

Analysis of Gravity Recovery and Climate Experiment (GRACE) data

Andreas Madsen – s123598
Frederik Wolgast Rørbech - s123956

June 24th, 2014

Abstract

Understanding the effect of global warming is important for predicting the future of the globe. As the temperature of the globe increases, ice melts and water flows into the oceans. This report analyses local mass losses on the Earth's surface which could be caused by glaciers turning into water. Several different approaches such as Ordinary Least Squares, autocorrelated OLS, basis expansion and Principle Component Analysis were used.

These methods lead to estimate of velocity and acceleration of the mass loss in which it was found that the West and East coast of Greenland as well as Western Antarctica are losing mass the fastest as well as also having the largest acceleration of mass loss. Surprisingly while the largest velocity in Greenland was found on the East Coast, the West Coast was found to have a higher acceleration.

Unsupervised clustering methods were also used and a Gaussian mixture model chained with kernel PCA gave amazing results. The results corresponds well to the velocity estimates and also revealed other interesting geophysiological phenomena.

Contents

1	Introduction	3
2	Problem definition	3
3	Data	3
3.1	Data example	4
4	Theory	6
4.1	SVD	6
4.2	OLS	7
4.3	The design matrix	9
4.4	Phase and Amplitude	10
4.5	OLS with autocorrelated residuals	10
4.6	Least angular regression (LAR)	12
4.7	Basis expansion	12
4.8	Time Series Analysis	13
4.9	Ljung-Box test	14
4.10	K-means clustering	15
4.11	Gaussian Mixture Model (GMM)	16
4.12	Kernel PCA	17
5	Results	21
5.1	OLS	21
5.2	Glacial Isostatic Adjustment	33
5.3	Time series analysis - ARIMA	34
5.4	Least angular regression (LAR)	40
5.5	OLS with autocorrelated residuals	41
5.6	Splines	42
5.7	k-means clustering	45
5.8	GMM clustering	47
6	Conclusion	49
7	Appendix	51
7.1	Time series analysis results	51

1 Introduction

Lately, the climate has become a very hot topic. In recent times most natural science researchers have reached a consensus; the globe is heating up and the primary catalyst of this process is human carbon dioxide emission. One of the consequences of a warmer Earth is melting ice at various locations. This leads to rising oceans which could cause a lot of damage to lowlands. In Denmark an example of the latter would be the marsh area in West Jutland. Thus, an important question is exactly how the process of ice melting is evolving.

Our data source is Gravity and Climate Experiment (GRACE) [1]. GRACE is a government funded research project which began collecting data in 2002, its mission is to track the gravitational changes on Earth. The data is captured using two satellites trailing each other while orbiting the Earth. By measuring the distance between the satellites one can estimate the strength of the gravitational field. Since gravity is caused by mass, the data can be interpreted as reductions or increases in mass, which may be caused by ice melting.

This report will seek to uncover locations, which are experiencing a significant mass gain/loss, using mathematical models. In addition to analysis of the spatial variance, variation in the time domain will also be analyzed. Finally, results and their uncertainties will be commented on.

2 Problem definition

Specifically this report will focus on analyzing local mass losses on the surface of the Earth, partially those changes near Greenland and the Antarctica.

- Is the changes accelerating or decelerating?
- Does it fluctuate?
- Does ice at the coasts of the Antarctica melt as fast as ice at the coasts of Greenland?

3 Data

The GRACE dataset can be downloaded from GRGS [1]. This report is based on the Equivalent Water Height (denoted EWH) dataset from release 2 in the GRGS format with a 10-day interval.

This dataset contains quite a few text files, each containing information in both its filename and its content. The filenames have the format:

```
1 || grid.water.10day_model_minus_RL02MF.19202_19211.txt  
2 || grid.water.10day_model_minus_RL02MF.19212_19221.txt  
3 || ...
```

In the filename, the last two numbers (e.g. 19202_19211) are important. The first number denotes the start date and the second number the end date for the file content. Each number specifies the number of days since “1950-01-01” [2].

The actual content of each text file should be read as a “Space Separated Values” format. When this is done one will have a 6480×10 matrix. This matrix can then be `reshape`’ed row-wise into a 180×360 matrix. The result is a matrix with decreasing latitude on the rows and increasing longitude on the columns [3].

3.1 Data example

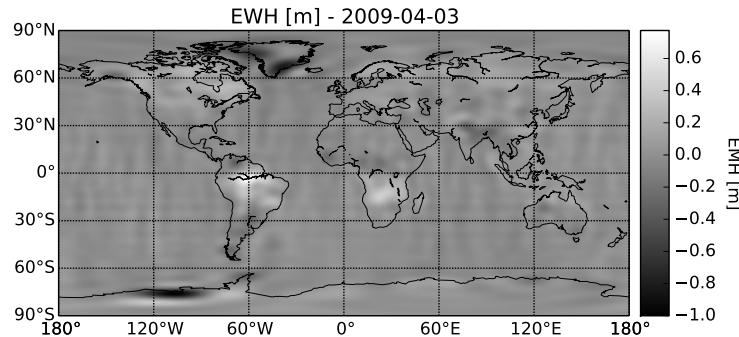


Figure 3.1: Plot of the data, 10 day period starting on 3 April 2009

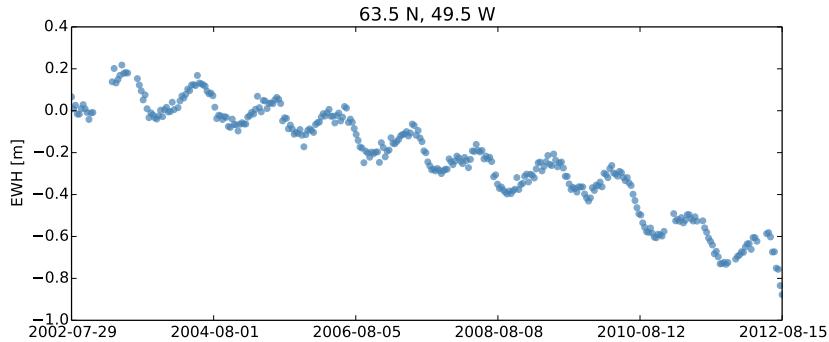


Figure 3.2: Plot of EWH at 63.5 N 49.5 W, west coast of Greenland.

From Figure 3.1 local mass losses at Greenland and the South Pole are seen. In Figure 3.2, the mass loss at Greenland is seen with a yearly periodic trend. For both locations, the ice melting may be caused by ice melting. However it is also possible that these changes are caused by post glacial rebound, in partially because no glacial isostatic adjustment (GIA) have been performed on this version of the GRACE dataset. In fact it is quite complicated to do sufficient adjustments to conclusively say something about ice melting [4]. When

that is said, attempts to analyze the GRACE data with the prospect of saying something about the ice melting patterns, will still be made.

From Figure 3.1 a mass increment in South America can also been seen, this is caused by the rain season.

4 Theory

4.1 SVD

Singular Value Decomposition (SVD) is a useful method for factorizing matrices. According to the SVD-theorem a matrix X can be expressed as

$$X = U\Sigma V^T. \quad (4.1)$$

If the matrix X has the size $m \times n$, the factorized components have the following properties [5]:

- The columns in U are the eigenvectors of XX^T , with the size $m \times m$.
- The columns in V are the eigenvectors of X^TX , with the size $n \times n$.
- Σ is a diagonal matrix, consisting of the square root of the corresponding eigenvalues to the eigenvectors in U and V . It has the size $m \times n$. The square root eigenvalues are also called “singular values”.

In the case where X has many more rows than columns, the U matrix will contain more eigenvectors than are strictly needed to reconstruct X and the Σ matrix will have lots of zero-rows. A so called *economy-size* SVD is therefore often used. The difference being that all the redundant eigenvectors in U are removed, so it has the size $m \times n$ and the zero-rows from Σ are also removed, so its size becomes $n \times n$.

An important property of SVD is that the V matrix contains all eigenvectors and is therefore an orthogonal matrix, for which it is known that

$$Q^T = Q^{-1} \quad QQ^T = I \quad Q^T Q = I, \text{ where } Q \text{ is an orthogonal matrix.} \quad (4.2)$$

In the normal SVD (not *economy-size*) U is also a complete orthogonal matrix. However in *economy-size* some of the columns have been removed, so it is only known that $U^T U = I$.

It is standard procedure to rearrange the matrices in $U\Sigma V^T$, so the largest values in Σ appears first in the diagonal. The reason being that V is a change of basis matrix – it transforms elements in the original space to a new orthogonal space. Here the first axis has the largest variance, the second axis has the second highest variance and so on. The variance described by each axis, is given by the diagonal elements in Σ^2 . Typically one calculates a percentage called “variance explained by principal components” using

$$\rho_{jj} = \frac{\Sigma_{jj}^2}{\sum_{i=1}^n \Sigma_{ii}^2}. \quad (4.3)$$

It is often seen that most of the variance in X can be described using very few eigenvectors, thus one can reduce the dimensionality of a dataset to the most “principal” components. The above mentioned analysis goes under the umbrella term “Principal Component Analysis (PCA)”, where a “principal component” (loadings) should be understood as the axes in the transformed space (i.e., the eigenvectors).

4.2 OLS

OLS (Ordinary Least Squares) regression is used for finding the best linear unbiased estimator(BLUE) of Y base on one or more exogenous variables X . This is done by a matrix-vector product between a β vector and X . Since X can be custom tailored for various purposes it is often referred to as the design matrix. The model is

$$Y = X\beta + \epsilon \quad , \text{ where } E[\epsilon] = 0 \text{ and } D[\epsilon] = D[Y] = \sigma^2 I. \quad (4.4)$$

In the equation above the β -vector is the unknown parameter which needs to be determined while ϵ is a vector containing the residuals which the model cannot account for. Often the purpose of OLS is to predict Y , but in this case we want to analyze the individual β_i elements.

4.2.1 The solution to the OLS problem

As the name sugest, the OLS method minimizes the sum of squared residuals ($\epsilon^T \epsilon$). It is seen directly that $\epsilon = Y - X\beta$ and thus the following is obtained

$$\begin{aligned} \epsilon^T \epsilon &= (Y - X\beta)^T (Y - X\beta) \\ &= (Y^T - \beta^T X^T)(Y - X\beta) \\ &= Y^T Y - \beta^T X^T Y - Y^T X \beta + \beta^T X^T X \beta \\ &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta. \end{aligned} \quad (4.5)$$

One can now differentiate with respect to the β vector

$$\frac{\partial \epsilon^T \epsilon}{\partial \beta} = -2X^T Y + 2X^T X \beta = 2(-X^T Y + X^T X \beta). \quad (4.6)$$

Now $\epsilon^T \epsilon$'s minimum is given by solving for $\frac{\partial \epsilon^T \epsilon}{\partial \beta} = 0$

$$\begin{aligned} \frac{\partial \epsilon^T \epsilon}{\partial \beta} = 0 &\Rightarrow 2(-X^T Y + X^T X \hat{\beta}) = 0 \\ X^T X \hat{\beta} &= X^T Y \\ \hat{\beta} &= (X^T X)^{-1} X^T Y. \end{aligned} \quad (4.7)$$

The solution above is formally correct [6, p. 12], but if X is badly conditioned ($X^T X$) $^{-1}$ might not be numerically stable (i.e., if the columns in X are highly correlated leading to a near singular $X^T X$) [7, p. 8]. To avoid this problem X should be factorised and the solution reformulated; in this report we will exclusively do this using SVD

$$\begin{aligned} (U\Sigma V^T)^T (U\Sigma V^T) \hat{\beta} &= (U\Sigma V^T)^T Y \\ V\Sigma^2 V^T \hat{\beta} &= V\Sigma U^T Y \\ (V\Sigma^{-2} V^T) V\Sigma^2 V^T \hat{\beta} &= (V\Sigma^{-2} V^T) V\Sigma U^T Y \\ \hat{\beta} &= V\Sigma^{-1} U^T Y. \end{aligned} \quad (4.8)$$

It should be noted, that when multiple $\hat{\beta}$ vectors need to be calculated (one vector for each spatial location on the surface) optimization is possible, by arranging Y as a matrix with each column corresponding to a location. $\hat{\beta}$ will then be a matrix containing all solutions instead of a vector containing only one solution.

4.2.2 The “Hat” matrix

In the special case of the GRACE data, the X matrix is identical for every position. This can be exploited by constructing a hat matrix H which only depends on X and projects Y onto \hat{Y} (puts the hat on Y).

$$\begin{aligned}\hat{Y} &= X\hat{\beta} \Rightarrow \hat{Y} = X V \Sigma^{-1} U^T Y \\ \hat{Y} &= H Y \quad , \text{ where } H = X(X^T X)^{-1} X^T = X V \Sigma^{-1} U^T.\end{aligned}\tag{4.9}$$

As earlier with β , the hat matrix can be calculated for all Y vectors by vertically stacking Y to form a matrix.

An important property of the H matrix is that it is idempotent ($H^2 = H$) and symmetrical ($H^T = H$). This is because H is a projection matrix which projects Y onto \hat{Y} . Projecting Y onto \hat{Y} and then projecting again onto \hat{Y} , will obviously not change anything, because one is already in the \hat{Y} -plane, spanned by the columns of X .

4.2.3 Root Mean Squared Error

The “Root Mean Squared Error” (RMSE) is an indicator of how good an Y estimate is. It can be calculated as

$$\hat{\sigma} = \sqrt{\frac{(Y - \hat{Y})^T (Y - \hat{Y})}{n - m}},\tag{4.10}$$

where m is the number of parameters (elements in β) and n is the number of observations (elements in Y). RMSE is an estimate for the standard deviation of Y [8, theorem 3.4] thus the symbol $\hat{\sigma}$.

4.2.4 The variance of \hat{Y}

The dispersion (variance-covariance) of \hat{Y} can be calculated as

$$D[\hat{Y}] = D[HY] = HD[Y]H^T = \sigma^2 H^2 = \sigma^2 H.\tag{4.11}$$

The variance of \hat{Y}_i is given by the diagonal elements in $D[\hat{Y}]$

$$\text{Var}[\hat{Y}_i] = \sigma^2 H_{ii}.\tag{4.12}$$

Because σ^2 is a scalar, the diagonal in H is important to examine, since it can reveal potential elements (in our case points of time) with high variance in the predictions.

4.2.5 The dispersion of $\hat{\beta}$

The dispersion of $\hat{\beta}$ is calculated as [8, theorem 3.2]:

$$D[\hat{\beta}] = \sigma^2(X^T X)^{-1} = \sigma^2 V \Sigma^{-2} V^T \quad (4.13)$$

Since σ^2 is a scalar and dependent of Y , a similar spatial independent expression can be made, by removing the scalar σ^2 factor, thus looking exclusively at $V \Sigma^{-2} V^T$.

