively simple to implement MCMC sampler is the Metropolis-Hastings algorithm. The Metropolis-Hastings algorithm generally performs well for a small number of parameters in θ. For a large number of parameters, Gibbs samplers or Hamiltonian samplers are more suitable (Hogg and Foreman-Mackey, 2018). Note that regardless of which sampler is used, MCMC samplers are only ever correct 'in the limit', meaning they only sample the parameters' true distributions when an infinite number of samples are taken. Hence, it should be realized that the distributions retrieved by MCMC samplers are merely estimates of the true distributions. In research, a trade-off has to be made between the accuracy of this estimation and the computational runtime of the MCMC algorithm.

1.3 Frequentist and Bayesian uncertainties

In this thesis, the Frequentist and Bayesian approaches to fitting state-space models to time series of ice mass and volume changes due to ice dynamics in Antarctic ice drainage basins are investigated. The research is limited to three ice basins: basins 6, 13 and 22, which show three distinct behaviors in their data. The mass and volume data for basins 6, 13 and 22 are discussed in chapter 2 and shown in figures 2 and 3, respectively. By investigating specifically basins 6, 13 and 22, the behavior of Frequentist and Bayesian state-space models for distinct types of ice-dynamical mass and volume changes is seen.

The first goal of this thesis is to compare the uncertainties of the trends, which are integrated random walk processes, in Frequentist and Bayesian state-space models of ice mass and ice volume in Basins 6, 13 and 22. The trend is specifically of interest, because it is expected that when the ice mass decreases, the ice volume would also decrease, but current models don't always show such a relationship (Willen et al., 2021).

Following the Frequentist approach, the states are generally estimated using the Kalman Filter and Smoother. Uncertainties of the states are then represented by a confidence interval. When following the Bayesian approach, the simulation smoother is used to estimate the conditional distribution of the states at every epoch given the data and parameters. The uncertainty of the states is then represented by a highest density interval.

Confidence intervals and highest density intervals are different concepts, both representing uncertainty in different ways. A 95% confidence interval, for example, delineates a range of values within which the true value will lie in 95% of repeated experiments. A 95% highest density interval, on the other hand, represents the most narrow range of values within which 95% of the posterior distribution's probability density is located.

To properly compare the results of the Frequentist and Bayesian approaches, confidence intervals cannot be directly compared to highest density intervals. To compare the Frequentist and Bayesian uncertainties in this thesis, instead of using the Kalman Filter and Smoother to estimate the states for the Frequentist approach, the simulation smoother is also used for the Frequentist approach. As a result, both the Frequentist and Bayesian states have highest density intervals, allowing for comparison between the two approaches.

For the Frequentist approach, each iteration of the simulation smoother uses the same maximum likelihood parameter estimates. Because the same parameter estimates are used each iteration of the simulation smoother, stochasticity that may be present in the parameters is not taken into account when estimating the states following the Frequentist approach. For the Bayesian approach, each iteration of the simulation smoother uses a different set of parameters sampled from $P(\theta|y)$. As a result, the stochastic nature of the parameters is accounted for when following the Bayesian approach. How much impact the stochasticity of model parameters has on the trends of ice mass and volume models is investigated by comparing the widths of the Frequentist and Bayesian 95% highest density intervals.

1.4 Combining trends of ice mass and ice volume models

Besides comparing Frequentist and Bayesian uncertainties in the trends of ice-dynamical ice mass and volume models, this thesis also aims to answer how the trends of GRACE-based ice mass models can be combined with those of altimetry-based ice volume models. This has been deemed difficult due to unknown errors in satellite altimetry data and the need to model surface mass balance anomalies and firn thickness changes to isolate the ice-dynamical ice mass and volume change, with unknown errors in the models used for this (Willen et al., 2021).

In order to combine the trends of ice-dynamical ice mass and volume models, the volume data are first converted to mass by multiplying with the mean density of ice: 0.917 Gt/km^3 (Rumble, 2024). This can be done because the ice-dynamical volume data have been corrected for glacial isostatic adjustment and firn thickness changes, thereby removing volume changes caused by changes on the surface layers of the Antarctic ice sheet consisting of snow and firn. This means it can be assumed that all remaining volume change is of ice.

To combine the estimated trends in GRACE-based and altimetry-based state-space models for ice-dynamical ice mass of basins 6, 13 and 22, epoch-wise weighted averaging is done. The weights are based on the uncertainties in the GRACE-based and altimetry-based trends. This is done for models fitted following the Frequentist approach and the Bayesian approach. For the Frequentist approach, the weights at each epoch are equal to the reciprocal of the variances of the trends. For the Bayesian approach, the weights at each epoch are equal to the squared widths of the 95% highest density intervals of the trends.

1.5 Research Questions

This thesis aims to answer the following research questions:

1. How do the uncertainties of the trends in Frequentist and Bayesian state-space models of ice-dynamical ice mass in Antarctic ice basins compare to each other?
2. How can the trends of gravimetry-based and altimetry-based state-space models of ice dynamical ice mass in Antarctic ice basins be combined?

A general (sub)goal of the project is to gain experience with the Bayesian approach to time series modelling. To this end, the importance and impact of different choices made when working with MCMC methods are investigated. These include the choice of sampler, priors, initialization, number of chains and samples, burn-in period length and the potential thinning of samples. Therefore, this thesis will also answer the following sub-question:

3. How do choices that have to be made when working with Bayesian Markov Chain Monte Carlo methods for state-space modelling affect the model?

1.6 Outline

In chapter 2, the data used in this thesis are explained. An overview of the study area, being Antarctic ice drainage basins 6, 13 and 22, is also given in this chapter. The GRACE gravimetry data for these basins and the models used to correct the GRACE data so that only ice-dynamical mass change remains, are explained as well. Furthermore, the multi-mission satellite altimetry data and models used to correct the altimetry data so that only ice-dynamical volume change remains, are described.

The dynamic state-space models introduced in section 1.1 are explained in more detail in chapter 3. This includes the state-space model equations, an explanation of the software used for generating state-space models and an overview of the different components that can be added to them.

The Frequentist approach to fitting state-space models to data is explained in chapter 4. The process of using maximum likelihood estimation to estimate the model parameters is expounded upon. Additionally, the Kalman Filter and Smoother to estimate the model states are explained. Frequentist diagnostics used to determine how well a model fits the data are explained as well. Finally, the Frequentist model selection process is outlined.

In chapter 5, the Bayesian method of estimating the model parameters is further explained. The process of estimating the marginal posterior distibution for the model parameters $P(\theta|y)$ using Markov Chain Monte Carlo methods is explained. The use of the simulation smoother to subsequently estimate the conditional distribution of the states at every epoch given the model parameters and data is also described. MCMC sampling diagnostics and posterior predictive checks are explained. Finally, informative and non-informative priors are discussed.

Chapter 6 shows the chosen models following the Frequentist model selection process outlined in chapter 4. The equations for the ice mass and volume models are shown in this chapter. Additionally, the priors and sampling settings chosen for the Bayesian approach are noted.

The chosen models are fitted to the data with both the Frequentist and the Bayesian approach. The Frequentist and Bayesian results of fitting the ice-dynamical mass models to the data are shown in chapter 7. The difference in uncertainty of the trends in the Frequentist and Bayesian ice mass models is determined in this chapter in order to answer the first research question.

In chapter 8, the trends of GRACE-based and altimetry-based state-space models for the ice mass are combined. To this end, the altimetry-based volume models of chapter 6 are converted to mass. The combination is done for both the Frequentist and Bayesian approach.

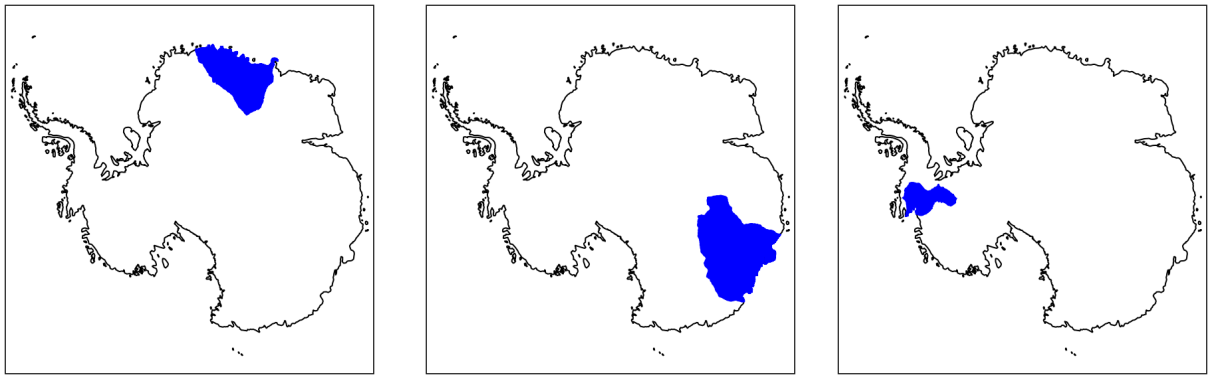
Chapter 9 gives a discussion on the choices that have to be made when working with Bayesian MCMC methods. This is used to answer the third sub-question. Finally, in chapter 10, the conclusion of this research project is given. Recommendations for future research are also noted in chapter 10

2 Data and study region

The Antarctic ice sheet is divided up into 27 different ice drainage basins (NASA, 2024). Three of these basins are of specific interest in this thesis because they each show distinct behavior in their ice mass and volume data. These focus basins are basins 6, 13 and 22. Section 2.1 shows the locations of the focus basins. The GRACE gravimetry data for these basins and the models used to correct the GRACE data so that only the ice mass changes caused by ice dynamics remains, are explained in section 2.2. Subsequently, the multi-mission satellite altimetry data, and the models used to correct the altimetry data for firn thickness changes so that only the ice-dynamical ice volume change remains, are described in section 2.3.

2.1 Antarctic ice drainage basins

Basins 6, 13 and 22 are of special interest, because they differ in the dominance of long-term and short-term changes in ice mass and volume (Willen et al., 2021). The mass and volume change in basin 22 is dominated by ice dynamics, whereas the mass and volume change in basin 13 is primarily caused by changes in the surface climate. The ice mass and volume in basin 6 is influenced by a combination of changes in the surface climate and ice dynamics. These behaviors can be seen when looking at the data in figures 2 (mass data) and 3 (volume data) in upcoming sections 2.2 and 2.3. The locations of the focus basins are shown in figure 1 below.



(a) Basin 6 (b) Basin 13 (c) Basin 22

Figure 1: Antarctic ice drainage basins 6, 13 and 22.

Beside the 27 basins defined by NASA, 7 other aggregations of GRACE & altimetry data are provided in the dataset used in this thesis. These include the Antarctic peninsula, East Antarctica, West Antarctica and the entire Antarctic ice sheet. Thus, data for a total of 34 basins exist. The locations of all 34 basins are shown in appendix A.

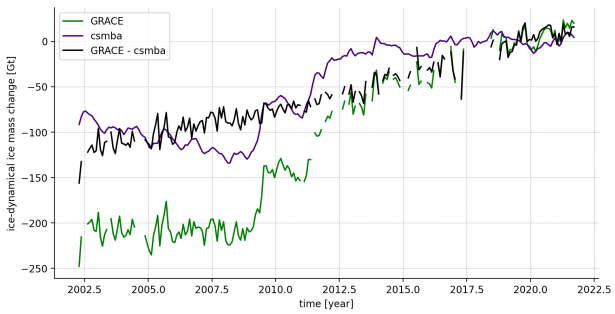
2.2 GRACE gravimetry data

The Gravity Recovery and Climate Experiment (GRACE) was a joint mission of NASA and the German Aerospace Center, active from 2002 up until 2017. The follow-on mission GRACE-FO started in 2018 and is active to this day. The way GRACE and GRACE-FO measure mass changes on the Earth is explained in appendix B.

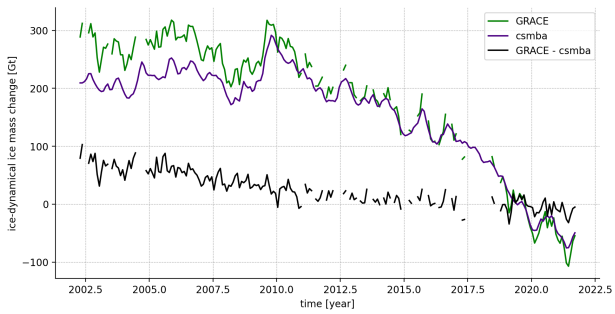
For the Antarctic ice drainage basins that are investigated in this thesis, aggregations of gridded data from GRACE and GRACE-FO are available (Groh and Horwath, 2021). These data represent the total mass change at an area. The data are corrected for glacial isostatic adjustment with a glacial isostatic adjustment model by Ivins and James (2005).

After correcting the aggregated GRACE data for glacial isostatic adjustment, the data still contain both the mass changes related to surface mass balance changes and the mass changes related to ice dynamics. As this thesis is focused on modelling ice dynamics, the data have to be corrected for mass changes caused by surface mass balance anomalies, so that only the ice-dynamical mass change remains. This has been done in previous research by subtracting the cumulated surface mass balance anomalies of the RACMOC2.3p2 model from the GRACE data (van Wessem et al., 2018).

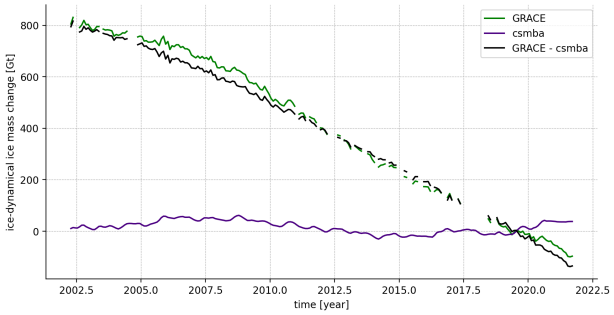
Preprocessed GRACE data as described above were used in an earlier study about seperating short- and long-term mass changes in Antarctic ice basins (Willen et al., 2021). Updated versions of these data have been provided by dr. M.O. Willen for use in this thesis and are plotted for basins 6, 13 and 22 in figure 2.



(a) Basin 6



(b) Basin 13



(c) Basin 22

Figure 2: Time series of monthly GRACE data corrected for glacial isostatic adjustment and cumulated surface mass balance anomalies for Antarctic ice drainage basins 6, 13 and 22 from 2002 up to 2021, representing ice-dynamical ice mass change (black). Cumulated surface mass balance anomalies model for the same time period (purple). GRACE data corrected for glacial isostatic adjustment, but not for cumulated surface mass balance anomalies (green). The ice mass data are given in gigatonnes.

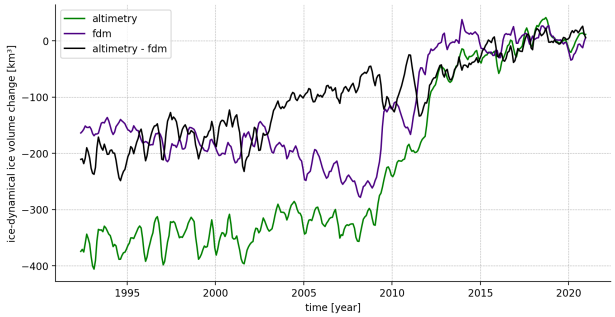
Figure 2 shows that ice mass changes in basin 6 are caused by a combination of ice dynamics and surface mass balance changes. The ice-dynamical mass change of basin 6 also shows an increasing trend. The ice mass change in basin 13 is primarily caused by surface mass balance changes, with the ice dynamical mass change being comparatively stable. Conversely, the ice mass change in basin 22 is dominated by ice dynamics, with little impact of surface mass balance changes. The ice-dynamical mass change of basin 22 shows a strong downward trend. Because basins 6, 13 and 22 have three distinct behaviors relating to mass change, Frequentist and Bayesian state-space models for their ice-dynamical mass data can be compared to see how well the modelling approaches perform for different types of trends in mass data.

2.3 Multi-mission satellite altimetry data

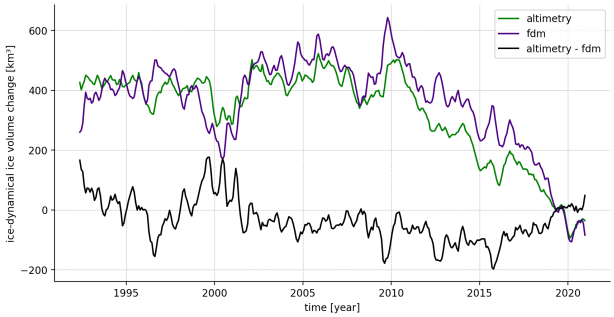
Altimetry satellites measure the elevation of the Earth's surface. A brief explanation of satellite altimetry is given in appendix B. Altimetry measurements over the Antarctic are aggregated to the Antarctic ice drainage basins shown in figure 1. By multiplying the aggregated altimetry data by the surface areas of the basins, ice volume data are obtained. Such ice volume time series have been created in an earlier study by Nilsson et al. (2022).

The multi-mission satellite altimetry dataset from Nilsson et al. (2022) contains the elevation change of Antarctic ice basins from 1985 to 2020. This is achieved by combining data from seven different satellite altimetry missions. The first, and oldest, of these missions is the GEOdetic SATellite (GEOSAT), which was launched by the United States' Navy in 1985 and was operational up until 1989. The second and third of these missions are the European Remote Sensing satellites (ERS-1 and ERS-2), which were launched by ESA in 1991 and 1995, respectively. ERS-1 operated until 1996 and ERS-2 operated until 2003. The fourth mission is ESA's Environmental Satellite (Envisat). Envisat was launched in 2002 and decommissioned 2012. However, only Envisat data from 2002 to 2010 are used, due to changes in Envisat's orbit initiated in October of 2010. The fifth and sixth missions whose data are used in the multi-mission altimetry dataset are NASA's Ice, Cloud, and land Elevation Satellites (ICESAT and ICESAT-2). ICESAT was operational from 2003 to 2009, and its follow-on mission ICESAT-2, was launched in 2018 and is active to this day. The final mission whose data is incorporated in the dataset is ESA's CryoSat-2, which launched in 2010 and is also active to this day.

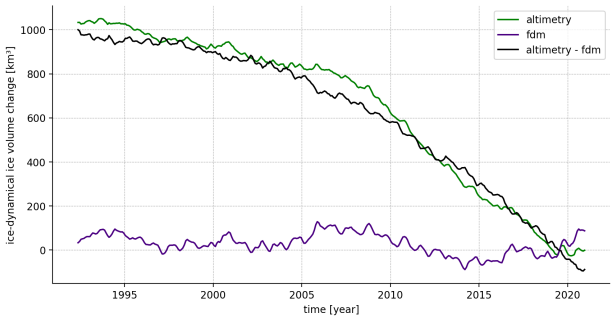
Altimetry measurements are more sensitive to changes in the surface mass balance than gravimetry measurements. This is due to the relatively low density of snow and firn compared to ice. Surface mass balance changes occur in the snow and firn layers. Due to the low density, a change in snow and firn thickness affects the height, as measured by altimetry, more severely than it affects the mass, as measured by gravimetry. Hence, correcting the altimetry data for changes in the surface climate, such as firn thickness changes, is important to be able to determine volume changes due to ice dynamics. To this end, the multi-mission altimetry dataset by Nilsson et al. was corrected in previous research by subtracting the surface height changes provided by the IMAU Firn Densification Model v1.2A from the data (Veldhuijsen et al., 2023). Additionally, the altimetry data are corrected for glacial isostatic adjustment with the same glacial isostatic adjustment model as the mass data (Ivins and James, 2005). The data for basins 6, 13 and 22 are shown below in figure 3



(a) Basin 6



(b) Basin 13



(c) Basin 22

Figure 3: Time series of monthly satellite altimetry data corrected for surface height changes of the IMAU-FDM v1.2A firn densification model, for Antarctic ice drainage basins 6, 13 and 22 from 2002 up to 2021, representing ice-dynamical volume change (black). Modelled volume change due to firn thickness changes for the same time period (purple). Original altimetry data that are not corrected for firn thickness changes (green). The ice volume data are given in cubic kilometers.

Figure 3 shows that ice volume changes in basin 6 are caused by a combination of changes in the surface climate and ice dynamics. There is a clear increasing trend in the ice volume of basin 6, with a large jump occurring between the years 2010 and 2013. For basin 13, ice volume changes are primarily caused by changes on the surface. Comparatively, the ice dynamical volume change has no clear trend. The ice volume change in basin 22 is dominated by ice dynamics, with little influence of changes in the surface climate, like firn thickness changes. Basin 22 encompasses the Pine Island glacier, which is the fastest retreating glacier of western Antarctica (Shepherd et al., 2001). This explains the large decrease seen in both the ice mass time series of basin 22 in figure 2(c) and the ice volume time series of basin 22 in figure 3(c).

Because basins 6, 13 and 22 have three distinct behaviors relating to volume change, Frequentist and Bayesian state-space models for their ice-dynamical volume data can be compared to see how well the modelling approaches perform for different types of trends in volume data.

To convert the ice-dynamical volume data to ice-dynamical mass, the volume data can be multiplied by a density of 0.917 Gt/km^3. This can only be done because the ice-dynamical volume data is corrected for firn thickness changes, meaning the remaining volume change is of ice, which has a density of 0.917 Gt/km^3 (Rumble, 2024).

3 State-space models

The ice-dynamical mass and volume time series of basins 6, 13 and 22 shown in figures 2 and 3 are approximated using state-space models. This chapter explains the state-space models introduced in section 1.1 in more detail, starting with the equations for a linear Gaussian state-space model in section 3.1. The different components that a state-space model can contain are explained in section 3.2.

3.1 State-space equations

State-space models contain model parameters and states. Parameters refer to constants that describe how the model evolves over time, such as the variance of a cyclic component in the model. States are unobserved variables that change with time, such as the amplitude of a cyclic component.

State-space models consists of the observation equation, state equation and initial state conditions. Equations 1, 2 and 3 describe the observation equation, state equation and the initial state condition for a linear Gaussian state-space model, sometimes referred to as a Dynamic Linear Model (Durbin and Koopman, 2012; Laine, 2020). State-space models are also able to account for non-linear conditions and non-Gaussian distributions, but this is outside the scope of this thesis.

Linear Gaussian State-Space Model

$$y_n = H_n x_n + \epsilon_n \qquad \epsilon_n \sim N(0, V_n) \tag{1}$$

$$x_{n+1} = A_n x_n + \omega_n \qquad \omega_n \sim N(0, Q_n) \tag{2}$$

$$x_1 \sim N(m_1, P_1) \tag{3}$$

In the equations above, y_n is the vector of observations at epoch n. For univariate state-space models y_n is a scalar. There are N epochs in the time series, thus $n = 1, 2, 3, ..., N$. The model states in state vector x_n are transformed into the observations in y_n by design matrix H_n. The model states in x_n change per epoch according to transition matrix A_n. A Gaussian irregular term ϵ_n, representing noise in the data, is added to the observation equation (1). A Gaussian disturbance vector ω_n, representing noise in the state evolution, is added to the state equation (2). The irregular terms ϵ_n are scalars for univariate models and are independent and identically distributed (IID), having Gaussian distributions centered at zero. Similarly, the disturbance vectors ω_n are also IID with Gaussian distributions centered at zero. ϵ_n and ω_n have variance-covariance matrices V_n and Q_n, respectively. The initial state condition shown in equation 3 describes the initial state vector x_1 as a Gaussian distribution with a vector of means m_1 and variance-covariance matrix P_1. The initial state vector x_1 contains the states at epoch $n = 1$, before the first observations in vector of observation y_1 has become available. The initial state condition is not always known, but is required for the model. Initialization choices are discussed in section 4.2.

3.2 Model components

State-space models are structural time series models, meaning they decompose the data into multiple components. The *Statsmodels* Python package is used to generate state-space models in this thesis. The unobserved components function of *Statsmodels* allows for the following components to be implemented in a state-space model (Statsmodels documentation, 2024b):

- Level: Long term behavior of the time series data.
- Slope: The change in level between consecutive epochs.
- Cycle: Cyclic fluctuations in the data with either known or unknown periods.
- Irregular: ϵ_n as given in equation 1, describing random noise in the observations.
- Autoregressor: Autoregressive component describing temporal dependence within the data, which could be unmodelled signal or correlated observation noise.
- Regressor: a variable that is external to the system being modelled, but still impacts the behavior or outcome of the system. In the case of modelling ice mass and volume of Antarctica, temperature or precipitation could be considered exogenous variables, for example.

The level, slope and cycle components can be either static or dynamic. The addition of model components increases the amount of parameters and states in the model. Adding level and slope components creates states for the level and slope. If the level and slope components are dynamic, the level and slope disturbance variances become additional model parameters. Adding a cycle component creates two new states. These represent the value of the cycle and an auxiliary term that allows for the recursive description of the cycle component, as

described in section 3.2.4. of Durbin and Koopman (2012). If a cycle component is dynamic, its disturbance variance becomes an additional model parameter. A cycle component can have either a known or unknown frequency. If the frequency is unknown, the unknown cycle frequency becomes an additional model parameter. The variance of the irregular term is also a model parameter. The addition of an autoregressive component adds a number of states for the autoregressive processes, equal to the order of the autoregressor. Additionally, multiple parameters are added: the disturbance variance of the autoregressor and a number of autoregressive coefficients, also equal to the order of the autoregressive component. Regressors are not considered in this thesis.

4 Frequentist approach

The corrected ice-dynamical ice mass and volume data shown in chapter 2 are modelled using the linear Gaussian state-space models introduced in chapter 3. In this chapter, the Frequentist approach to fitting state-space models is explained.

The Frequentist approach to state-space modelling entails estimating model parameters with maximum likelihood estimation and estimating model states using the Kalman Filter and Smoother. Maximum likelihood estimation of model parameters is described in section 4.1. Estimation of model states using the Kalman Filter and Smoother is explained in section 4.2. Frequentist diagnostics used to determine the goodness of fit of a model and identify potential improvements are outlined in section 4.3. The Frequentist process of model selection is described in section 4.4.

4.1 Maximum likelihood estimation of model parameters

The Frequentist approach to time series modelling employs maximum likelihood Estimation to estimate the model parameters. Maximum likelihood estimation is considered a standard approach to parameter estimation and inference in statistics (Myung, 2003). Maximum likelihood estimation, as the name implies, maximizes the marginal data likelihood $P(y_{1:N}|\theta)$. The likelihood function is obtained as a byproduct of the Kalman Filter explained in section 4.2.

Often, the log-likelihood function is used in maximum likelihood estimation. The log-likelihood is the natural logarithm of the likelihood function. Taking the natural logarithm simplifies calculations and provides numerical stability by avoiding potential underflow or overflow errors caused by multiplication or division of very small or very large values. The maximum likelihood estimates $\hat{\theta}$ are found when the log-likelihood function is maximized. This results in equation 4 (Myung, 2003).

Maximum likelihood estimates

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ln\left(P\left(y_{1:N}|\theta\right)\right) \tag{4}$$

The model parameters θ in parameter space Θ are optimized to find the maximum likelihood estimates $\hat{\theta}$. This is an iterative process where parameter values are changed every iteration until a converged solution is found for which the log-likelihood function $\ln\left(P\left(y_{1:N}|\theta\right)\right)$ is maximized. The log-likelihood function is maximized when the prediction errors are minimized. The prediction errors are computed at each epoch by the Kalman Filter; see equations 7 and 8 in section 4.2.

