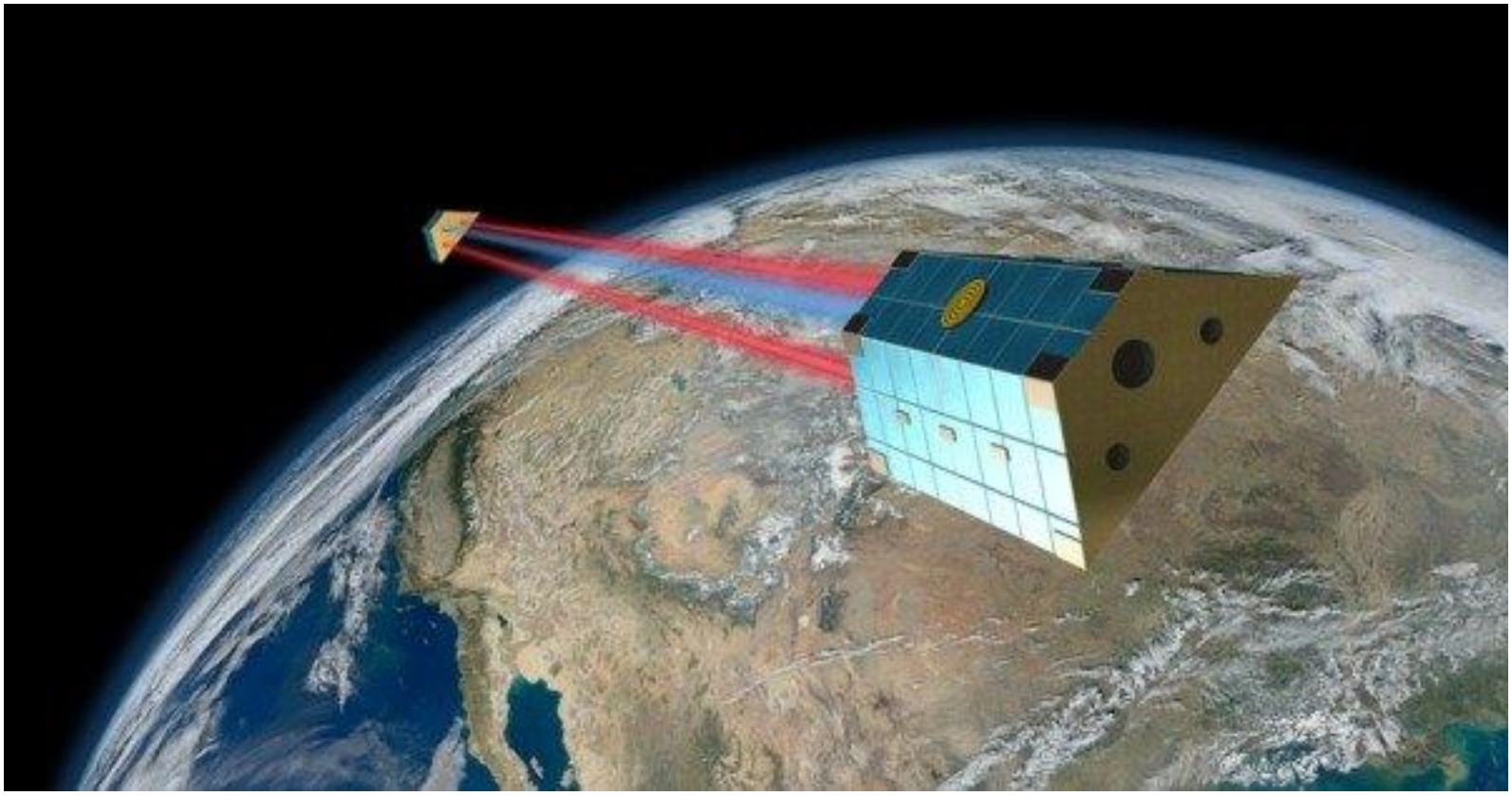
Analysis of noise in K-Band Ranging data



An investigation of noise in KBR ranging data of NASA's and GFZ's GRACE Follow-On mission. Detection of outliers and estimation on the noise power spectral density.

Bram Eijgenraam (4490886)