4.2.6 p-values for $\hat{\beta}$

Assuming the residuals are normally distributed, the p-values for OLS parameters can be calculated using the student's t-distribution, with the t-score [8, p. 172]

$$t = \frac{\hat{\beta}_i}{SD[\hat{\beta}_i]} \quad \text{where: } SD[\hat{\beta}_i] = \sqrt{\text{Cov}[\hat{\beta}]_{ii}} = \hat{\sigma} \sqrt{(V \Sigma^{-2} V^T)_{ii}}. \quad (4.14)$$

Now by plugging the t-score into the Student's Cumulative distribution function (Φ_t) with $N - p$ degrees of freedom, we get

$$p = 2 \cdot \Phi_t(\text{abs}(t), N - p) \quad (4.15)$$

The null-hypothesis is that $\beta_i = 0$ and the alternative hypothesis is $\beta_i \neq 0$.

4.3 The design matrix

In OLS regression, a linear combination of the columns in the design matrix X , is used to predict Y . As a starting point, the only values in X are a column of ones (to fit an intercept) and a column containing the observed time t

$$X = [\mathbf{1} \quad \mathbf{t}]. \quad (4.16)$$

In the above case one would predict \hat{Y} as $\beta_1 + \beta_2 t$ where β_1 will correspond to the intercept and β_2 will correspond to the speed of mass gain. From this it follows naturally to include the acceleration by adding a column

$$X = [\mathbf{1} \quad \mathbf{t} \quad \frac{1}{2}\mathbf{t}^2] \quad (4.17)$$

From Fourier analysis it is known that functions can be approximated by infinite sums of the trigonometric functions sines and cosines. Trigonometric functions are periodic in their behavior and thus are able to model periodic signals in the function that one wishes to approximate.

In the GRACE dataset, however, the measurements are recorded with a frequency of $\frac{1}{10}$ days. The Nyquist-Shannon sampling theorem states that one cannot find periodic signals above the Nyquist frequency which is given as half

of the sample frequency. Therefore an infinite sum is not suited for approximating the function. Instead it is used, that the maximal frequency in the data has a period of a year ($T = 365.242$ days) while the minimal frequency will have 18 periods per year since $365.242/18 \approx 20.29$ days. Hence it follows that the final design matrix will be given as

$$X = \begin{bmatrix} 1 & t & \frac{1}{2}t^2 & \cos\left(\frac{2\pi}{365.242}t\right) & \sin\left(\frac{2\pi}{365.242}t\right) & \cdots & \cos\left(\frac{2\pi}{\frac{365.242}{18}}t\right) & \sin\left(\frac{2\pi}{\frac{365.242}{18}}t\right) \end{bmatrix}.$$

From this it's seen that the angular frequency is $\omega_i = \frac{2\pi}{365.242/i} = \frac{2\pi i}{T}$

4.4 Phase and Amplitude

The $\hat{\beta}$ parameter in the OLS problem contains the coefficients to the linear combinations of $\cos(\omega_i t)$ and $\sin(\omega_i t)$, there are used to predict Y . Since there is a cosine and sine for each frequency in the design matrix X , one will end up with a linear combination of frequency function pairs. This makes interpretation difficult. Instead one should convert the β parameters for the trigonometric functions into amplitudes and phases corresponding to each pair of periodic functions such that only a cosine $A_i \cos(\omega_i t - \phi_i)$ is needed. This is done by using Ptolemy's theorem

$$A_i \cos(\omega_i t - \phi_i) = A_i \cos(\phi_i) \cos(\omega_i t) + A_i \sin(\phi_i) \sin(\omega_i t). \quad (4.18)$$

Comparing with the linear combination from OLS

$$\hat{Y} = \cdots + \beta_{c,i} \cos(\omega_i t) + \beta_{s,i} \sin(\omega_i t) + \cdots \quad (4.19)$$

It is seen that

$$\beta_{c,i} = A_i \cos(\phi_i) \quad \text{and} \quad \beta_{s,i} = A_i \sin(\phi_i). \quad (4.20)$$

By dividing these two equations with each other, ϕ_i can be calculated as

$$\frac{A_i \sin(\phi_i)}{A_i \cos(\phi_i)} = \frac{\beta_{s,i}}{\beta_{c,i}} \Rightarrow \phi_i = \arctan\left(\frac{\beta_{s,i}}{\beta_{c,i}}\right). \quad (4.21)$$

To isolate A_i , square both equations and add them together

$$A_i^2 \cos(\phi_i)^2 + A_i^2 \sin(\phi_i)^2 = \beta_{c,i}^2 + \beta_{s,i}^2 \Rightarrow A_i = \sqrt{\beta_{c,i}^2 + \beta_{s,i}^2}. \quad (4.22)$$

The result can be plotted with a circular color scale, using ϕ_i (the hue) and A_i (the intensity).

4.5 OLS with autocorrelated residuals

In OLS it is assumed that the residuals between different observations are uncorrelated. However since our data is a time series this assumption does most likely

not hold. This can also be confirmed with the Durbin-Watson test statistic [9, p. 173]. This test statistic is calculated using

$$d = \frac{\sum_{i=2}^n (\hat{\epsilon}_i - \hat{\epsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\epsilon}_i^2} \quad (4.23)$$

which lies between 0 and 4. If d is close to 0 the residuals $\hat{\epsilon}_i$ and $\hat{\epsilon}_{i-1}$ are correlated positively, if close to 4 the residuals are negatively correlated. The purpose of this method is to reduce the correlation between the residuals, the Durbin-Watson test statistic should then become 2.

When the residuals are correlated, one gets bad estimation of the variance of the residuals $\hat{\sigma}_\epsilon^2$. This in turn leads to bad inference statistics, such as misleading p-values.

To correct for the correlated residuals, one can use a general least squares (GLS) regression. This uses a Σ matrix there will correct for the correlated residuals

$$\min_{\beta, \rho} (Y - X\beta)^T \Sigma^{-1} (Y - X\beta), \quad (4.24)$$

where Σ^{-1} is given as

$$\Sigma^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \quad (4.25)$$

The optimization problem (4.24) is nonlinear and there is no closed form solution to this problem. However when keeping ρ constant the β parameters can be estimated using General Least Squared, which is similar to OLS but with a constant Σ matrix and have the solution [8, p. 38]

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y. \quad (4.26)$$

This should then be rewritten using SVD, to account for a near singular $X^T \Sigma^{-1} X$ matrix.

Similarly when β is kept constant, ρ can be estimated with [9, p. 178]

$$\hat{\rho} = \frac{\sum_{i=2}^n \hat{\epsilon}_i \hat{\epsilon}_{i-1}}{\sum_{i=2}^{n-1} \hat{\epsilon}_i^2}. \quad (4.27)$$

A practical way of solving the problem with respect to both β and ρ , looks as follows:

1. initialize by letting $\rho = 0$
2. Iterate:

- (a) Keep $\hat{\rho}$ constant and estimate β using as a WLS problem.
- (b) Keep $\hat{\beta}$ constant and estimate ρ using (4.27).
- (c) Repeat until convergence of ρ or until some predetermined upper iteration boundary is met.

4.6 Least angular regression (LAR)

The LAR model is almost equivalent to the linear regression, except that instead of only minimizing the sum of squares, an ℓ^1 norm penalty is added on the $\hat{\beta}$ -vectors coefficients.

$$\min_{\hat{\beta}} (Y - \hat{Y})^T (Y - \hat{Y}) \quad \text{subject to} \quad \|\hat{\beta}\|_1 \leq s \quad (4.28)$$

The LAR algorithm [10] solves this problem by initially letting all the β coefficients be zero. It then finds the attribute, b_1 , with the highest absolute correlation with the dependent variable Y and increases b_1 (or decreases depending on the sign of the correlation) until it reaches a point where another attribute b_2 has as much correlation with the residuals $R = Y - \hat{Y}$ as b_1 has. At this point the algorithm then increases/decreases both b_1 and b_2 in their joint direction until another attribute b_i has the highest residual correlation. This process can then be continued until there is no benefit in increasing any of the b_i s - that is, the full LAR solution is equivalent to the linear regression. It should also be noted that if a coefficient crosses 0 in a iteration (step), that coefficient should be set to 0, and the regression direction should subsequently be recomputed.

So why might one use LAR instead of for example the LASSO model? When using the LASSO one has to chose a specific value of s to get one solution. However, in the LAR model you actually get the full solution path which means that you can find all the LASSO solutions instead of having to do multiple LASSO with different regularization parameters (s).

4.7 Basis expansion

Basis expansion is done by adding more columns to design matrix. Let us first motivate why this might be needed.

4.7.1 Motivation

It has previously been described how the design matrix X was constructed such that we had

$$X = \begin{bmatrix} 1 & t & \frac{1}{2}t^2 & \cos\left(\frac{2\pi}{365.242}t\right) & \sin\left(\frac{2\pi}{365.242}t\right) & \dots & \cos\left(\frac{2\pi}{365.242}t\right) & \sin\left(\frac{2\pi}{365.242}t\right) \end{bmatrix}.$$

The problem with the above X is that the periodic columns assumes the pattern accounts for the entire period. This causes problems in areas such as

Denmark, where the winter comes later than “usual”, or the summer is much colder/warmer etc..

This leads to higher variance of the residuals which decreases the accuracy of the velocity and acceleration parameter of the model. By adding columns to the design matrix, such that the periodic pattern is only assumed extend over one period, the accuracy can be improved.

4.7.2 The Hinge-function

One can achieve this basis expansion is by virtue of the hinge-function

$$f(x - \zeta)_+ = \begin{cases} f(x - \zeta) & \text{if } x - \zeta \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall x \in \mathbb{R}, \zeta \in \mathbb{R} \quad (4.29)$$

ζ denotes the “knot” between the zero part and the f part. By choosing f and adding the hinge function to the design matrix, a basis expansion where terms are only active for some parts of the time series is obtained.

When f is a polynomial one achieve a piecewise polynomial which is called a spline. In this case each hinge function is guaranteed to be continuous and it can be shown that the spline’s derivatives are continuous to order $M - 2$ where M is the degree of polynomial f [6, p. 144]. When dealing with polynomials, splines which are continuous up to and including its second order derivatives are called *cubic splines* and they generate nicely looking curves when performing the regression. [6, p. 143].

4.7.3 Knots and trigonometric functions

To model each year with different behavior, in essence all that is needed is to create 9 knots (10 intervals requires 9 splits) and construct the hinge functions. However, since the functions being used are sines and cosines, there is no guarantee of continuity at the knots (only continuity if the trigonometric function evaluates to zero at the knot). In grim cases, when multiple discontinuous functions are combined, one can end up with cusps, that is a steep descent/ascend followed by the opposite movement in rapid succession. If this phenomenon becomes too large, one should consider rejecting the model, this will be discussed in the result section.

4.8 Time Series Analysis

In time series analysis one attempts to find a model for predicting future observations (this is different form the OLS analysis which tries to estimate the trends in the data, such as velocity and acceleration). Once such a model is found, it should be validated by analyzing the residuals. These residuals must be approximately white noise distributed [8, p. 130], otherwise the model is invalid.

4.8.1 ARIMA

ARIMA is the name of the stochastic model used. The theory is very complex so it will not be discussed in detail here, instead we refer to [8, p. 130].

In short ARIMA uses the previous values with some weight to predict future values. Which values are used is denoted by the following notation:

$$(p, d, q) \times (P, D, Q)_s \quad (4.30)$$

p is the highest lags of actual measurements used and q is the maximum amount of lags in residuals used. The d part indicates how many difference operators which should be used to transform the data. (P, D, Q) is completely similar, but steps not by one lag but by s lags (s = season length). This allows for seasons trends such as an yearly pattern.

4.8.2 AutoCorrelation function (ACF) and Partial AutoCorrelation Function (PACF)

ACF and PACF are measures of correlation between different time lags and are used to make a qualified guess about how the prediction model should look like. For more detail on how these are estimated, please see [8, p. 146].

When ACF and PACF have been estimated, rules [8, table 6.1] for how the stochastic model should look can be applied. Typically when dealing with complex models, this becomes an iterative and subjective process - different analysts might reach different models. The iterative step consists of estimating parameters, then calculate ACF and PACF, interpret them, estimate parameters, ... , and so it continues until the ACF and PACF suggest that no more lag terms(weights) are needed.

4.9 Ljung-Box test

The null hypothesis for the Ljung-Box is “The data is independently distributed”. In the ARIMA case it is applied to model residuals; so in other words it tests whether residuals at different lags have any correlation with each other. Thus a low p-value means that the residuals aren’t white noise.

The Ljung-Box is a χ^2 test with the statistical value:

$$L = n \cdot (n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k} \quad (4.31)$$

n is the number of observations. The parameter h is the highest lag in the ACF which should be considered. The χ^2 statistics, have $h - (p + q + P + Q)$ degrees of freedom.

4.10 K-means clustering

Clustering is the process of grouping or dividing a set of objects into disjunct subsets, such that objects in each cluster are similar. All clustering methods are unsupervised learning methods, thus there is no “correct” answer. It have been shown that given a set of objects, humans also differ in their choice of clusters.

Since similarity is an imprecise term, one uses a distance function to use as a Dissimilarity Measure d .

There are 4 conditions a distance function d must satisfy, they are all derived from the norm definition [11, p. 30]:

$$(i) d(p_1, p_2) \geq 0 \quad \forall p_1, p_2 \in V \quad (4.32)$$

$$(ii) d(p_1, p_2) \leq d(p_1, p_3) + d(p_3, p_2) \quad \forall p_1, p_2, p_3 \in V \quad (4.33)$$

$$(iii) d(p_1, p_2) = d(p_2, p_1) \quad \forall p_1, p_2 \in V \quad (4.34)$$

$$(iv) d(p_1, p_2) = 0 \Leftrightarrow p_1 = p_2 \quad \forall p_1, p_2 \in V \quad (4.35)$$

Given a distance measure the total cluster variance C^* can be calculated as

$$C^* = \sum_{k=1}^K N_k \sum_{i \in C_k} D(x_i, \mu_k). \quad (4.36)$$

Here μ_k denotes the k^{th} cluster center (also called a centroid), N_k the number of observations in cluster k and C_k denotes the subset with all the points in the cluster k .

In this analysis the Euclidean distance $D(x_i, x_j) = \|x_i - x_j\|_2$ has been chosen as the dissimilarity measure. K-means is then the chosen algorithm for minimizing C^* given K clusters. The k-means method as the dataset size is (*lat·lon, days*) = (64800, 341) and k-means is fairly quick to converge and calculate.

K-means makes a single big assumption about the clusters being hyper dimensional sphere. But beyond this K-means is a simple iterative algorithm:

1. Initialize cluster centroids
2. Iterate until centroid convergence:
 - (a) Given the current set of centroids, reassign each observation to the closest centroid using the distance function.
 - (b) Using the current cluster assignment $C_k \forall k$, minimize C^* by recomputing the centroids for each cluster as the mean of points in each cluster.

4.10.1 Gap-statistics

The big issue with clustering is that there is no obvious way of selecting the amount of clusters K . In this analysis the Gap-statistics [6, p. 519] as proposed by Hastie, Tibshirani and Firedman is used.

First calculate the cluster dissimilarity using K clusters [12]

$$W_K = \sum_{k=1}^K \frac{1}{2N_k} \sum_{x_i \in C_k} \sum_{x_j \in C_k} \|p_i - p_j\|_2^2 = \sum_{k=1}^K \sum_{x_i \in C_k} \|p_i - \mu_k\|_2^2. \quad (4.37)$$

Given the quantity W_k one can then calculate a gap G by simulating b datasets from a random uniform distribution and calculating the gap as

$$G(k) = E[\log(W_k)] - \log(W_k), \quad (4.38)$$

where the expectation $E[\log(W_k)]$ can be estimated by the mean over the b simulated datasets. That is the gap statistics tries to avoid overfitting by comparing the cluster gain on a dataset where there are no clusters (uniformly distributed).