Usually, the log-likelihood is a non-convex function of θ (Klees, 2024). This means the function has local maxima in addition to just one global maximum. The non-convex nature of the log-likelihood requires more advanced optimization routines to find the model parameters' maximum likelihood estimates $\hat{\theta}$ than would be needed for convex functions (Klees, 2024). Maximum likelihood estimation requires initial values for θ from which the optimization routines start iterating. If improper initial values are chosen, the maximum likelihood estimation may not converge. Frequently, multiple different initial values for θ are chosen and the corresponding maxima are compared to ensure the global maximum is found instead of a local maximum (Klees, 2024). In cases of log-likelihood functions that are quite flat, different initial values deviate in the found maximum likelihood estimates. In such occasions, the model usually has to be adjusted by using different components and thus changing the parameters in θ (Klees, 2024).

In this thesis, the *Statsmodels* Python module is used to generate state-space models and to estimate the model parameters using maximum likelihood estimation (Fulton, 2015; Seabold and Perktold, 2010). To obtain the maximum likelihood estimates $\hat{\theta}$, *Statsmodels* uses numerical optimization algorithms from the *Scipy* Python package (Virtanen et al., 2020). By default, *Statsmodels* uses the limited-memory Broyden-Fletcher-Goldfarb-Shanno optimization algorithm when estimating state-space model parameters through maximum likelihood estimation (Fulton, 2017). An introductory explanation of the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm is given by Lam (2020).

When the maximum likelihood estimation is complete, the state-space model parameters $\hat{\theta}$ for which the log-likelihood is maximized are found. These parameters are considered to be deterministic, meaning the outcomes of the parameters are fixed.

4.2 Kalman Filter & Smoother to estimate the model states

Generally, estimation of the state vector x_n at every epoch n is performed using the Kalman Filter and Smoother when following the Frequentist approach (Kalman, 1960).

The Kalman Filter finds optimal estimates of the states x_n at epoch n based on all data up to that epoch. These estimates are referred to as the filtered states $\tilde{x}_n$. After running the Kalman Filter, the Kalman Smoother can be used to provide optimal estimates for the states x_n at every epoch based on all available data. These estimates are called the smoothed states $\hat{x}_n$.

4.2.1 Kalman Filter

Recall the linear Gaussian state-space representation from chapter 3. There, the initial state conditions x_1 at epoch $n = 1$ are described in equation 3. If these are known, the Kalman Filter uses the initial state conditions to start estimating the states at subsequent epochs. If the initial state conditions aren't known, as if often the case in practice, the Kalman Filter has to be initialized using an exact diffuse Gaussian distribution centered at zero with infinite variance or an approximate diffuse distribution centered at zero with a very large variance.

The Kalman filter starts from the initial state conditions and computes the filtered states $\tilde{x}_n$ at epoch n based on observations $y_{1:n} = y_1, ..., y_n$. The Kalman Filter estimates the filtered states for every epoch until the final epoch N has been reached. The definitions of the filtered states $\tilde{x}_n$ and the error variance-covariance matrix of the filtered states P_n are given in equations 5 and 6, respectively. The Kalman Filter equations shown in this section are retrieved from section 4.3 of Durbin and Koopman (2012).

Definitions of the filtered states and their error variance-covariance matrix

$$\tilde{x}_n = \mathbb{E}(x_n|y_{1:n}) \tag{5}$$

$$P_n = \mathrm{Var}(x_n|y_{1:n}) \tag{6}$$

To estimate $\tilde{x}_n$ and P_x, the Kalman Filter first computes the prediction errors v_n, the corresponding variance-covariance matrix F_n and the Kalman gain K_n (Durbin and Koopman, 2012). The prediction errors v_n and their variance-covariance matrix F_n are computed using equations 7 and 8, respectively. The prediction errors v_n are often referred to as the innovations.

Prediction errors and their variance-covariance matrix

$$v_n = y_n - H_n A_{n-1} \tilde{x}_{n-1} \tag{7}$$

$$F_n = H_n P_n H_n^\top + V_n \tag{8}$$

Recalling the state-space representation denoted in chapter 3, H_n and V_n are the design matrix and the variance-covariance matrix of the irregular term, respectively. A_n is the transition matrix of the state equation. After determining F_n, The Kalman gain K_n can be computed using equation 9.

Kalman gain

$$K_n = A_n P_n H_n^\top (F_n)^{-1} \tag{9}$$

The Kalman gain K_n can be seen as a weight given to the observation at epoch n. The greater K_n is, the more the filtered state $\tilde{x}_n$ will conform to observation y_n. When the prediction errors v_n, their variance-covariance matrix P_n and the Kalman gain K_n at epoch n are all computed, the filtered states $\tilde{x}_{n+1}$ and the state error variance-covariance matrix P_{n+1} at epoch $n + 1$ can be computed using equations 10 and 11, respectively.

Filtered states at epoch $n + 1$ and their error variance-covariance matrix

$$\tilde{x}_{n+1} = A_n \tilde{x}_n + K_n v_n \tag{10}$$

$$P_{n+1} = A_n P_n (A_n - K_n H_n)^\top + Q_n \tag{11}$$

In equation 11, Q_n is the variance-covariance matrix of the disturbance vector ω_n at epoch n. The Kalman Filter iteratively computes the filtered states $\tilde{x}_n$ for all epochs in the observational record.

4.2.2 Kalman Smoother

As mentioned earlier, the Kalman Filter computes the filtered states $\tilde{x}_n$, which are estimates of the states x_n given the data up to epoch n. To obtain estimates of the x_n at every epoch using all available data, both past and future, the Kalman Smoother is used. These estimates are referred to as the smoothed states $\hat{x}_n$. The definitions for the smoothed states $\hat{x}_n$ and their error variance-covariance matrix M_n are given in equations 12 and 13, respectively. All Kalman Smoother equations shown in this section are retrieved from section 4.4 of Durbin and Koopman (2012).

Definitions of the smoothed states and their error variance-covariance matrix

$$\hat{x}_n = \mathbb{E}(x_n|y_{1:N}) \tag{12}$$

$$M_n = \mathrm{Var}(x_n|y_{1:N}) \tag{13}$$

In equations 12 and 13, $y_{1:N}$ is the full data record $y_{1:N} = y_1, ..., y_N$. The Kalman Smoother goes backwards from the final epoch N to the first epoch and computes $\hat{x}_n$ and M_n at every epoch. To do this, the smoothing gain L_n must first be computed using equation 14.

Smoothing gain

$$L_n = A_n - K_n H_n \tag{14}$$

L_n is used to compute the vector r_{n-1} at epoch $n-1$, containing the weighted sum of prediction errors. The matrix U_{n-1} at epoch $n-1$, consisting of the weighted sum of inverse variances, is also computed using L_n. r_N and U_N at the final epoch N, from which the Kalman Smoother starts, are equal to zero (Durbin and Koopman, 2012). r_{n-1} is determined using equation 15. U_{n-1} can be computed using equation 16.

Weighted sum of prediction errors and the weighted sum of inverse variances

$$r_{n-1} = H_n^\top F_n^{-1} v_n + L_n^\top r_n \tag{15}$$

$$U_{n-1} = H_n^\top F_n^{-1} H_n + L_n^\top U_n L_n \tag{16}$$

r_n and U_n are used to compute the smoothed states $\hat{x}_{n-1}$ and their error variance-covariance matrix M_{n-1} with equations 17 and 18, respectively.

Smoothed states and their error variance-covariance matrix

$$\hat{x}_n = \tilde{x}_n + P_n r_{n-1} \tag{17}$$

$$M_n = P_n - P_n U_{n-1} P_n \tag{18}$$

Using the smoothed states $\hat{x}_n$ and the maximum likelihood estimates for the parameters $\hat{\theta}$, the smoothed observed states can be computed using the observation equation (1).

4.3 Frequentist diagnostics

To determine whether the addition of a new model component actually improves the model, diagnostics and model selection criteria are viewed in this thesis. Diagnostics are used to indicate whether or not the model is consistent with the data. Diagnostic metrics are quantitative measures of model performance. Conversely, diagnostic figures are qualitative, visual indicators of how well the model performs. There may be multiple models models that are consistent with the data. Model selection criteria help to find which of these models is best.

4.3.1 Akaike and Bayesian information criteria

Model selection criteria include the Akaike and Bayesian information criteria (AIC and BIC). The AIC and BIC scores give a model's relative measure of goodness of fit, while penalizing for increased model complexity. This penalty for adding more model components prevents choosing models that overfit the data. When comparing two models, the model with lower AIC and BIC scores is considered the better model. A better model in this case means the increase in goodness of fit is greater than the increase in complexity. The AIC and BIC are computed using equations 19 and 20 below.

Akaike and Bayesian Information Criteria

$$AIC = 2k - 2\ln(P(y_{1:N}|\hat{\theta})) \tag{19}$$

$$BIC = k\ln(N) - 2\ln(P(y_{1:N}|\hat{\theta})) \tag{20}$$

In equations 19 and 20 above, k is the number of model parameters (i.e., the length of θ). N is the length of the total data record $y_{1:N} = y_1, ..., y_N$ and $P(y|\hat{\theta})$ is the maximized likelihood (i.e., the likelihood given the maximum likelihood estimates $\hat{\theta}$). The AIC penalizes increased model complexity based only on the number of parameters in the model. The BIC on the other hand, penalizes increased model complexity based on both the number of parameters in the model, as well as the length of the data record that is being modelled. As a result, the BIC generally penalizes increased model complexity more severely than the AIC.

4.3.2 Ljung-Box, Jarque-Bera and heteroskedasticity tests

The prediction errors (innovations) v_n, computed with equation 7, are normally and independently distributed if the model is correct. For the linear Gaussian state-space models used in this thesis, the prediction errors should follow a zero-mean Gaussian distribution. Hence, the standardized prediction errors should be $NID(0, 1)$ if the model is correct. This characteristic of the standardized prediction errors makes them ideal to use when computing diagnostic metrics and creating diagnostic figures.

For each model, the Ljung-Box test, Jarque-Bera test and a test on heteroskedasticity are performed. The Ljung-Box test checks whether or not the standardized prediction errors of the model contain significant autocorrelation. In this thesis, the Ljung-Box test is performed for the first 10 lags. The Jarque-Bera test checks if the standardized prediction errors of the model follow a standard Normal distribution. Finally, the heteroskedasticity test checks whether or not the variance of all standardized prediction errors is the same. The test statistics of these tests are computed, alongside their p-values. The p-value is the probability of finding a statistic equal to the calculated statistic, under the null hypothesis that there is no autocorrelation in the residuals (Ljung-Box test), the residuals are normally distributed (Jarque-Bera test) and the residuals are homoskedastic, meaning they have constant variance (heteroskedasticity test). For each test, the corresponding p-value should ideally be greater than 0.05.

4.3.3 Diagnostic figures

Several diagnostic figures are made, which are visual indicators of how well the model performs. To check if the standardized prediction errors are $NID(0, 1)$, a histogram is created of the standardized prediction errors alongside their kernel density estimate and a $N(0, 1)$ density for reference. A QQ plot for the standardized prediction errors is also made. If the dots in the QQ plot are all on the diagonal, the standardized prediction errors have a perfect $N(0, 1)$ distribution. Another diagnostic plot is the correlogram, which can be used to asses whether the prediction error sequence is autocorrelated.

Lomb-Scargle periodograms of the prediction errors are created for each model. Periodograms show whether a cycle component can be added to the model and indicate the frequency of these cycles. Periodograms are plotted up to the Nyquist frequency, which is half the sampling frequency. As both the GRACE and altimetry datasets contain monthly data, the sampling frequency is equal to 12 cycles/year. Thus, the Nyquist frequency is 6 cycles/year.

Confidence intervals can be computed for peaks in a periodogram under the null hypothesis that there is no peak. If a significant annual cycle is present in the prediction errors, for example, a large peak in the periodogram can be seen at a frequency of 1 cycle/year. If this peak is outside of the 95% confidence interval, for example, the null hypothesis can be rejected with 95% certainty. This is an indication that the model could be improved by adding an annual cycle component.

If the periodogram has a large peak at a frequency that has a physical meaning, a cycle component with that frequency can be added. The aforementioned peak at a frequency of 1 cycle/year, indicating an annual cycle, is an example of this. If a peak is found at a frequency that is not related to some known physical process or is surrounded by several other peaks, which could indicate spectral leakage, a cycle component of unknown frequency can be added to the model. The frequency of the cycle then becomes an additional model parameter.

If a periodogram contains multiple peaks, cycle components should be added one-by-one, starting with the largest peak. This is because the addition of one cycle component could remove multiple peaks in the periodogram. This can happen due to spectral leakage, which may bias frequencies and generate non-physical peaks in the periodogram. Peaks at frequencies close to 0 can be disregarded. These represent slowly varying signals (i.e., a trend), which may point to a suboptimal trend model. Such signals may bias frequencies of periodic signal constituents due to spectral leakage. This is also a reason why periodograms are made for the prediction errors of a model instead of for the original data itself. A periodogram of data that still contain a trend (i.e., data that is not detrended), would show very dominant peaks at small frequencies and may, as a result, not clearly show the peaks related to cycles in the data.

To further assess a state-space model, a correlation matrix is generated. This matrix indicates how large the correlation is between different model states. If different states are strongly correlated, this may point to a problem with the identifiability of the model. The phase offsets of cycle components are plotted as well. Ideally, these phase offsets vary slowly over time. The correlation matrix and phase offsets are, strictly speaking, not diagnostic figures, as they are not related to the (standardized) prediction errors. Nevertheless, they are important to check, as they may indicate something has gone wrong in the model.

4.4 Model selection process

The diagnostics explained in section 4.3 are used to determine if a model fits the data well. Model selection is an iterative process where state-space models with different model components are generated and fitted to the data. The goodness of fit of these different models are compared to each other, while accounting for model complexity. The models are fitted to the corrected ice mass and ice volume data (black lines) shown in figures 2 and 3, respectively. Model selection is done separately per basin, for both the ice mass and ice volume time series.

To start, integrated random walks are used to model slow variations in the ice-dynamical mass and volume data for both basins 6, 13 and 22. Integrated random walk models are smooth dynamic linear trend models, containing a static level with a dynamic slope, in addition to an irregular component to account for noise in the data. This integrated random walk model provides a baseline to compare to other models of greater complexity. If models with greater complexity (i.e., more components) have greater p-values for the Ljung-Box, Jarque-Bera and heteroskedasticity tests and lower AIC and BIC scores than the integrated random walk model, the models of greater complexity are considered an improvement. Of the diagnostic tests, the Ljung-Box test is considered especially important, as autocorrelation in the prediction errors is unwanted. As noted in section 4.3, a model is considered to perform well for a diagnostic test if the corresponding p-value is greater than 0.05.

A first step towards improving the integrated random walk models is to compute the Lomb-Scargle periodogram of the prediction errors, as explained in section 4.3. If they contain a statistically significant peak, a cycle component is added to the model. As noted there, it is important that cycles are added one-by-one, to avoid adding non-physical cycles to the model. It is also important to check the amplitude of an added cycle component. If this amplitude is very small relative to its uncertainty (i.e., confidence interval), the addition of a cycle component may not be a significant contributor to the observed states. After adding a cycle, the Ljung-Box, Jarque-Bera and heteroskedasticity tests should be checked to determine if the new model with the added cycle is consistent with the data. If so, the AIC and BIC model selection criteria should be compared to the earlier model without the cycle to see whether the addition of the cycle component has actually improved the model.

After all relevant cycles in the data are added to the model, autoregressive components may be added, representing either correlated observation noise or unmodelled slowly-varying signal. However, the integrated random walk is already expected to represent all slowly varying signal. If the autoregressor also captures slowly varying signals, which the integrated random walk is supposed to represent, this is unwanted. In such cases, the autoregressor might capture some of the slowly varying signals meant for the integrated random walk, causing difficulty in distinguishing between the two. Therefore, adding an autoregressor to the model should be a last resort, used only when strong autocorrelations persist in the standardized prediction errors and no other components (such as cycles or regressors) remain to be added to the model in order to reduce these autocorrelations.

When adding an autoregressor to the model, the autoregressive coefficients become additional model parameters. It is imperative that autoregressive coefficients does not approach 1, as this means that the autoregressor approaches a random walk. This is unwanted, as no distinction can be made between signal and noise for a random walk component. Again, the AIC, BIC and diagnostic tests should be checked and compared to earlier models, to determine whether the addition of an autoregressive component actually improved the model.

5 Bayesian approach

The maximum likelihood estimates for the model parameters found when following the Frequentist approach, are considered to be deterministic. Subsequently, these deterministic parameter estimates are used to estimate the states. This means that potential stochasticity of model parameters is not accounted for, which could lead to overconfident small uncertainties when following the Frequentist approach.

To account for stochasticity of the estimated model parameters, the Bayesian approach is followed. As noted in section 1.2, the marginal posterior distribution for parameters $P(\theta|y)$ can be estimated using MCMC methods when following the Bayesian approach. Samples from $P(\theta|y)$ are used in the simulation smoother to obtain the conditional distribution of the states at each epoch given the data and parameters $P(x_{1:N}|\theta, y)$. Because samples of the parameters from $P(\theta|y)$ are used to estimate $P(x_{1:N}|\theta, y)$, stochasticity of the estimated parameters is taken into account when estimating the states in the Bayesian approach. Hence, it is expected that the uncertainties of the states found by the Bayesian approach are larger, but more realistic, compared to those of the Frequentist approach.

As noted in section 1.2, many different MCMC samplers exist. The specific MCMC sampler used in this thesis is the Hamiltonian No-U-Turn sampler. Section 5.1 explains the concept of Hamiltonian MCMC methods and the No-U-Turn sampler's implementation of them. The simulation smoother is explained in section 5.2. Afterwards, section 5.3 describes several sampling diagnostics that can be used to check if the sampling has performed well. Posterior predictive checks, used to determine how well a Bayesian model fits to the data, are explained in section 5.4. Finally, informative and non-informative priors are discussed in 5.5.

5.1 Hamiltonian Markov Chain Monte Carlo

As noted in section 1.2, many different MCMC methods exist. For models with a relatively large amount of parameters, Hamiltonian MCMC samplers are considered the state of the art (Hogg and Foreman-Mackey, 2018). Inherent to the Bayesian approach, is the use of priors. Starting with initial values taken from prior distributions, the parameter space is explored by generating candidate values. These candidate values are checked against the Metropolis-Hastings acceptance criterion. If accepted, the candidate value is added to the list of samples. If rejected, the candidate value is thrown out and the most recently accepted value is repeated in the list of samples. This updating of the sample list is what creates the Markov Chain in MCMC samplers: a sample only depends on the previous sample. By generating a large list of samples, an estimation of a model parameter's marginal distribution given the data $P(\theta|y)$ is obtained.

While Hamiltonian methods use the Metropolis-Hastings acceptance criterion to determine the acceptance of candidate values, they differ from the Metropolis-Hastings approach in the way the candidate values themselves are generated. The Metropolis-Hastings algorithm randomly samples candidate values from a proposal distribution. Hamiltonian samplers instead use Hamiltonian dynamics to compute new candidate values (Betancourt, 2018; Brooks et al., 2011). This process is explained further in appendix C.

The candidate values for θ generated by Hamiltonian MCMC samplers are evaluated using the Metropolis-Hastings acceptance criterion (Hogg and Foreman-Mackey, 2018). The Metropolis-Hastings acceptance criterion determines whether candidate values θ_{new} are accepted over the previous accepted value θ_{old}, starting from the initial priors. The acceptance ratio α is determined using equation 21.

Metropolis-Hastings acceptance ratio

$$\alpha = \frac{f(\theta_{new})}{f(\theta_{old})} \tag{21}$$

In equation 21, f is a function proportional to the probability density function of the distribution that is sampled. Hence, the acceptance ratio α is equal to the ratio of the probability densities with the new candidate value θ_{new} and with the previously accepted value θ_{old}. To decide whether or not the candidate value is accepted, a random value is drawn from a uniform distribution between 0 and 1. If this value is smaller than or equal to α, the candidate value is accepted, θ_{new} replaces θ_{old} and is added to the list of samples. If the uniform random value is greater than α, the candidate value is rejected and θ_{old} is repeated in the list of samples.

α would always be equal to 1 if the Hamiltonian dynamics in the MCMC run were to be solved analytically (Hoffman and Gelman, 2011). This would lead to every candidate value being accepted. However, solving Hamilton's equations analytically is rarely possible, which is why Hamilton's equations are solved numerically in Hamiltonian MCMC methods. Errors introduced by numerical approximation cause α to be smaller than 1, meaning not all candidate values will be accepted in practice.

Regardless, Hamiltonian MCMC samplers explore the parameter space more efficiently than other MCMC methods that randomly generate candidate values from a proposal distribution, such as the Metropolis-Hastings algorithm (Brooks et al., 2011). That is, Hamiltonian methods generate candidate values that are more likely to be accepted than those generated by the Metropolis-Hastings algorithm. Hamiltonian MCMC samplers especially perform better than samplers like Metropolis-Hastings when sampling from complex distributions (Hoffman and Gelman, 2011; Hogg and Foreman-Mackey, 2018).

5.1.1 No-U-Turn sampler

In this thesis, the No-U-Turn sampler is used to estimate the marginal distribution of the model parameters given the data $P(\theta|y)$. The No-U-Turn sampler is considered the state of the art implementation of Hamiltonian MCMC (Hoffman and Gelman, 2011; Hogg and Foreman-Mackey, 2018).

To numerically integrate Hamilton's equations, most Hamiltonian samplers use a step size Δt and a number of steps m that have to be specified by the user (Hoffman and Gelman, 2011). Every step Δt, Hamilton's equations are numerically integrated. A new candidate value for θ is then generated every m steps of size Δt. Notably, the numerical integration is not done with larger steps of size $m\Delta t$ to instantly arrive at new candidate values after every iteration of the numerical integration. If this were done, the errors caused by the numerical approximation would be significantly larger than when smaller steps of Δt are used every iteration of the numerical approximation.

Tuning Δt and m is important to ensure the sampler performs well. If step size Δt is too small, computational efficiency suffers and consecutive samples may be correlated. If the number of steps m is too large, Hamiltonian samplers may loop back to the first sample and retrace its steps (making a U-turn), unnecessarily increasing computational cost.

To tune Δt and m, analysts usually perform several preliminary Hamiltonian MCMC runs. After each run, they look at diagnostics, such as if there is unwanted autocorrelation between samples, to determine if Δt and m have to be changed. These sampling diagnostics are explored in section 5.3. The need to tune the step size Δt and the number of steps m may make Hamiltonian MCMC difficult to use effectively for those without expertise (Hoffman and Gelman, 2011).

To avoid potential issues related to fine tuning Δt and m and avoid the sampler looping back to the first sample and retrace its steps, the Hamiltonian No-U-Turn sampler is used in this thesis. The No-U-Turn sampler does not require the user to specify the step size Δt and number of steps m. Instead, they are automatically chosen per iteration. This means the user of the No-U-Turn sampler does not need to manually tune the sampler at all. Δt and m are chosen in such a way that the sampler will never loop back to the first sample and retrace its steps. Hence the name of the No-U-Turn sampler: no U-turns are made during sampling.

In this thesis, the *PyMC* Python package is used for the Bayesian approach to state-space modelling (Grabowski, 2023). The *PyMC* implementation of the No-U-Turn sampler for state-space models is used to obtain the marginal distributions of the model parameters given the data $P(\theta|y)$.

5.2 Simulation smoother

After estimating the marginal distribution of the parameters given the data $P(\theta|y)$ using the No-U-Turn sampler, the simulation smoother is used to estimate the conditional distribution of the states given the parameters and data $P(x_{1:N}|\theta, y)$ for each epoch n. The simulation smoother from the *Statsmodels* Python package is used in this thesis.

The simulation smoother is sometimes called the forward filtering and backward sampling method, referring to its use of the Kalman Filter and subsequent backward sampling. Note that while the simulation smoother does make use of the Kalman Filter, it does not use the Kalman Smoother to generate estimates of the states at every epoch, as is done in the Frequentist approach. Instead, after the Kalman Filter is used, backward sampling is done. This entails sampling the states at each epoch from the conditional distribution of the states given the parameters and data $P(x_{1:N}|\theta, y)$. This is done, starting from final epoch N and going back to the first epoch, hence being referred to as 'backwards' sampling.

Recall that the No-U-Turn sampler provides an estimation of the model parameters' marginal distribution given the data $P(\theta|y)$. Multiple sets of parameters θ are sampled from $P(\theta|y)$ and used in different iterations of the simulation smoother to generate samples of the states x_n at every epoch. Because multiple different samples of θ are used to generate the samples of x_n, stochasticity of model parameters is accounted for when sampling the states when using the simulation smoother. As a result, the uncertainties in the observed states of the Bayesian approach to state-space modelling are larger than those of the Frequentist approach. In chapter 7, the difference in the uncertainties of the trends found by both approaches is quantified.

5.3 Sampling diagnostics

To evaluate how well a MCMC sampling has performed, sampling diagnostics are used. These are diagnostic figures including trace plots, rank plots, plots of the estimated model parameters' marginal densities, autocorrelation plots, plots of the effective sample size and energy plots. It can never be said for certain that a MCMC sampling provides reliable results, as MCMC samplers only find the parameters' true distributions when an infinite amount of samples are taken (Hogg and Foreman-Mackey, 2018). Nevertheless, the sampling diagnostics may be used to identify if something went wrong during the sampling. The sampling diagnostics are generated using the *ArviZ* Python package (Kumar et al., 2019).

5.3.1 Trace

Usually, multiple Markov chains are ran simultaneously when doing MCMC sampling (Hogg and Foreman-Mackey, 2018). These chains independently explore the parameter space. Trace plots are used to evaluate how well an MCMC chain is mixed. A chain is mixed well when it does not reject or accept too many candidate values in a row. When a disproportionate amount of candidate values are rejected, many iterations are required to generate enough samples to estimate the parameter's distribution. On the other hand, if nearly all candidate values are accepted, barely any values are repeated in the list of samples and again, no good estimate of the parameter's distribution can be obtained.

Usually, it takes some iterations for a chain to converge to conditions where it mixes well. Because of this, a burn-in period is often used. A burn-in period refers to a number of samples at the start of a chain that are discarded. Only after the burn-in period, when the chain has hopefully converged to conditions where it mixes well, are samples actually recorded. Determining the required length of a burn-in period requires some preliminary MCMC runs, where the trace plots are examined to determine if the chains are mixed well.

Trace plots show the iteration number on the horizontal axis and the sampled parameter value on the vertical axis. Long horizontal parts in a trace plot, where the chain stays at the same value, should be avoided. This indicates too many candidate values are rejected in a row. Additionally, a large amount of consecutive samples in one direction indicate that too many consecutive values are accepted, which should also be avoided. To solve such issues in trace plots, a longer burn-in period or different priors can be chosen.

Figure 4 shows an example of a trace plot. This figure shows the trace plot of the irregular variance parameter for the chosen ice mass model of basin 6. For this trace plot, one chains of 1000 samples is shown. The trace plot shows that the chain is mixed well, as there are no long horizontal parts and no trend can be distinguished in the trace plot.