Department of Geoscience and Remote Sensing

Supervised by Dr. P.G. Ditmar

Delft University of Technology

Table of contents

Abstract

The GRACE Follow-On mission is using classical K-Band ranging (KBR) and a new laser-ranging interferometry (LRI) method. The latter gives ranging data two orders of magnitude more accurate compared to the classical K-Band ranging data *(Dahl C et al. 2016).* This gives a new opportunity for analyzing KBR noise by defining KBR noise as the difference between the KBR and LRI ranging data.
In order to get a realization of KBR noise, the KBR and LRI epochs should be aligned and outliers have to be removed.
An interpolation on the LRI data to make the LRI epochs aligned with the KBR epochs did not give sufficient results. Therefore only the overlapping epochs were used as input for the next step of outlier detection, a total of 2,250,000 epochs.
The outlier detection method starts by taking the absolute difference between the KBR and LRI range-rates. The largest differences were investigated where the original range-rate was compared to an estimation of the true value. This 'true' value is found by the use of a third degree polynomial function through the six range-rates that lie next to the suspected outlier. The outlier detection method removes LRI range-rates that differ more than $2.0*10^{-8}$ m/s from the estimated true value and KBR range-rates that differ more than $7.7*10^{-7}$ m/s. 3764 LRI epochs and 362 KBR epochs were labelled as outliers.
After the outlier detection a histogram was made for the KBR range-rate noise. It was found that in order to get a normal distribution of KBR range-rate noise, the interval had to be in the range of $<-3*10^{-7}$ m/s, $3*10^{-7}$ m/s>. 59,000 still fell out of this noise range. 54,000 of these points formed 10 major clusters together where a subtraction of the KBR and LRI range-rates did not give a realization of noise, but a trend similar to the original signal and are therefore likely related to a clocking error. These 54,000 points were therefore also removed from the dataset.
To the remaining KBR range-rate noise values a threshold of 3σ was applied, removing an additional 4972 noise values that were larger than $4.03*10^{-7}$ m/s.
Finally , an estimation of the PSD of the KBR range-rate noise was made and compared to an old PSD image of December 2008 of the GRACE mission. The PSD plot of this project shows three peaks between 10^{-4} Hz and 10^{-3} Hz. These peaks are not present in the PSD plot found in literature of the GRACE mission. Except from the second and the third peak, the PSD values of the KBR range-rate noise of this project were lower, up to two orders of magnitude in the lower frequency range (around 10^{-4} Hz) and one order of magnitude in the higher frequency range (around 10^{-1} Hz).

1. introduction

This report is about the analysis of noise in the K-band ranging data of NASA's and GFZ's 'GRACE Follow-On' mission. Launched in May 2018, it is a successor of the GRACE mission (2002-2017), and is planned to do measurements for 5 years *(n.d. 2017).*

One goal of this new mission is to accurately map changes in the Earth's gravity field and the other one to model the static gravity field. The orbit and design are similar to its predecessor but the main new feature is the addition of laser-ranging interferometry. This means that the range between the two twin satellites is measured by the classical K-Band ranging and the new laser-ranging interferometry method. The latter is considered to be an experimental type of data collection. However, it gives ranging data two orders of magnitude more accurate compared to the classical K-Band ranging data *(Dahl C et al. 2016).* From now on, they will be simply referred to as 'KBR' and 'LRI'.

The first part in this report gives a brief description of the KBR and LRI method, as well as the equipment that is used. Next, the data is visualized and a preliminary analysis is carried out.

Since the range accuracy of LRI data is 2 orders of magnitude more accurate, KBR noise is defined to be the difference between the KBR and LRI data. The KBR and LRI range-rates take measurements on different time intervals however, so after the preliminary analysis, it is tried to perform an interpolation on the LRI data to make the time intervals aligned. This interpolation was not accurate enough, so it was decided to only use the data at the overlapping epochs.

Next, the KBR and LRI outliers are detected and removed. This is done by first finding the largest absolute noise values, so by subtracting The KBR and LRI range-rates from each other of the overlapping epochs. The corresponding KBR and LRI range-rates around these epochs are now analyzed and a outlier detection model is introduced to determine if a certain point can be labelled as an outlier or not.

After outlier removal there were 59,000 epochs that fell out of the expected noise range interval of $<-3*10^{-7}$ m/s, $3*10^{-7}$ m/s> and an investigation on these points is carried out in this section.

Finally, the power spectral density of the KBR Noise is estimated using Welch's method and it is compared to results published in literature.

2. Ranging instruments

2.1: K-Band Ranging

The objectives of K-Band Ranging is ultra-precise satellite-to-satellite tracking in a low orbit. K-Band has a frequency of about 24 GHz. Ka Band is also used which has a frequency of around 32 GHz.
The measurement concept used is Dual One Way Ranging (DOWR). This means that the two satellites transmit a carrier signal and the phase of the carrier generated by the other satellite is measured relative to the transmitted signal. The sum of the generated phases is proportional to the range change between the satellites *(Kramer H.J. n.d.).*

The KBR system consists of the following elements:

- **USO** – Ultra Stable Oscillator, it serves as frequency and clock reference
- **MWA** – Microwave Assembly – up-converts the ultra-stable oscillator signal to K and Ka-Band for transmission, and down-converts the received K and Ka-Band signals to baseband frequencies of 670 kHz and 500 kHz
- **The horn**
- **IPU** – Instrumentation Processing Unit, it provides the digital signal processing functions for the K- and Ka-band signals and also for the GPS signals.