The amount of clusters K is then minimized under the condition

$$K^* = \arg \min_K \{K \mid G(K) \geq G(K+1) - s'_{K+1}\}. \quad (4.39)$$

Here s'_{K+1} is the standard error on the b samples calculated as

$$s'_K = \text{SD}[\log(W_k)] \sqrt{1 + \frac{1}{b}}. \quad (4.40)$$

4.11 Gaussian Mixture Model (GMM)

The GMM is a more advanced clustering model than K-Means. The main advantage is that GMM allows for hyper elliptical clusters. This uses gaussian kernels with its the shape described in a covariance matrix. A result similar to K-means could be obtained by forcing this covariance matrix to be the identity matrix.

Restrictions on the covariance matrix (i.e. shared covariance, diagonal covariance, spherical covariance etc.) can easily be applied in GMM and is quite common. In this analysis however, no covariance restrictions will be used.

In the GMM the assumption is that data comes from a single density function. The density function is assumed to be a combination (mixture) of K Gaussian PDFs where K is finite and denotes the number of mixture components (i.e. clusters).

Each mixture component has a centroid (the mean), a covariance matrix and a mixing weight. The sum of mixing weights across components has to be one for the GMM to constitute an actual pdf.

To estimate the model parameters several different methods exists. The most common and is the expectation maximization (EM) algorithm which is quite complex and thus wont be described here. To the curious reader we recommend [6, p. 214,272,463]).

In practice if K is large and the vector space X is high dimensional, estimation of the model parameters will take too much computing power, and even if one

get model parameters to converge, the degrees of freedom will be low. In our case the input space would be 341-dimensional (341 observations in time per location) and thus a dimensionality reduction of some kind is needed.

4.11.1 Dimensionality reduction

In many cases when dealing with high dimensional data, most of the data lies on a lower dimensional manifold. Different methods exists to try to identify such manifolds, but in this analysis the previously described technique PCA will be used. This is done by selecting only the most important principal components from the PCA, thus forcing the data onto a lower dimensional manifold. The GRACE data contains quite a bit of noise and one might hope that the noise will be primarily contained in its own principal components. Hopefully those PCs will only account for a small amount of the variance in the data. Thus when selecting only the most significance PCs some of the noise will be "lost". It should be noted that the standard PCA method wasn't used, instead a more complex method called Kernel PCA is used.

Using kernel PCA combined with the more flexible GMM over K-Means, will hopefully lead to better clustering.

4.12 Kernel PCA

Before delving into kernel methods, the standard PCA method will briefly be recapped. PCA as introduced in this report was done via SVD ($X = U\Sigma V^T$). The columns of U is the eigenvectors of XX^T . The projected space is then $Z = XV = U\Sigma$. Looking at the dependencies of U , it is apparent that the principal component scores, will end up being linear combinations of basis in original input space X . In most cases this suffices but what if there are no linear relationships in the X space, standard PCA won't be appropriate.

4.12.1 Motivating example

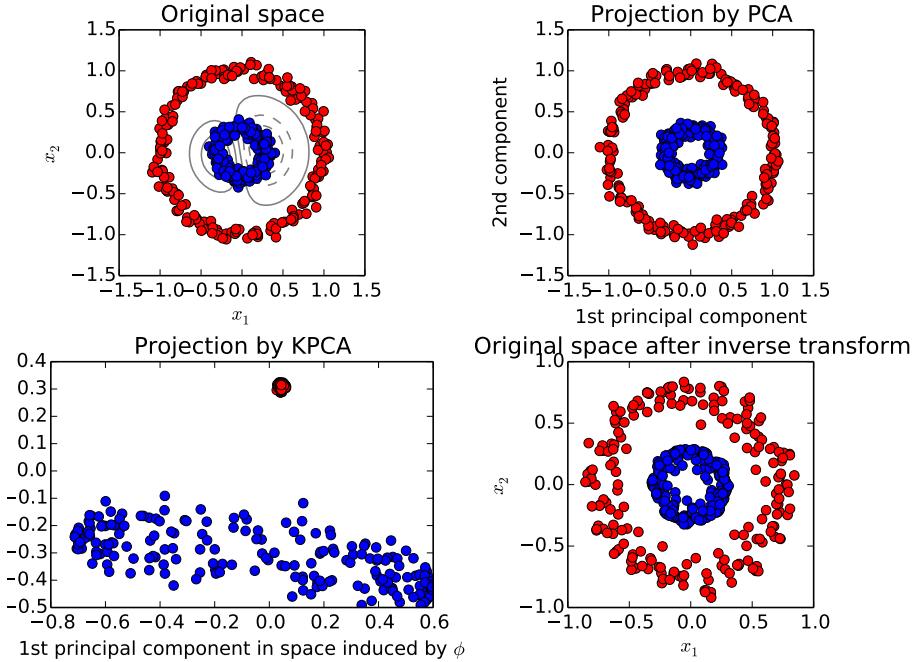


Figure 4.1: Motivating kernel PCA example. Image courtesy of scikit-learn. License: BSD 3 clause. Authors: Mathieu Blondel and Andreas Mueller.

As is seen in Figure 4.1, classes that wasn't linearly separable in the original space can be become linearly separable in the kernel space. This in turn makes clustering much easier and hopefully improves results.

4.12.2 The Kernel Trick

In kernel PCA, instead of working on XX^T , a nonlinear mapping $\Phi : X \rightarrow Y$ is used, such that the SVD is carried out on $\Phi(X)\Phi(X)^T = K(X, X^T)$. Specifically one define the inner product as a function of (x_i, x_j) , thus $\Phi(X)\Phi(X)^T$ is calculated, but without the need for the nonlinear mapping function Φ . This inner product function is the kernel.

For something to be a valid kernel, all the kernel needs to satisfy is to be function of an inner product in some vector space. For the above to become more clear, let us give an example of a polynomial kernel in a 2-dimensional space. in the following X is a vector and X' is simply some other vector in the same space as X :

$$\begin{aligned} K(X, X') &= (1 + X^T X')^2 = (1 + x_1 x'_1 + x_2 x'_2)^2 \\ &= 1 + x_1^2 x'^2_1 + 2x_1 x'_1 + 2x_2 x'_2 + 2x_1 x'_1 x_2 x'_2 \end{aligned} \tag{4.41}$$

For the above to actually be a actual kernel there would have to be some transformed space in which the above was an inner product. From the coefficients above it can be deduced that the basis in this transformed space must be given as

$$(1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2) \quad (4.42)$$

Now consider if one used a power of 100 instead of 2. Calculating the kernel in the non transformed space is easy; $(1 + X^T X')$ is just a number and raising it to the power of 100 can be done quickly. On the other hand if one were to explicitly calculate the kernel as the inner product in the transformed space, a huge vector would have to be computed, transposed and be subjected to an the inner product in this space. This is clearly not very efficient.

The shortcut to define K instead of Φ is called the Kernel Trick.

4.12.3 The Radial Basis Function (RBF) kernel

The RBF kernel is commonly used when the amount of samples is much larger than the amount of dimensions in the original space.

The RBF kernel is defined as

$$K(x, x') = \exp(-\gamma \|x - x'\|_2^p) \quad (4.43)$$

The above can easily be calculated. The following shows that the above is indeed an inner product. It is not a complete proof as $\gamma = 1, p = 2$ and x is a scalar not a vector.

$$\begin{aligned} K(x, x') &= \exp(-\|x - x'\|_2^2) = \exp(-(x - x')^2) \\ &= \exp(-x^2)\exp(-x'^2)\exp(2xx') \end{aligned} \quad (4.44)$$

since $\exp(2xx')$ can be Taylor expanded to $\sum_{k=0}^{\infty} \frac{2^k (x)^k (x')^k}{k!}$ it follows that:

$$K(x, x') = \sum_{k=0}^{\infty} \left(\sqrt{\frac{2^k}{k!}} (x)^k \exp(-x^2) \right) \left(\sqrt{\frac{2^k}{k!}} (x')^k \exp(-x'^2) \right) \quad (4.45)$$

From the above it is seen that for any k , the term in the right parenthesis is exactly equal to the left parenthesis, if x' was substituted with x . This means that in an expanded vector space, this correspond to the the inner product from the ℓ^2 Hilbert space. Do also note that since the sum is infinite, the transformed vector space is of infinite dimensional space and is thus impossible to calculate without the kernel trick.

4.12.4 Mercer's Condition

Proving that something is a kernel, is actually easier than shown above. It is done using Mercer's Condition, though it does not say anything about the size of the transformed space when Φ is applied.

The theorem states, that given all real valued square integrable functions g ($g \in \{\mathbb{R} \rightarrow \mathbb{R}\} \cap L^2(\mathbb{R})$). That is the following must be true:

$$\int_{-\infty}^{\infty} g(x)^2 dx < \infty \quad (4.46)$$

Then for all these g functions, if the following condition holds:

$$\int \int K(x, y)g(x)g(y)dxdy \geq 0 \quad (4.47)$$

Then K is a valid kernel.

4.12.5 Final notes of kernel PCA

In kernel PCA the columns of U is the eigenvectors of $K(X, X^T)$ and the columns of V the eigenvectors of $K(X^T, X)$. The size of XX^T and X^TX is going to be $n \times n$ and $p \times p$ respectively where n is the number of samples and p the dimensions in transformed space. Thus when dealing with kernel PCA the calculation of V is omitted and calculation of U is expensive, but possible.

In the case of the GRACE data the size of X is (64800×341) . Memory wise this results in a U matrix of size (64800×64800) with float32 numbers, that is approximately 15.64 GB. Furthermore kernel PCA also involves quite a lot of simple computing. So for practical purposes one should use a computer cluster (e.g. the DTU HPC cluster).

5 Results

5.1 OLS

The first part of the OLS analysis will examine the EWH at two specific locations. The second part will then focus on the entire world. The main purpose of the OLS analysis is to estimate the velocity and acceleration of the mass changes.

5.1.1 Selected locations

The two selected locations are:

- Greenland (63.5° N, 49.5° W)
- Western Antarctica (74.5° S, 87.5° W)

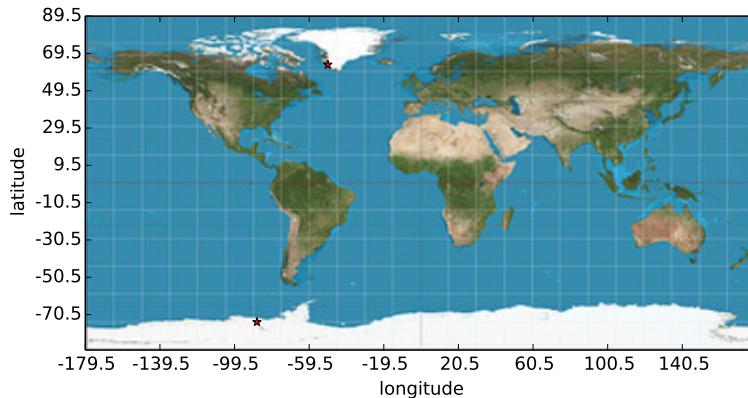


Figure 5.1: The two selected locations, marked with a red star

5.1.1.1 Greenland At the west coast of Greenland a strong period trend can be observed, probably caused by the the ice melting over the summer and reappearing during the winter. This trend is seen in Figure 5.2. In order to estimate the velocity and acceleration of the EWH accurately, it is important that the periodic trend is caught by the $\sin(\cdot)$ and $\cos(\cdot)$ terms so that seasonal patterns do not interfere with the estimates.

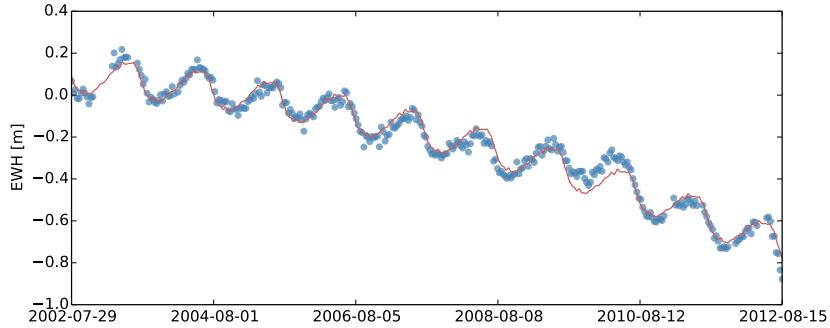


Figure 5.2: Measurements are blue, the OLS fit is red.

From Figure 5.2 it's seen that the period trend is caught by the model, however for some seasons (particularly between 2008 and 2010) the fit is not very good. This is even more clear when looking at the residuals in Figure 5.3. From this it is clear that the residuals are far from white noise, which was one of the OLS assumptions.

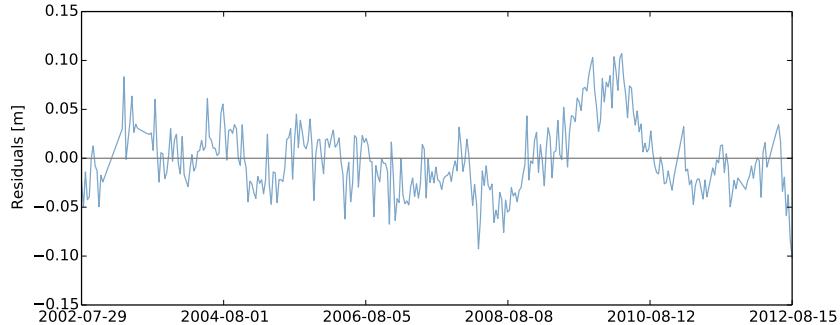


Figure 5.3: The OLS residuals are blue.