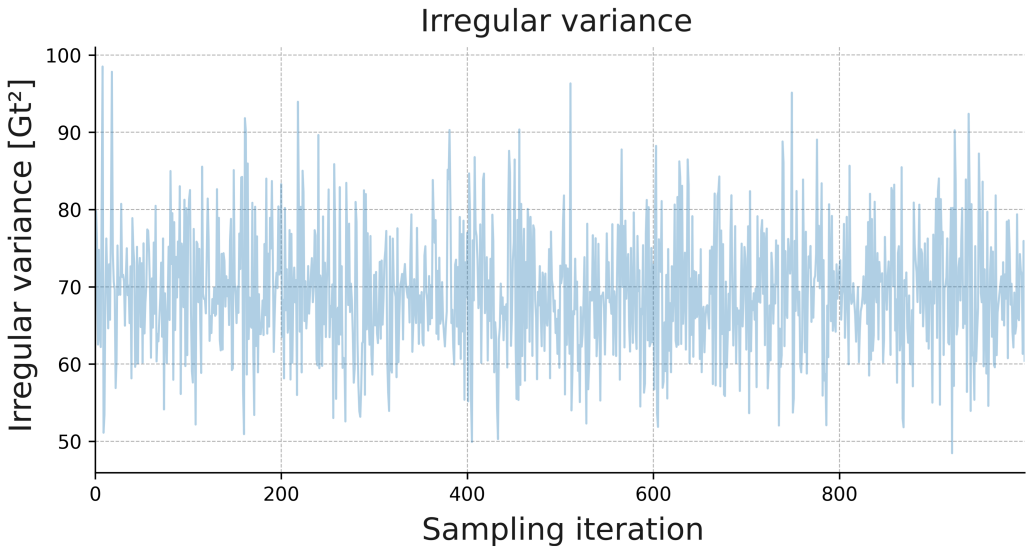


Figure 4: Trace plot for the sampling of the irregular variance parameter of the ice mass model for basin 6, showing one chains of 1000 samples.

5.3.2 Rank

Rank plots have similar use to trace plots. That is, to check if MCMC chains are mixed well. There are situations where trace plots fail to properly convey improper mixing of a chain. For example, this can happen when sampling from a multi-modal distribution, and two chains stay within different high probability regions of

the distribution. The chains may look to be mixed properly in a trace plot, but because they don't transition between different modes of the distribution that is being sampled, they actually are not. Hence, it is important to examine both trace and rank plots to see if chains are properly mixed.

Rank plots were introduced by Vehtari et al. (2021). For each chain and each parameter that is being sampled, a rank plot displays a histogram of the chain's samples, ranked over all chains. This means that the samples of all chains are put together, and the sample with the smallest value gets rank 1 while the largest sample gets the maximum rank, which is equal to the total number of samples of all chains together. After ranking all samples, separate histograms are made for each chain of the ranked samples belonging to each chain. If the chains sample from the same distribution, each chain's samples should be similar to each other. Thus, when ranking the samples of all chains together, it should not be the case that many successive ranked samples come from a single chain.

If all chains have sampled from the same distribution, the rank histograms of each chain should be uniform. If one chain has sampled from a different distribution than the others, the rank of that chain will deviate from uniformity. In the example of two chains being stuck in different modes of a multi-modal distribution when sampling, this will result in rank plots that are not uniform. Chains are mixed well if the rank plots of all chains look uniform.

Figure 5 shows an example of a rank plot. This figure shows the rank plot of the irregular variance parameter for the ice mass model of basin 6. For this rank plot, 4 chains of 1000 samples are shown. It can be seen that the rank histograms for the different individual chains are relatively uniform, indicating decent mixing of the chains.

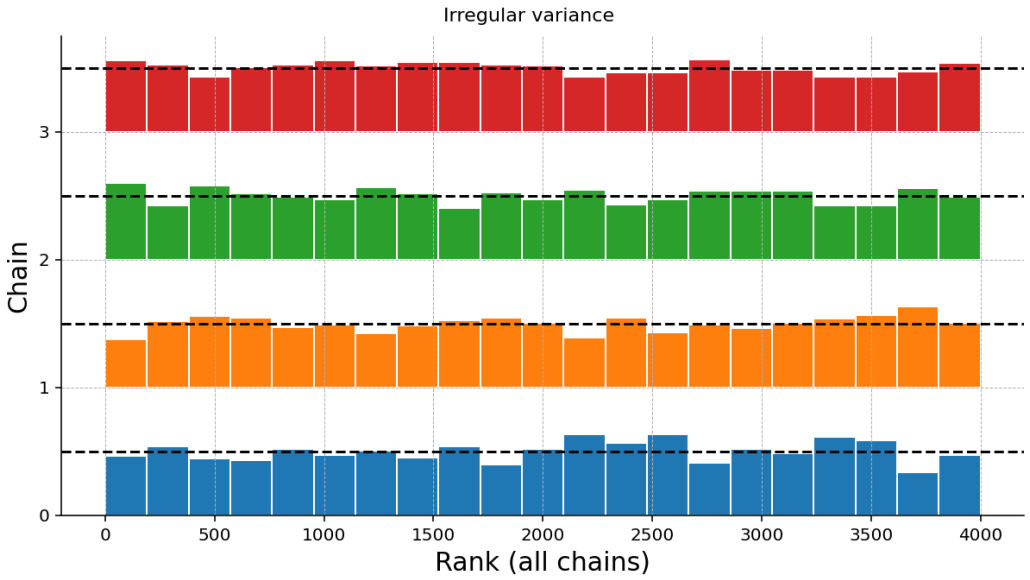


Figure 5: Rank plot for the irregular variance for the ice mass model of basin 6, sampled with 4 chains of 1000 samples each.

5.3.3 Parameters' marginal posterior distributions

Plots of the model parameters' estimated marginal probability distributions $P(\theta|y)$ are important figures to check whether the MCMC sampling has performed well. Such plots are made by either plotting histograms of the samples from $P(\theta|y)$ or by plotting their kernel density estimate. The estimated marginal posterior distributions can be plotted for each parameter that is being sampled. They can be made separately per chain, which is an additional way to check if the chains have sampled from the same distribution. Alternatively, posterior density plots can be made by combining all samples of different chains. If the chains are mixed well, this will lead to a more accurate estimate, as more samples are used in the estimation.

If analysts already have an idea of how the model parameter's distribution that is being sampling should look, plots of the estimated posterior density can easily be used to confirm if sampling went well. If the estimated density looks as expected, the sampling has probably gone well. In most cases, however, a Bayesian analysis is done precisely because it is, in fact, not known how the posterior density that is sampled from looks. In such cases, the estimated posterior density plots give information about the stochasticity of the model parameters. Statistics such as the mean or mode, can be retrieved from the parameters' estimated posterior distributions.

Figure 6 shows an example of an estimated marginal posterior distribution. This figure shows the kernel density estimate of the posterior density of the irregular variance for the ice mass model of basin 6, while indicating its mode and 95% highest density interval.

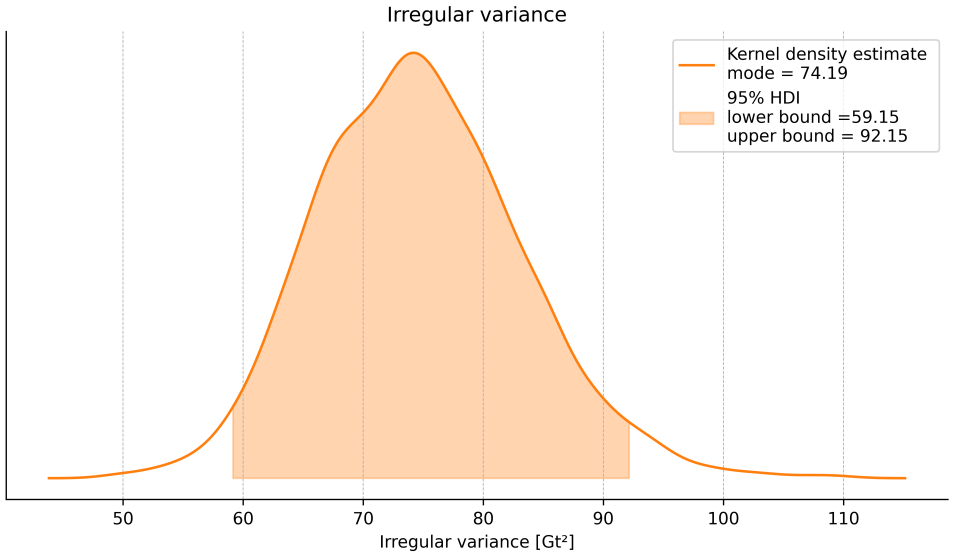


Figure 6: Estimated marginal posterior distribution of the irregular variance for the ice mass model of basin 6, indicating the mode and the 95% highest density interval.

5.3.4 Autocorrelation

Autocorrelation plots are diagnostic figures that show correlation between successive samples. They are made for each parameter that is being sampled. Autocorrelation in successive MCMC samples is unwanted, as this means these samples are not independent. Autocorrelated samples slow down convergence to good mixing conditions. Autocorrelation plots can be made separately per chain. Alternatively, a single combined autocorrelation plot can be made for each parameter, by combining samples of all chains into one single chain.

Figure 7 shows an example of an autocorrelation plot. This figure shows 10 lags of the autocorrelation of the irregular variance parameter samples for the ice mass model of basin 6. It shows that, after lag 2, no significant autocorrelation remains between the samples.

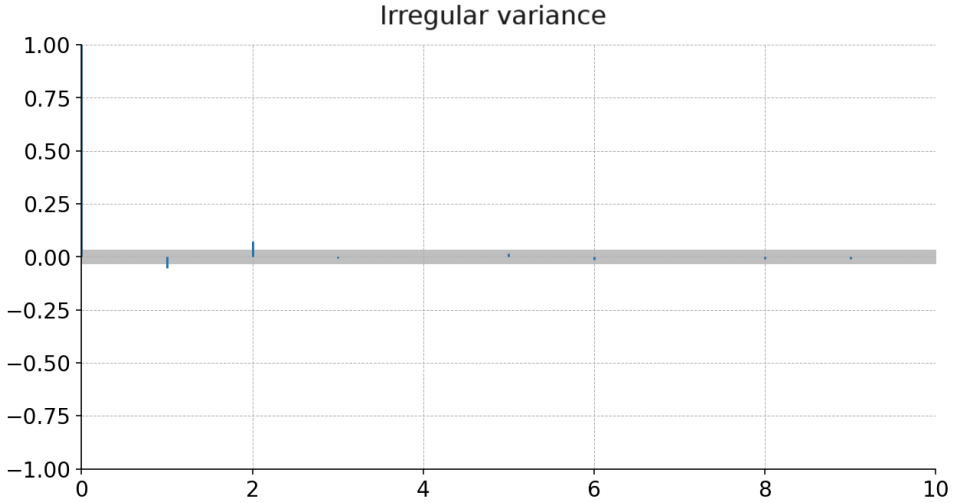


Figure 7: Autocorrelation plot (10 lags) for the irregular variance parameter samples of the ice mass model for basin 6.

If too many autocorrelated samples are present in a MCMC chain, thinning can be performed. Thinning refers to the process of removing samples at a certain interval. As a result, the number of samples is reduced. When thinning, longer MCMC runs would have to be run to obtain a sufficient amount of independent samples.

Thinning an MCMC chain is computationally inefficient, as many generated samples are simply thrown out. Additionally, thinning greatly reduces the precision of the estimated marginal posterior distribution for the parameters (Link and Eaton, 2012). Hence, it is generally not recommended to thin MCMC chains, even if there are autocorrelated samples.

Significant autocorrelation in MCMC chains could indicate that the model is not suitable for the data. Thus, another way of dealing with a significant amount of autocorrelated samples in a chain is to make changes to the model and start another MCMC run.

5.3.5 Effective sample size

The effective sample size is another important sampling diagnostic when working with MCMC methods. Like autocorrelation plots, plots of the effective sample size give information about the independence of samples in a MCMC chain, albeit in a different way. The effective sample size refers to the amount of independent samples that would convey the same amount of information about the sampled distribution as the amount of samples in an MCMC chain (Roy, 2020). The effective sample size can differ between model parameters. This is because different model parameters have different distributions. The more complex a parameter's distribution, the smaller the effective sample size will be. That is, more independent samples are necessary to explore the full parameter space.

Ideally, the effective sample size would be equal to the amount of samples in a chain. Due to autocorrelation in a MCMC chain, the effective size is generally smaller than the sample size of a MCMC chain. In rare situations, where negative autocorrelation is present in a MCMC chain, the effective sample size can be greater than the sample size of a MCMC chain. While this technically seems like a good thing, negative autocorrelation is likely to cause bad mixing of a chain, and is thus not desirable.

When computing the effective sample size, a distinction is usually made between the bulk and the tail effective sample size (Vehtari et al., 2021). The bulk effective sample size refers to the number of independent samples obtained from high-likelihood regions of the distribution that is being sampled, i.e., the bulk of the distribution. This is defined as the samples within the 95% highest density interval (Vehtari et al., 2021). Conversely, the tail effective sample size refers to the number of independent samples obtained from the low-likelihood regions of the distribution that is being sampled, i.e., the tail of the distribution. This is defined as the samples outside of the 95% highest density interval (Vehtari et al., 2021). High bulk and tail effective sample sizes relative to the number of samples in a chain, indicate that the chain efficiently explores the high-likelihood and low-likelihood regions of the sampled distribution, respectively.

Figure 8 shows an example of an effective sample size plot. This figure shows the effective sample size for the irregular variance parameter of the ice mass model of basin 6, indicating both the bulk and tail effective sample sizes.

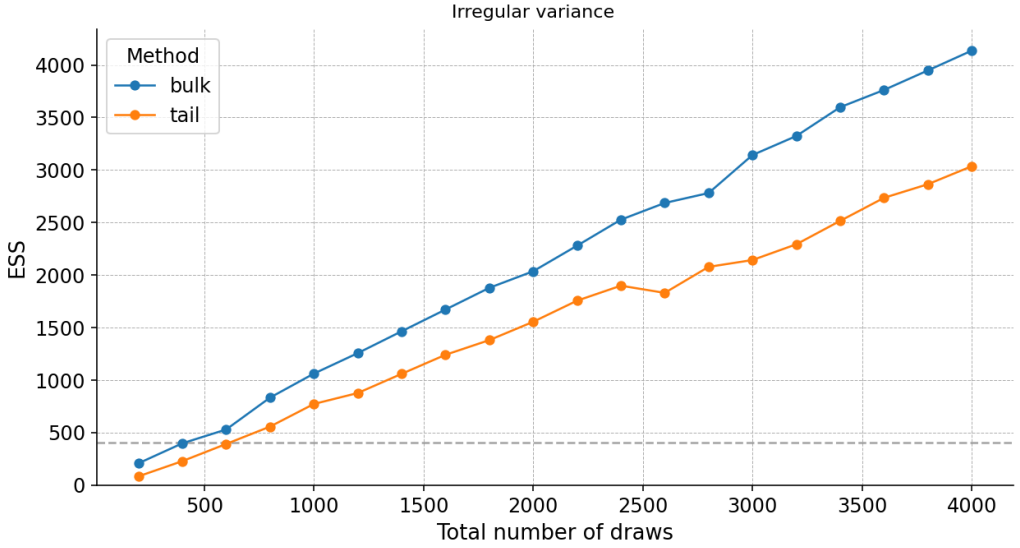


Figure 8: Plot of the effective sample size of the irregular variance parameter of the ice mass model for basin 6.

5.3.6 Energy

Energy plots are diagnostic plots that show the energy transition distribution and marginal energy distribution of Hamiltonian MCMC algorithms (Betancourt, 2018). Energy plots aid in determining whether exploration of the parameter space by Hamiltonian MCMC methods, such as the No-U-Turn sampler, has gone well. Energy plots consist of two parts: the energy transition and the marginal energy.

In Hamiltonian dynamics, which are used in Hamiltonian MCMC to compute candidate values, the Hamiltonian represents the total energy, defined as the sum of the potential and kinetic energies. In Hamiltonian MCMC, the potential energy is equal to minus the natural logarithm of marginal posterior for parameters $P(\theta|y)$ and the kinetic energy corresponds to minus the natural logarithm of the probability density function of a standard normal distribution, as explained in appendix C.

The energy transition distribution is the distribution of the Hamiltonian for a constant potential energy and a changing kinetic energy. The marginal energy distribution is the distribution of the Hamiltonian when both the potential and the kinetic energy are updated with iterations.

If the marginal distribution is significantly more wide than the energy transition distribution, this indicates the Hamiltonian MCMC sampler will take a long time to explore the parameter space. Ideally, the energy transition and marginal energy distributions are well-matched. If so, the Hamiltonian MCMC sampler will efficiently explore the parameter space.

To quantify the mismatch between the energy transition and marginal energy distribution, the Bayesian fraction of missing information is computed. For a perfect match between the energy transition and marginal energy distribution, a Bayesian fraction of missing information of 1 is found. Empirically, the Bayesian fraction of missing information should at least be greater than 0.3 for the Hamiltonian MCMC run to have performed well (Betancourt, 2018).

Figure 9 shows an example of an energy plot. This figure shows the energy plot of the MCMC run for the ice mass model of basin 6. It shows that the marginal energy distribution and the energy transition distribution are quite well-matched, indicating the Hamiltonian MCMC sampler was able to efficiently explore the parameter space. Additionally, the Bayesian fraction of missing information for the 4 chains that were ran is given. These are consistently close to 1, indicating the MCMC run performed well.

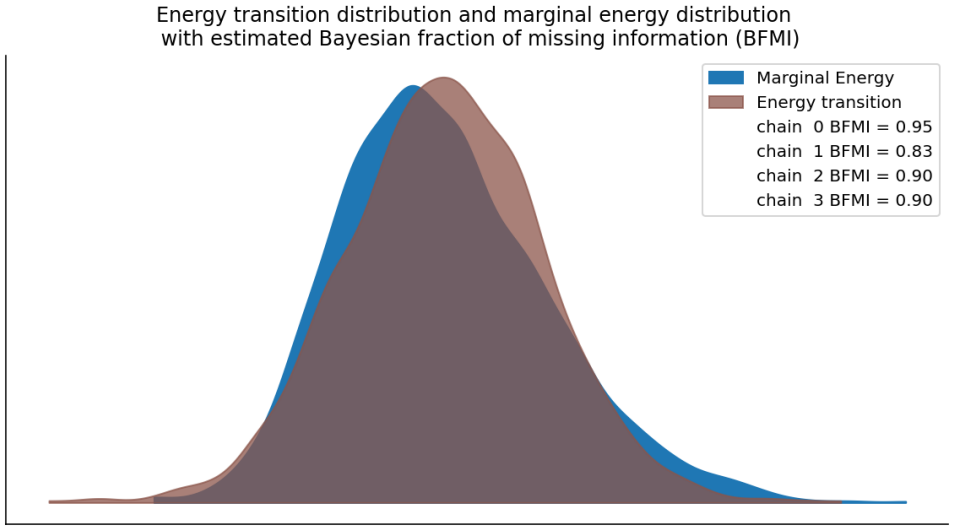


Figure 9: Energy plot of an MCMC run with 4 chains for the ice mass model of basin 6.

5.4 Posterior predictive checks

It is possible for a model that is not appropriate for the data being approximated, to still converge to a solution when MCMC sampling (Klees, 2023). For example, a state-space model that contains a cycle component, when no actual cycle is present in the data, might still converge to a solution. In such cases, the diagnostics described in section 5.3 may not indicate anything has gone wrong. To check the suitability of a model, posterior predictive checks are used.

For posterior predictive checks, random sets of parameters are sampled from the estimated marginal posterior distribution for parameters $P(\theta|y)$ and used in the simulation smoother to estimate the conditional distribution of the states at each epoch given the parameters and the data $P(x_n|\theta, y)$. The sampled parameters and states

are then used in the observation equation (1) to predict the observations. These predicted observations, referred to as the observed states, should fit well to the actual data.

Three distinct types of posterior predictive checks exist, which differ in the data they use. These posterior predictive checks are referred to as the posterior predicted, posterior filtered and posterior smoothed checks. This is similar to the Kalman Prediction, Filtering and Smoothing explained in section 4.2. Posterior predicted checks use data up to the previous epoch. Data up to the current epoch are used for posterior filtered checks. Finally, smoothed posterior checks use the full data record. Below, the types of data used by the different posterior predictive checks are listed.

- Posterior predicted: $y_{1:n-1}$
- Posterior filtered: $y_{1:n}$
- Posterior smoothed: $y_{1:N}$

Where N is the total amount of epochs $n = 1, 2, 3, ..., N$.

To compute the observed states when following the Bayesian approach, the smoothed posterior checks are used, as these contain the most data information. If there are no systematic discrepancies between the data and the smoothed posterior checks, it can be inferred that the model is appropriate and the MCMC sampling has gone well.

5.5 Informative and non-informative priors

When fitting state-space models to data following the Bayesian approach, priors are needed. Priors are used by MCMC samplers to obtain estimates of the model parameters' marginal distributions given the data $P(\theta|y)$ according to Bayes' theorem.

Priors encode known information about the distribution that is being sampled, before any data are available. For instance, a prior can be constraining a parameter to only take on non-negative values, if it is known that the parameter can not be negative in reality. This is the case for irregular and disturbance variance parameters, for example. A distinction is made between two broad types of priors: informative and non-informative priors.

Non-informative priors contain very limited information about the parameter that is being sampled. If a parameter is likely to be in a range between two values $[a, b]$, a non-informative prior distribution could be a Uniform distribution between a and b, for example. If the amount of data is large relative to the model complexity and the data is of high quality, non-informative priors may be sufficient for the MCMC sampler to estimate a parameter's distribution (Klees, 2023). However, if this is not the case, the use of non-informative priors could lead to convergence issues. The use of informative priors may avoid this.

Informative priors introduce more stringent constraints on the parameters that are being sampled than non-informative priors, thereby providing more a priori information about model parameters. Using the example of a parameter that is likely to be in a range $[a, b]$, an example of an informative prior is a Gaussian distribution with mean $\frac{(a+b)}{2}$ and standard deviation $\frac{(b-a)}{2}$. This choice would leave just a 0.32 probability that the parameter is outside of the range $[a, b]$ (Klees, 2023). For parameters with only positive values, a recommended prior is a Half Student's T-distribution with 3 or 4 degrees of freedom and a standard deviation that is likely to be larger than the true standard deviation of the parameter's distribution. In this thesis, informative priors will be used when fitting the models to the data following the Bayesian approach. Section 6.3 discusses the chosen priors for the mass and volume models described in sections 6.1 and 6.2, respectively.

6 Model selection

To model ice-dynamical mass and volume changes, the corrected GRACE and altimetry data are used, respectively. These are the black lines in figures 2 and 3. The models that best fit these data, are chosen following the Frequentist model selection process described in section 4.4. The chosen ice mass models for basins 6, 13 and 22 are discussed in section 6.1 and the chosen ice volume models for basins 6, 13 and 22 are discussed in section 6.2. Priors and sampling settings used to fit these models when following the Bayesian approach are explained in sections 6.3 and 6.4, respectively.

6.1 Ice mass models

The chosen ice mass models for basins 6, 13 and 22 have the same four components. These are a static level, dynamic slope, an annual cycle and a cycle component with a period of 0.44 years. The cycle component with a period of 0.44 years represents a solar annual (SA) tidal alias. As the level is static, it does not have a disturbance variance. The slope is dynamic, so there is a disturbance variance for the slope. For basins 6 and 22, the cycles are static, whereas the cycles are dynamic for basin 13. The observation equations of the ice mass models contain irregular terms ϵ_n, meant to represent noise in the data.

6.1.1 Addition of cycles

Lomb-Scargle periodograms of the prediction errors of integrated random walk models, indicate two clear cycles in the ice mass time series of basins 6, 13 and 22. These are the aforementioned annual cycle and the SA tidal alias cycle with a period of 0.44 years. As an example, the Lomb-Scargle periodogram on the prediction errors of an integrated random walk ice mass model for basin 6 is shown below in figure 10. Two dominant peaks in the periodogram can be found at frequencies of 1 and $\frac{1}{0.44} = 2.27$ cycles/year. These represent the annual and SA tidal alias cycles, respectively.

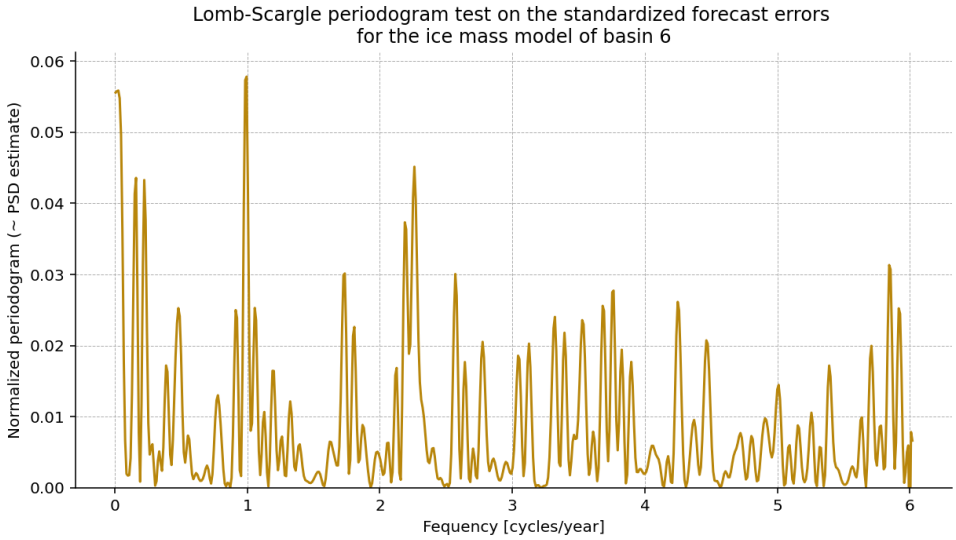


Figure 10: Lomb-Scargle periodogram test on the prediction errors of an integrated random walk model for the ice-dynamical ice mass of basin 6, showing two clear peaks at frequencies of 1 and 2.27 cycles/year, respectively representing the annual and SA tidal alias cycles. Similar results are found for basins 13 and 22.

The addition of cycle components to the model leads to increased model complexity, especially if the cycles are dynamic. On the other hand, the addition of cycles can strongly improve the goodness of fit and may also improve the model's performance for the diagnostic tests discussed in section 4.3. Table 1 shows the AIC, BIC and p-values for the Ljung-Box, Jarque-Bera and heteroskedasticity tests for three different ice mass models each for basins 6, 13 and 22. One where the model has no cycles and is just an integrated random walk, one where the model contains static cycles and one where the model contains dynamic cycles, meaning they have disturbance variances.

Basin	Model	AIC	BIC	Ljung-Box p-value (lag 1)	Jarque-Bera p-value	Heteroskedasticity p-value
6	No cycles	1391.124	1404.945	0.05	0.00	0.57
6	**Static cycles**	**1374.411**	**1402.054**	**0.27**	**0.00**	**0.78**
6	Dynamic cycles	1376.215	1410.768	0.41	0.00	0.65
13	No cycles	1490.408	1504.229	0.00	0.05	0.20
13	Static cycles	1473.074	1500.716	0.00	0.05	0.68
13	**Dynamic cycles**	**1469.554**	**1504.107**	**0.06**	**0.18**	**0.26**
22	No cycles	1302.545	1316.367	0.14	0.00	0.04
22	**Static cycles**	**1264.392**	**1292.035**	**0.52**	**0.00**	**0.52**
22	Dynamic cycles	1262.549	1297.102	0.61	0.00	0.30

Table 1: AIC, BIC scores and p-values for the Ljung-Box, Jarque-Bera and heteroskedasticity tests for three different ice mass models for basins 6, 13 and 22: an integrated random walk with no cycles, an integrated random walk with two static cycles and an integrated random walk with two dynamic cycles. The models are considered to perform well for the Ljung-Box, Jarque-Bera and heteroskedasticity tests if the corresponding p-values are greater than 0.05. The rows corresponding to the chosen models are in bold.