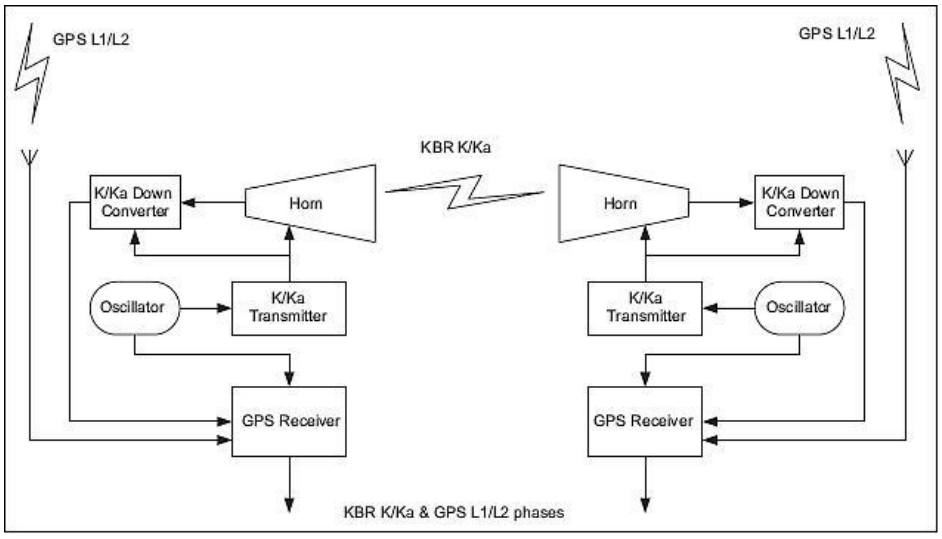


Figure 1: Schematic overview of KBR components (Jeongrae Kim & Seung Woo Lee 2009)

2.2: Laser Ranging Interferometer (LRI)

LRI is an experimental technology to improve the satellite-to-satellite tracking solution provided by K-Band Ranging. The basic principle of laser interferometry is that a single incoming light beam is split by a beam splitter. These two beams travel then a different path before they arrive at the same detector. The different paths cause a change in phase, which creates an interference fringe pattern (Renishaw plc. n.d.).From this pattern it is then possible to evaluate the displacement.

For the Grace Follow-on mission there is a master spacecraft and slave spacecraft with its laser offset phase-locked to the beam coming from the master spacecraft. Then the beam is sent back to the master spacecraft where the phase change is measured by interfering the local spacecraft beam with the incoming beam *(Abich K. et al., 2015).*By this method an improvement of ranging accuracy by a factor of around 100 is achieved.

The LRI system consists of the following elements:

- **An Optical Bench Assembly –** routes, detects and points the laser optical beams
- **Optical Bench Electronics** – provides power to the steering mirror and photoreceiver and signal conditioning between the photoreceiver and laser ranging processor
- **A Triple Mirror Assembly** – routes the beam around the MWI.
- **Optical Baffles** – prevent obstruction of the laser beams and control scattered light effects in the interferometer
- **A Light Path Closure** – protects the LRI during spacecraft integration and covers the triple mirror assembly to avoid contamination from the spacecraft
- **The Laser**
- **The Laser Ranging Processor**
- **An Optical Cavity** – stabilizes the laser light wavelength

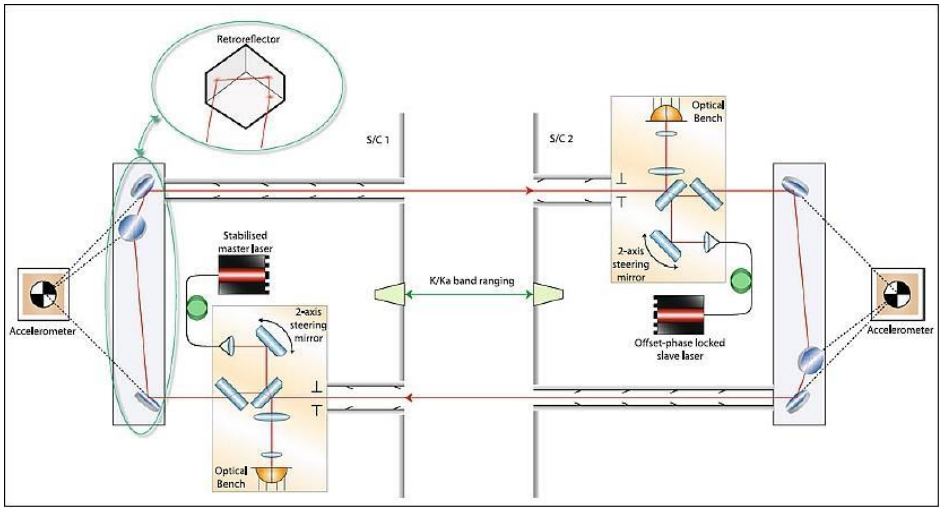


Figure 2: Schematic overview of LRI components (Kramer, H.J. n.d.)

3. Data visualization and preliminary analysis

I downloaded the KBR and LRI datasets from NASA's 'Physical Oceanography Distributed Active Archive Center' (PODAAC). They are part of the Level-1B Data Products *(GRACE FO 2019)*. I downloaded also a third dataset containing the position of the leading satellite C which will later be further elaborated. The time span of the downloaded data is from January 1 2019 to November 21 2019. They are stored in separate text files which all cover a period of 1 day. The selection of this time span is arbitrary but to obtain representative results a large dataset is desired. The first step was to write a script that merges these files and saves only the numerical values from each file. I wrote this code in MATLAB , and both KBR and LRI data have been saved as an .MAT file. These merged files can now easily be opened and read in a new script.

Let's briefly discuss the data structure. The first column is the GPS time. The first of January 2000 11:59:47 UTC is set here as the reference but in this project the reference point has been changed to zero. The second column of the KBR and LRI datafiles consists of the biased ranges. The third column is the range range-rate (first time -derivative of biased range) and the fourth column the range-acceleration (second derivative of biased range). Column 5 to column 11 give corrections for the ranges, range-rates and range accelerations. A full description of the KBR and LRI datasets can be found in the appendix. Both the range-rate and range -acceleration data would fit for an outlier detection analysis. It has been decided to only use the range-rate data in this project.