In addition to the yearly season, when looking at the estimated OLS coefficients (Table 5.1) there also seems to be a half yearly season. Though strangely enough it is not very visible on the OLS fit (Figure 5.2). The remaining seasons are with 95% confidence not significantly different from zero and for the purpose of getting a better estimate, it might be worth iteratively removing these parameters.

name	estimate	p-value	name	estimate	p-value
<i>intercept(1)</i>	$-1.77 \cdot 10^{-01}$	0.000	$\sin(9 \cdot \omega \cdot t)$	$8.48 \cdot 10^{-04}$	0.772
<i>vel.(t)</i>	$-2.24 \cdot 10^{-04}$	0.000	$\cos(10 \cdot \omega \cdot t)$	$1.54 \cdot 10^{-03}$	0.599
<i>acc.(0.5 · t²)</i>	$-8.27 \cdot 10^{-08}$	0.000	$\sin(10 \cdot \omega \cdot t)$	$-5.77 \cdot 10^{-04}$	0.843
<i>cost(1 · ω · t)</i>	$-1.35 \cdot 10^{-02}$	0.000	$\cos(11 \cdot \omega \cdot t)$	$-5.95 \cdot 10^{-04}$	0.839
<i>sin(1 · ω · t)</i>	$-8.28 \cdot 10^{-02}$	0.000	$\sin(11 \cdot \omega \cdot t)$	$-1.47 \cdot 10^{-03}$	0.615
<i>cos(2 · ω · t)</i>	$-5.86 \cdot 10^{-03}$	0.045	$\cos(12 \cdot \omega \cdot t)$	$1.78 \cdot 10^{-04}$	0.952
<i>sin(2 · ω · t)</i>	$-1.41 \cdot 10^{-02}$	0.000	$\sin(12 \cdot \omega \cdot t)$	$5.37 \cdot 10^{-04}$	0.854
<i>cost(3 · ω · t)</i>	$-5.01 \cdot 10^{-03}$	0.089	$\cos(13 \cdot \omega \cdot t)$	$2.57 \cdot 10^{-03}$	0.383
<i>sin(3 · ω · t)</i>	$-5.69 \cdot 10^{-03}$	0.053	$\sin(13 \cdot \omega \cdot t)$	$-6.35 \cdot 10^{-04}$	0.827
<i>cos(4 · ω · t)</i>	$-5.49 \cdot 10^{-03}$	0.060	$\cos(14 \cdot \omega \cdot t)$	$3.16 \cdot 10^{-04}$	0.914
<i>sin(4 · ω · t)</i>	$-1.66 \cdot 10^{-03}$	0.573	$\sin(14 \cdot \omega \cdot t)$	$-2.99 \cdot 10^{-04}$	0.919
<i>cos(5 · ω · t)</i>	$-2.19 \cdot 10^{-03}$	0.455	$\cos(15 \cdot \omega \cdot t)$	$-7.23 \cdot 10^{-04}$	0.804
<i>sin(5 · ω · t)</i>	$-1.93 \cdot 10^{-03}$	0.511	$\sin(15 \cdot \omega \cdot t)$	$3.33 \cdot 10^{-04}$	0.910
<i>cos(6 · ω · t)</i>	$4.29 \cdot 10^{-04}$	0.883	$\cos(16 \cdot \omega \cdot t)$	$-8.07 \cdot 10^{-04}$	0.785
<i>sin(6 · ω · t)</i>	$1.43 \cdot 10^{-03}$	0.627	$\sin(16 \cdot \omega \cdot t)$	$-6.93 \cdot 10^{-04}$	0.811
<i>cos(7 · ω · t)</i>	$9.55 \cdot 10^{-04}$	0.744	$\cos(17 \cdot \omega \cdot t)$	$-1.26 \cdot 10^{-03}$	0.665
<i>sin(7 · ω · t)</i>	$1.09 \cdot 10^{-04}$	0.970	$\sin(17 \cdot \omega \cdot t)$	$-7.28 \cdot 10^{-04}$	0.803
<i>cos(8 · ω · t)</i>	$-3.08 \cdot 10^{-04}$	0.916	$\cos(18 \cdot \omega \cdot t)$	$-4.93 \cdot 10^{-04}$	0.864
<i>sin(8 · ω · t)</i>	$-2.37 \cdot 10^{-04}$	0.935	$\sin(18 \cdot \omega \cdot t)$	$-9.94 \cdot 10^{-06}$	0.997
<i>cos(9 · ω · t)</i>	$-2.02 \cdot 10^{-03}$	0.491			

Table 5.1: Parameter estimates $\hat{\beta}$ and their p-values for Greenland.

5.1.1.2 Western Antarctica Just like with Greenland there is a clear drop in mass as seen in Figure 5.4. Interestingly enough, there is no seasonal trend or at least it is not very strong. Also the existence of high frequency terms in the OLS regression, is more apparent. Though from the actual data it does not look like such periodic trends exist. This suggests that too many seasonal parameters have been used, thus causing some overfitting.

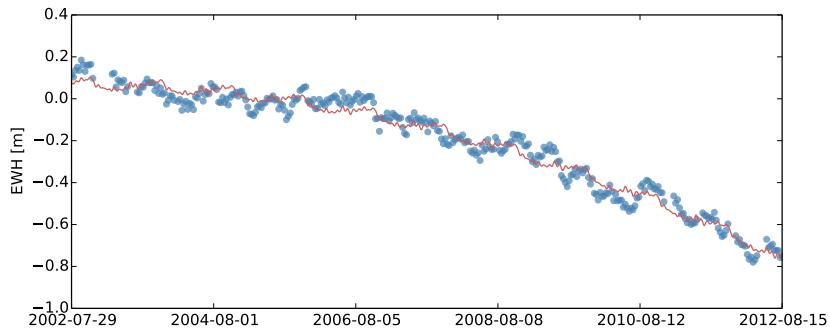


Figure 5.4: Measurements are blue, the OLS fit is red.

When looking at the residuals in Figure 5.5, it is again clear that the residuals are far from being white noise. There also seems to be some periodic trend in

the residuals; that is the residuals appear to have a seasonal pattern. However, the beginnings and the endings of these seasons are quite hard to make out.

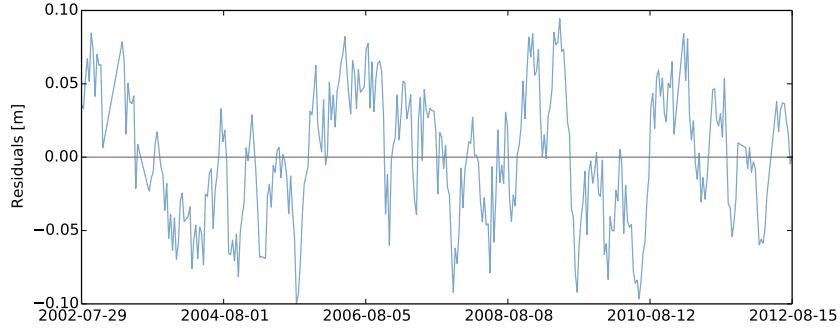


Figure 5.5: The OLS residuals are blue.

Just as with Greenland many of the OLS terms are with 95% confidence not significantly different from zero. Interestingly enough there is a yearly season when looking at Table 5.2, though it's very hard to make out from the OLS fit in Figure 5.4.

name	estimate	p-value	name	estimate	p-value
<i>intercept(1)</i>	$-1.38 \cdot 10^{-01}$	0.000	$\sin(9 \cdot \omega \cdot t)$	$2.05 \cdot 10^{-03}$	0.568
<i>vel.(t)</i>	$-2.26 \cdot 10^{-04}$	0.000	$\cos(10 \cdot \omega \cdot t)$	$-4.54 \cdot 10^{-04}$	0.900
<i>acc.(0.5 · t²)</i>	$-1.24 \cdot 10^{-07}$	0.000	$\sin(10 \cdot \omega \cdot t)$	$-1.26 \cdot 10^{-03}$	0.726
$\cos(1 \cdot \omega \cdot t)$	$1.69 \cdot 10^{-02}$	0.000	$\cos(11 \cdot \omega \cdot t)$	$-1.21 \cdot 10^{-04}$	0.973
$\sin(1 \cdot \omega \cdot t)$	$1.42 \cdot 10^{-02}$	0.000	$\sin(11 \cdot \omega \cdot t)$	$-1.07 \cdot 10^{-03}$	0.766
$\cos(2 \cdot \omega \cdot t)$	$-7.19 \cdot 10^{-03}$	0.046	$\cos(12 \cdot \omega \cdot t)$	$-1.20 \cdot 10^{-03}$	0.740
$\sin(2 \cdot \omega \cdot t)$	$2.82 \cdot 10^{-03}$	0.440	$\sin(12 \cdot \omega \cdot t)$	$-8.20 \cdot 10^{-04}$	0.819
$\cos(3 \cdot \omega \cdot t)$	$-4.36 \cdot 10^{-03}$	0.227	$\cos(13 \cdot \omega \cdot t)$	$-5.73 \cdot 10^{-03}$	0.114
$\sin(3 \cdot \omega \cdot t)$	$-1.23 \cdot 10^{-03}$	0.732	$\sin(13 \cdot \omega \cdot t)$	$-8.50 \cdot 10^{-04}$	0.812
$\cos(4 \cdot \omega \cdot t)$	$1.78 \cdot 10^{-03}$	0.619	$\cos(14 \cdot \omega \cdot t)$	$-6.55 \cdot 10^{-04}$	0.855
$\sin(4 \cdot \omega \cdot t)$	$2.81 \cdot 10^{-03}$	0.437	$\sin(14 \cdot \omega \cdot t)$	$2.25 \cdot 10^{-03}$	0.533
$\cos(5 \cdot \omega \cdot t)$	$2.45 \cdot 10^{-03}$	0.496	$\cos(15 \cdot \omega \cdot t)$	$-4.32 \cdot 10^{-04}$	0.904
$\sin(5 \cdot \omega \cdot t)$	$2.53 \cdot 10^{-03}$	0.484	$\sin(15 \cdot \omega \cdot t)$	$3.05 \cdot 10^{-04}$	0.933
$\cos(6 \cdot \omega \cdot t)$	$-1.60 \cdot 10^{-03}$	0.654	$\cos(16 \cdot \omega \cdot t)$	$3.51 \cdot 10^{-04}$	0.923
$\sin(6 \cdot \omega \cdot t)$	$-8.66 \cdot 10^{-04}$	0.811	$\sin(16 \cdot \omega \cdot t)$	$-3.73 \cdot 10^{-04}$	0.917
$\cos(7 \cdot \omega \cdot t)$	$8.95 \cdot 10^{-04}$	0.803	$\cos(17 \cdot \omega \cdot t)$	$1.22 \cdot 10^{-03}$	0.735
$\sin(7 \cdot \omega \cdot t)$	$-3.27 \cdot 10^{-03}$	0.364	$\sin(17 \cdot \omega \cdot t)$	$-1.11 \cdot 10^{-03}$	0.758
$\cos(8 \cdot \omega \cdot t)$	$1.62 \cdot 10^{-03}$	0.652	$\cos(18 \cdot \omega \cdot t)$	$-1.92 \cdot 10^{-04}$	0.957
$\sin(8 \cdot \omega \cdot t)$	$-1.51 \cdot 10^{-03}$	0.674	$\sin(18 \cdot \omega \cdot t)$	$7.18 \cdot 10^{-04}$	0.844
$\cos(9 \cdot \omega \cdot t)$	$3.76 \cdot 10^{-04}$	0.917			

Table 5.2: Parameter estimates $\hat{\beta}$ and their p-values for Western Antarctica.

5.1.2 World view

5.1.2.1 Parameters As mentioned earlier, the velocity and acceleration of the mass changes are the most important parameters in the OLS regression.

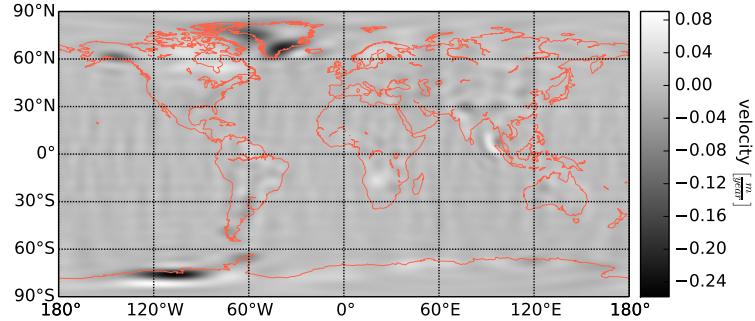


Figure 5.6: Estimated velocity (t) parameters

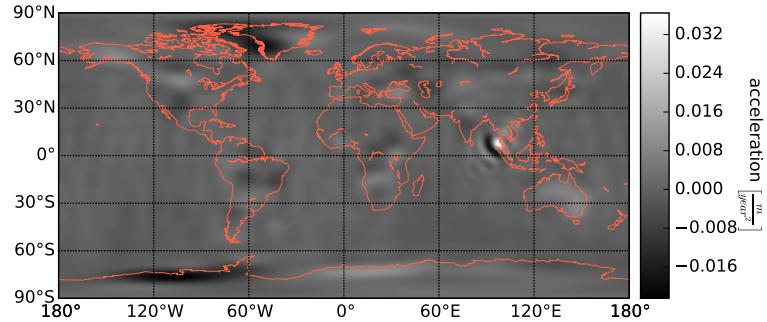


Figure 5.7: Estimated acceleration ($\frac{1}{2}t^2$) parameters

The first two figures (Figure 5.6 and Figure 5.7) show that, generally there is a mass loss at both Greenland and the Western Antarctica and that it is accelerating. However interestingly enough the East Coast of Greenland do not show any acceleration, maybe even a slight deceleration of the mass loss (a positive value).

On both figures there is also a circular pattern at Thailand (10 N, 95 E), which is caused by the infamous earthquake of the 26th December 2004, resulting in significant changes in the mass distribution, thus affecting the data from GRACE.

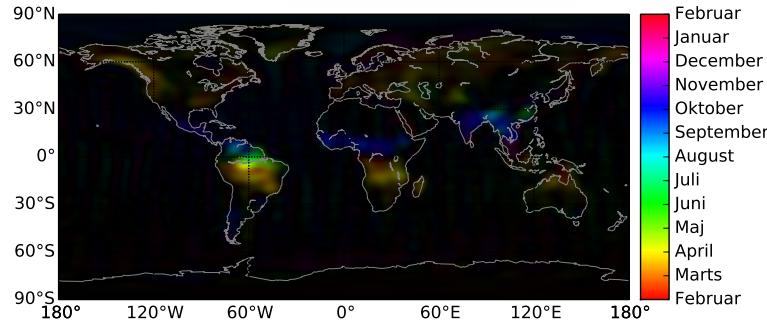


Figure 5.8: Intensity-Hue plot. Estimated phase controls the hue while the amplitude sets the intensity.

The seasonal trends do not reveal any information about mass loss due to ice melting, since this is something which will continue to have effect for many years. However it does affect the estimation of the velocity and acceleration. Also having more season related terms in the OLS regression than necessary, will result in fewer degrees of freedom.

In Figure 5.8 it is seen from the intensity, that the Western Antarctica do not have a strong yearly seasonal trend unlike Greenland. It also appears that it is primarily the East Coast of Greenland which has a yearly seasonal trend. The strongest seasonal trend appears in South America (more specifically the Amazon Basin), and this is due to the rain season. This rain also takes quite some time to reach the ocean, thus causing the phase gradient.

5.1.2.2 Performance As a measure of model performance the root mean squared error has been calculated for each position.

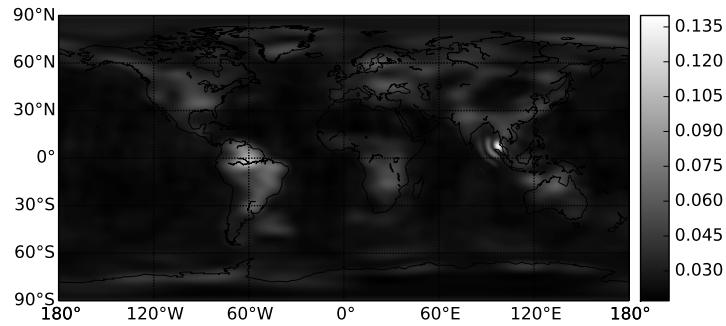


Figure 5.9: RMSE for each position.