Looking at the AIC, BIC and p-values in table 1 alone is not enough to make a well-informed decision about what types of mass models to choose for basins 6, 13 and 22. The addition of cycles, and the static or dynamic nature of the cycles, may affect the level and slope components of the models. The level component describes the trend in the ice mass data and the slope component is defined as the change in level per epoch. The Frequentist and Bayesian trends in the ice mass models of basins 6, 13 and 22 are used to answer the research questions of section 1.5. Hence, it is imperative to check how the level and slope are affected by different models. For model selection, the Frequentist level and slope are looked at specifically, because model selection is done following the Frequentist model selection process explained in section 4.4.

For basins 6, 13 and 22, appendix D shows the Frequentist level and slope components of three different models; integrated random walks with either no cycles, static cycles or dynamic cycles. No significant differences are seen between the three models for the level and slope components. Thus, when only interested in the level and slope, an integrated random walk model without cycles may be sufficient based on the results shown in appendix D. However, table 1 shows that the addition of cycles does improve the overall models, leading to observed states that better fit the data than models without cycles, while also improving the diagnostic metrics.

For the ice mass models of basin 6, an integrated random walk model with a static annual cycle and a static SA tidal alias cycle is chosen. Table 1 shows that such a model results in lower AIC and BIC scores than a model without cycles or with dynamic cycles, while performing well for the Ljung-Box and heteroskedasticity tests.

For the ice mass model of basin 13, an integrated random walk model with dynamic annual and SA tidal alias cycles is chosen. Table 1 shows that this model is the only model where the Ljung-Box p-value at lag 1 is greater than 0.05. As mentioned in section 4.4, the Ljung-Box test is considered particularly important, to avoid autocorrelation in the prediction errors. The model with dynamic cycles also performs well for the Jarque-Bera and heteroskedasticity tests.

For the ice mass model of basin 22, the same model is chosen as for the ice mass model of basin 6: an integrated random walk with static annual and SA tidal alias cycles. Table 1 shows that this model has the lowest BIC score, with only a slightly larger AIC score than the model with dynamic cycles. The model with static cycles for basin 22 performs well for both the Ljung-Box and heteroskedasticity tests.

Note that for basins 6 and 22, no model is found that results in a p-value for the Jarque-Bera test that is greater than 0.05. This means none of the investigated models obtain standardized prediction errors that are standard normally distributed. In future research, other models, such as state-space models that are not limited to linear Gaussian conditions, might be worth investigating.

6.1.2 Model equations

Recall the state-space representation described in section 3.1. The state vector x_n at epoch n for the ice mass models is described in equation 22. The ice mass models' parameters θ are given in equation 23.

State vector

$$x_n = \begin{bmatrix} l_n & r_n & c_{1,n} & c^*_{1,n} & c_{2,n} & c^*_{2,n} \end{bmatrix}^\top \tag{22}$$

Parameter vector

$$\theta = \begin{bmatrix} \sigma^2_\epsilon & \sigma^2_r & \sigma^2_{cycle,1} & \sigma^2_{cycle,2} \end{bmatrix}^\top \tag{23}$$

The states and parameters in equations 22 and 23 are defined as follows.

- l_n: level at epoch n [Gt]
- r_n: slope at epoch n, representing the change in level between consecutive epochs [Gt]
- $c_{1,n}$: annual cycle component at epoch n [Gt]
- $c^*_{1,n}$: auxiliary term for the annual cycle component at epoch n [Gt]
- $c_{2,n}$: SA tidal alias cycle component at epoch n [Gt]
- $c^n_{2,n}$: auxiliary term of the SA tidal alias cycle component at epoch n [Gt]
- σ^2_ϵ: variance of the irregular component [Gt^2]
- σ^2_r: disturbance variance of the slope [Gt^2]
- $\sigma^2_{cycle,1}$: disturbance variance of the annual cycle component (only for basin 13) [Gt^2]
- $\sigma^2_{cycle,2}$: disturbance variance of the SA tidal alias cycle component (only for basin 13) [Gt^2]

For the ice mass models of basins 6 and 22, static cycles are used. This means that the cycle components for basins 6 and 22 have disturbance terms of zero. For the ice mass model of basin 13, which contains dynamic cycles, $\sigma^2_{cycle,1}$ and $\sigma^2_{cycle,2}$ are unknown parameters to be estimated.

The ice mass models for basins 6, 13 and 22 have design matrix H_n shown in equation 24 and transition matrix A_n shown in equation 25. The variance-covariance matrix of the irregular term V_n and the variance-covariance matrix of the disturbance term Q_n are shown in equations 26 and 27, respectively.

Design matrix

$$H_n = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{24}$$

Transition matrix

$$A_n = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\psi_1) & \sin(\psi_1) & 0 & 0 \\ 0 & 0 & -\sin(\psi_1) & \cos(\psi_1) & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos(\psi_2) & \sin(\psi_2) \\ 0 & 0 & 0 & 0 & -\sin(\psi_2) & \cos(\psi_2) \end{bmatrix} \tag{25}$$

Variance-covariance matrix of the irregular term

$$V_n = \begin{bmatrix} \sigma^2_\epsilon \end{bmatrix} \tag{26}$$

Variance-covariance matrix of the disturbance term

$$Q_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_r & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{cycle,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{cycle,1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{cycle,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{cycle,2} \end{bmatrix} \tag{27}$$

In equation 25, $\psi_1 = \frac{2\pi dt}{T_1}$ and $\psi_2 = \frac{2\pi dt}{T_2}$, where dt is the sampling period. As the ice mass time series consist of equally-spaced monthly data, $dt = \frac{1}{12}$ year. T is the period of the cycle. For the annual cycle, $T_1 = 1$ year and for the SA tidal alias cycle, $T_2 = 0.44$ years.

In equation 27, $\sigma^2_{cycle,1}$ and $\sigma^2_{cycle,2}$ can be seen on the diagonal of Q_n. For basins 6 and 22, the cycles are static and these cycle disturbance variances are zero. For basin 13, the cyles are dynamic and the cycle disturbance variances are parameters to be estimated.

The design matrix H_n is part of the observation equation (1). Only the level l_n and the cycle components ($c_{1,n}$ and $c_{2,n}$) directly influence the observations. Hence, the design matrix has ones for the entries corresponding to these states in x_n and zeroes everywhere else. This results in the observed states $\hat{y}_n$ described in equation 28

Observed states

$$\hat{y}_n = l_n + c_{n,1} + c_{n,2} + \epsilon_n \qquad \epsilon_n \sim N(0, \sigma^2_\epsilon) \tag{28}$$

The transition matrix A_n is part of the state equation (2), which describes how the states x_n evolve over time. The level l_n evolves as described in equation 29.

Level

$$l_{n+1} = l_n + r_n \tag{29}$$

The level l_n has no disturbance variance in equation 29. This can also be seen in Q_n, where the first row, corresponding to l_n, is filled with zeroes. The slope r_n is dynamic, describing how the level evolves. Hence, r_n evolves based on its disturbance variance σ_r^2. This means the slope state evolves as described in equation 30.

Slope

$$r_{n+1} = r_n + \omega_{r,n} \qquad \omega_{r,n} \sim N(0, \sigma_r^2) \tag{30}$$

The cycle components and their auxiliary terms are c_n and c_n^*, respectively. c_n is determined using equation 31 and c_n^* is computed using equation 32.

Cycle components and their auxiliary terms

$$c_{n+1} = c_n \cos(\psi) + c_n^* \sin(\psi) + \omega_{c,n} \qquad \omega_{c,n} \sim N(0, \sigma_c^2) \tag{31}$$

$$c_{n+1}^* = -c_n \sin(\psi) + c_n^* \cos(\psi) + \omega_{c^*,n} \qquad \omega_{c^*,n} \sim N(0, \sigma_c^2) \tag{32}$$

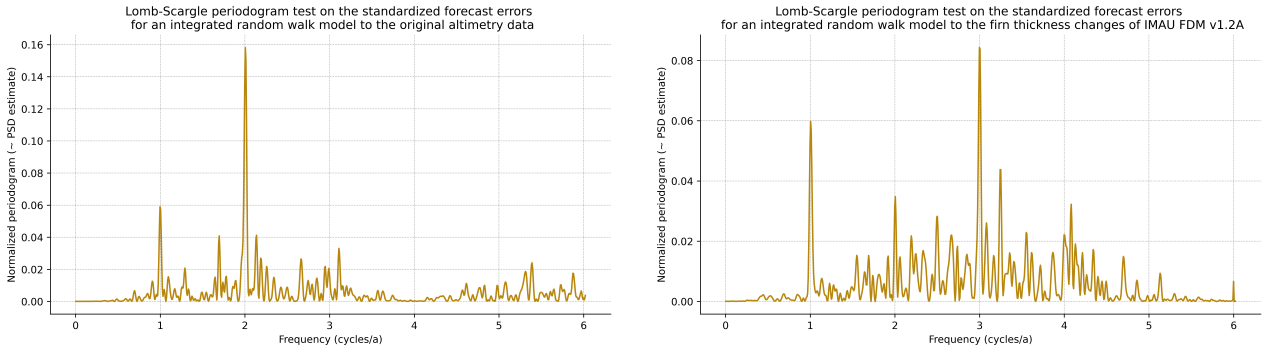
6.2 Ice volume models

When modelling the ice-dynamical volume data for basins 6, 13 and 22, overfitting issues occur. If a state-space model contains a dynamic component, the observed states will replicate the volume data exactly. This occurs both when following the Frequentist and Bayesian approaches and makes model selection difficult.

6.2.1 Overfitting issues with the altimetry data

The altimetry data contain significant autocorrelated uncertainties, likely caused by a combination of errors in the original altimetry measurements and errors in the firn densification model used to correct the measurements for firn thickness changes so that only ice-dynamical volume changes remain.

When comparing the periodograms of the prediction errors of integrated random walk models for the original altimetry data (green lines in figure 3) to those of the prediction errors of integrated random walk models for the modelled firn thickness changes (purple lines in figure 3, something unexpected is seen. Figure 11 shows these two periodograms. The periodogram for the firn thickness change model for basin 6 shown in figure 11(b) shows a clear cycle with a frequency of 3 cycles/year. Such a cycle is not present in the periodogram for the original altimetry data shown in figure 11(a). This is unexpected, as the original altimetry data should contain both the ice-dynamical volume changes and volume changes due to changes on the surface, like firn densification. The GRACE data for basin 6 also do not contain a clear cycle at a frequency of 3 cycles/year, as seen in the periodogram of figure 10. Thus, it is likely that the problem is with the firn densification model. The same differences are found for basins 13 and 22.



(a) Periodogram for original volume data of basin 6

(b) Periodogram for modelled firn thickness changes of basin 6

Figure 11: Lomb-Scargle periodogram tests on the prediction errors of integrated random walk models to the original altimetry data of basin 6 (left) and to the firn thickness changes for basin 6 of the IMAU firn densification model v1.2A (right). The periodogram for the firn thickness changes shows a clear peak at 3 cycles/year, which is not seen in the periodogram for the original altimetry data. This indicates that the modelled firn thickness changes contain cyclic behavior not present in the original altimetry data.

The modelled firn thickness changes are subtracted from the original altimetry data in an effort to isolate the ice-dynamical volume change. This causes errors, because the cycle with a frequency of 3 cycles/year in the modelled firn thickness changes is not present in the original data. Thus, subtracting the modelled firn thickness changes from the original altimetry data will result in residual errors.

It is likely that that the firn densification model is not accurately predicting the true firn thickness changes that are present in the original altimetry data. Currently, research is being done on the accuracy of firn thickness changes obtained by different firn densification models (Kappelsberger et al., 2023). This research is not finalized as of writing this thesis, so this thesis is limited to the IMAU Firn Densification Model v1.2A (Veldhuijsen et al., 2023).

As a result of the aforementioned errors in the altimetry data, the state-space models are not able to properly decompose the volume data into different components. To still approximate the data, the disturbance variance parameter of a dynamic component in the model, such as the slope of an integrated random walk or a dynamic cycle, is estimated by both the Frequentist and Bayesian approaches to be very large. This allows for large fluctuations in the corresponding component between epochs. For example, if a dynamic cycle is added to the model, the cycle disturbance variance is estimated to be very large. As a result, the cycle component is able to change so much between consecutive epochs, that the observed states completely replicate the data. This leads to a cycle that fluctuates erratically and does not look like a typical sinusoidal wave. An example of this is given in figure 12. Similar issues occur for the slope components, which also fluctuate very erratically and thus do not represent typical slowly varying behavior of a slope.

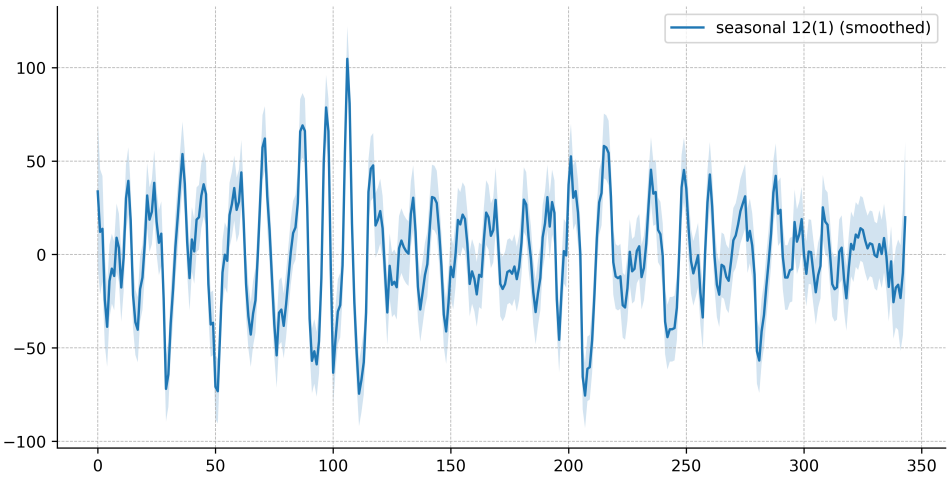


Figure 12: Example of an erratic annual cycle component found when added to an integrated random walk model fitted to the ice-dynamical volume data of basin 13. The maximum likelihood estimate of the disturbance variance of the dynamic cycle component is so large that it fluctuates heavily between epochs in an effort to get close to the data, effectively overfitting the data. As a result, the cycle component does not look like a typical sinusoidal cycle.

To avoid the overfitting issues while still using models that contain a dynamic slope, the choice is made to fix the slope disturbance variances of the ice volume models for basins 6, 13 and 22 to the maximum likelihood estimates of the slope disturbance variances for the corresponding ice mass models. To do this, the ice-dynamical volume data are first converted to mass by multiplying with the mean density of ice: 0.917 Gt/km^3 (Rumble, 2024). This conversion can be done, the ice-dynamical volume data are of the ice layer volume, after the snow and firn layer volumes have been removed. Thus, it can be assumed all volume in the data is of ice. Such a conversion cannot be done for the original volume data, which consists of both the volume of the ice layer and the volume of the snow and firn layers.

Any cycles added to the volume models are static, to avoid erratic cycles like the example shown in figure 12. Fixing the slope disturbance variances with no other dynamic components in the models, results in strongly autocorrelated prediction errors. To account for this, autoregressive components are added to the volume models.

Because the models tend to overfit the volume data, the Frequentist diagnostics described in section 4.3, such as the Ljung-Box or Jarque-Bera tests, cannot be used properly for the volume model selection. Instead, model selection is done based on visual inspection. For this, the correlogram diagnostic figure is especially important, to decide what order autoregressive component is needed to avoid strongly autocorrelated prediction errors.

6.2.2 Model equations

The chosen volume models for basins 6, 13 and 22 consist of a static level, a dynamic slope, one static cycle and a third order autoregressive component (AR(3) process). For basin 6, the cycle component represents a semi-annual cycle. For basins 13 and 22, the cycle components represent annual cycles. These cycles are identified from Lomb-Scargle periodogram tests on the prediction errors of integrated random walk models, just as for the ice mass models.

The state vector x_n at epoch n for the ice volume models is given in equation 33. The ice volume models' parameters θ are given in equation 34.

State vector

$$x_n = \begin{bmatrix} l_n & r_n & c_n & c_n^* & \varphi_n & \varphi_{n-1} & \varphi_{n-2} \end{bmatrix}^\top \tag{33}$$

Parameter vector

$$\theta = \begin{bmatrix} \sigma_r^2 & \sigma_{AR}^2 & a_1 & a_2 & a_3 \end{bmatrix}^\top \tag{34}$$

The states and parameters in equations 33 and 34 are defined as follows.

- l_n: level at epoch n [km^3]
- r_n: slope at epoch n, representing the change in level between consecutive epochs [km^3]
- c_n: coefficient of a cycle component at epoch n [km^3]
- c_n^*: auxiliary term for a cycle component at epoch n [km^3]
- φ_n: AR(3) process at epoch n [km^3]
- φ_{n-1}: AR(3) process at epoch $n-1$ [km^3]
- φ_{n-2}: AR(3) process at epoch $n-2$ [km^3]
- σ_r^2: disturbance variance of the slope [km^6]
- σ_{AR}^2: variance of the white noise process which drives the AR(3) process [km^6]
- a_1: first autoregressive coefficient of the AR(3) process [-]
- a_2: second autoregressive coefficient of the AR(3) process [-]
- a_3: third autoregressive coefficient of the AR(3) process [-]

The design matrix H_n of the volume models is given in equation 35. The transition matrix A_n is shown in equation 36. The variance-covariance matrix of the disturbance term is shown in equation 37. As the volume models do not contain an irregular component ($\epsilon_n = 0$), there is no variance-covariance matrix for the irregular term, like there was for the ice mass models.

Design matrix

$$H_n = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{35}$$

Transition matrix

$$A_n = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\psi) & \sin(\psi) & 0 & 0 & 0 \\ 0 & 0 & -\sin(\psi) & \cos(\psi) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_1 & a_2 & a_3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{36}$$

Variance-covariance matrix of the disturbance term

$$Q_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_r^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{AR}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{37}$$

In equation 36, $\psi = \frac{2\pi dt}{T}$, where dt is the time between consecutive epochs. Like the mass data, the volume data consist of equally-spaced monthly data, so $dt = \frac{1}{12}$ year. T is the period of the cycle. For the volume model of basin 6, the cycle is semi-annual, so $T = \frac{1}{2}$ year. For basins 13 and 22, the cycle component is annual, so $T = 1$ year. The level and cycle components of the volume models are static, so they have a disturbance variance of zero.

The observed states $\hat{y}$ of the volume models are computed with the observation equation (1). With the design matrix H_n defined in equation 35, the observed states $\hat{y}$ become as described in equation 38.

Observed states

$$\hat{y}_n = l_n + c_n + \varphi_n \tag{38}$$

The autoregressive process φ_n is seen as noise, thus it is not included when plotting the observed states. As a result, a model fit is obtained that approximates the data, but does not overfit.

The static level l_n at each epoch n is computed using equation 39.

Level

$$l_{n+1} = l_n + r_n \tag{39}$$

The dynamic slope r_n is computed at each epoch using equation 40.

Slope

$$r_{n+1} = r_n + \omega_{r,n} \qquad \omega_{r,n} \sim N(0, \sigma_r^2) \tag{40}$$

The static cycle components of the volume models and their auxiliary terms are computed with equations 41 and 42, respectively.

Cycle component and its auxiliary terms

$$c_{n+1} = c_n \cos(\psi) + c_n^* \sin(\psi) \tag{41}$$

$$c_{n+1}^* = -c_n \sin(\psi) + c_n^* \cos(\psi) \tag{42}$$

Finally, the third order autoregressive process is updated using equation 43.

AR(3) process

$$\varphi_{n+1} = a_1\varphi_n + a_2\varphi_{n-1} + a_3\varphi_{n-2} + \omega_{AR,n} \qquad \omega_{AR,n} \sim N(0, \sigma_{AR}^2) \tag{43}$$

6.3 Priors

When following the Bayesian approach, priors are needed for the parameters that are being sampled. For the ice mass models, these parameters are the irregular variance, the slope disturbance variance and, for basin 13, also the disturbance variances for the annual and SA tidal alias cycles. For the ice volume models, these are the slope disturbance variance, the autoregressive component's disturbance variance and the autoregressive coefficients.

For the ice mass model, Half Students' T-distributions are used for the irregular and disturbance variance parameters, as recommended in section 5.5. These have 3 degrees of freedom and a standard deviation of 10^3 Gt. Such a large standard deviation is chosen, as it is unlikely that the square roots of the irregular and disturbance variances are that large. The observed mass and volume changes in basins 6, 13 and 22 shown in figures 2 and 3, respectively, indicate that such large deviations are not present.

For the ice volume models, the prior distributions are a bit more complicated due to the use of autoregressive components and the aforementioned overfitting issues. As noted earlier, the disturbance variance of the slope components of the ice volume models of basins 6, 13 and 22 (converted to mass) are fixed to the maximum likelihood estimates of the slope disturbance variances for the corresponding ice mass models. This is done to avoid overfitting and applies to both the Frequentist and Bayesian approaches.

When adding an autoregressive component to a volume model, the autoregressive coefficients become additional model parameters. The number of coefficients is equal to the order of the autoregressive process added to the model, in this case being 3 for all volume models (a_1, a_2 and a_3). However, sampling from the marginal posterior distribution of the autoregressive coefficients causes issues in the *PyMC* Python package, leading to extremely slow sampling rates. The reason for this is unknown.

The Frequentist approach does not have issues when estimating the autoregressive coefficients. Hence, when following the Bayesian approach, the autoregressive coefficients of the volume models are fixed to the maximum likelihood estimates found by the Frequentist approach. The variances of the white noise processes driving the autoregressive components added to the volume models for basins 6, 13 and 22 are not fixed. Instead, Half Student's T-distributions with 3 degrees of freedom and standard deviations of 10^3 Gt^2 are used as their priors when following the Bayesian approach (Gt^2 instead of km^6, as the volume is converted to mass).

6.4 Sampling settings

Several settings have to be chosen when sampling with MCMC methods. First, the amount of chains in an MCMC run have to be decided upon. Then, the amount of samples in each chain need to be selected. The burn-in period length, being the amount of samples to discard from the start of each chain, has to be determined as well. Often, 4 chains are used (Schuett, 2021). The number of samples in these chains and the burn-in period length are chosen depending on the length of the data record that is modelled and the required accuracy. More chains and a larger number of samples per chain increase the accuracy of the MCMC sampler, but increase computational cost. In this thesis, 4 chains of 1500 samples and a burn-in period of 500 samples are used. This means 1000 samples are recorded for each parameter in each chain, providing 4000 samples for each parameter. When subsequently running the simulation smoother to sample from the conditional distribution of the states given the parameters and data $P(x_{1:N}|\theta, y)$, 1000 simulations are done. Each of these 1000 simulations use a different set of parameters, randomly drawn from the 4000 parameter samples obtained earlier. This results in 1000 samples of each state at each epoch, that are obtained while taking the stochasticity of model parameters into account. The different MCMC settings are discussed in more detail in chapter 9.

7 Differences in the uncertainties of the Frequentist and Bayesian approaches

To answer the first research question, the uncertainties found by the Frequentist and the Bayesian approaches have to be compared. The manner in which this is done is explained in section 7.1. Due to the overfitting issues discussed in section 6.2, the volume models are not used for the comparison of the Frequentist and Bayesian uncertainties. Instead, we limit this chapter to the mass models of section 6.1. The Frequentist and Bayesian results of fitting the mass models to the ice-dynamical mass data of section 2.2 are shown in section 7.2. The Frequentist and Bayesian uncertainties are also quantified there.

7.1 Confidence intervals and highest density intervals

Recall from section 4.2 that generally, the Kalman Filter and Smoother are used to estimate the states when following the Frequentist approach. The Kalman Filter and Smoother expect that the model parameters are given, for which the deterministic maximum likelihood parameter estimates can be used. This leads to estimates of the states with uncertainties represented by confidence intervals. When following the Bayesian approach, samples from the marginal posterior distribution for the parameters $P(\theta|y)$ are obtained with the No-U-Turn sampler. Subsequently, different sets of parameters are drawn from these samples and used in the simulation smoother to sample the states at each epoch from the conditional distribution of the states given the parameters and data $P(x_{1:N}|\theta, y)$. The uncertainty is then represented by a highest density interval.

Comparing the Frequentist and Bayesian uncertainties is difficult because of the different ways these uncertainties are represented. Namely, confidence intervals and highest density intervals, respectively. While similar in the sense that they represent uncertainty, confidence intervals and highest density intervals are fundamentally different concepts. A 95% confidence interval, for example, delineates a range of values centered at the estimated mean within which the true value will lie in 95% of repeated experiments. A 95% highest density interval, on the other hand, represents the most narrow range of values within which 95% of the posterior distribution's probability density is located.

To compare the Frequentist and Bayesian uncertainties, instead of using the Kalman Filter and Smoother to estimate the states with the maximum likelihood estimates for the Frequentist approach, the simulation smoother is used. In the Bayesian approach, the simulation smoother is used to sample from the conditional distribution of the states at each epoch given the parameters and data $P(x_{1:N}|\theta, y)$, using different sets of parameters θ sampled from the posterior of the parameters given the data $P(\theta|y)$ at each iteration. For the Frequentist approach, instead of using different sets of parameters θ in each iteration of the simulation smoother, the same maximum likelihood parameter estimates $\hat{\theta}$ are used in each iteration. As a result, both the Frequentist and Bayesian approaches provide samples of the conditional distribution of the states given the parameters and data with highest density intervals, enabling the comparison between the two approaches.

Because the Frequentist approach uses the same maximum likelihood estimates each iteration of the simulation smoother to estimate the states, stochasticity of the model parameters are not taken into account when sampling the states. Conversely, the Bayesian approach uses different sets of parameters, sampled from $P(\theta|y)$, at each iteration of the simulation smoother, thus taking stochasticity of the parameters into account when sampling the states. As a result, the Bayesian highest density intervals for the states will be larger, but more realistic compared to the Frequentist highest density intervals. The width of the Frequentist and Bayesian 95% highest density intervals are compared in this chapter. This is done specifically for the level and slope components of the ice mass models for basins 6, 13 and 22 described in section 6.1.

7.2 Frequentist and Bayesian ice mass model results

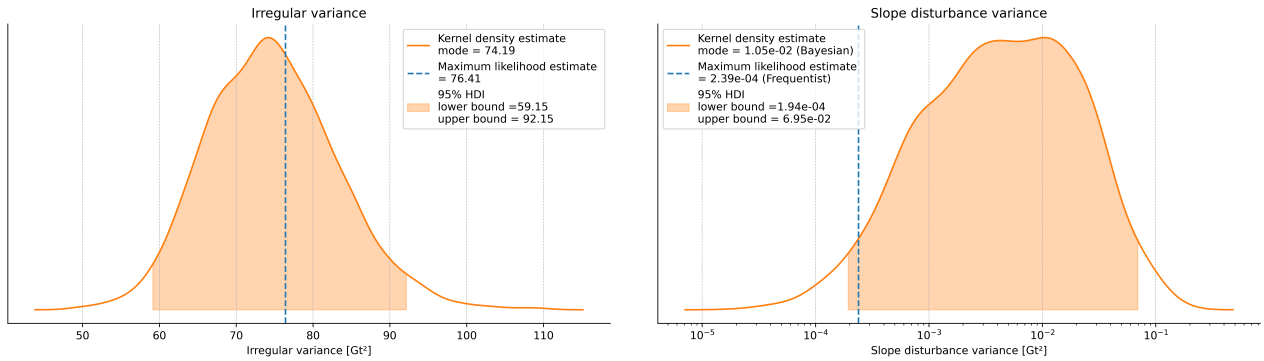
For the ice mass models of basins 6, 13 and 22, the Bayesian kernel density estimates of the marginal posterior distributions for the model parameters are shown, with a vertical line indicating the Frequentist maximum likelihood estimates of the parameters.