To correctly use this ranging data it is required to first compute the corrected range-rates.
According to the Level-1 Data Product User Handbook *(Wen et al. 2019)*, this is done by the following formula:

$$Corrected\ KBR\ range\ rate\ = range\ rate + lighttime\ rate + antenna\ phase\ center\ offset$$

$$Corrected\ LRI\ range\ rate\ = range\ rate + lighttime\ rate$$

Where the light time-rate is the seventh column of both the KBR and LRI dataset.
The antenna phase center offset is the tenth column of the KBR dataset (see appendix).

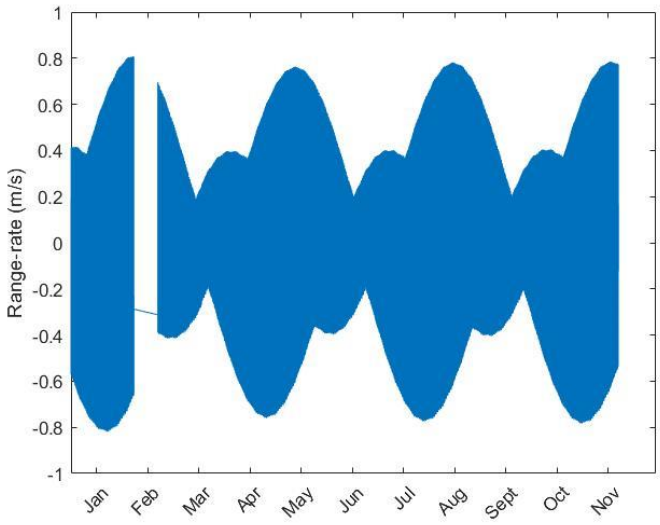


Figure 3: Corrected KBR range-rates

Figure 3 shows the corrected KBR range-rates as a function of time in months. On first sight there do not seem to be any blunders or large outliers in the data. It can be directly seen however that there is a gap in the first part. The datafiles from February 8 to February 20 are empty and those of February 7 and February 21 miss more than half of the data. This gives a gap of 340 hours or 14.2 days. This gap has been reported on in the Grace Follow-on newsletter *(Webb F et al. 2019)*. On February 7, the On-Board Computer (OBC) on GF-2 (Trailing satellite), was automatically shut-down. It is explained that the cause of this is likely related to a bus communication fault. All science instruments were powered off as the default response to this event.

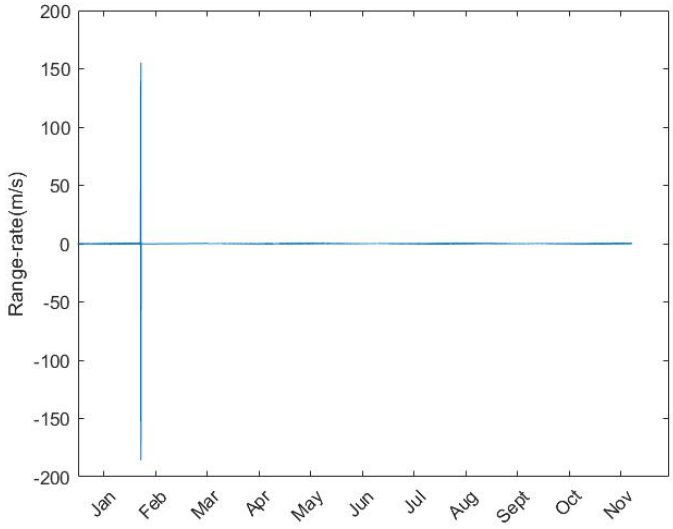


Figure 4: Corrected LRI range-rate

Figure 4 shows the corrected LRI range-rates, which gives a rather unexpected result.
It turns out that the dataset of February 6 contains a lot of blunders, ranging from roughly -150 to +150 m/s. These blunders are not specifically mentioned in literature (not found at least), but were measured just before the shut-down on February 7. The original GPS time interval is 602746304 s-602769598 s. What is striking here is the fact that the end of this interval is also the last measurement of February 6 and that from the first measurement of February 7 all the measurements fall in the correct value range again and don't show any strange behavior. All the range-rates that fall into the interval mentioned above are removed from the dataset, which is about 30% of the measurements made on February 6.

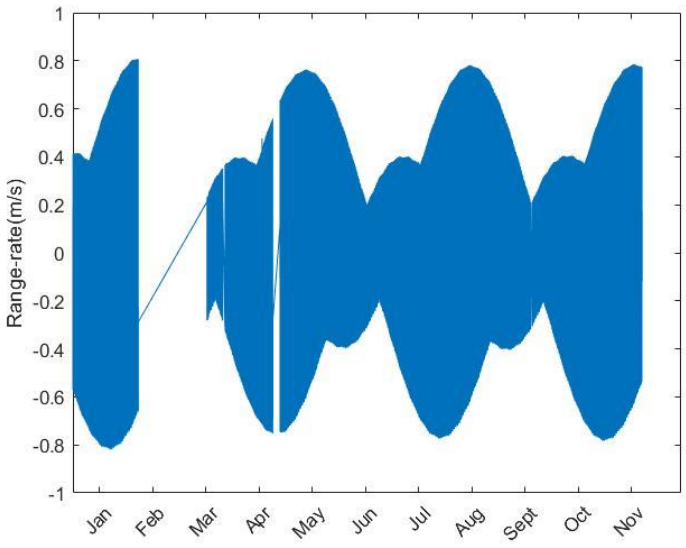


Figure 5: LRI range-rate after removal of February 6 time series

Figure 5 visualizes the range-rate without the data of February 6. The result now looks similar to the range-rate of the KBR dataset of figure 3 except that there is an even bigger gap in the data. From February 7 to March 17 the datafiles are empty. As explained before, on February 7 there was a shut-down of the On-Board Computer (OBC) on GF-2. In the GRACE Follow-On newsletter it is also stated that on March 18 2019, the LRI on GF-2 was restarted. This is in line with what can be seen in figure 5.