From 5.9 it is seen that the hardest places to fit are South America and Thailand. In South America it is especially around the Amazon River that the RMSE is high. This suggests that it is caused by changes in the rain seasons, since this

is also the place with the highest amplitude of the yearly seasonal term. It's a bit odd that this is so hard to fit, since the model does support seasonal trends. One explanation could be that this OLS model assumes that the seasons are equally long and with identical phases each year, which is not very likely. The circular pattern around Thailand is no surprise, since the OLS model was never meant to fit this distortion.

Neither of these issues are related to the ice melting, which makes them less important. However the East Coast of Greenland and the Western Antarctica also show a relative higher RMSE, when comparing to the ocean or the nearby land, which is less than ideal. The fact that it is primarily the East Coast of Greenland and not the west coast, suggests that the issues are related to the seasons, just like with the Amazon Basin. This was also something observed, when looking at the specific location's time series.

5.1.2.3 Standard diagnostics Recalling that $\text{Var}[\hat{Y}_i] = \sigma^2 H_{ii}$ and $D[\hat{\beta}] = \sigma^2 V \Sigma^{-2} V^T$ the diagonal of H and $V \Sigma^{-2} V^T$ are both important to examine. This will identify any days and parameters with high variance (the reader should remember that this applies to all positions).

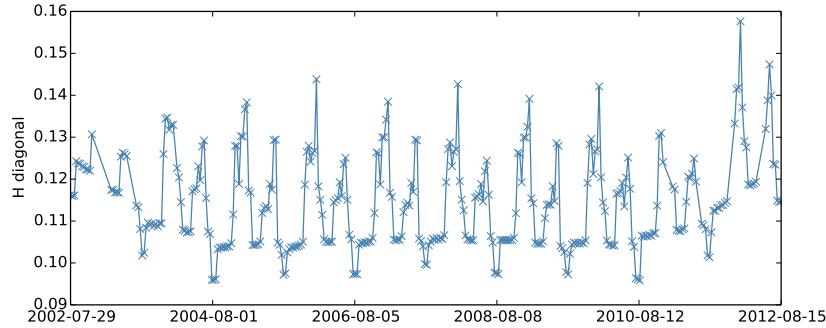


Figure 5.10: The diagonal of H will reveal any days with a high variance.

The variances shown in Figure 5.10 are not very big, the highest is 0.16 and recalling that the 95% confidence interval can be approximated by $\hat{Y}_i \pm 2 \cdot \hat{\sigma} \sqrt{\text{Var}[\hat{Y}_i]}$. The maximal confidence interval becomes $\hat{Y}_i \pm 2 \cdot 0.135 \cdot \sqrt{0.16} = \hat{Y}_i \pm 0.108$, which is small when comparing to the EWH values, which range from 1 to -3. What is odd about $\text{diag}(H)$ is that there appears to be some parts in each year which are more difficult to predict than others.

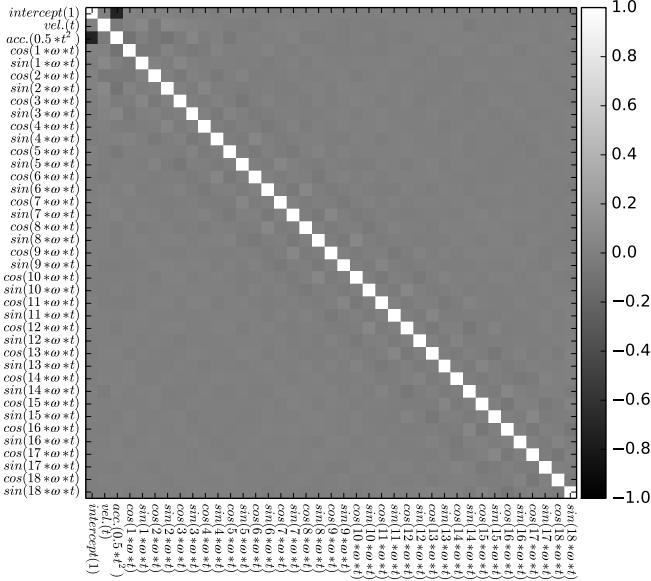


Figure 5.11: Correlation matrix reveals correlated parameter.

From Figure 5.11 it is seen that the parameters are nicely uncorrelated and while the seasonal parameters have a slightly high variance when comparing to their values; the variation of the velocity and the acceleration parameters are practically zero.

name	$\text{Var}[\hat{\beta}_i]$
$vel.(t)$	$2.91390457 \cdot 10^{-9} [\frac{m}{\text{day}^2}]$
$acc.(0.5 \cdot t^2)$	$1.29525397 \cdot 10^{-14} [\frac{m^2}{\text{day}^4}]$

Table 5.3: Variance of the velocity and acceleration parameters for all positions

5.1.2.4 p-values Just as for the selected positions, inspecting the p-values can tell which OLS terms are important.

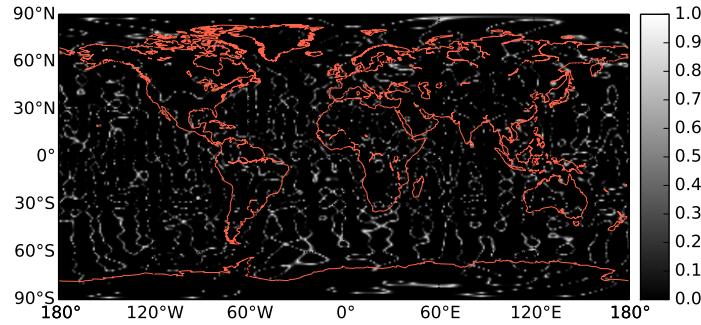


Figure 5.12: p-values for the velocity parameter.

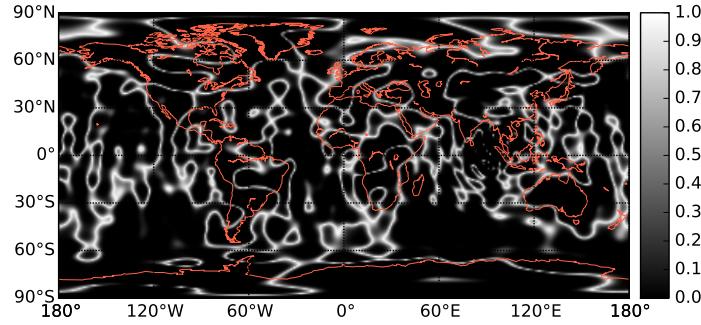


Figure 5.13: p-values for the acceleration parameter.

There is not much surprise in either Figure 5.12 or Figure 5.13. Both the velocity and the acceleration appears to be significant pretty much everywhere. The places where it is not, is where there is a sign shift thus causing the p-value gradient. This is particularly apparent with the acceleration. One odd detail is that also on the ocean, the velocity and acceleration parameters appears to be significantly different from zero. In the ocean there is no reason to expect a steady mass change, though when looking at actual values it's seen that there is a slight slope. This quite small slope then becomes significant, due to the very small variance of the velocity and acceleration parameters.

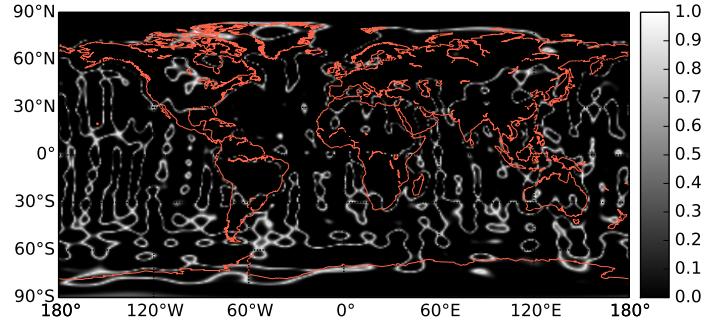


Figure 5.14: p-values for the yearly seasonal cosine parameter.

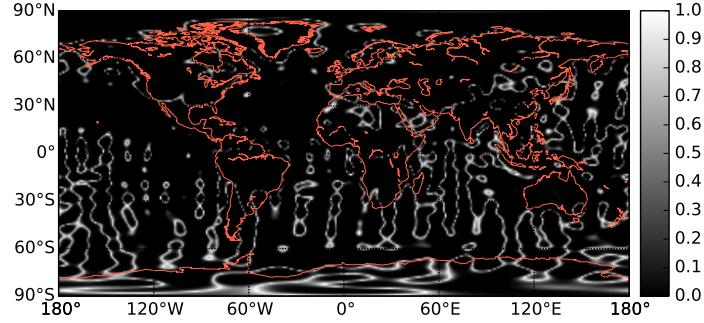


Figure 5.15: p-values for the yearly seasonal sine parameter.

Also the yearly seasonal parameters are significant pretty much everywhere. For these parameters it makes more sense that they are significant in the ocean.

5.1.2.5 PCA of residuals By computing the principal component for the OLS residuals, it is sometimes possible to discover patterns in the residuals. As with all PCA one the first couple of principal components are valuable to look at.

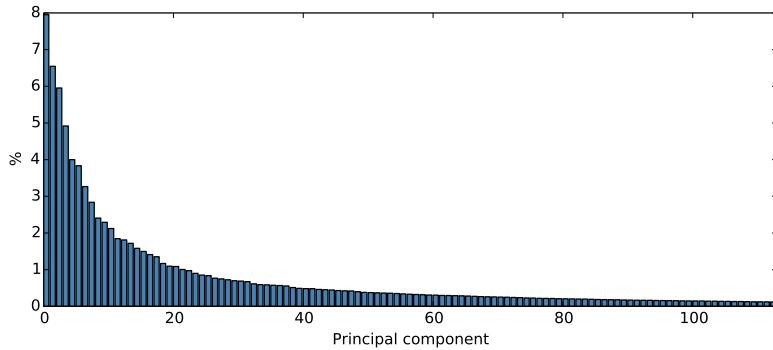


Figure 5.16: Variance explained by each principal component

If the PCA had been used on the original GRACE data, first couple of PCs would likely have explained most of the variance. However since this PCA is done on the residuals, the primary data is hopefully noise, thus no particularly direction shows a lot of variance. The shape of Figure 5.16 indicates that there is a lot of noise, but that there also exists residuals there isn't just noise.

Plotting the first principal component score and loading, one gets:

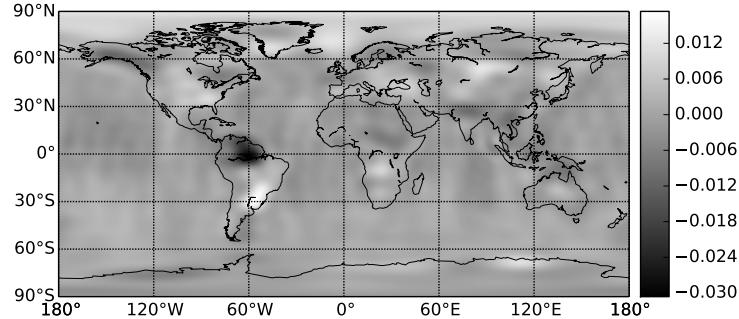


Figure 5.17: PC score for the first principal component

From Figure 5.17 its seen that there is some residuals in South America which are unlikely to be caused by noise. This could be because of the rainy seasons, since those are poorly modeled using this OLS regression. If that is the case one should expect Figure 5.18 to show some yearly periodic trend. Unfortunately this does not seem to be the case.

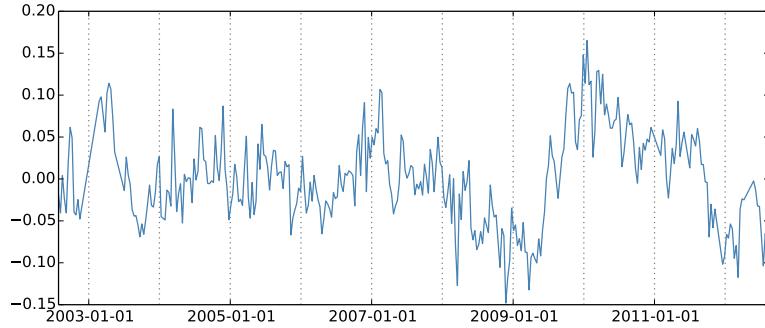


Figure 5.18: PC loadings for the first principal component

Plotting the second principal component score and loading, one gets:

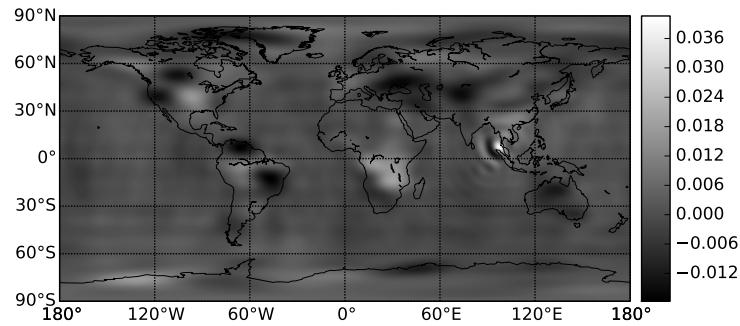


Figure 5.19: PC scores for the second principal component

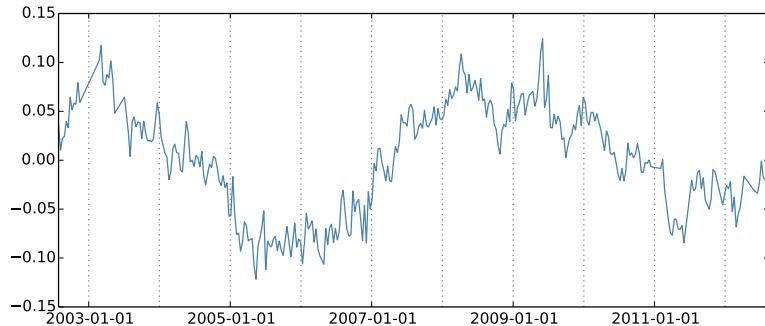


Figure 5.20: PC loading for the second principal component

In Figure 5.19 a strong circular pattern appears. This is caused by the earthquake near Thailand in 2004 and can't be modeled by this OLS regression,

so those residuals are expected. It is a bit strange that Fig. 5.20 indicates that the residuals not only comes from 2002 to 2004 but also around 2006 and 2009.

More importantly Fig. 5.19 also show some residuals near west Greenland. This is not ideal since this is one of the areas this report attempts to analyze.

5.2 Glacial Isostatic Adjustment

The GRACE data that has been used in this report has not been corrected for Glacial Isostatic Adjustment (GIA), also called post-glacial rebound.

GIA is an geophysical effect that makes land either rise, fall or shift horizontally. It can be observed at all locations which during the last ice age was covered by a thick layer of ice. The cause is the sheer weight of the ice compressed the crust of the Earth so much, that when the ice retracted and the downward force was removed the crust started to rebound. It is estimated that these movements will take many thousand years to subside.