Furthermore, both the mean Frequentist and Bayesian observed states $\hat{y}$ of the ice mass models for basins 6, 13 and 22 are shown together in the same figures with 95% highest density intervals. The mean observed states $\hat{y}$ are obtained by using the mean of the states sampled from the conditional distribution of the states given the parameters and data $P(x_{1:N}|\theta, y)$ as input in the observation equation (1).

Additionally, the Frequentist and Bayesian model components are also plotted together. The mean width of the Frequentist and Bayesian 95% highest density intervals of the observed states, the level and the slope for the ice mass models are computed. The ratios between the widths of the Frequentist and Bayesian 95% highest density intervals are also determined. By comparing these, the first research question can be answered.

7.2.1 Basin 6

The Bayesian kernel density estimates of the marginal posterior distribution for the model parameters of the ice mass model for basin 6, with vertical lines indicating the parameters' Frequentist maximum likelihood estimates, are shown in figure 13. The ice mass model for basin 6 contains two parameters: the irregular variance and the slope disturbance variance.



(a) Irregular variance

(b) Slope disturbance variance

Figure 13: Estimated marginal posterior distributions for the model parameters of the ice mass model for basin 6, with 95% highest density intervals and vertical lines indicating the parameters' maximum likelihood estimates.

Figure 13 shows that the maximum likelihood estimate of the irregular variance is a bit off from the mode of the estimated marginal posterior distribution for the irregular variance, but still lies in a high likelihood region. The maximum likelihood estimate of the slope disturbance variance is far off from the mode of the estimated marginal posterior distribution of the slope disturbance variance. However, both the Frequentist and Bayesian approaches estimate the slope disturbance variance to be close to zero.

The difference between the maximum likelihood parameter estimates and the mode of the estimated marginal posterior distributions for the model parameters is caused by the priors. If non-informative uniform priors were used, the maximum likelihood estimate and the mode of the estimated marginal posterior distribution should be the same.

Figure 14 shows the Frequentist and the Bayesian observed states of the ice mass model for basin 6, with corresponding 95% highest density intervals. Due to the stochasticity of parameters in the Bayesian approach that is not present in the Frequentist approach, the Bayesian 95% highest density interval is larger than the Frequentist 95% highest density interval.

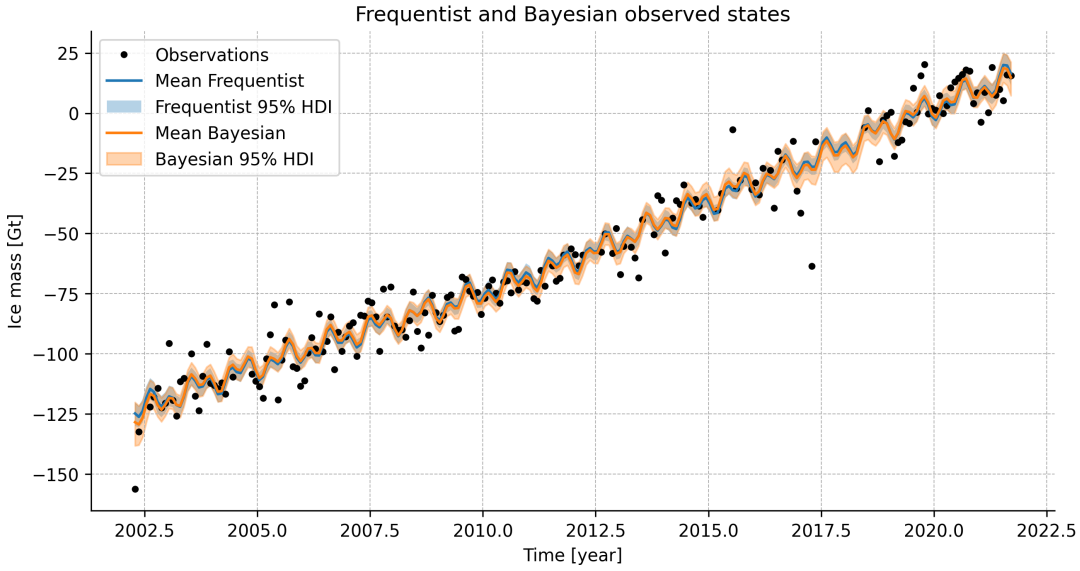
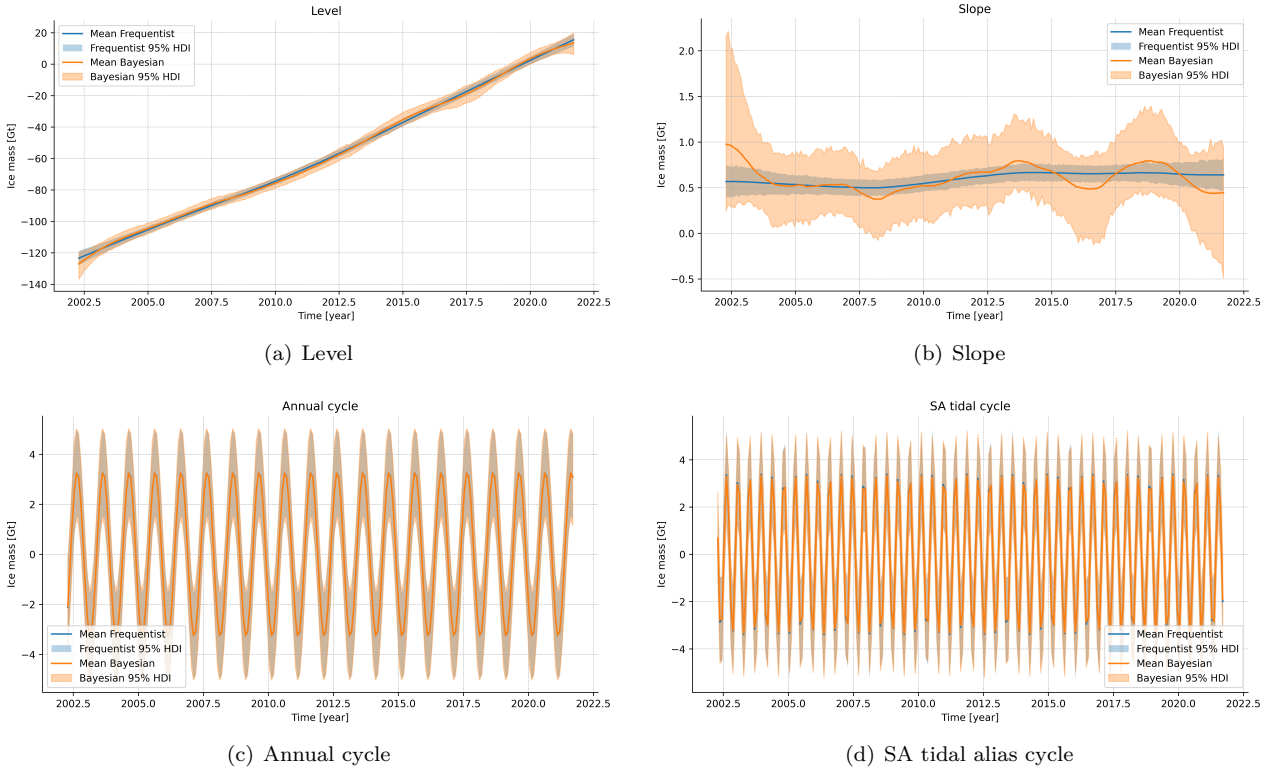


Figure 14: Frequentist (blue) and Bayesian (orange) observed states for the ice-dynamical mass time series of basin 6 (black).

Figure 15 below shows the Frequentist and Bayesian model components of the ice mass model for basin 6.



(a) Level

(b) Slope

(c) Annual cycle

(d) SA tidal alias cycle

Figure 15: Mean Frequentist (blue) and mean Bayesian (orange) model components for the ice mass model of basin 6, with their corresponding 95% highest density intervals.

Figure 14 shows that the Frequentist and Bayesian results are quite similar. However, as expected, the Frequentist and Bayesian 95% highest density intervals are different. This is also the case for the Frequentist and Bayesian model components shown in figure 15. The difference in the Frequentist and Bayesian 95% highest density intervals is caused by the fact that stochasticity of the model parameters is taken into account when sampling the states with the Bayesian approach, which is not the case for the Frequentist approach, where the model parameters are considered deterministic.

The largest difference between the Frequentist and Bayesian 95% highest density intervals is found in the slope component, as can be seen in figure 15(b). This is likely because the slope is the only dynamic component in the model. The slope also does not directly influence the observed states, but only influences how the level evolves. The level component is static, but because the slope is dynamic, the Frequentist and Bayesian levels are slightly different. The Frequentist and Bayesian cycles are nearly identical because the cycles of the ice mass model for basin 6 are static.

Table 2 below shows the mean width of the Frequentist and Bayesian 95% highest density intervals for the observed states and for the level and slope components for the ice mass model of basin 6.

	Average Frequentist 95% HDI	**Average Bayesian 95% HDI**	**Units**
Observed states	7.16	9.17	Gt
Level	5.05	7.63	Gt
Slope	0.21	0.83	Gt

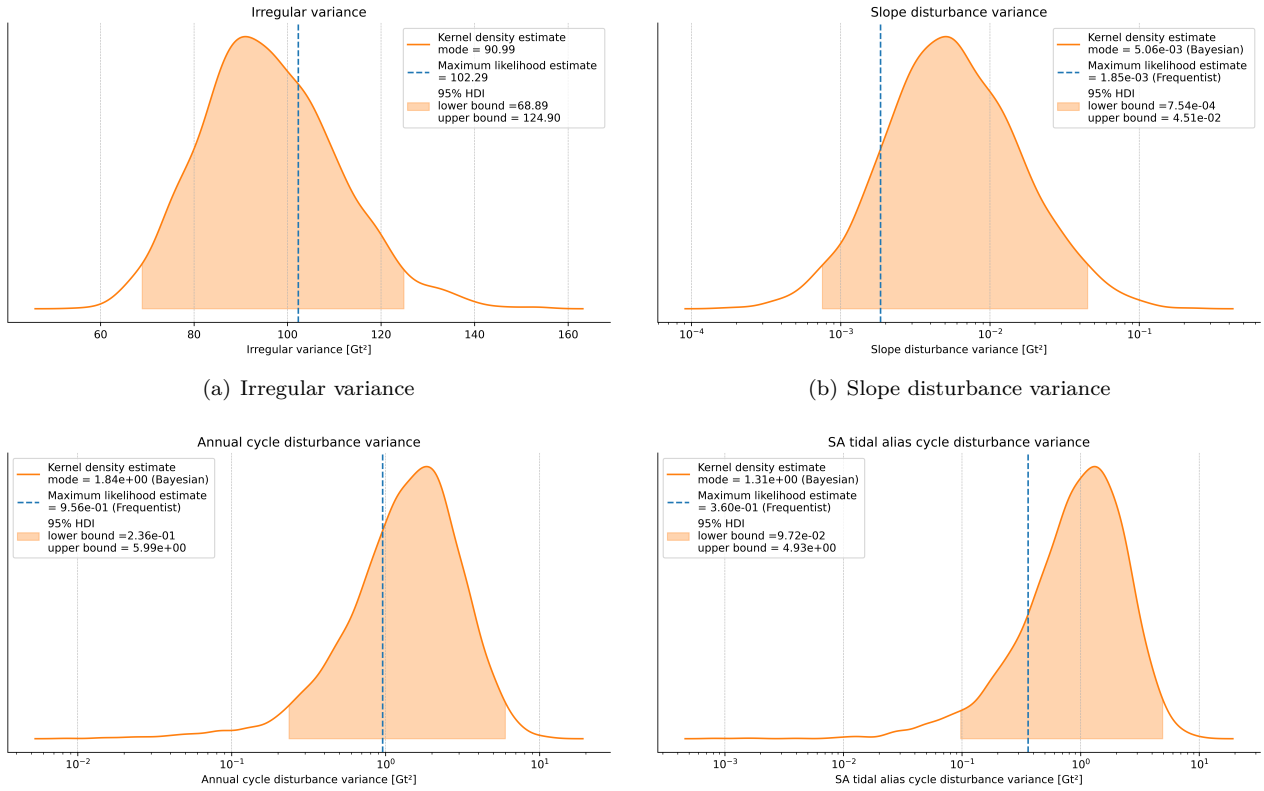
Table 2: Average Frequentist and Bayesian 95% highest density intervals (HDI) widths for the observed states, level and slope of the ice mass model for basin 6.

From table 2, it can be inferred that, on average, the Bayesian 95% highest density interval of the observed states is 1.28 times as large as the Frequentist 95% highest density interval for the ice mass model of basin 6. For the level component, this ratio is 1.51. For the slope component, the size disparity between the Frequentist

and Bayesian 95% highest density intervals is the largest. The Bayesian 95% highest density interval for the slope is 4.03 times as large as the Frequentist 95% highest density interval.

7.2.2 Basin 13

For the ice mass model of basin 13, the kernel density estimates of the marginal posterior distributions for the model parameters are shown in figure 16. Vertical lines are plotted representing the maximum likelihood estimates of the model parameters. As the ice mass model of basin 13 contains dynamic cycles, the model has four parameters: the irregular variance, slope disturbance variance, annual cycle disturbance variance and the SA tidal alias cycle disturbance variance.



(a) Irregular variance

(b) Slope disturbance variance

(c) Annual cycle variance

(d) SA tidal alias cycle disturbance variance

Figure 16: Estimated marginal posterior distributions for the model parameters of the ice mass model for basin 13, with 95% highest density intervals and vertical lines indicating the parameters' maximum likelihood estimates.

Figure 16 shows that the Frequentist maximum likelihood estimates are consistently off from the modes of the Bayesian estimated marginal posterior distributions of the parameters, but do lie in their relatively high likelihood regions. Figure 17 shows the Frequentist and the Bayesian observed states of the ice mass model for basin 13, with corresponding 95% highest density intervals.

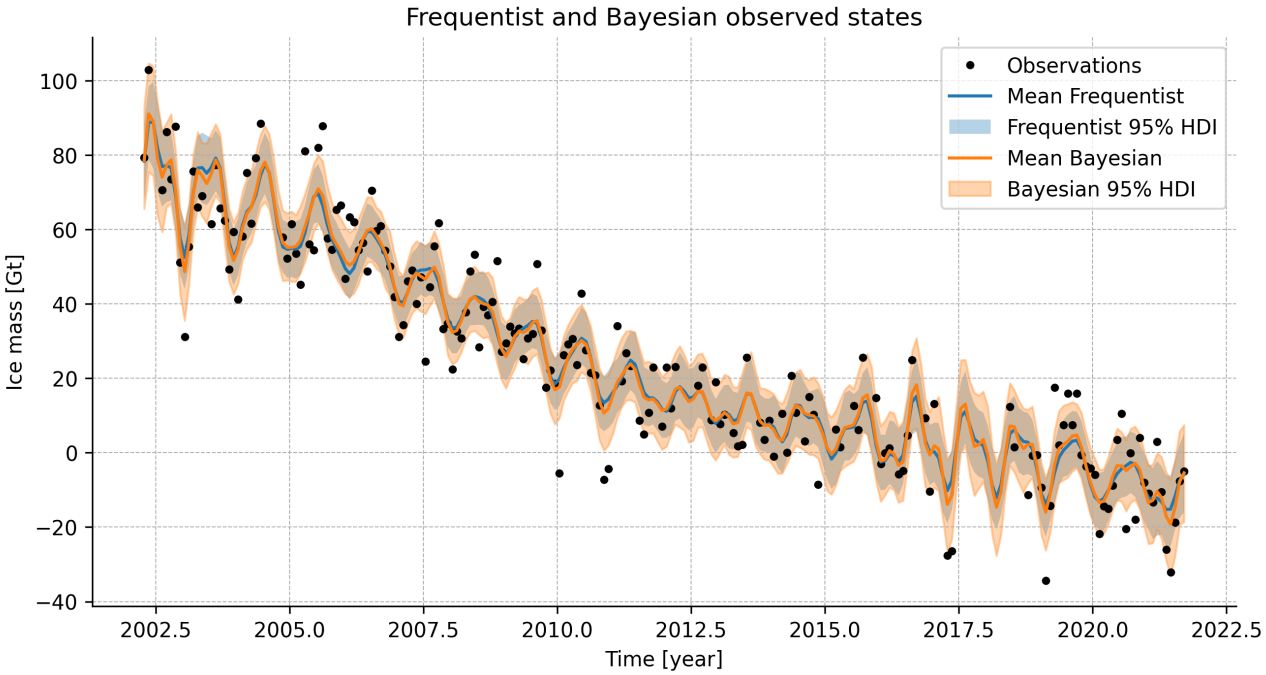
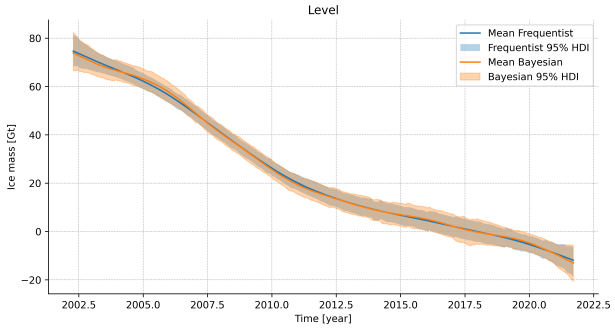
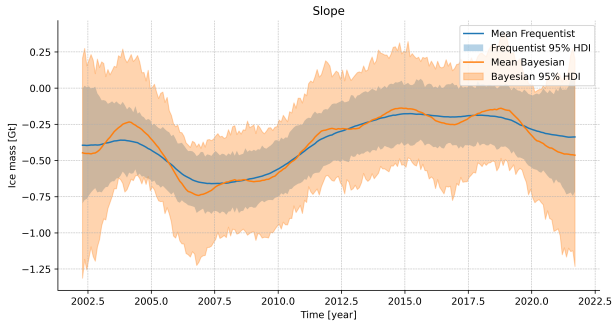


Figure 17: Frequentist (blue) and Bayesian (orange) observed states for the ice-dynamical mass time series of basin 13 (black).

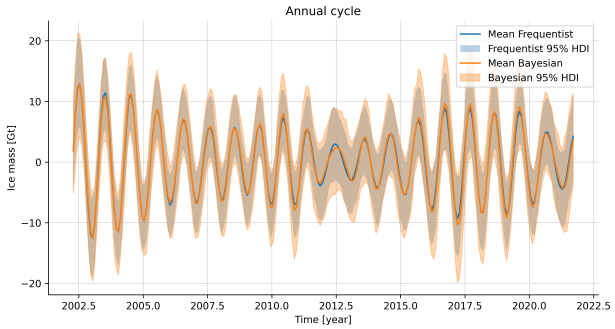
Figure 18 shows separate plots of the Frequentist and Bayesian model components.



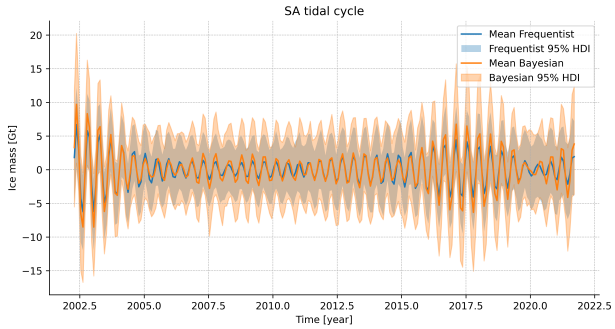
(a) Level



(b) Slope



(c) Annual cycle



(d) SA tidal alias cycle

Figure 18: Mean Frequentist (blue) and mean Bayesian (orange) model components for the ice mass model of basin 13, with their corresponding 95% highest density intervals.

The cycles of the ice mass model for basin 13 are dynamic. As seen in figures 18(c) and 18(d), the amplitudes of these dynamic cycles fluctuate quite heavily with time. The slope component is generally very small, meaning the change in level between consecutive epochs is not very large.

Table 3 below shows the mean width of the Frequentist and Bayesian 95% highest density intervals for the observed states and for the level and slope components for the ice mass model of basin 13.

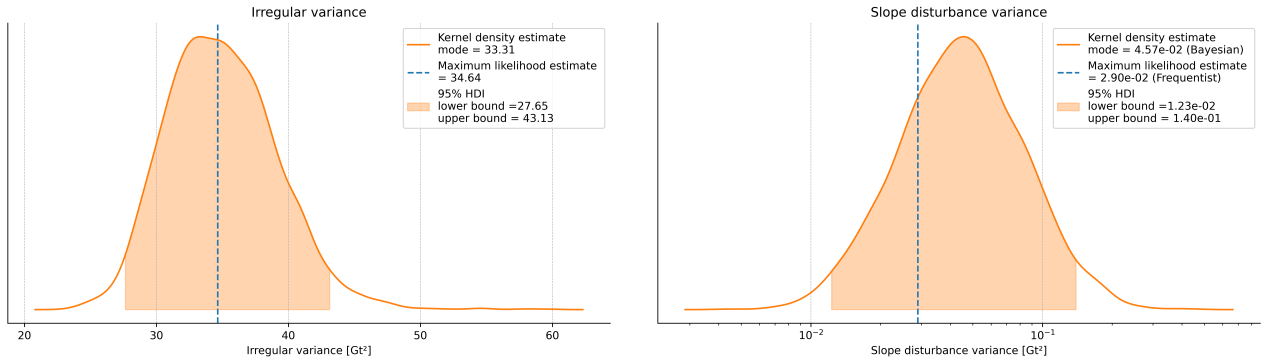
	Average Frequentist 95% HDI	**Average Bayesian 95% HDI**	**Units**
Observed states	15.93	19.90	Gt
Level	7.15	8.83	Gt
Slope	0.45	0.85	Gt

Table 3: Average Frequentist and Bayesian 95% highest density intervals (HDI) widths for the observed states, level and slope of the ice mass model for basin 13.

Table 3 shows that the Bayesian 95% highest density interval for the observed states is, on average, 1.25 times larger than the Frequentist 95% highest density interval. For the level, the mean Bayesian 95% highest density interval is 1.24 times as large as the Frequentist 95% highest density interval. For the slope, this ratio is 1.93. The differences between the Frequentist and Bayesian 95% highest density intervals for the level and slope components of the ice mass model for basin 13 are smaller than for the ice mass model of basin 6.

7.2.3 Basin 22

Kernel density estimates of the marginal posterior distribution for the model parameters of the ice mass model for basin 22, obtained by the Bayesian approach, are displayed in figure 19. In figure 19, the maximum likelihood estimates for the parameters, obtained by the Frequentist approach, are represented by vertical lines. The ice mass model for basin 22, like the ice mass model for basin 6, contains 2 parameters: the irregular variance and the slope variance.



(a) Irregular variance

(b) Slope disturbance variance

Figure 19: Mean Frequentist (blue) and mean Bayesian (orange) model components for the ice mass model of basin 22, with their corresponding 95% highest density intervals.

Figure 19 shows that the modes of the estimated marginal posterior distribution for the irregular variance of the ice mass model for basin 22 lies close to the maximum likelihood estimates of the irregular variance. For the slope component, the maximum likelihood estimate is again smaller than the mode of the estimated marginal posterior distribution of the slope disturbance variance. Figure 20 shows the Frequentist and the Bayesian observed states of the ice mass model for basin 22, with corresponding 95% highest density intervals.

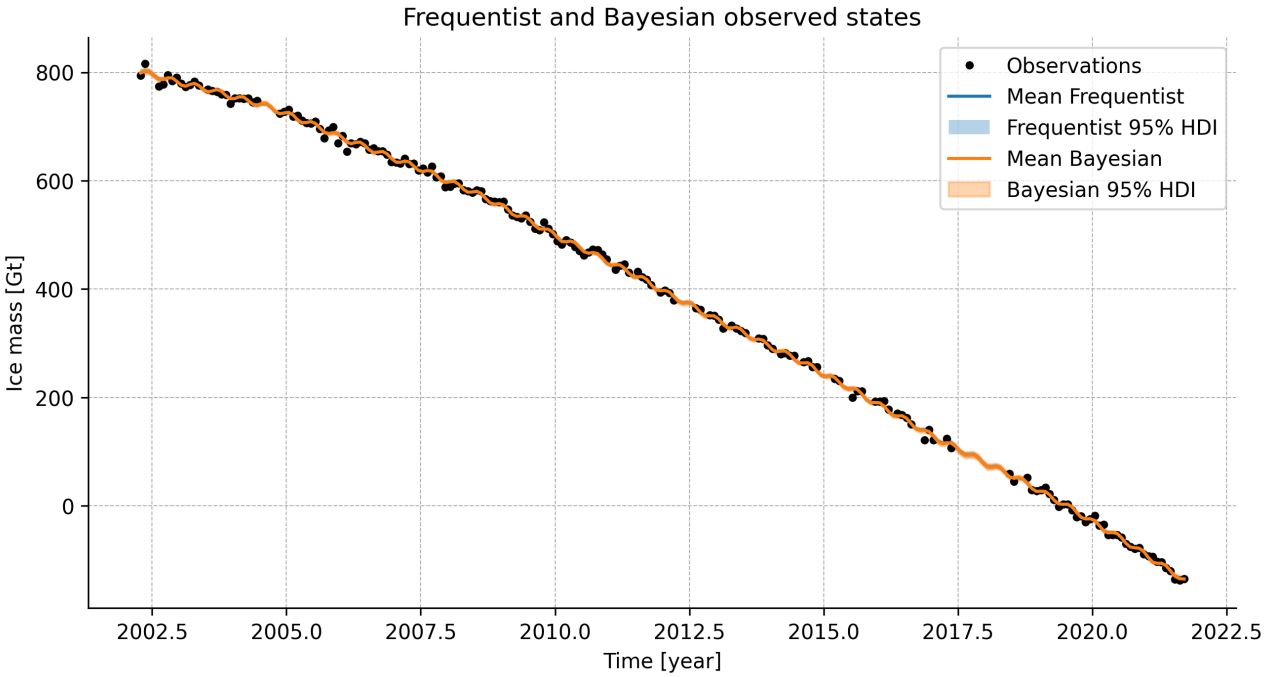
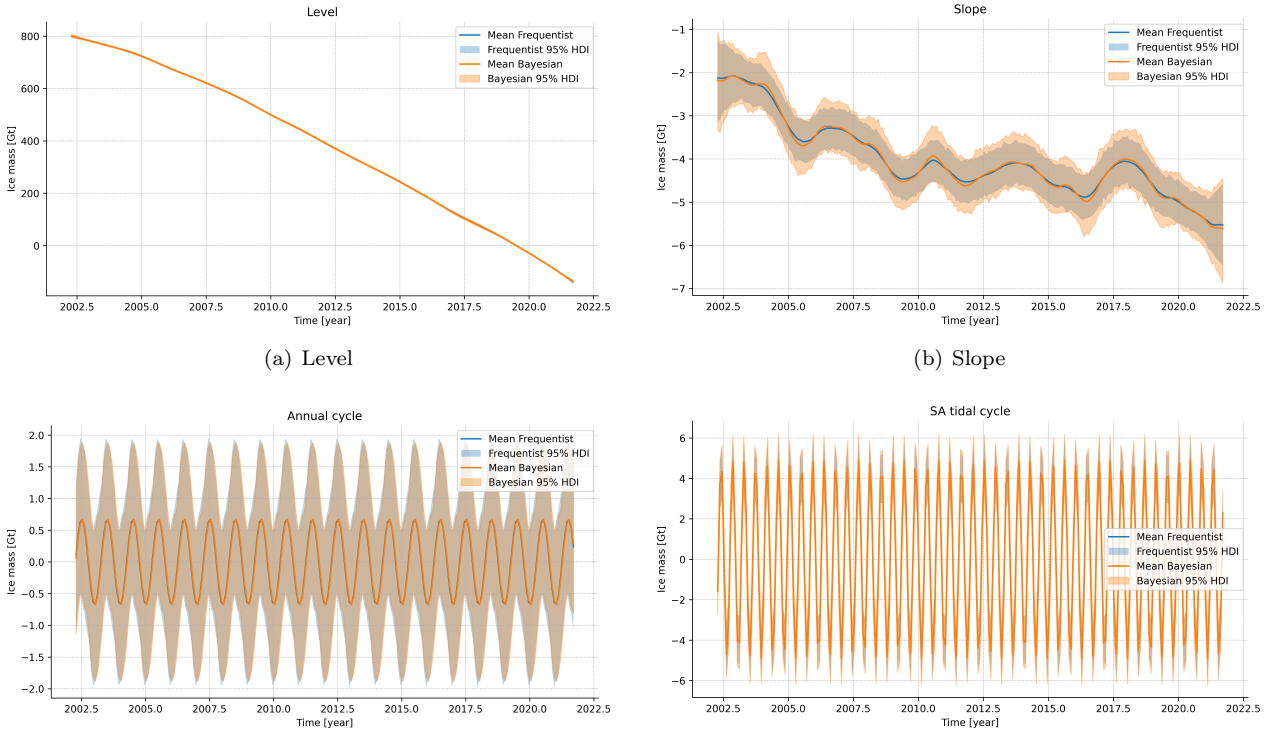


Figure 20: Frequentist (blue) and Bayesian (orange) observed states for the ice-dynamical mass time series of basin 22 (black).