4. LRI range-rate interpolation

The KBR dataset contains measurements made every 5 seconds and the LRI dataset consists of measurements made every 2 seconds. So on an interval from 0 to 10 seconds the KBR dataset got 3 datapoints at 0,5, 10 seconds and the LRI dataset 6 datapoints at 0, 2, 4, 6, 8, 10 seconds.
It has been verified that these 10 second intervals are the same, so the KBR and LRI measurements at t=0,10,20...sec indicate the same epochs for both datasets. Noise can be defined as the difference between the KBR and LRI data as mentioned before. In order to compute this difference it is necessary to have measurements at the same epochs. When an interpolation is carried out on the KBR dataset, the stochastic properties of the noise are changed, which will make further analysis less reliable. It has therefore been decided to interpolate the LRI dataset.

A polynomial interpolation has been used. The built-in function in MATLAB called 'polyfit' is used in order to do this interpolation. Depending on the degree chosen, the function will return a number of coefficients which can be used to recreate the original signal. For this interpolation, a certain interval length is used as input to create the polynomial. On the chosen time interval, some points are left out which will be used as test data.

The number of test data samples on a chosen time interval is defined as the number of points on that interval divided by 4, decimal values will be rounded downwards. This way there is at least 75% training data available for the model. After determining the number of test data samples, their locations for each polynomial are randomly picked from the second index to the second last (using the first and final point is extrapolation). So if one chooses for example to use a time interval of 20 seconds which contains 10 data points (n=10), there are 2 test data samples randomly determined from n=2 to n=9 and for the next polynomial there will be 2 new locations randomly picked.

The training data is used to estimate the model parameters. Then these parameters are used to estimate the range-rates at the test data epochs. The interpolation error is defined as the absolute difference between the original range-rates and the estimated range-rates at the test data epochs.
The number of analyzed time intervals from which the average is taken is set to 10,000. One can choose to create more than 10,000 time intervals for a more accurate error estimation, but it is assumed that 10,000 time intervals are enough to get a decent approximation.

Different interval lengths and degrees of the polynomials have been tried. Table 1 gives a schematic overview of the results. The empty spaces indicate that either the degree is larger than the interval length or that it is obvious that increasing/decreasing the degree will not give better results.

Table 1: Mean absolute interpolation error (m/s) as function of polynomial interval length and degree

	2nd degree pol.	**3rd** degree pol.	**4th** degree pol.	**5th** degree pol.	**6th** degree pol.
l=5 1 test sample	$1.35*10^{-8}$	$1.97*10^{-9}$	-	-	-
l=6 1 test sample	$2.53*10^{-8}$	$1.86*10^{-9}$	-	-	-
l=7 1 test sample	$3.51*10^{-8}$	$2.17*10^{-9}$	-	-	-
l=8 2 test samples	-	$2.96*10^{-9}$	$8.74*10^{-9}$	-	-
l=9 2 test samples	-	$4.08*10^{-9}$	$9.86*10^{-9}$	-	-
l=10 2 test samples	-	$5.98*10^{-9}$	$1.07*10^{-8}$	$1.62*10^{-8}$	-
l=11 2 test samples	-	-	$1.48*10^{-8}$	$1.93*10^{-8}$	-
l=12 3 test samples	-	-	$2.04*10^{-8}$	$2.66*10^{-8}$	
l=15 3 test samples	-	-	$4.06*10^{-8}$	$5.25*10^{-8}$	$6.19*10^{-8}$
l=20 5 test samples	-	-	-	$1.27*10^{-7}$	$1.57*10^{-7}$

LRI range-rate accuracy is of order 10^{-9} m/s which means that the interpolation error should be at most 10^{-10} m/s. Based on the results of table 1, it can be concluded for now that the mean absolute error is too large in order to perform a decent interpolation. A third-degree polynomial for 6 points is the most accurate fit that has been achieved. The corresponding error of $1.86*10^{-9}$ m/s is still too large however.

Considering the KBR range-rate accuracy of 10^{-7} m/s it could be argued that the error is acceptable however. To check whether this is true it would be interesting to investigate this in a future analysis. In this project only the data at overlapping epochs is investigated in the following processing steps.

5. Noise preliminary analysis

Let's first introduce a third dataset : The GPS navigation dataset (shortly mentioned in section 1 of the preliminary analysis). The reason to include this dataset is to investigate if there is any geographical relationship between the outliers. The GPS dataset contains the orbit of the leading satellite C in the Earth-Fixed Frame, i.e., the x-, y-, and z-coordinates every second (GPS data for satellite D is also available but using data of only one satellite will be sufficient). The first measurement has been set as a zero reference point, just like the KBR and LRI dataset. The x and y coordinates have been converted to latitude and longitude in degrees. A full description of the data can be found in the appendix.