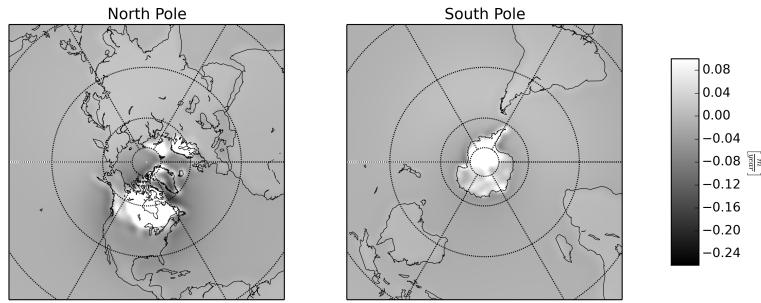


Figure 5.21: Glacial Isostatic Adjustment. Data is based on ICE-5G and the specific source is NASA [13].

The GRACE data only measures the gravity and because of the geophysical effect the rebound will be part of the observations. Thus one can not directly conclude from the GRACE data, if changes are caused by recent ice melting or GIA. Removing the GIA signal is therefore important, if one wants to assert that ice indeed is melting.

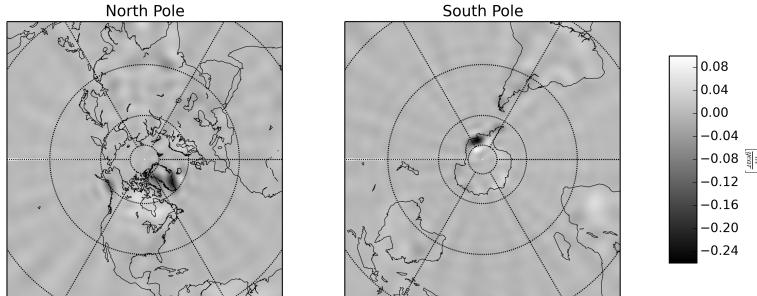


Figure 5.22: Estimated velocity from GRACE data, without correction for GIA.

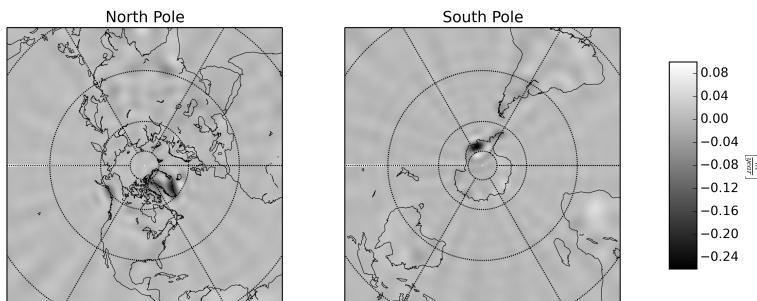


Figure 5.23: Estimated velocity from GRACE data, with the GIA signal removed. A big improvement around North America can be observed.

Even though there appears to be a significant velocity in EWH after GIA has been removed from the velocity estimate, one still has to be careful about making conclusions. According to NASA [4], in order to sufficiently correct for GIA, region specific averaging kernels are needed to properly account for GIA as well as noise from nearby land hydrology.

NASA specifically state [4] that the GRACE data (with standard GIA applied) is not suited for Cryospheric studies (ice mass changes) and thus one should keep these reservations in mind when viewing the results.

For the remaining report no correction for GIA has been used. Only this section uses the standard GIA correction.

5.3 Time series analysis - ARIMA

To get an indication of whether its possible to use time series analysis on the GRACE data, a single position (63.5 N 49.5 W, west coast of Greenland) has been selected.

To analyze the data using an ARIMA model an equidistant dataset is required. For example it would otherwise not be possible to solve the Yule-Walker equa-

tions [8, s. 122]. In the original GRACE dataset some values are missing, thus they should be interpolated (linear interpolation was used). Also in order to get an indication of the model performance, the last 36 observations (corresponding to one year) have been separated for model validation.

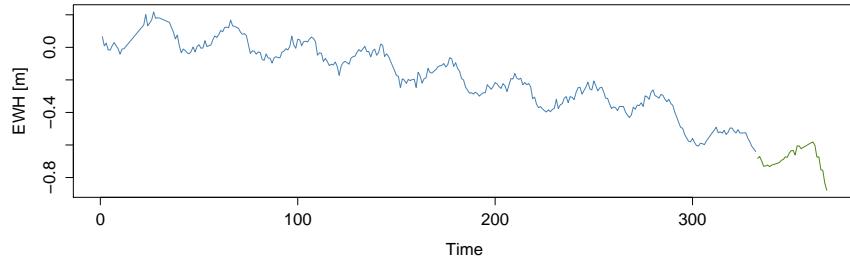


Figure 5.24: GRACE data at 63.5 N 49.5 W, where missing values are interpolated. Blue is the training data and green is the test data.

The time series in Figure 5.24 is clearly not stationary, thus it is necessary to consider the time series difference.

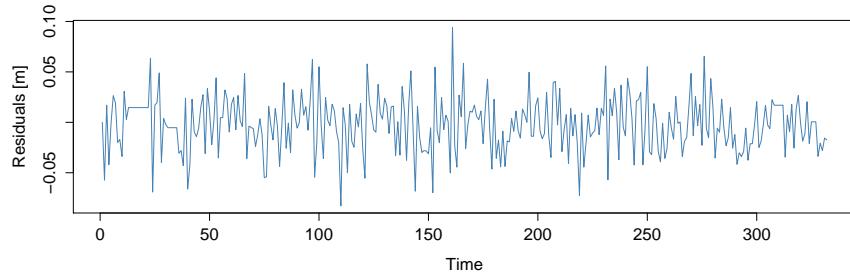


Figure 5.25: The ARIMA(0, 1, 0) \times (0, 0, 0)₃₆ residuals.

On Figure 5.25 there are some seasonal periods where the mean and variance are not the same as the remaining period, thus it is not completely stationary. Taking also the seasonal difference (assuming the season is 36 observations, a year) gives however a very stationary output.

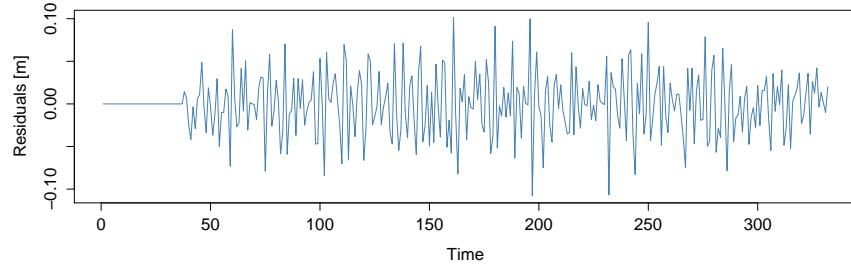


Figure 5.26: The ARIMA $(0, 1, 0) \times (0, 1, 0)_{36}$ residuals.

On Figure 5.26 its seen that the first 37 observations act strange, this is because there aren't enough past observations to estimate the EWH correctly, thus they should be excluded from further analysis.

To determine the AR and MA terms in the ARIMA model, the ACF and PACF should be estimated using the residuals from Figure 5.26.

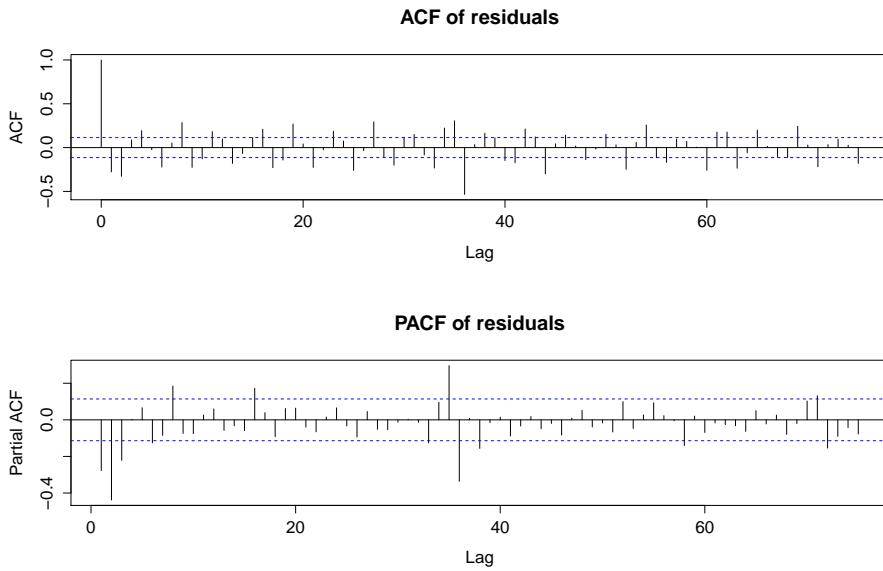
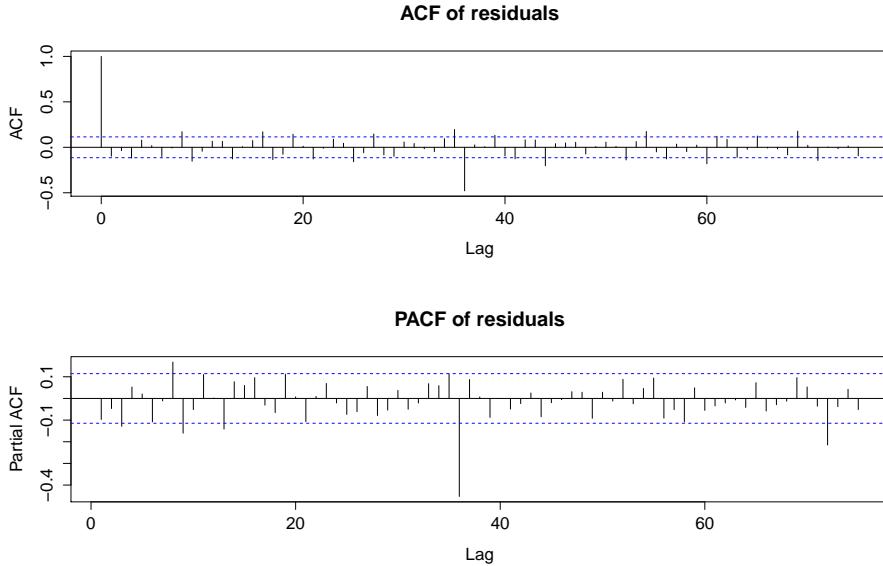
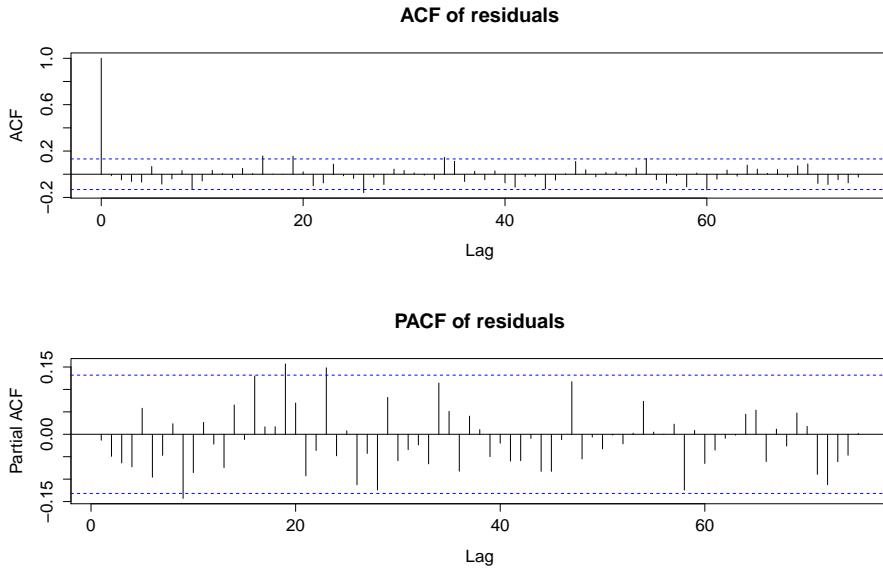


Figure 5.27: ACF and PACF for ARIMA $(0, 1, 0) \times (0, 1, 0)_{36}$ residuals

Using the PACF in Figure 5.27 and the rules for the AR term [8, Table 6.1] AR(2), seams like a good guess for the non-seasonal AR term.

Figure 5.28: ACF and PACF for $\text{ARIMA}(2,1,0) \times (0,1,0)_{36}$ residuals

This clearly fitted the non-seasonal trend. From Figure 5.27 it might have looked like there was a AR(3) or MA(2) term, but the AR(2) is the simplest of those and fit the trend just fine. The seasonal part is now extremely apparent in Figure 5.28, where it looks like either a SAR(2)- or a SAR(1)-term.

Figure 5.29: ACF and PACF for $\text{ARIMA}(2,1,0) \times (2,1,0)_{36}$ residuals

From just looking at the estimated ACF and PACF in Figure 5.29, ARIMA(2, 1, 0) \times (2, 1, 0)₃₆ seems like a good choice. To finally validate the model, a good start is to look at the residuals.

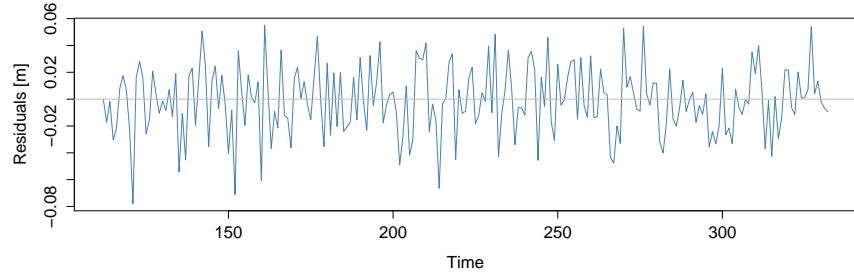


Figure 5.30: ARIMA(2, 1, 0) \times (2, 1, 0)₃₆ residuals. The first 111 residuals have been skipped since they cannot be estimated correctly.

Figure 5.30 looks stationary, there are no outliers nor seasonal trends. To validate the model further the Ljung-Box test can be used. This however requires the residuals to be normally distributed, a QQ-plot (Figure 5.31) shows that this is the case:

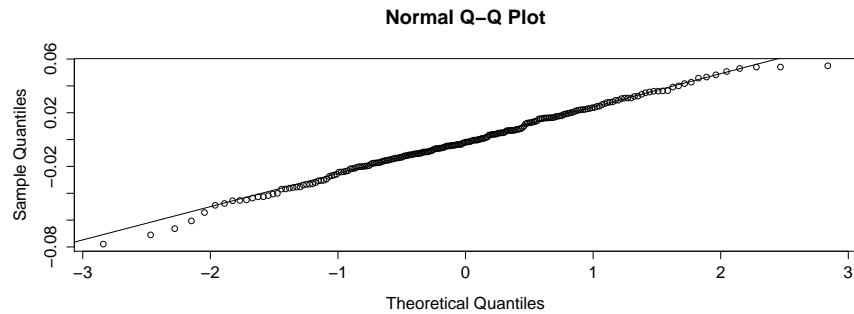


Figure 5.31: QQ-plot for the ARIMA(2, 1, 0) \times (2, 1, 0)₃₆ residuals.