The Frequentist and Bayesian model components of the ice mass model for basin 22 are shown in figure 21 below.



(a) Level

(b) Slope

(c) Annual cycle

(d) SA tidal alias cycle

Figure 21: Mean Frequentist (blue) and mean Bayesian (orange) model components for the ice mass model of basin 22, with their corresponding 95% highest density intervals.

The amplitude of the annual cycle in figure 21(c) is small relative to its uncertainty: the 95% highest density interval is approximately twice as large as the amplitude. This suggests that no annual cycle should be included in the ice mass model for basin 22. However, when comparing the AIC and BIC values of a model with and without this annual cycle, the model with an annual cycle scores better. Thus, the choice is made to keep the annual cycle, even if it may not seem to be a significant contributor to the observed states. However, in future research this annual cycle could be left out.

Similarly to the cycles of the ice mass model for basin 6, the cycles of the ice mass model for basin 22 are static. As a result, their Frequentist and Bayesian 95% highest density intervals are identical. For the observed states and the level, no significant difference between the Frequentist and Bayesian results can be observed visually from figures 20 and 21(a). Figure 21(b) does indicate that there is a slight difference between the Frequentist and Bayesian 95% highest density intervals for the slope component.

Table 4 gives the mean width of the Frequentist and Bayesian 95% highest density intervals for the observed states and for the level and slope components for the ice mass model of basin 22.

	Average Frequentist 95% HDI	**Average Bayesian 95% HDI**	**Units**
Observed states	7.26	7.83	Gt
Level	6.47	7.06	Gt
Slope	1.03	1.31	Gt

Table 4: Average Frequentist and Bayesian 95% highest density intervals (HDI) widths for the observed states, level and slope of the ice mass model for basin 22.

Table 4 shows that the difference between the widths of the Frequentist and Bayesian 95% highest density intervals for the ice mass model of basin 22, are small compared to those of the ice mass models of basins 6 and 13. The average width of the Bayesian 95% highest density interval for the observed states is only 1.08 times larger than the average width of the Frequentist 95% highest density interval. For the level component, this ratio is 1.09. For the slope component, the Bayesian 95% highest density interval is, on average, 1.27 times larger than the Frequentist 95% highest density interval. These relatively small differences are likely seen because the mass data in basin 22 are dominated by mass changes caused by ice dynamics, with little influence of cumulated surface mass balance anomalies, as discussed in chapter 2.

Ice dynamics result in long-term, slow-moving mass changes (Willen et al., 2021). The mass data for basin 22, primarily consisting of ice dynamical mass changes, is simpler and more predictable than the mass data for basins 6 and 13, which contain strong fluctuations due to changes in the surface mass balance. As the data being modelled shows a relatively simple and predictable pattern, the Frequentist and Bayesian approaches find very similar results for the ice mass model for basin 22.

8 Combination of the trends in GRACE-based and altimetry-based models

To answer the second research question, the level components of the ice mass and ice volume models described in chapter 6 are combined. This is done for both the level components obtained by the Frequentist approach and the level components obtained by the Bayesian approach. To combine the level of ice mass and ice volume models, the volume data are first converted to mass by multiplying them with the mean density of ice, which is 0.917 Gt/km^3 (Rumble, 2024). The combinations are made for the time period where both the GRACE and altimetry datasets have observations. The GRACE-based and altimetry-based levels are shown with respect to the first observations taken at the same time, to reduce potential biases between the GRACE and altimetry data affecting the combination.

Combining the levels of GRACE-based models and altimetry-based is done using weighted averaging. This process is explained in section 8.1. The Frequentist combinations for basins 6, 13 and 22 are shown in section 8.2. The Bayesian combinations are shown in section 8.3. The Frequentist and Bayesian combinations are compared in section 8.4. Finally, a brief discussion is given in section 8.5, comparing the combination approaches of this chapter with those of previous research.

8.1 Weighted averaging

To combine the GRACE-based level with the altimetry-based level, weighted averaging is used. The weighted averaged level $\bar{l}_n$ at each epoch can be computed using equation 44.

Weighted averaged level

$$\bar{l}_n = \frac{w_{1,n} l_{1,n} + w_{2,n} l_{2,n}}{w_{1,n} + w_{2,n}} \tag{44}$$

In equation 44, $l_{1,n}$ is the level of the GRACE-based model at epoch n. The weight given to $l_{1,n}$ is $w_{1,n}$. The level of the altimetry-based model at epoch n is $l_{2,n}$, which is given weight $w_{2,n}$.

For the Frequentist approach, the weights are equal to the reciprocal of the variance. For the Bayesian approach, the weights are equal to the reciprocal of the squared width of the 95% highest density interval (HDI). The reciprocal of specifically the squared 95% HDI width is used for the Bayesian approach because it is analogous to the Frequentist approach, as the reciprocal of the variance (σ^2) is equal to the reciprocal of the squared standard deviation (σ). This choice ensures consistency in the weighting scheme between the two approaches.

8.2 Frequentist combination

For the Frequentist combination, the smoothed levels of the GRACE-based and altimetry-based models are used. These are obtained by using the Kalman Filter and Smoother, explained in section 4.2. The uncertainties of the Frequentist levels in this section are represented by 95% confidence intervals. Note that this is different from the Frequentist levels shown in chapter 7. There, the Frequentist results were obtained by running the simulation smoother with the maximum likelihood parameter estimates, to facilitate a comparison between the Frequentist and Bayesian approaches by representing both Frequentist and Bayesian results with 95% highest density intervals. This is not necessary for the combination between the levels of GRACE-based and altimetry-based models, as these are both obtained by the same modelling approach. Hence, the more classical Frequentist approach is followed here, using the Kalman Filter and Smoother to estimate the states.

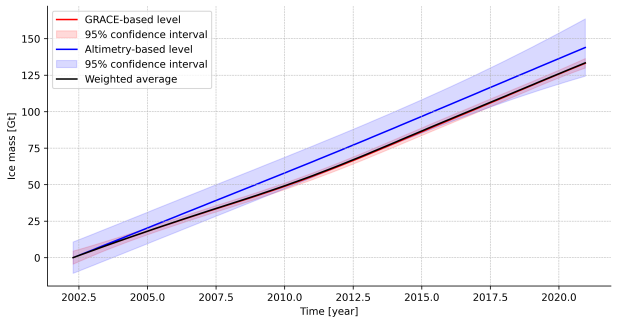
As mentioned in section 8.1, the weights for the Frequentist combination are equal to the reciprocal of the variance. Thus, the Frequentist weights for the GRACE-based level and the altimetry-based level are defined in equations 45 and 46, respectively.

Frequentist weights

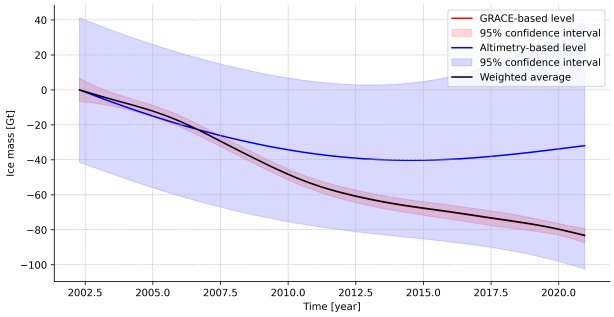
$$w_{1,n} = \frac{1}{\mathrm{Var}(l_{1,n})} \tag{45}$$

$$w_{2,n} = \frac{1}{\mathrm{Var}(l_{2,n})} \tag{46}$$

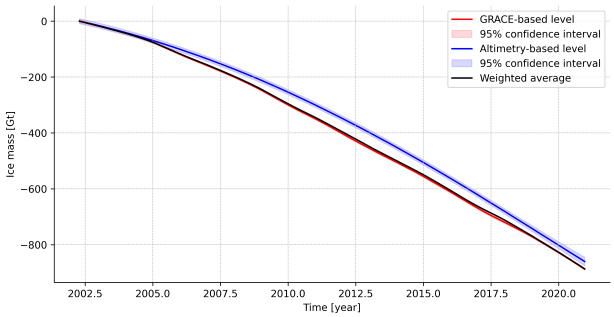
The Frequentist levels of the corresponding GRACE-based models and altimetry-based models for basins 6, 13 and 22, with 95% confidence intervals are shown in figure 22, alongside their weighted averages.



(a) Basin 6



(b) Basin 13



(c) Basin 22

Figure 22: Frequentist levels of the GRACE-based ice mass models (red) and altimetry-based (blue) ice mass models for basins 6, 13 and 22. The uncertainties of the levels are represented by 95% confidence intervals. A weighted average of the two, with weights equal to the reciprocal of the level's variances, is shown as well (black). The ice mass is given in gigatonnes.

Figure 22 shows that the 95% confidence interval, and thus also the variance, of the GRACE-based level is consistently significantly smaller than the variance of the altimetry-based level. As a result, the weighted average aligns closely with the level of the GRACE-based model. This is the case for both basins 6, 13 and 22.

For basins 6 and 22, the GRACE-based and altimetry-based levels align relatively well, showing similar behavior over time. For basin 13, however, the GRACE-based and altimetry-based levels behave quite differently. This discrepancy is likely caused by the fact that the mass changes in basin 13 are not strongly affected by ice dynamics, but primarily by changes in the surface climate, as discussed in chapter 2. As the ice-dynamical mass changes are not very significant for basin 13, measurement errors and errors in the models used to correct the original GRACE and altimetry data for changes in the surface climate have a significant effect on the ice-dynamical data being modelled. These different errors likely cause the discrepancy between the GRACE-based and altimetry-based levels seen in figure 22(b). The ice masses of basins 6 and 22 are more strongly impacted by ice dynamics, hence why the GRACE-based and altimetry-based levels align more closely for those basins.

8.3 Bayesian combination

For the Bayesian combination, the mean simulated levels of the GRACE-based and altimetry-based models are used, obtained by the simulation smoother. To combine the GRACE-based and altimetry-based levels, weighted averaging is done with weights based on the reciprocal of the squared width of the 95% highest density intervals. The width of a 95% highest density interval is equal to the difference between the upper and lower bound of the 95% highest density interval. The Bayesian weights are defined in equations 47 and 48.

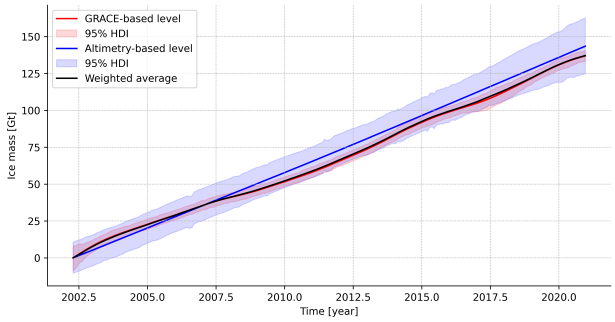
Bayesian weights

$$w_{1,n} = \frac{1}{(b_{1,n}^{\text{upper}} - b_{1,n}^{\text{lower}})^2} \tag{47}$$

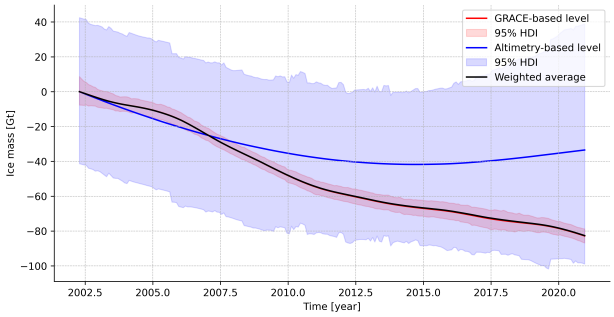
$$w_{2,n} = \frac{1}{(b_{2,n}^{\text{upper}} - b_{2,n}^{\text{lower}})^2} \tag{48}$$

In equation 47, $b_{1,n}^{\text{upper}}$ and $b_{1,n}^{\text{lower}}$ are the upper and lower bounds of the 95% highest density interval of the GRACE-based level at epoch n, respectively. For equation 48, $b_{2,n}^{\text{upper}}$ and $b_{2,n}^{\text{lower}}$ are the upper and lower bounds of the 95% highest density interval of the altimetry-based level at epoch n, respectively.

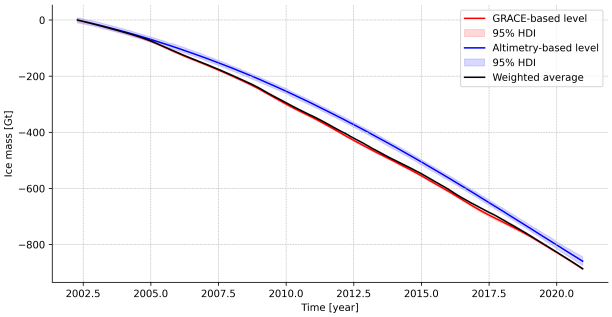
Figure 23 shows the Bayesian levels of the GRACE-based models and altimetry-based models for basins 6, 13 and 22. The uncertainties are represented by 95% highest density intervals. The weighted average is shown as well.



(a) Basin 6



(b) Basin 13



(c) Basin 22

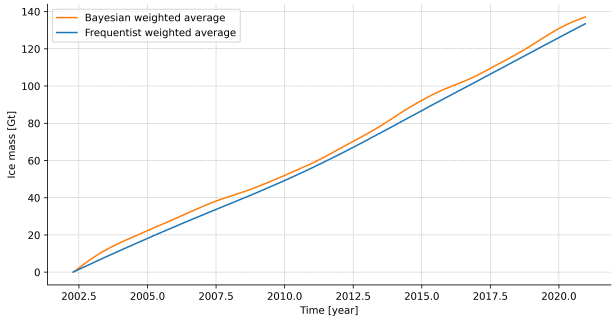
Figure 23: Bayesian levels of the GRACE-based ice mass models (red) and altimetry-based (blue) ice mass models for basins 6, 13 and 22. Their uncertainties are represented by 95% confidence intervals. A weighted average of the two, with weights equal to the reciprocal of the squared widths of the 95% highest density intervals, is shown as well (black). The ice mass is given in gigatonnes.

Figure 23 shows that the weighted average of the Bayesian GRACE-based and altimetry-based levels aligns very closely to the GRACE-based level for both basins 6, 13 and 22. This is because the 95% highest density interval of the altimetry-based level is consistently significantly larger than the 95% highest density interval of the GRACE-based level.

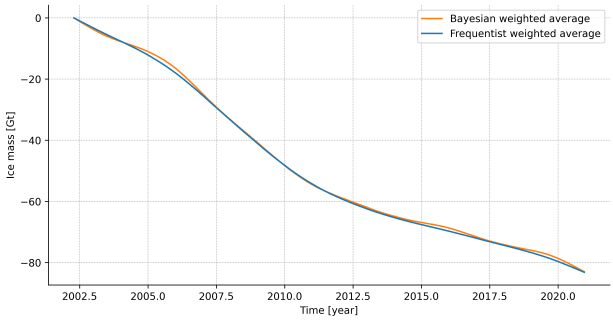
As for the Frequentist combination discussed in section 8.2, the GRACE-based and altimetry-based levels align quite well for basins 6 and 22, but not so well for basin 13. Again, this is likely caused by the fact that mass changes in basin 13 are predominantly caused by changes in the surface climate, with comparatively little effect of ice dynamics. For basins 6 and 22, ice dynamics play a more dominant role.

8.4 Comparing Frequentist and Bayesian combinations

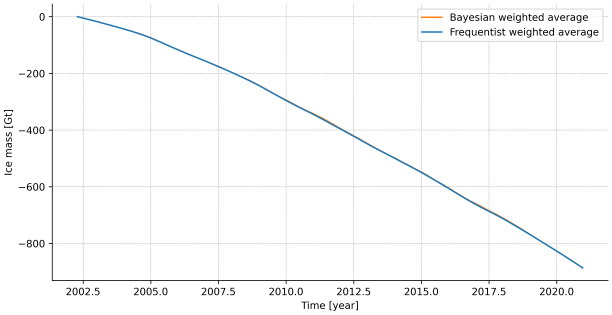
Figure 24 below shows the Frequentist and Bayesian combinations of the GRACE-based and altimetry-based ice mass models for basins 6, 13 and 22 together.



(a) Basin 6



(b) Basin 13



(c) Basin 22

Figure 24: Frequentist (blue) and Bayesian (orange) combinations of the GRACE-based and altimetry-based level components of ice mass models for basins 6, 13 and 22.

Figure 24(a) shows a bias between the Frequentist and Bayesian combinations of the level for basin 6. This is not seen for basins 13 and 22 (figures 24(b) and 24(c) respectively). For these basins, the Frequentist and Bayesian weighted averaged levels are very similar.

As discussed in section 7.2, the average Bayesian 95% highest density intervals for the GRACE-based level of basin 6 is 1.51 times larger than the corresponding average Frequentist 95% highest density interval. The weighted average of the level closely aligns with the GRACE-based level, so this difference between the Bayesian and Frequentist 95% highest density intervals of the GRACE-based level for basin 6 also affects the Frequentist and Bayesian weighted averaged levels for basin 6. This could explain the bias between the Frequentist and Bayesian weighted averaged levels seen in figure 24(a).

8.5 Comparison to previous research

In this thesis, the slope disturbance variance of the altimetry-based model is fixed to the same value as found by maximum likelihood estimation for the GRACE-based model. This is similar to what is done in previous research by Willen et al. (2021), where the slope disturbance variances of the GRACE-based and altimetry-based models were also set to be equal.

The slopes of the GRACE-based and altimetry-based models are uncorrelated in this thesis, meaning they can be entirely different. Take for example the slopes of the models for basin 13, shown in figure 22(b). There, it can be seen that the levels of the GRACE-based and altimetry-based models move in different directions at the end of the time series, meaning they have different slopes. Logically, it would follow that a decrease in ice mass comes with a decrease in ice volume. Thus, this anomaly indicates the GRACE and altimetry data for basin 13 do not agree with each other.

Such behavior was remedied in the research by Willen et al. (2021) by having the change in the slope components of GRACE-based and altimetry-based models be fully correlated. As a result, the levels found by Willen et al. (2021) always move in the same direction, although biases may exist between the GRACE-based and altimetry-based levels. Correlated changes in the slopes of the GRACE-based and altimetry-based models also affect the level uncertainty. Due to this correlation and the fact that the slope disturbance variances of the GRACE-based and altimetry-based models are set to be equal, the slope and level uncertainties of the two models will be more similar than when they are not correlated in any way as is the case in this thesis. As a result, a weighted average between the GRACE-based and altimetry-based levels, with weights based on the level uncertainties, would likely be more in the middle of the two, instead of aligning closely to the GRACE-based dataset as found with the approach taken in this thesis.

9 Discussion of Markov Chain Monte Carlo settings

In section 1.5, it was noted that a subgoal of this thesis is to gain experience with the Bayesian approach to time series modelling using state-space models. The Bayesian approach entails the use of MCMC methods, for which choices have to be made, such as the type of sampler to use and the priors for the model parameters. Several of these have already been addressed in the previous chapters. This chapter provides an overview of these choices and describes how they may impact an MCMC run when working with state-space models. For sampling diagnostics that can be viewed after performing an MCMC run, refer to section 5.3.

Section 9.1 summarizes the main types of MCMC samplers that can be used when following the Bayesian approach and when they are suitable. The choice of priors for different types of model parameters is discussed in section 9.2. An MCMC run has to be initialized. Initialization choices are denoted in section 9.3. Section 9.5 is about the number of chains and samples to use in an MCMC run. Choices regarding the burn-in period of each chain are discussed in section 9.4. Finally, the option of thinning MCMC samples is discussed in section 9.6.

9.1 Type of sampler

There are many types of MCMC samplers that can be used to sample from model parameters' posterior distributions. Each sampler does this in different ways and are, as a result, suitable in different situations. Note that all MCMC samplers are only ever correct 'in the limit', i.e., when an infinite number of samples are taken (Hogg and Foreman-Mackey, 2018). In practice, a tradeoff has to be made between computational costs and the accuracy of an MCMC run.

For a small number of model parameters, the Metropolis-Hastings sampler is efficient and relatively easy to implement (Hogg and Foreman-Mackey, 2018). Ensemble samplers, which use multiple random walks to explore the parameter space, are also well suited for a small number of parameters, but are more complicated to implement. For large numbers of parameters, Metropolis-Hastings or ensemble samplers are less suited, leading to inefficient sampling due to large autocorrelation times (Carpenter, 2017).

For large numbers of parameters, Gibbs samplers or Hamiltonian methods perform well (Hogg and Foreman-Mackey, 2018). Hamiltonian MCMC samplers are considered the state of the art for models with a large amount of parameters (Hogg and Foreman-Mackey, 2018). Hamiltonian samplers use Hamiltonian dynamics to efficiently explore the parameter space, which is explained in more detail in appendix C. The No-U-Turn sampler used in this thesis is a Hamiltonian MCMC method. The Python code made for this thesis is stored on a GitLab repository, as explained in appendix E. This code provides an example usage of Hamiltonian MCMC.

9.2 Priors

Inherent to the Bayesian approach is the use of priors for the model parameters. Prior distributions contain information about the model parameters before data are available. For example, if a parameter only takes on non-negative values, such as a variance parameter, a prior distribution can be chosen that only takes on values greater than zero.

9.2.1 Prior recommendations from literature

Non-informative priors can be chosen for parameters, such as uniform distributions between values a and b. When enough data is available and the model complexity is not very large, non-informative priors may be sufficient to estimate a parameters' posterior distribution (Klees, 2023). In many cases, the use of non-informative priors leads to converge issues, however. To prevent this, informative priors are used.

For parameters that lie between values a and b, an informative prior is to use a Gaussian distribution with a mean of $\frac{(a+b)}{2}$ and a standard deviation of $\frac{(b-a)}{2}$ (Klees, 2023). For parameters that take on only non-negative values, the recommended informative prior is a half Student's T-distribution with 3 or 4 degrees of freedom and a standard deviation that is likely to be larger than the true standard deviation of the parameter's distribution.

9.2.2 Results with different priors

Recall from section 6.1 that the mass model for basin 6 contains 2 parameters: the irregular variance and the slope disturbance variance. In literature, it is recommended to use half Student's T-distributions with either 3 or 4 degrees of freedom for variance parameters (Klees, 2023). In this thesis, half Student's T-distributions with 3 degrees of freedom and standard deviations of 1000 Gt^2 were chosen. There are other possibilities, however.

Figure 25 below shows the level of the ice mass model for basin 6 looks when using different priors. Specifically, non-informative Uniform distributions between 0 and 1000 Gt^2, Half normal distributions centered at 0 with a

standard deviation of 1000 Gt^2 and Inverse Gamma distributions with a mean of 1 and a standard deviation of 1000 Gt^2. All these chosen priors have in common that they do not allow for negative values, as the variance parameters cannot be negative in reality. The impact of using either 3 or 4 degrees of freedom in the Half Student's T-distributions as priors (both with standard deviations of 1000 Gt^2) is also shown in figure 25.

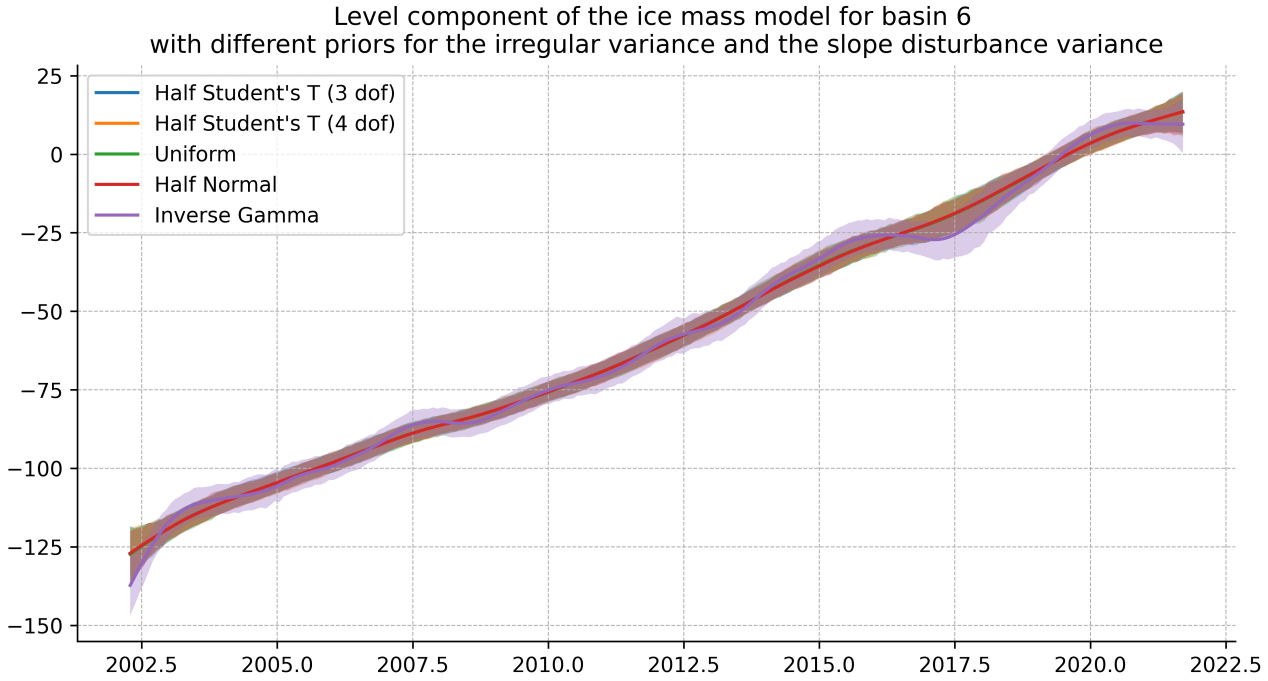


Figure 25: The level of the ice mass model for basin 6 with different priors: Half Student's T with 3 degrees of freedom (blue), Half Student's T with 4 degrees of freedom (orange), Uniform (green), Half Normal (red) and Inverse Gamma (purple). Only the Inverse Gamma prior results in significant differences in the level when compared to the other priors.