In order to produce a realization of noise in KBR data, the time series of the KBR, LRI and GPS dataset must fully coincide. There are some gaps in the data (especially for the KBR and LRI data) and it is therefore needed to perform an intersection.

The intersection is done on the KBR, LRI and GPS navigation dataset. The KBR and LRI time series overlap at t=0,10,20, etc., which means that at most 1/2 of the KBR and 1/5 of the LRI data could be used in the final range-rates that are used for the noise analysis.
Table 2 gives an overview of the original amount of datapoints that would be used with full overlap, and the actual amount of overlapping datapoints (GPS data included) which are used for the next step.

Table 2: Amount of KBR and LRI datapoints before and after intersection

Data type	Original amount of datapoints	Amount of overlapping datapoints	Ratio of original amount of datapoints
KBR	2683059	2254329	0.84
LRI	2411726	2254329	0.93

Now that the time series of the datasets are aligned, the next step is to take the difference between the KBR and LRI range-rate data to get a realization of noise in KBR data.

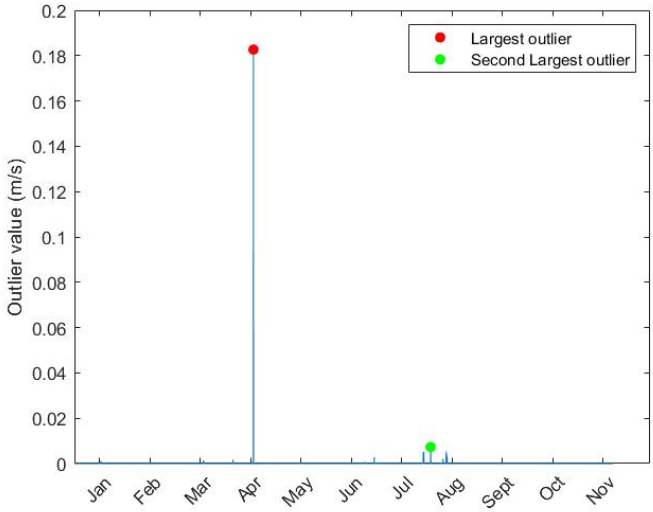


Figure 6: Absolute difference between KBR and LRI range-rates

Figure 6 shows the absolute difference between the KBR and LRI range-rates as a function of time. If both datasets were correct and didn't contain any outliers, one would see a visualization of the KBR range-rate noise in the graph above. This is not the case however. KBR range-rate data is of order 10^{-7} m/s accuracy and LRI data of order 10^{-9} m/s. The large values that are shown in figure 6, indicate that at some epochs one of the datasets (or perhaps both), contain outliers that cause the large differences. The red dot indicates the largest difference present. Let's take a closer inspection at this point:

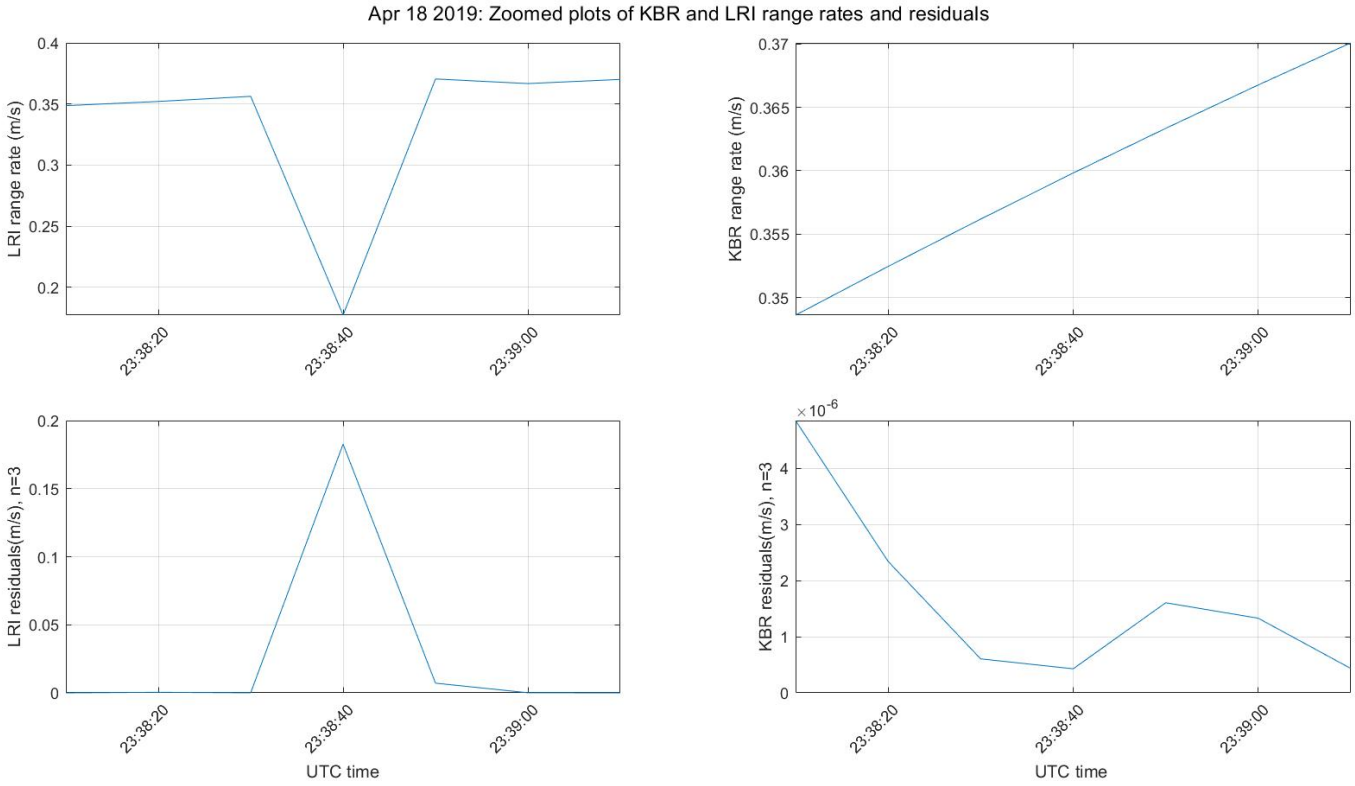


Figure 7:
Upper panel: Zoomed plots of KBR and LRI range-rates and residuals (April 18, 2019). Left top displays the LRI range-rate, the right top the KBR range-rate.
Lower panel: A third-degree polynomial has been fit through the KBR range-rate data on a time interval of 70 seconds (7 datapoints). This polynomial is subtracted from the original LRI and KBR range-rates. The resulting LRI residuals are shown at the Left bottom and the resulting KBR residuals at the right bottom.

In figure 7 it is clearly visible in the left top image that the LRI range-rates show a sudden drop at the epoch where the largest outlier value occurs. A cycle slip can be the explanation of this error. This specific event has not been found back in literature. A power loss or failure of the receiver software can both be causes of this phenomenon. The KBR range-rates in the right top does not show any strange behavior. All the subplots display data for 7 epochs, with the epoch containing the suspected outlier in the middle. A third degree polynomial fit has also been performed on both KBR and LRI range-rates. The choice of performing a third degree polynomial on 7 points is based on table 1. The polynomial fit for l=5 or l=6 is slightly better, but in order to include more points to be inspected, l=7 has been chosen. The corresponding residuals of this fit ($|\boldsymbol{y}\text{-}\hat{\boldsymbol{y}}|$), are displayed in the lower two subplots. The peak in the lower left image gives a clear indication that an outlier is present in the LRI dataset.

Let us also take a look at the next largest outlier value.

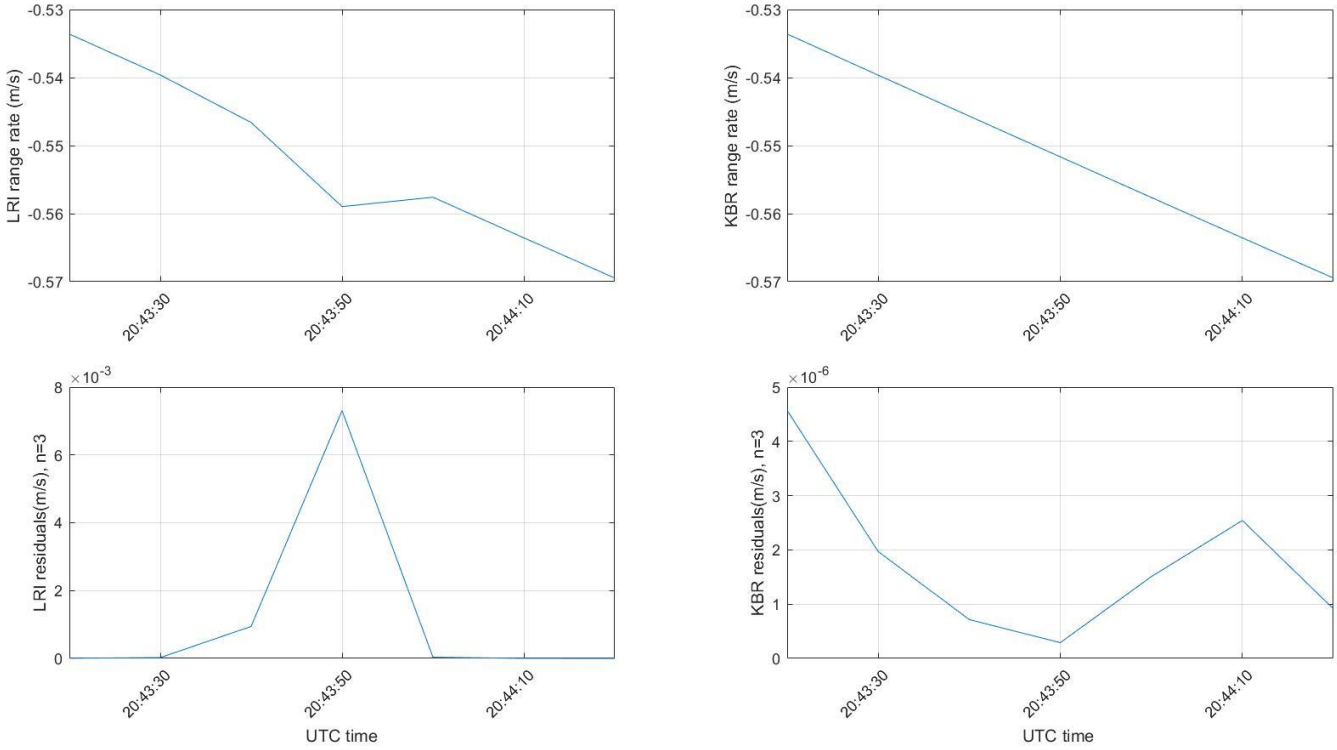


Figure 8: Same as Fig.8, but for an outlier on August 3 2019

From figure 8 it can be seen that the large outlier value is again caused by an outlier in the LRI dataset. In fact, a visual inspection of the first ten largest noise values clearly shows that all of them are due to outliers in the LRI range-rates.