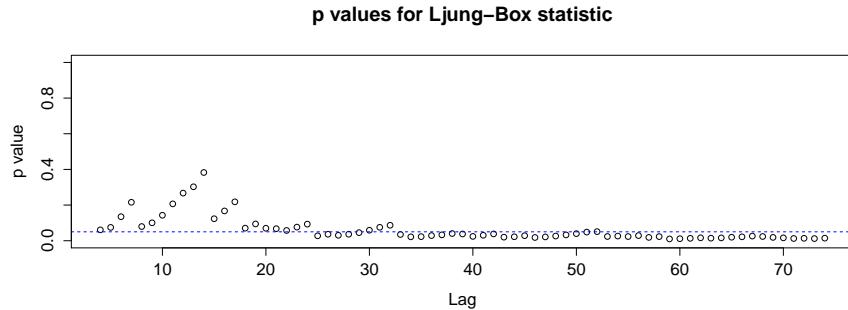


Figure 5.32: Ljung-Box test for the $\text{ARIMA}(2, 1, 0) \times (2, 1, 0)_{36}$ residuals.

From the Ljung-Box test (Figure 5.32) the p-value for the first many lags looks good, however after 25 it can be with 95% confidence statically significant concluded that the residuals are correlated. In terms of pure time series analysis is makes the model quite useless, however it is still valuable to do the cross validation.

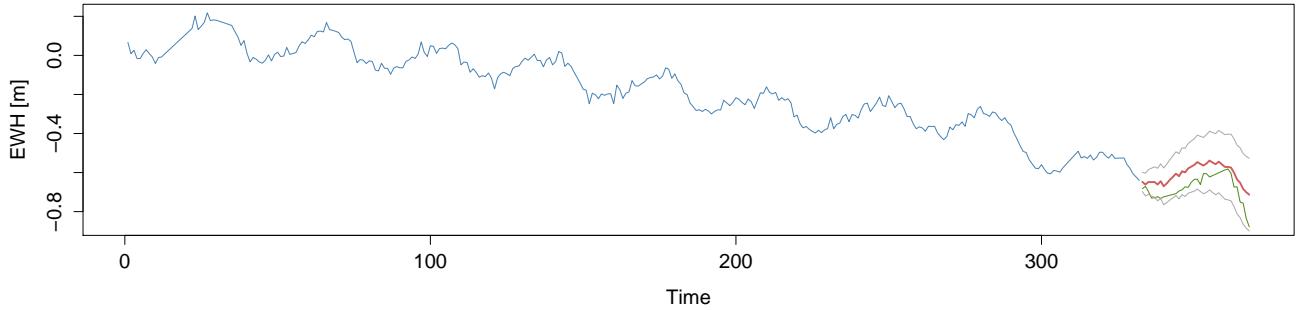


Figure 5.33: Forecast on Cross Validation. Blue is the training data, green is the test data. Red is then the predicted test values with its 95% confidence interval marked with gray lines.

From Figure 5.33 it quite clear that the test data is consistently bellow the expectation line (red). While it is still inside the 95% confidence interval, this high correlation in the error between lags indicates that the model is not particular useful.

5.4 Least angular regression (LAR)

The full solution path is in this case very hard to interpret, so instead the relative coefficients are shown. That is the coefficients at an iteration divided by the final value (the plain OLS case).

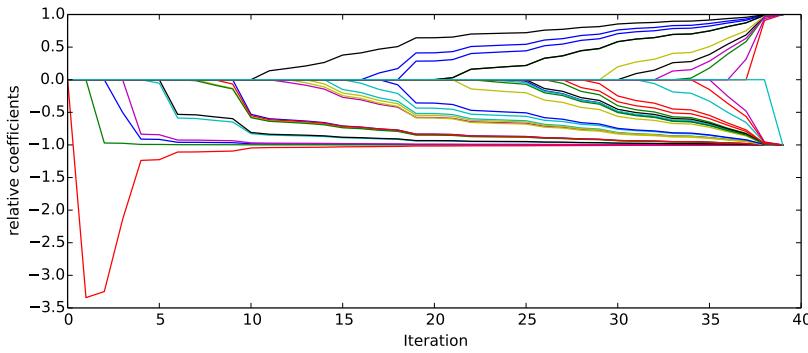


Figure 5.34: The full solution path shown with relative coefficients. Position is 63.5 N 49.5 W, west coast of Greenland.

From Figure 5.34 it appears that the first 7 coefficients are the more relevant.

name	coeffecients
intercept (1)	$-1.69 \cdot 10^{-1} [m]$
velocity (t)	$-2.23 \cdot 10^{-4} [m/day]$
acceleration ($0.5 \cdot t^2$)	$-9.16 \cdot 10^{-8} [m/day^2]$
$\cos(2\pi/365.2 * t)$	$-8.06 \cdot 10^{-3}$
$\sin(2\pi/365.2 * t)$	$-7.67 \cdot 10^{-2}$
$\sin(2\pi/182.6 * t)$	$-7.59 \cdot 10^{-3}$
$\sin(2\pi/121.7 * t)$	$-8.67 \cdot 10^{-5}$

Table 5.4: LAR coefficients in the seventh iteration

Using these seven coefficients the following line \hat{y} can be calculated:

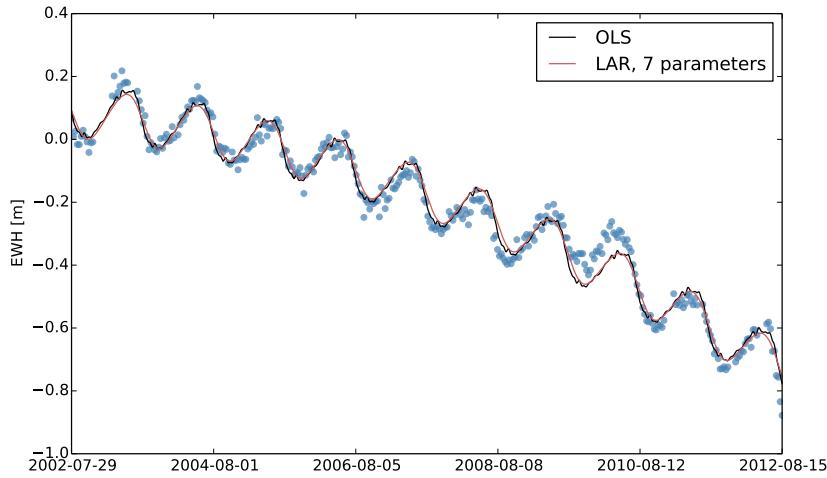


Figure 5.35: OLLS and LAR (seven parameters) compared. Position is 63.5 N 49.5 W, west coast of Greenland.

5.4.1 How many frequencies are actually needed

This was done for one position on the globe only. If one were to apply this to other positions one might get different lasso-paths and perhaps also a slight variation in the number of coefficients needed. For example for modeling the ocean one would not need as many coefficients, while at places with heavy rain season such as South America, more coefficients might be needed.

5.5 OLS with autocorrelated residuals

Since the OLS model with autocorrelated residuals, requires an equidistant time series a linear interpolation, was used to fill in the gaps in the GRACE data.

The difference between OLS with and without autocorrelated residuals, was then examined for a multitude of different locations on the globe. In this examination the result was generally the same, the estimated \hat{y} was not changed very much, but the variance of the residuals were quite a bit smaller.

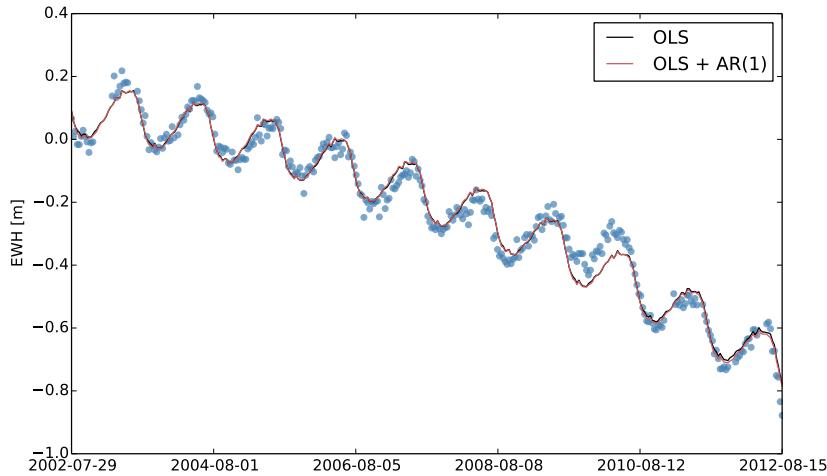


Figure 5.36: OLS and OLS with autocorrelated residuals. Position is 63.5 N 49.5 W, west coast of Greenland.

	OLS	OLS+AR(1)
ρ	0	0.773
$\hat{\sigma}_\epsilon^2$	$1.35 \cdot 10^{-3}$	$5.64 \cdot 10^{-4}$
MSE	43.027	43.482
Durbin-Watson	0.477	2.328

Table 5.5: Summarized difference between OLS with and without autocorrelated residuals

From Figure 5.36 it is seen that the \hat{y} are almost identical, which is because the $\hat{\beta}$ parameters are almost identical. One could have hoped for a better result here, but the reason for the poor fit is probably that the season length are not the same each year and correcting for autocorrelated residuals thus does not do much of a difference here.

In Table 5.5 it is seen that the MSE ($\sum_{i=1}^n (y_i - \hat{y}_i)^2$, not interpolated) is similarly unchanged. The estimated variance and the Durbin-Watson test statistics both, however, improved. This suggests that the OLS overestimated the noise variance of the data. The result should be that the p-value are more correct and in this case more parameters should be significantly different from zero.

5.6 Splines

Using nine knots with a year in between and letting the sine and cosine functions enter at each knot, one can allow for different seasons. The columns in X can be visualized as in Figure 5.37.

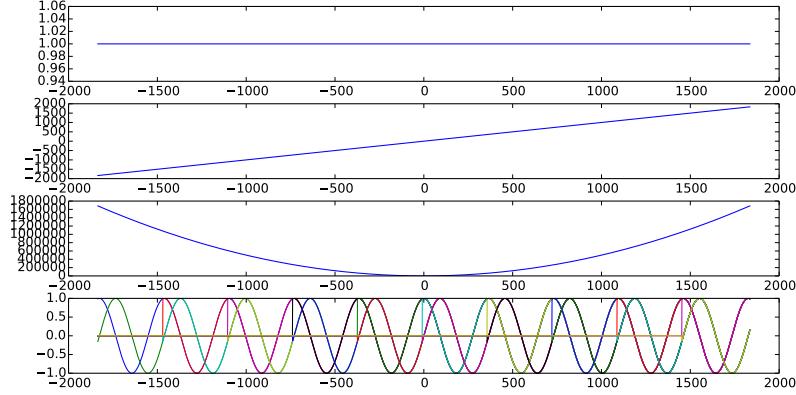


Figure 5.37: Constant, trend and acceleration in the three uppermost plots. Sine and cosine hinge functions in the bottom plot.

A RMSE diagnostic shows that the residuals are indeed smaller (max standard OLS was 0.135).

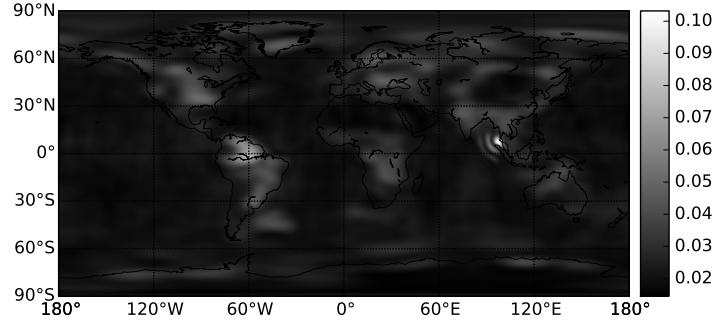


Figure 5.38: RMSE for each position using basis expansion.

Looking at the west coast of Greenland now using a basis expansion, the overall fit (Figure 5.39) looks improved. For some reason 2009 still causes some issues, which is partially seen in the residuals (Figure). Another thing to notices is that the curves are a lot more smooth, but this is simply because only the first 3 frequencies (year, half year and quarter year) was used. This was to prevent the degrees of freedom to drop dramatically, dude to the otherwise huge amount of parameters.

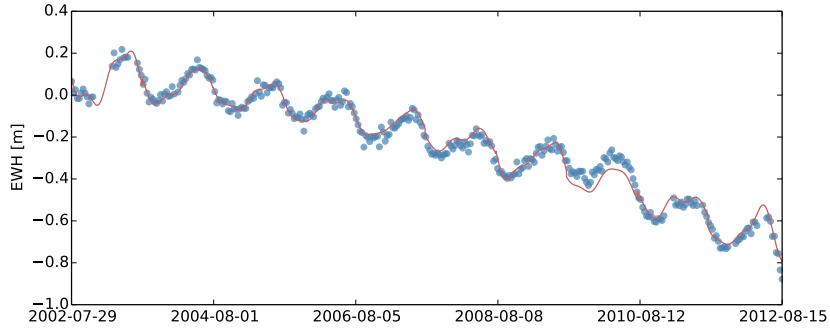


Figure 5.39: Measurements are blue, the OLS fit is red.

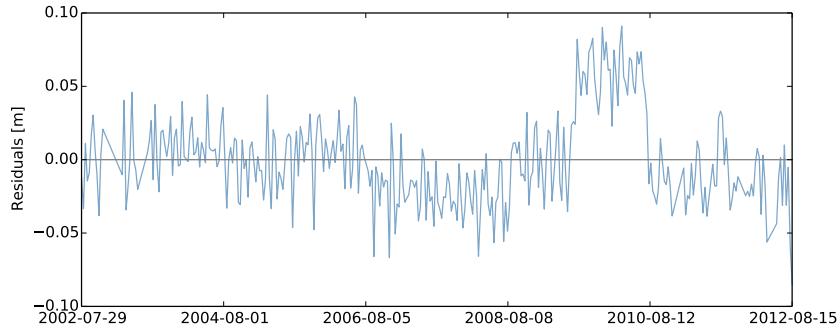


Figure 5.40: The OLS residuals are blue.

Unfortunately for the south pole, the resulting fits do display some cusps and overfitting behavior. This is particularly seen at 2006 and 2008. Here the residuals haven't improved much, however interestingly enough there now appear to be a two year seasonal trend, between 2003 and 2007.

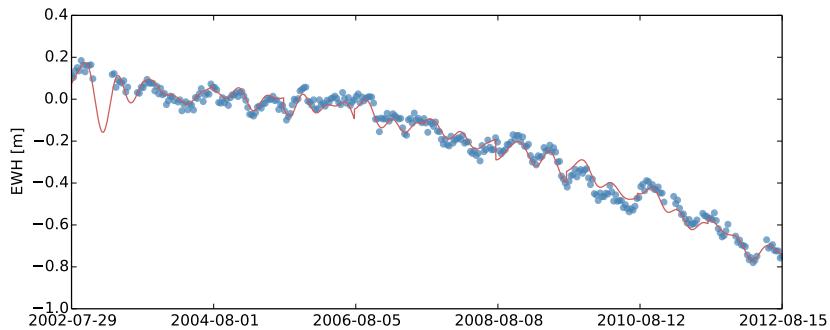


Figure 5.41: Measurements are blue, the OLS fit is red.