Figure 25 shows no significant differences between the Half Student's T, Uniform and Half Normal priors. The Inverse Gamma prior does result in a level divergent from the levels found with other priors.

When enough high-quality data is available, the priors should have little impact on the results (Klees, 2023). This could explain why the Half Student's T, Uniform and Half Normal priors all provide nearly identical results. However, this does not explain why the Inverse Gamma prior leads to a different level. It is known that Inverse Gamma priors can cause problems (Gelman, 2006; Klees, 2023). This may explain the anomaly of the level found with Inverse Gamma priors, although further research would be required to ascertain this.

Figure 25 also shows no significant difference between 3 and 4 degrees of freedom for the Half Student's T prior. Because of this, the choice was made to use 3 degrees of freedom for the results shown in chapters 7 and 8.

9.3 Initialization

An MCMC run has to be initialized. This entails selecting initial values from which to start the sampling process.

When working with state-space models, the initial states for the Kalman Filter need to be set. In the *PyMC-experimental* Python package, which is the experimental development version of *PyMC*, it is possible to set priors for the initial states (Grabowski, 2023). This functionality is not yet present in the main *PyMC* package, which is used in this thesis. Instead, the *Statsmodels* Python module is used to generate state-space models whereby the Kalman Filter can be initialized with either exact diffuse or approximate diffuse initialization. The *Statsmodels* state-space models are then used in the *PyMC* package, where priors only have to be given to the model parameters and not for the initial states, as these already have either the exact or approximate diffuse initialization from *Statsmodels*.

For the model parameters, initial values are usually chosen in high likelihood regions of the prior distributions. For example, the means of the prior distributions are often used as initial values for the parameters when sampling. To add randomness and thus avoid potential biases introduced by the priors, uniform noise in the range of [-1, 1], referred to as "jitter", is often added to these means (PyMC documentation, 2021).

When working with Hamiltonian MCMC methods, as is done in this thesis, the mass matrix needs to be initialized as well. The mass matrix guides the sampler to efficiently explore the parameter space, by adjusting step sizes in different directions (Brooks et al., 2011). The mass matrix is usually initialized as an identity matrix and subsequently adapts a diagonal matrix based on the variance of tuning samples (PyMC documentation, 2021). This tuning takes place during the burn-in period.

9.4 Burn-in period

The initial values discussed in section 9.2, are unlikely to immediately lead to samples that are representative of the posterior distribution that is being sampled. Depending on the priors and initialization, it may take several samples before the samples arrive in high-likelihood regions of the sampled posterior distribution. Because of this, the first few samples are thrown out. The number of initial samples that are thrown out is referred to as the burn-in period. Only samples taken after the burn-in period has passed are stored. In this thesis, a burn-in period length of 500 samples is used.

A burn-in period is not always strictly necessary. For example, if the prior distributions are very close to the posterior distributions that are being sampled, the samples will likely start out in high-likelihood regions of the posterior distributions. However, in practice it is often unknown how representative a prior distribution is of the true distribution (hence why it is being sampled in the first place). Therefore, it is generally advised to use a burn-in period.

The Bayesian level of the ice mass model for basin 6 obtained when using no burn-in period, a burn-in period of 100 samples and a burn-in period of 500 samples is shown in figure 26. For these levels, the same Half Student's-T priors with 3 degrees of freedom and a standard deviation of 1000 Gt^2 are used.

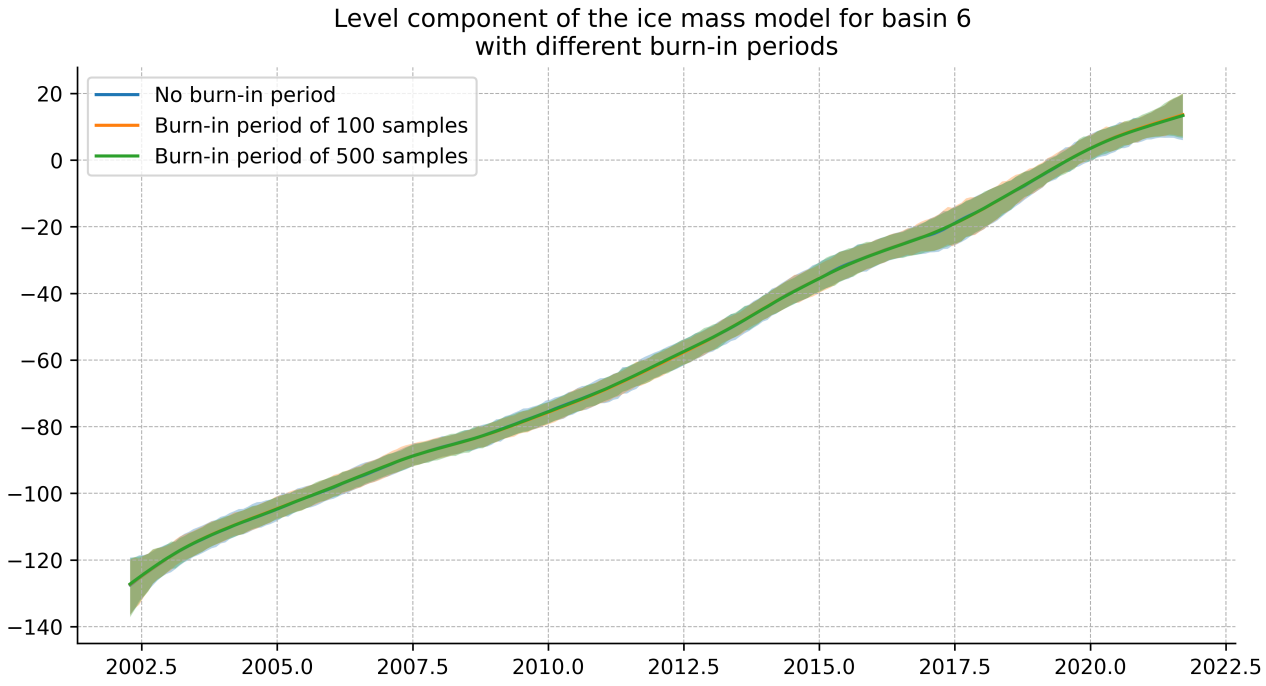


Figure 26: The level of the ice mass model for basin 6 with different burn-in period lengths: No burn-in period (blue), a burn-in period of 100 samples (orange) and a burn-in period of 500 samples (green). No significant differences are seen between the levels obtained using different burn-in period lengths.

Figure 26 shows the burn-in period length has no effect on the level. Thus, in hindsight, the analyses done in this thesis would likely have been fine without a burn-in period. This would have saved computational time.

9.5 Number of chains & samples

Usually, several chains are ran in an MCMC run. Each chain independently explores the parameter space. These chains should be mixed well, meaning different chains should not be stuck in different regions of the posterior distribution that is being sampled. Sampling diagnostics to check this are discussed in section 5.3. Often, 4 chains are used in an MCMC run (Schuett, 2021). More chains may be chosen, increasing the accuracy of the sampling if the chains are mixed well, at the cost of increased computational time. Each chain of an MCMC run can be ran parallel on separate processor cores of a computer. An even number of chains are often used, because modern computer processors generally have an even number of cores.

Each chain of an MCMC run has a certain length, which is equal to the burn-in period plus the number of samples that are actually stored after the MCMC run. Again, a tradeoff has to be made between the accuracy of each chain versus the computational time. In this thesis, 4 chains of 1500 samples are used for the Bayesian results. With a burn-in period of 500 samples, this means 1000 samples are recorded per chain, leading to 4000 recorded samples of each parameter in total.

9.6 Thinning of samples

If the samples in an MCMC chain are strongly autocorrelated, thinning can be done. Thinning refers to the process of removing samples at a certain interval. As a result, the number of samples is reduced. Thinning of an MCMC chain is computationally inefficient, as many generated samples are simply thrown out. Additionally, thinning greatly reduces the precision of the estimated marginal posterior distribution for the parameters (Link and Eaton, 2012). Hence, it is generally not recommended to thin MCMC chains even if there are autocorrelated samples in them.

Significant autocorrelation between samples in MCMC chains often indicate that the model isn't suitable to fit the data. Thus, another way of dealing with a significant amount of autocorrelated samples in a chain is to make changes to the model and start another MCMC run.

10 Conclusion and recommendations

This chapter answers the three research questions posed in section 1.5. Sections 10.1, 10.2 and 10.3 answer research questions 1, 2 and 3, respectively. Recommendations for future research are given in section 10.4.

10.1 First research question

The first research question was as follows.

1. How do the uncertainties of the trends in Frequentist and Bayesian state-space models of ice-dynamical ice mass in Antarctic ice basins compare to each other?

The results in section 7.2 showed significant deviations in the uncertainties of the trends found by the Frequentist and Bayesian approaches. The trends of the ice-dynamical ice mass models were described by the level and slope components, where the level showed the long-term behavior of the ice-dynamical ice mass data and the slope was defined as the change in level between consecutive epochs. The ratios between the average Bayesian uncertainty and the average Frequentist uncertainty for the level and slope components of the ice mass models for basins 6, 13 and 22 are summarized in table 5.

Basin	**Component**	**Ratio Bayesian/Frequentist**
6	Level	1.51
6	Slope	4.03
13	Level	1.24
13	Slope	1.93
22	Level	1.09
22	Slope	1.27

Table 5: Ratio between the average width of the Bayesian and Frequentist 95% highest density intervals of the level and slope components of basins 6, 13 and 22. For example, the Bayesian uncertainty of the level component for basin 6 is, on average, 1.51 times larger than the Frequentist uncertainty.

Table 5 shows that the difference between the Bayesian and Frequentist trend uncertainty is largest for basin 6. For basin 13 and especially for basin 22, the difference is less significant. This is likely because the original GRACE data for basin 6 consists of a combination of surface mass balance changes and ice dynamics. The interactions between surface mass balance changes and ice dynamics introduce additional complexities and sources of error into the modeling process for basin 6. Conversely, the ice mass change of basin 13 is dominated by surface mass balance changes and the ice mass change of basin 22 is dominated by ice dynamics. For these basins, modelling the data is less complex, because the mass changes are primarily caused by only one process.

In conclusion, the uncertainties of the trends in Bayesian state-space models of ice-dynamical ice mass were consistently larger than those in Frequentist models. This discrepancy is caused by the fact that uncertainties in model parameters are propagated to the states when following the Bayesian approach, which does not happen in the Frequentist approach. The difference was more pronounced for basins where the interactions between surface mass balance changes and ice dynamics introduce additional complexities and sources of error into the modeling process.

10.2 Second research question

The second research question was as follows.

2. How can the trends of gravimetry-based and altimetry-based state-space models of ice-dynamical ice mass in Antarctic ice basins be combined?

The trends of gravimetry-based (GRACE-based) and altimetry-based state-space models of ice-dynamical ice mass in Antarctic ice basins were combined using weighted averaging. When following the Frequentist approach, the weights given to the GRACE-based level and the altimetry-based level were equal to the reciprocal of their respective variances. For the Bayesian approach, the weights at each epoch were equal to the reciprocal of the squared width of the 95% highest density intervals.

The results of the weighted averaging of the Frequentist GRACE-based and altimetry-based levels for basins 6, 13 and 22 were shown in figure 22. The same results for the Bayesian approach were shown in figure 23. For all basins, the weighted average aligned very closely to the GRACE-based level. This was the case for both the Frequentist and Bayesian approaches. This alignment of the weighted average with the GRACE-based level occurred because the uncertainty of the GRACE-based level is significantly smaller than the uncertainty

of the altimetry-based level. Such a large difference in uncertainty between the GRACE-based models and the altimetry-based models is an indication that the altimetry data is of lower quality than the GRACE data. This was also found while selecting models to the altimetry data in chapter 4.4. Due to strong fluctuations in the data, likely caused by measurement errors or errors in the firn densification model used to correct the data for firn thickness changes, the state-space models had difficulty decomposing the model into separate components.

In conclusion, gravimetry-based and altimetry-based state-space trends of ice-dynamical ice mass in Antarctic ice basins can be combined using weighted averaging, but due to significant uncertainties in the altimetry data, the weighted average consistently aligns closely to the GRACE-based trend. This is because the GRACE data are of significantly higher quality than the altimetry data.

That the altimetry dataset used in this thesis is of significantly lower quality (i.e., being much more uncertain) than the GRACE data is an important finding that has implications for future research using these data. If these altimetry data are used to extrapolate how the ice volume of the Antarctic will behave in the future, for example, this could lead to bad predictions if their uncertainty is not properly addressed.

10.3 Third research question

Finally, this thesis answers the following sub-question.

3. How do choices that have to be made when working with Bayesian Markov Chain Monte Carlo methods for state-space modelling affect the model?

When following the Bayesian approach to state-space modelling in this thesis, the Hamiltonian No-U-Turn MCMC sampler was used to sample from the model parameters' marginal posterior distributions $P(\theta|y)$. The Bayesian modelling approach was explained in chapter 5. A general overview of the choices that have to be made when working with MCMC methods for state-space modelling, was provided in chapter 9.

Depending on the number of parameters in a state-space model, different types of MCMC samplers are suitable. For a relatively small amount of parameters, Metropolis-Hastings or ensemble samplers are suited (Hogg and Foreman-Mackey, 2018). These samplers fail for larger sets of parameters (Carpenter, 2017). In such cases, Gibbs or Hamiltonian samplers perform better. Hamiltonian MCMC methods are considered the state of the art for models with a large number of parameters (Hogg and Foreman-Mackey, 2018).

Each MCMC method requires priors for their model parameters. Prior distributions contain information about model parameters before data have been viewed. It is generally recommended to use informative priors, as non-informative priors can lead to convergence issues when not enough data is available. If enough high-quality data is available, the choice of prior has little impact on the result. Figure 25 showed that most priors (both informative and non-informative) did not result in differences for the level in this thesis. Only the use of Inverse Gamma priors provided a different result.

MCMC samplers need to be initialized. Generally, initialization for parameters is done by taken values from high likelihood region of the prior distributions, such as the means. For state-space models, the states also need to be initialized. This can be done by initializing the Kalman Filter with either exact diffuse or approximate diffuse initialization. Python software is currently being worked on by the *PyMC-experimental* development team, where priors for the initial states can be specified before sampling, as an alternative to the diffuse initialization of the Kalman Filter.

When working with MCMC methods, it is recommended that a minimum of 4 chains are ran simultaneously. Each chain independently explores the parameter space by sampling from the marginal posterior distributions of the parameters. A burn-in period is often used, whereby an initial number of samples are thrown out in order to reduce the chance of potential biases in the samples caused by the priors and initialization. In this thesis, 4 chains of 1500 samples with a burn-in period of 500 were used when sampling. However, figure 26 showed that the burn-in period length could have been reduced for the analyses done in this thesis, which in turn would have reduced computational costs.

If the samples in an MCMC chain are strongly autocorrelated, a process referred to as thinning can be performed. When thinning an MCMC chain, samples are removed at a certain interval, so that the remaining samples are no longer autocorrelated. This method is not recommended, however, as it is computationally inefficient and actually reduces the precision of the estimated marginal posterior distributions for the parameters that are being sampled (Link and Eaton, 2012). Instead of thinning MCMC chains, it is recommended to make changes to the model when significant autocorrelations are found between samples.

10.4 Recommendations

This thesis focuses on only basins 6, 13 and 22. However, as noted in chapter 2 and shown in appendix A, 34 total basins exist. The research done in this thesis could be expanded to all 34 basins, to more strongly support the conclusions found in this thesis. Expanding the research done in this thesis to all 34 basins would require going through the model selection process for all basins separately, which is relatively time-consuming. However, this could be sped up by reusing the Python code made for this thesis.

The Python code made for this thesis is stored at the Delft University of Technology's GitLab, as described in appendix E. This code may be freely reused and could prove useful in any future projects relating to Frequentist and Bayesian approaches to state-space modelling, even if unrelated to the modelling of ice dynamics. The code is intentionally made reusable for other types of time series data. The code made for this thesis does have limitations. For instance, it currently only allows for univariate state-space models. Additionally, the design matrices of the state-space models used in this thesis do not depend on time. However, this could be required for different state-space models in future research, but the current code does not account for this. In future research, custom state-space models could be made with the same *Statsmodels* Python package used in this thesis (Statsmodels documentation, 2024a).

Additionally, this thesis uses the *PyMC* Python package when working with the Bayesian approach. However, there are alternatives. For example, the Stan programming language allows for the use of several MCMC algorithms. The *PyMC-experimental* Python package is also worth considering in future research. Both Stan and *PyMC-experimental* provide the user more options when working with MCMC than *PyMC*. For example, priors for the initial states' means and variances can be provided in Stan and *PyMC-experimental*, which is currently not possible in *PyMC*.

On the topic of MCMC methods, the use of different MCMC samplers than the No-U-Turn sampler may be worth investigating in future research. The No-U-Turn sampler is used in this thesis as it is considered the state of the art for models with a relatively large number of parameters (Hogg and Foreman-Mackey, 2018). However, it could be that a simpler MCMC method like Metropolis-Hastings is sufficient for the state-space models used in this thesis and provides similar results to the No-U-Turn sampler while being significantly more computationally efficient. Due to time constraints, this was not investigated in this thesis.

When expanding the research done in this thesis to all 34 basins would also be interesting to investigate how different geographic locations on the Antarctic ice sheet behave. The ice mass and volume of Antarctic basins in different geographic regions may be influenced by changes in the surface climate and changes in ice dynamics differently, in turn causing differences between the Frequentist and Bayesian results. Evidence of this is already seen when looking at the results of this thesis, where the ice mass basin 6 was influenced strongly by both changes in the surface climate and ice dynamics. As a result, the difference between the Frequentist and Bayesian state-space model uncertainties for basin 6 were larger than for basins 13 and 22, where ice mass changes were predominantly caused by just one process (changes in the surface climate for basin 13 and ice dynamics for basin 22). By looking at more basins in different areas of Antarctica, these findings may be substantiated.

Currently, research is being done on the accuracy of firn thickness changes obtained by different firn densification models (Kappelsberger et al., 2023). In future research, it will be valuable to investigate the use of such different firn densification models than the ones used in this thesis. Different altimetry datasets may be investigated in general. Higher quality altimetry data would be valuable towards investigating how GRACE and altimetry data can be combined.

In this thesis, the data were not checked for outliers. In hindsight, when viewing the mass data of figure 2 and the volume data of figure 3 again, there do seem to be outliers. For instance, the GRACE data show potential outliers at the final observations of the original GRACE satellite mission, before the data gap between the GRACE and GRACE-FO observations. Outliers in the data may influence the model and could bias the observed states towards the outliers. In future research, outliers could be removed from the data. This can be done with hypothesis tests, such as the Z-score test.

In section 6.1, no state-space models were found for the ice mass data of basins 6 and 22 that passed both the Ljung-Box, Jarque-Bera and heteroskedasticity tests. As discussed in section 4.4, generally, all of these tests should be passed first, before checking the AIC and BIC scores during model selection. In future research, other types of models are worth investigating for basins 6 and 22. For instance, state-space models that are not limited to linear Gaussian conditions may perform better.

References

Albers, C. J., Kiers, H. A. L., and van Ravenzwaaij, D. (2018). Credible Confidence: A Pragmatic View on the Frequentist vs Bayesian Debate. *Collabra: Psychology*, 4(1):31.

Alley, R. B., Clark, P. U., Huybrechts, P., and Joughin, I. (2005). Ice-Sheet and Sea-Level Changes. *Science*, 310(5747):456–460.

Bayarri, M. J. and Berger, J. O. (2004). The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*, 19(1):58 – 80.

Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo.

Bronselaer, B., Winton, M., Griffies, S. M., Hurlin, W. J., Rodgers, K. B., Sergienko, O. V., Stouffer, R. J., and Russell, J. L. (2018). Change in future climate due to Antarctic meltwater. *Nature*, 564(7734):53–58.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Bulthuis, K., Arnst, M., Sun, S., and Pattyn, F. (2019). Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change. *The Cryosphere*, 13(4):1349–1380.

Carpenter, B. (2017). Ensemble Methods are Doomed to Fail in High Dimensions. Statistical Modeling, Causal Inference, and Social Science `https://statmodeling.stat.columbia.edu/2017/03/15/ensemble-methods-doomed-fail-high-dimensions/`.

Chen, J. L., Wilson, C. R., Blankenship, D. D., and Tapley, B. D. (2006). Antarctic mass rates from GRACE. *Geophysical Research Letters*, 33(11).

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82(2):339–350.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. Oxford University Press.

Frederikse, T. (2018). Sea-level changes on multiple spatial scales: estimates and contributing processes. *Dissertation (TU Delft), Delft University of Technology*.

Fulton, C. (2015). Estimating time series models by state space methods in Python: Statsmodels. `https://www.chadfulton.com/files/fulton_statsmodels_2017_v1.pdf`.

Fulton, C. (2017). Maximum likelihood estimation. State Space Estimation of Time Series Models in Python: Statsmodels documentation `https://www.chadfulton.com/fulton_statsmodels_2017/sections/4-maximum_likelihood_estimation.html`.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515 – 534.

Grabowski, J. (2023). 'pymc-experimental' now includes state space models. `https://discourse.pymc.io/t/pymc-experimental-now-includes-state-spaces-models/12773`.

Groh, A. and Horwath, M. (2021). Antarctic Ice Mass Change Products from GRACE/GRACE-FO Using Tailored Sensitivity Kernels. *Remote Sensing*, 13(9).

Groh, A., Horwath, M., Horvath, A., Meister, R., Sørensen, L. S., Barletta, V. R., Forsberg, R., Wouters, B., Ditmar, P., Ran, J., Klees, R., Su, X., Shang, K., Guo, J., Shum, C. K., Schrama, E., and Shepherd, A. (2019). Evaluating GRACE Mass Change Time Series for the Antarctic and Greenland Ice Sheet—Methods and Results. *Geosciences*, 9(10).

Hamilton, J. D. (1994). *Chapter 50 state-space models*, volume 4, pages 3039–3080. Elsevier.

Helm, V., Humbert, A., and Miller, H. (2014). Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2. *The Cryosphere*, 8(4):1539–1559.

Hoffman, M. D. and Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.

Hogg, D. W. and Foreman-Mackey, D. (2018). Data Analysis Recipes: Using Markov Chain Monte Carlo. *The Astrophysical Journal Supplement Series*, 236(1):11.

Ivins, E. R. and James, T. S. (2005). Antarctic glacial isostatic adjustment: a new assessment. *Antarctic Science*, 17(4):541–553.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

Kantz, H. and Schreiber, T. (2004). *Nonlinear time series analysis*. Cambridge university press.

Kappelsberger, M. T., Horwath, M., Buchta, E., Willen, M. O., Schröder, L., Veldhuijsen, S. B. M., Kuipers Munneke, P., and van den Broeke, M. R. (2023). How well can satellite altimetry and firn models resolve Antarctic firn thickness variations? *The Cryosphere Discussions*, 2023:1–47.

Klees, R. (2023). Lecture notes: MCMC methods in Time Series Analysis using State Space Models.

Klees, R. (2024). Lecture slides: Analysis of dynamic time series models. Delft University of Technology. `https://brightspace.tudelft.nl/d2l/le/content/593739/viewContent/3431220/View`.

Kumar, R., Carroll, C., Hartikainen, A., and Martin, O. (2019). ArviZ: a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, 4(33):1143.

Laine, M. (2020). *Introduction to Dynamic Linear Models for Time Series Analysis*, pages 139–156. Springer International Publishing, Cham.

Lam, A. (2020). BFGS in a Nutshell: An Introduction to Quasi-Newton Methods. `https://towardsdatascience.com/bfgs-in-a-nutshell-an-introduction-to-quasi-newton-methods-21b0e13ee504`.

Link, W. A. and Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3(1):112–115.

Menkveld, A., Koopman, S. J., and Lucas, A. (2007). Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods. *Journal of Business & Economic Statistics*, 25(2):213–225.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.

NASA (2024). Antarctic Drainage System Boundaries and Masks. `https://earth.gsfc.nasa.gov/cryo/data/polar-altimetry/antarctic-and-greenland-drainage-systems`.

Nilsson, J., Gardner, A. S., and Paolo, F. S. (2022). Elevation change of the Antarctic Ice Sheet: 1985 to 2020. *Earth System Science Data*, 14(8):3573–3598.

Pek, J. and van Zandt, T. (2020). Frequentist and Bayesian approaches to data analysis: Evaluation and estimation. *Psychology Learning & Teaching*, 19(1):21–35.

PyMC documentation (2021). PyMC NUTS initialization. pymc.init_nuts `https://www.pymc.io/projects/docs/en/stable/api/generated/pymc.init_nuts.html`.

Roy, V. (2020). Convergence Diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7(Volume 7, 2020):387–412.

Rumble, J. (2024). *Properties of Ice and Supercooled Water*. CRC Handbook of Chemistry and Physics, 105th Ed. Section 5: Properties of Water. CRC Press. ISBN 9781032655628.

Schuett, H. (2021). A Brief Explanation of MCMC Sampling. A Primer on Bayesian Inference for Accounting research. `https://hschuett.github.io/BayesForAccountingResearch/mcmc.html`.

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, 2010.

Shepherd, A., Wingham, D. J., Mansley, J. A. D., and Corr, H. F. J. (2001). Inland Thinning of Pine Island Glacier, West Antarctica. *Science*, 291(5505):862–864.

Statsmodels documentation (2024a). Custom state-space models. Statsmodels documentation `https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_custom_models.html`.

Statsmodels documentation (2024b). Statsmodels univariate unobserved components time series model. `https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.structural.UnobservedComponents.html`.

van Wessem, J. M., van de Berg, W. J., Noël, B. P. Y., van Meijgaard, E., Amory, C., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J. T. M., Lhermitte, S., Ligtenberg, S. R. M., Medley, B., Reijmer, C. H., van Tricht, K., Trusel, L. D., van Ulft, L. H., Wouters, B., Wuite, J., and van den Broeke, M. R. (2018). Modelling the climate and surface mass balance of polar ice sheets using RACMO2 – Part 2: Antarctica (1979–2016). *The Cryosphere*, 12(4):1479–1498.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718.

Veldhuijsen, S. B. M., van de Berg, W. J., Brils, M., Kuipers Munneke, P., and van den Broeke, M. R. (2023). Characteristics of the 1979–2020 Antarctic firn layer simulated with IMAU-FDM v1.2A. *The Cryosphere*, 17(4):1675–1696.

Velicogna, I. (2009). Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. *Geophysical Research Letters*, 36(19).

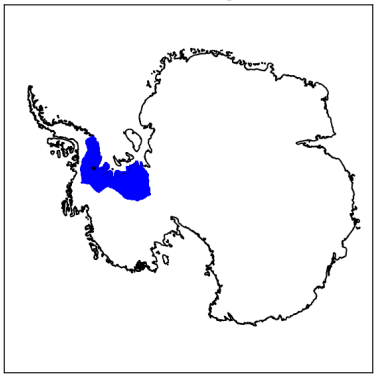
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., Van der Plas, J., and SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Willen, M. O., Broerse, T., Groh, A., Wouters, B., Kuipers Munneke, P., Horwath, M., van den Broeke, M. R., and Schröder, L. (2021). Separating Long-Term and Short-Term Mass Changes of Antarctic Ice Drainage Basins: A Coupled State Space Analysis of Satellite Observations and Model Products. *Journal of Geophysical Research: Earth Surface*, 126(6).

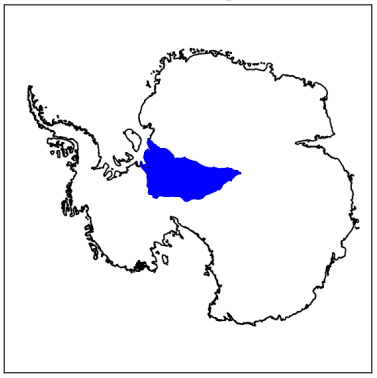
Zhang, B., Wang, Z., Yang, Q., Liu, J., An, J., Li, F., Liu, T., and Geng, H. (2020). Elevation Changes of the Antarctic Ice Sheet from Joint Envisat and CryoSat-2 Radar Altimetry. *Remote Sensing*, 12(22).

Appendix A: Location of all 34 Antarctic ice drainage basins

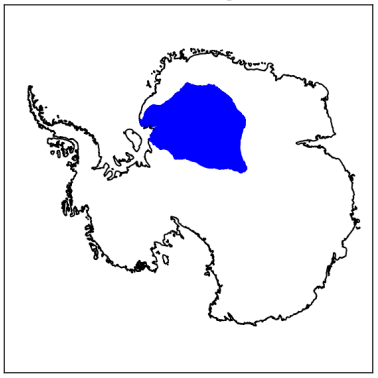
Antarctic ice drainage basin 1



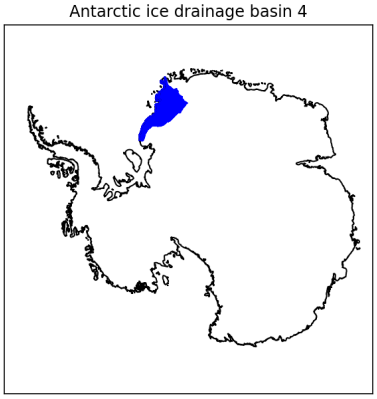
Antarctic ice drainage basin 2



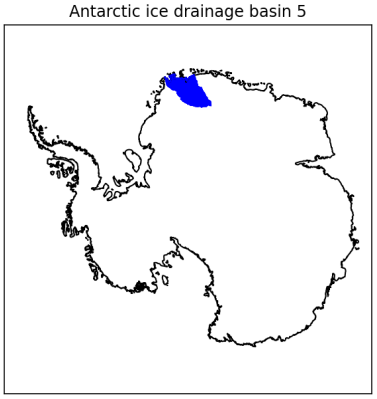
Antarctic ice drainage basin 3



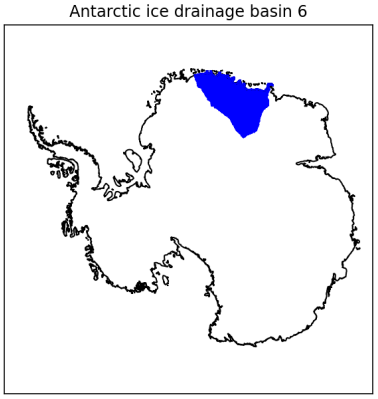
Antarctic ice drainage basin 4



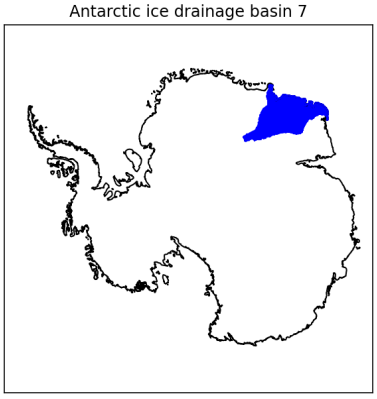
Antarctic ice drainage basin 5



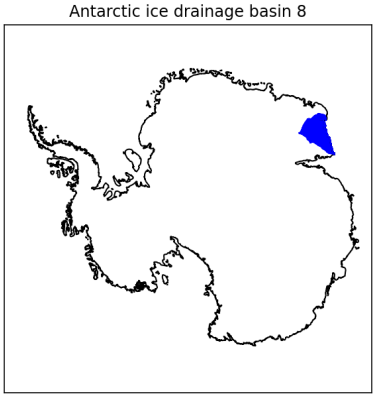
Antarctic ice drainage basin 6



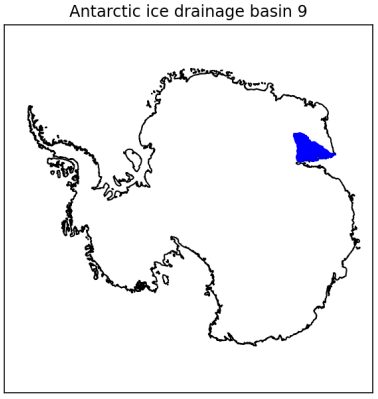
Antarctic ice drainage basin 7



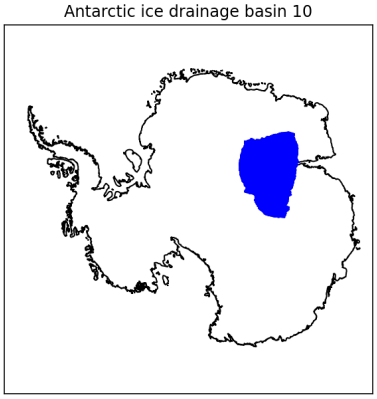
Antarctic ice drainage basin 8



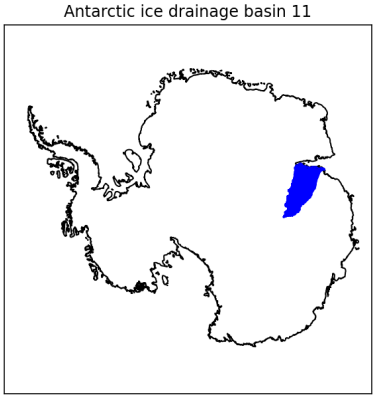
Antarctic ice drainage basin 9



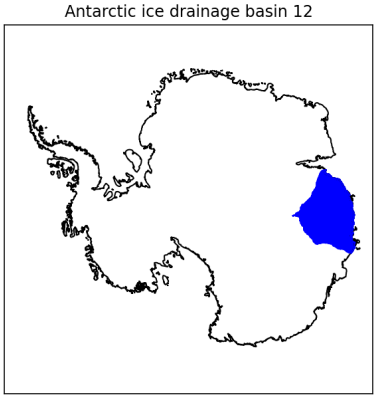
Antarctic ice drainage basin 10



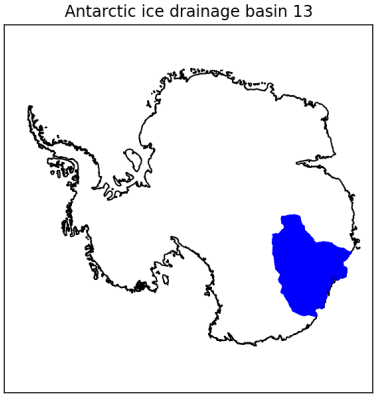
Antarctic ice drainage basin 11



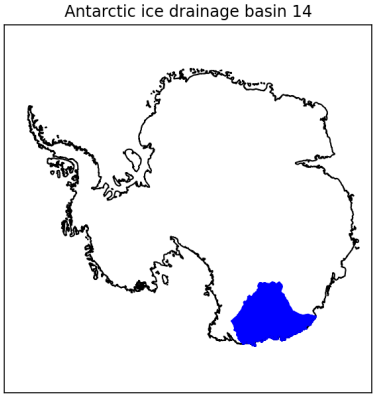
Antarctic ice drainage basin 12



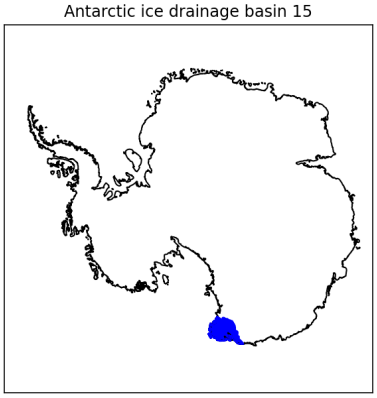
Antarctic ice drainage basin 13



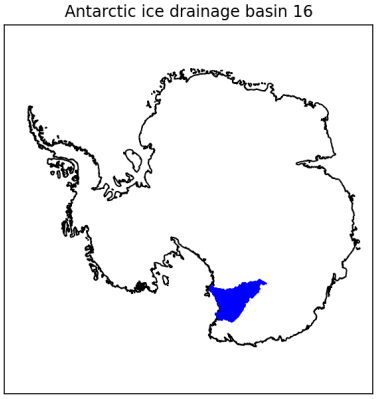
Antarctic ice drainage basin 14



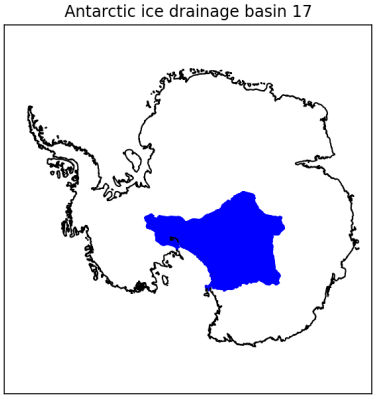
Antarctic ice drainage basin 15



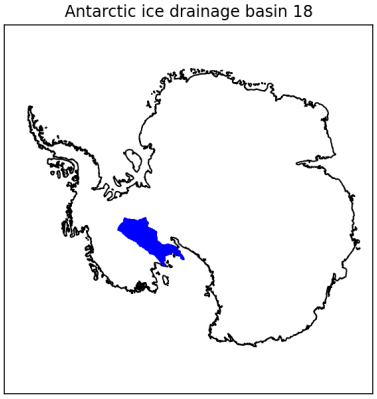
Antarctic ice drainage basin 16



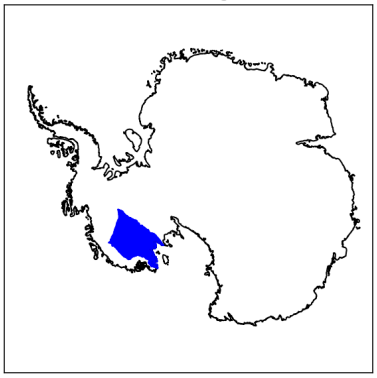
Antarctic ice drainage basin 17



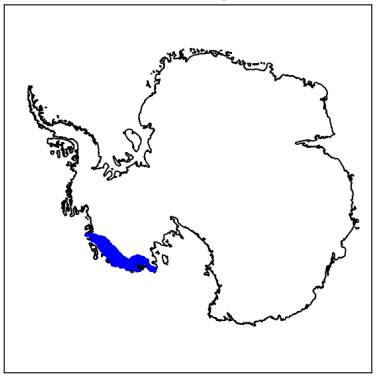
Antarctic ice drainage basin 18



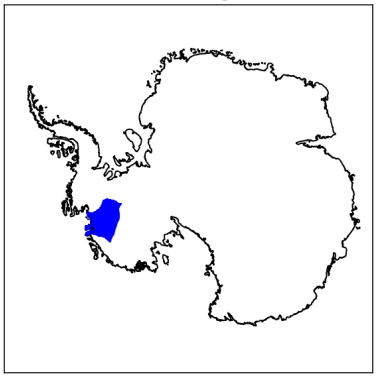
Antarctic ice drainage basin 19



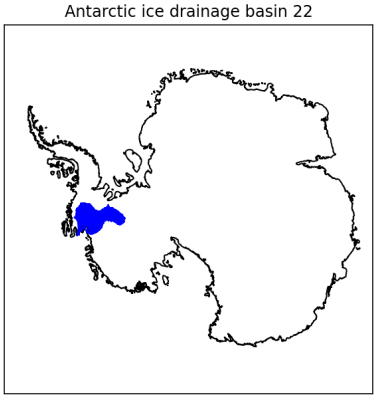
Antarctic ice drainage basin 20



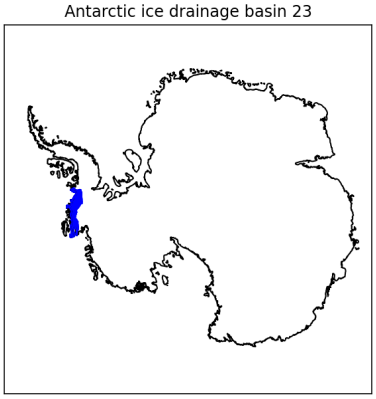
Antarctic ice drainage basin 21



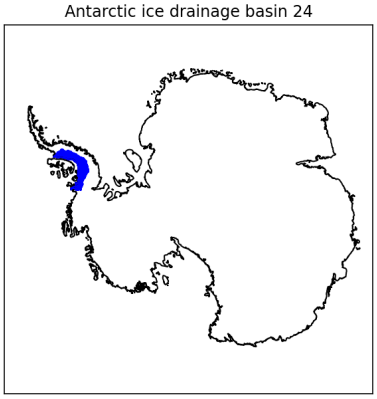
Antarctic ice drainage basin 22



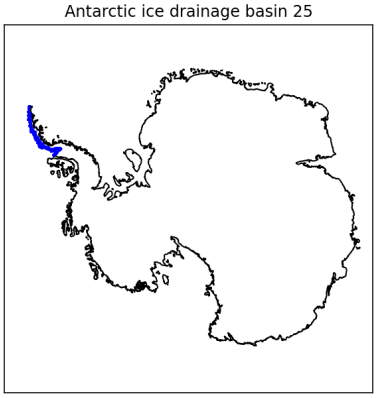
Antarctic ice drainage basin 23



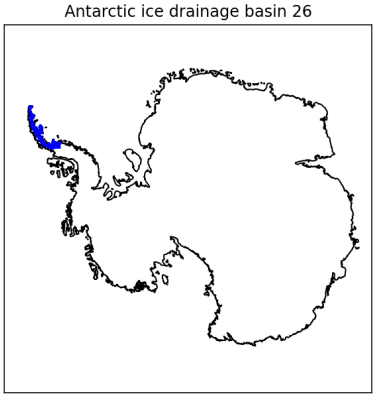
Antarctic ice drainage basin 24



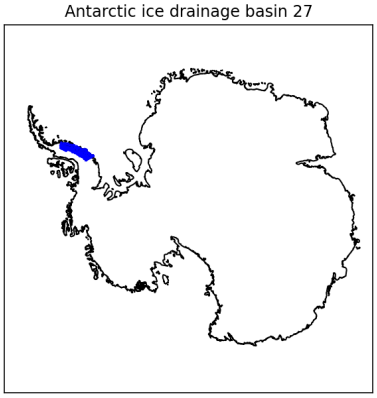
Antarctic ice drainage basin 25



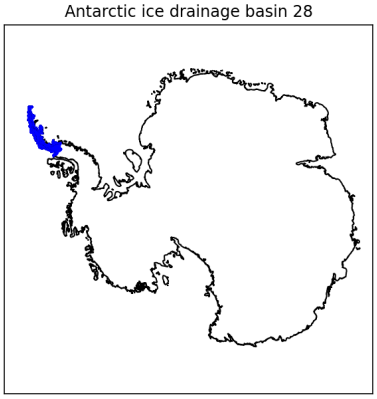
Antarctic ice drainage basin 26



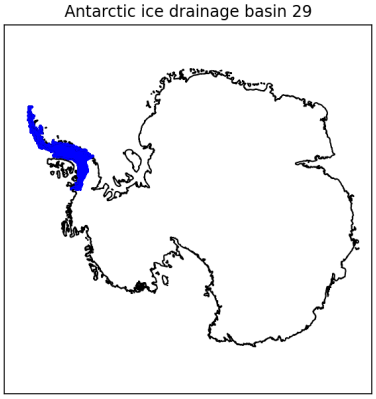
Antarctic ice drainage basin 27



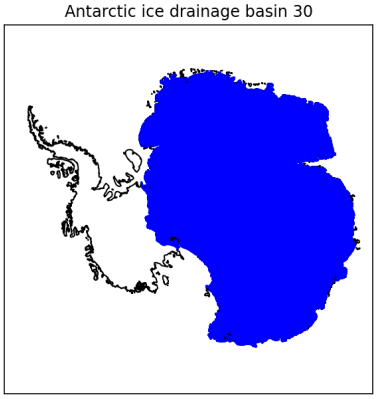
Antarctic ice drainage basin 28



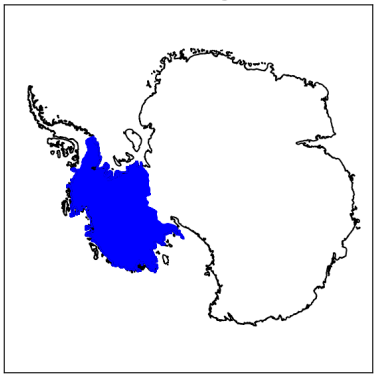
Antarctic ice drainage basin 29



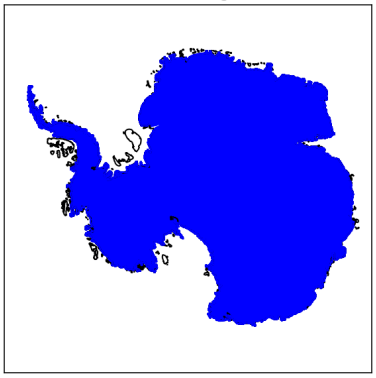
Antarctic ice drainage basin 30



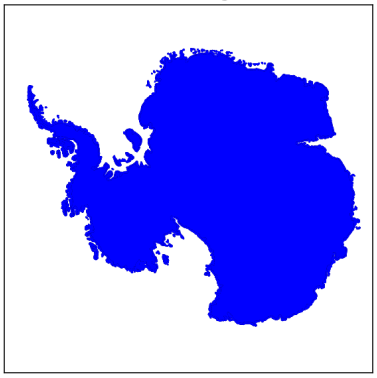
Antarctic ice drainage basin 31



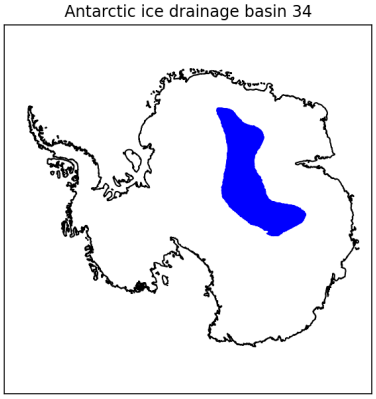
Antarctic ice drainage basin 32



Antarctic ice drainage basin 33



Antarctic ice drainage basin 34



Appendix B: Satellite gravimetry and altimetry

Gravimetry data from the GRACE and GRACE-FO satellites are used in this thesis, alongside altimetry data from a multi-mission satellite altimetry dataset. The data are explained in detail in chapter 2 This appendix briefly explains the concepts of satellite gravimetry and satellite altimetry.

Satellite gravimetry

Both the GRACE and GRACE-FO missions are comprised of two satellites in the same orbit, that continuously measure their locations and the distance between the two satellites. Regions of the planet where more mass is concentrated pull the leading satellite slightly away from the trailing satellite. When the trailing satellite reaches this same area, it is pulled closer to the leading satellite. These changes in distance are very small, but can be accurately measured by the microwave ranging systems on board of the twin satellites. Accelerations that are not caused by gravity, such as atmospheric drag, are measured by accelerometers on the satellites. This way, accelerations caused by changes in gravity due to mass anomalies can be clearly distinguished from other accelerations. Global Positioning System (GPS) receivers determine the position of the satellites over the Earth, so that the areas where these mass anomalies occur can be located.

Satellite altimetry

Altimetry satellites measure the surface elevation. This is done by sending out a radar or laser pulse and measuring the time it takes for the signal to arrive back to the satellite after reflecting from the Earth's surface. The distance the signal travelled can be inferred from this, which, in combination with the known altitude of the satellite's orbit, can be converted to surface height. By aggregating altimetry data over a certain surface area, such as the size of an Antarctic ice basin, and multiplying the elevation by that surface area, volume data are obtained.

Appendix C: Hamiltonian dynamics in MCMC

In mechanics, Hamiltonian dynamics are commonly used to describe physical systems, such as the movement of an object. Hamiltonian dynamics describe a system by a function of position vector q and momentum vector p, known as the Hamiltonian $H(q,p)$. The partial derivatives of the Hamiltonian $H(q,p)$ describe how position q and momentum p evolve over time, and are referred to as Hamilton's equations (Brooks et al., 2011). Hamilton's equations are given in equation 49 and 50.

Hamilton's equations

$$\frac{dq_n}{dt} = \frac{\delta H(q,p)}{\delta p_n} \tag{49}$$

$$\frac{dp_n}{dt} = -\frac{\delta H(q,p)}{\delta q_n} \tag{50}$$

The Hamiltonian $H(q,p)$ is the total energy in the system, defined as the sum of the potential energy E_{pot} and the kinetic energy E_{kin} (Brooks et al., 2011). The potential energy E_{pot} is a function of the position vector q and the kinetic energy E_{kin} is a function of the momentum vector p. The Hamiltonian is shown in equation 51.

Hamiltonian

$$H(q,p) = E_{pot}(q) + E_{kin}(p) \tag{51}$$

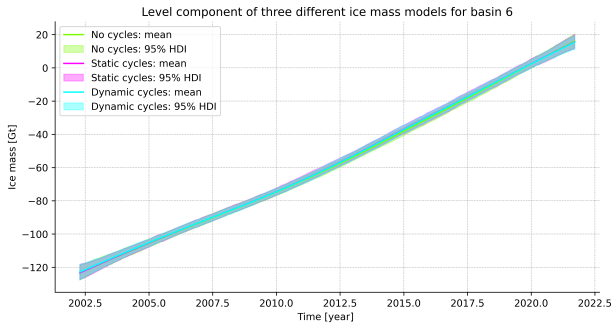
The use of Hamiltonian dynamics isn't limited to just physical systems. Hamiltonian MCMC methods use the same concept, but instead of describing a physical system, a probabilistic system is described.

When Hamiltonian MCMC methods are used to sample from the marginal distribution of model parameters given the data $P(\theta|y)$, the position vector q is equal to the vector of model parameters θ. The momenta in vector p are auxiliary variables used to guide the exploration of the parameter space Θ. The potential energy E_{pot} is equal to minus the natural logarithm of the marginal density of $P(\theta|y)$. Finally, the kinetic energy E_{kin} corresponds to minus the natural logarithm of the probability density function of a standard normal distribution (Brooks et al., 2011).

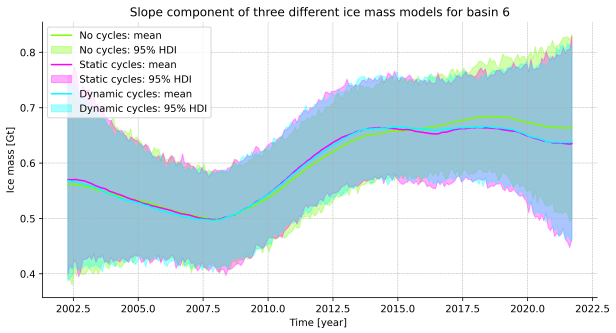
As the position vector q is analogous to the parameter vector θ in Hamiltonian MCMC, equation 49, gives the slope of θ. Candidate values for θ can be determined by numerically integrating equation 49 at each iteration. To do this, the corresponding auxiliary variable at each iteration is needed, obtained by integrating equation 50. In practice, the auxiliary variables are typically drawn from a standard normal distribution. Different Hamiltonian samplers use different numerical methods to generate candidate values for θ. The sampler used in this thesis is the Hamiltonian No-U-Turn sampler, explained in section 5.1.1.

Appendix D: Plots of the level and slope for model selection

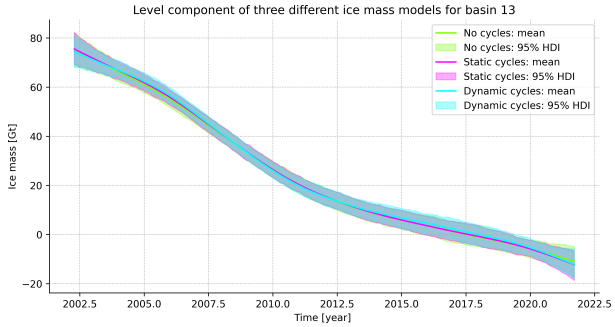
For basins 6, 13 and 22, figure 27 shows the Frequentist level and slope components of three different mass models; integrated random walks with either no cycles, static cycles or dynamic cycles.



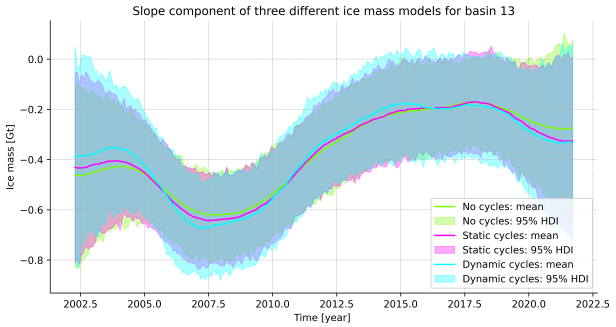
Basin 6: level



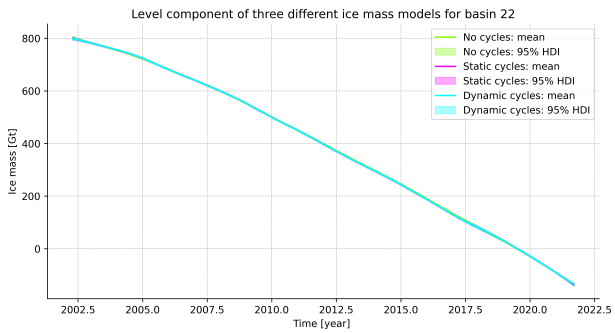
Basin 6: slope



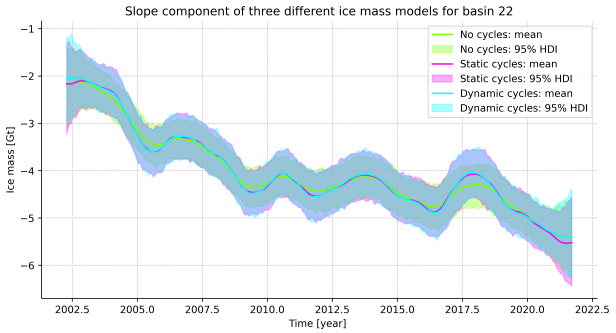
Basin 13: level



Basin 13: slope



Basin 22: level



Basin 22: slope

Figure 27: Mean Frequentist level and slope components, with 95% highest density intervals, for three different ice mass models for basins 6, 13 and 22; an integrated random walk (green), an integrated random walk with a static annual cycle and a static SA tidal alias cycle (magenta) and an integrated random walk with a dynamic annual cycle and a dynamic SA tidal alias cycle (cyan).

The uncertainties of the Frequentist levels and slopes shown in figure 27 are represented with 95% highest density intervals, instead of 95% confidence intervals. The difference between highest density intervals and confidence intervals, and the reason for using highest density intervals in this thesis, is explained when discussing the results in section 7.1. The 95% highest density intervals are centered around the mean level. Figure 27 shows no significant differences between the three types of models.

Appendix E: Code repository

The data and Python code used for this thesis can be found at: `https://gitlab.tudelft.nl/tsa-ssm/mcmc/`. The GitLab repository contains a README text file, which describes the files stored in the repository and contains an installation guide.