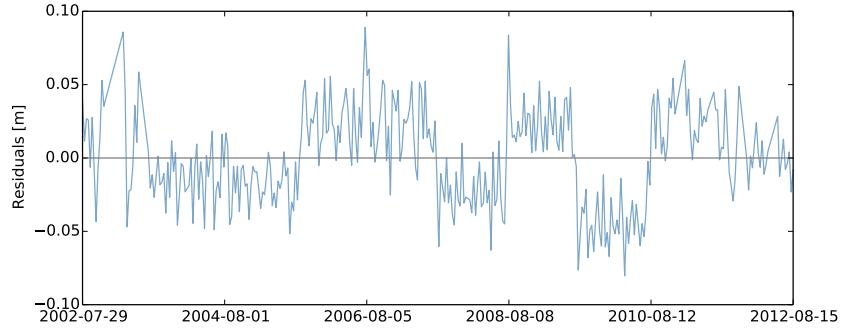


Figure 5.42: The OLS residuals are blue.

5.7 k-means clustering

5.7.1 Identifying the amount of clusters

Due to the size of the dataset (64800, 341), and the fact multiple simulated datasets of the same size would be needed to calculate the gap-statistic it was calculated on the HPC cluster that DTU offers for students and faculty. It should be noted that if such a setup had not been available, one could have used smaller subsamples of the data. 20 simulation samples of size (64800,341) were made using a multivariate uniform distribution. The following is the plot of the resulting gap statistic with its standard deviation.

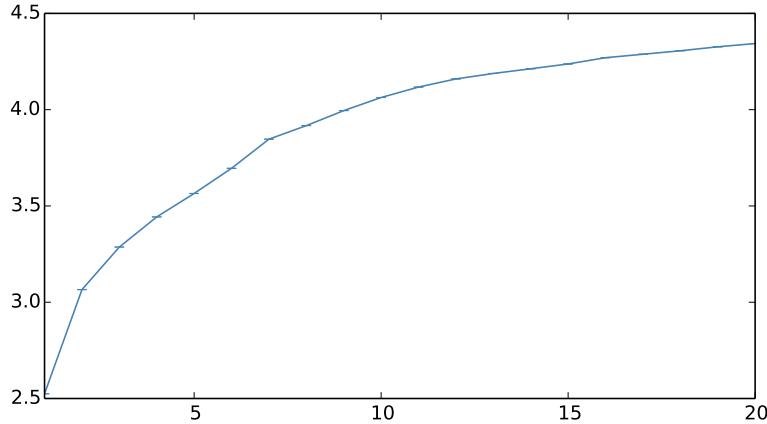


Figure 5.43: Gap-statistics with standard deviation. Note the standard deviation is very small.

From Figure 5.43 it is seen that according to the gap-statistics the optimal amount of cluster, is more than 20. Unfortunately such a high number of clusters are not suitable for visualization. Instead 7 clusters have been chosen; this is

based on the large slope change which can be observed in the gap-statistics graph. This is also seems like a suitable number of colors for visualization.

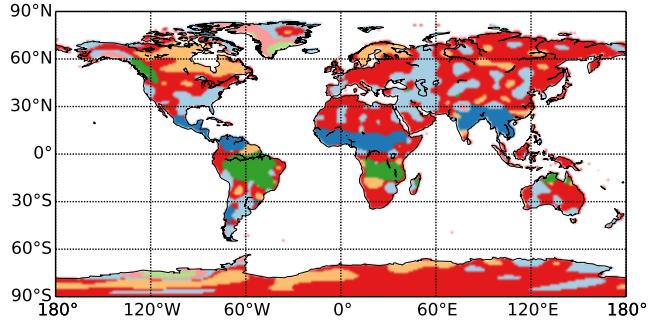


Figure 5.44: Each point belongs to the cluster with the closest centroid.

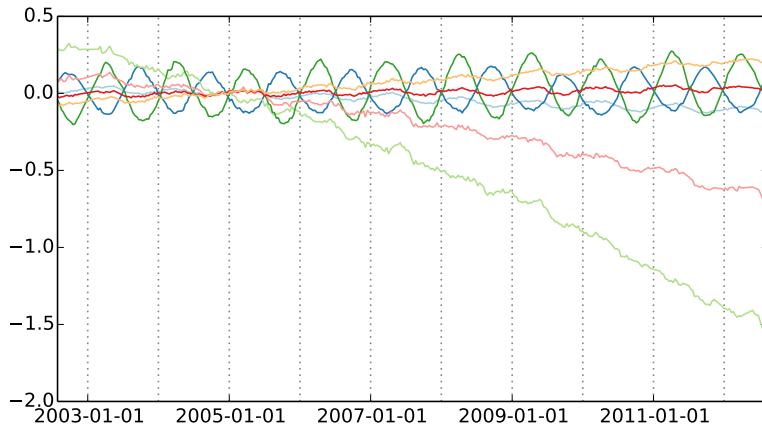


Figure 5.45: Cluster centroids. The colors correspond to those in Figure 5.44

Comparing the two plots above a few interesting insights are gained:

- Orange areas have a slight mass increase. At the south pole it appears that some of the mass loss at the edge actually moves inward towards the South Pole. This may be caused by post glacial rebound as no GIA has been performed.
- Green and blue correlates with extreme and regular seasonality (i.e. rain season in Amazon Basin).
- Light green and pink correspond with trending mass loss. The most significant locations appear to be located around the tip of the Western Antarctica along with Greenland's east coast.

5.8 GMM clustering

Like with K-means clustering, a matrix X with dimensions (64800×341) is used. Due to the high number of variables in GMM, the dimensionality was reduced using PCA and only selecting the first 10 PCs. The results were not ideal, so instead a kernel version of PCA was used.

- The first 10 principal components were used
- The “Radial basis function” was used as a kernel function
- The amount of degree is 3
- The gamma coefficient is set to $1/341$

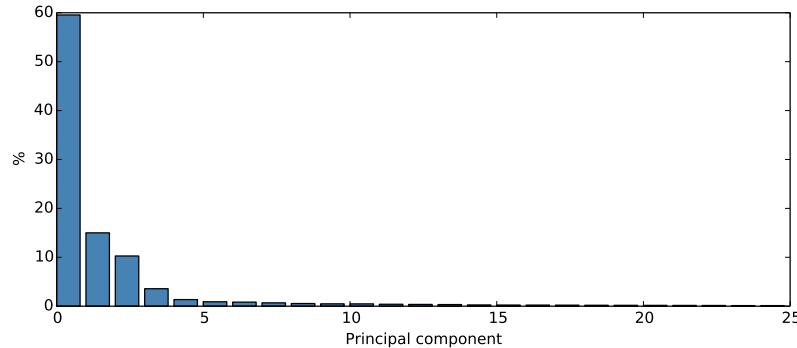


Figure 5.46: Variance explained for the first 25 PCs, calculated as eigenvalues in $K = H(X)H(X)^T$.

As seen in Figure 5.46, by selecting the first 10 PCs 93.18% of the variance in the data (in the kernel space) is explained.

Next the GMM with 6 clusters was computed using the projected data.

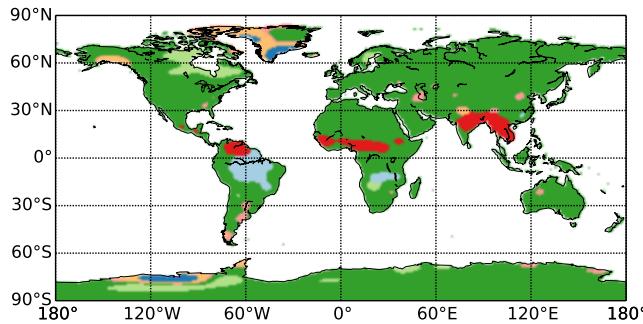


Figure 5.47: Each position is a point, the color then matches a given cluster.

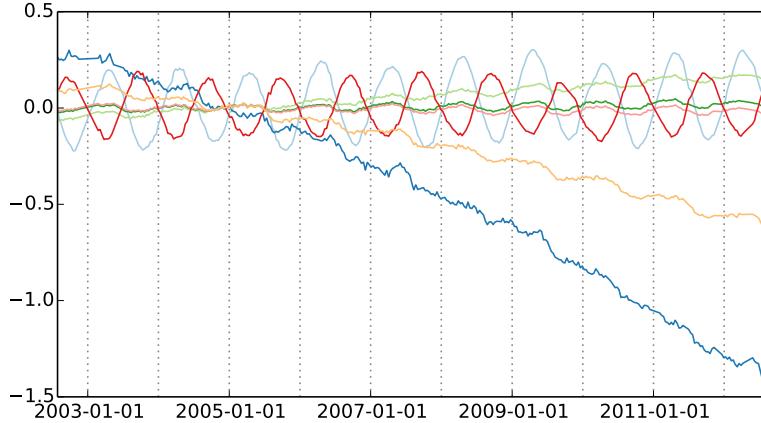


Figure 5.48: The mean vectors for each Gaussian component, projected back in the original space, using the inverse kernel PCA transformation.

Comparing the two plots above:

- Light green have a slight mass increase, this was also seen in K-means clustering. Since this is observed in North America, it is likely caused by post glacial rebound.
- Red and light blue correlates with extreme and regular seasonality, also very much like K-means.
- Blue and orange correspond with trending mass loss. With the same locations as in K-means.
- The big difference between GMM and K-means, is that the world graph is a lot less noisy. Only those positions with a significant different behavior than normal (green) have been caught by GMM.
- Finally the pink cluster is almost not noticeable and could be considered redundant.

6 Conclusion

The main purpose of this report was to describe mass losses or gains at the West Coast of Greenland and Western Antarctica. Both locations were found to be losing mass. Both location had periodic trends, but while West Greenland had soft oscillating curves Western Antarctica's mass loss was more jagged; this might suggests that ice is melting and breaking of in a more unstable pattern. However, overall it was found that ice at the two locations was melting with approximately the same speed and with a slight difference in acceleration.

Parameter/location	Western Antarctica	South Coast of Greenland
Velocity $\left[\frac{m}{day} \right]$	$-2.26 \cdot 10^{-4}$	$-2.24 \cdot 10^{-4}$
Acceleration $\left[\frac{m}{day^2} \right]$	$-1.24 \cdot 10^{-7}$	$-8.27 \cdot 10^{-8}$

Table 6.1: Velocity and acceleration for the two analysed locations

Of course these estimates depend on the selected locations and by choosing a different pair of locations, the estimates may not be exactly the same. To get a better understanding of the patterns plots of the velocity (Figure 5.6) and acceleration (Figure 5.7) in a world view, was also made. From these it is seen that Greenland and Antarctica are generally very similar in maximum magnitude of both acceleration and velocity. The most surprising observation is probably found in Greenland. Here the highest velocity is found on the east coast, but the highest acceleration is actually located on the west cost. This is not similar to Antarctica, where the highest acceleration and velocity is located at approximately the place.

Similar results were gained when using clustering algorithms to find similar locations. Here GMM with a Kernel PCA 5.8 seems to give a much less noisy result than K-means 5.7.

These results are of course influenced by the glacial isostatic adjustment. Later the ICE-5G dataset was used to correct for this (glacial isostatic adjustment), though NASA suggest [4] that this is not enough to get useful estimates.

Attempts to improving the variance of these estimates were also made. First by using a GLS model instead of a OLS model 5.5 and later by introducing basis expansion 5.6 to allow for more flexibility in the seasons. Both of these gave good results and it is even possible to combine the methods. However, in the case of the splines (technically it is not a spline, as it is not a polynomial), even though the obtained fit was theoretically good, several cusp and movements that to the human eye looked discontinuous existed. Thus it should not be recommended to use splines with cosine and sine functions, without doing something to improve the continuity.

Another way to improve the variance is to reduce the amount of OLS parameters. Here the p-values and the LAR 5.4 model suggests that only OLS parameters down to a period of $3 \cdot \omega$ is worth keeping. At least for those selected locations.

Finally a time series model was used in form of a seasonal ARIMA 5.3. The

6 CONCLUSION

results here were disappointing in that the residuals were far from being white noise distributed. This suggests that there are exogenous inputs which affects the system in a significant way. Finding and observing those variables would possibly also allow for a better OLS estimate.

7 Appendix

7.1 Time series analysis results

The first model is based on interpretations of graphs (R-output):

```

1 Series: xdata
2 ARIMA(3,1,3)(1,1,4) [36]
3
4 Coefficients:
5      ar1      ar2      ar3      ma1      ma2      ma3      sar1      sma1      sma2
6      -0.3860   0.1239   0.0471   0.2225   -0.2783   0.0035   0.0926   -0.9147   0.1808
7 s.e.   0.1556     NaN     NaN   0.1673     NaN     NaN     NaN     NaN   0.0508
8      sma3      sma4
9      0.0980   -0.1206
10 s.e.   0.0466   0.0749
11
12 sigma^2 estimated as 0.0005546: log likelihood=727.89
13 AIC=-1431.78    AICc=-1430.76    BIC=-1386.6
14
15 $AIC
16 [1] -6.435554
17
18 $AICc
19 [1] -6.427381
20
21 $BIC
22 [1] -7.315823
23
24 Warning message:
25 In sqrt(diag(x$var.coef)) : NaNs produced

```

References

- [1] NASA and DLR, *GRACE - Gravity Recovery and Climate Experiment*, 2012. [Online]. Available: <http://grgs.obs-mip.fr/grace/variable-models-grace-lageos/grace-solutions-release-02>.
- [2] GRGS, *GRACE - date format for GRGS files*, 2012. [Online]. Available: <http://grgs.obs-mip.fr/grace/variable-models-grace-lageos/formats#tab3>.
- [3] ——, *GRACE - grid format for GRGS files*, 2012. [Online]. Available: <http://grgs.obs-mip.fr/grace/variable-models-grace-lageos/formats#tab2>.
- [4] NASA, *GRACE monthly mass grids*, [Online; accessed 6-June-2014], 2014. [Online]. Available: <http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland/>.
- [5] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Pearson Education, Inc., 2006, ISBN: 0321420527.
- [6] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>.
- [7] A. A. Nielsen, *Least squares adjustment: Linear and nonlinear weighted regression analysis*, Richard Petersens Plads, Building 321, 2013. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?2804>.
- [8] H. Madsen, *Time Series Analysis*, ser. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2008, ISBN: 9781420059670.
- [9] N. Kousgaard, *Anvendt Regressionsanalyse For Samfundsvidenskaberne*, 1st ed. Akademisk Forlag, 1986, ISBN: 8750026143.
- [10] R. Tibshirani, *A simple explanation of the lasso and least angle regression*, 2014. [Online]. Available: <http://statweb.stanford.edu/~tibs/lasso/simple.html>.
- [11] O. Christensen, *Functions, Spaces, and Expansions*, 1st ed. Birkhäuser Verlag GmbH, 2010, ISBN: 9780817649791.
- [12] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a dataset via the gap statistic”, *J. R. Statist. Soc. B*, vol. 63, pp. 411–423, 2000. [Online]. Available: <http://www.stanford.edu/~hastie/Papers/gap.pdf>.
- [13] NASA, *Glacial isostatic adjustment*, [Online; accessed 23-June-2014], 2014. [Online]. Available: <http://grace.jpl.nasa.gov/data/pgr/>.