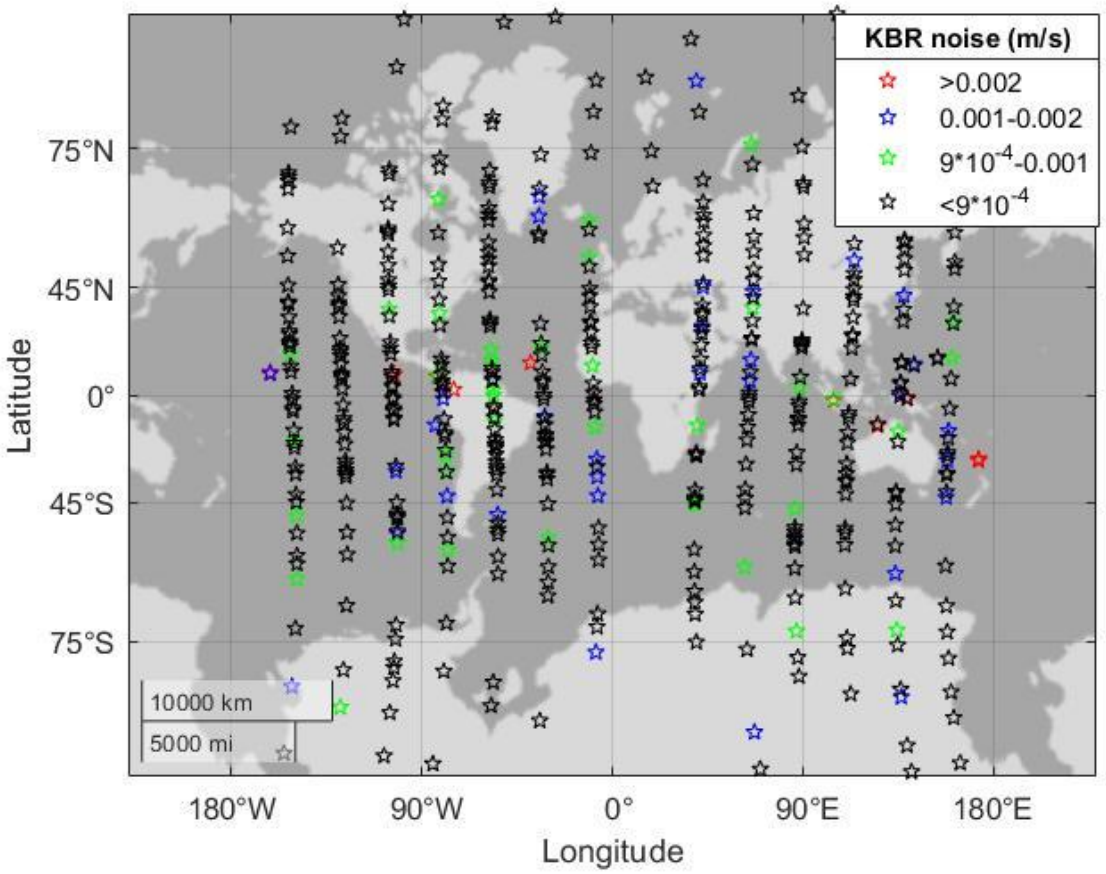


Figure 9: Geographical representation of largest range-rate outliers (m/s)

It is also interesting to investigate whether the range-rate outliers have any geographical correlation. The largest 500 outlier values are represented in figure 9. It can be seen that there is not a specific latitude range that dominates the location of the outliers. What is interesting to see however is that most outliers are located on tracks that have the same longitude. This might indicate that outliers on these tracks are part of the same satellite orbit.

The detection of outliers by visual inspection is inconvenient when there is a large dataset which may contain many outliers. It is therefore desired to create a code which can automatically label all the values which are considered to be an outlier. This will be discussed in the next chapter.

6. Outlier detection

6.1: LRI range-rate data

Let me introduce the outlier detection method used in this project. The idea is to first determine a polynomial function around the epoch of a suspected outlier. Then the difference between the polynomial and the actual value is computed. When this difference is too large the corresponding epoch is labelled as an outlier. The design matrix **A** looks as follows:

$$\mathbf{A}=\begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^n \\ 1 & t_2 & t_2^2 & \dots & t_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_m & t_m^2 & \dots & t_m^n \end{pmatrix},$$

where n is the polynomial degree, m the number of datapoints per polynomial and t denotes the time. For each polynomial, t_1 is set to zero and each timestep is equal to 10 seconds. The middle row is the location of the suspected outlier and this middle row contains only zero values. Since this is the location of a suspected outlier it has been decided not to use this epoch as input. When there is a large outlier it might have a big impact on the estimation of the polynomial function, which will then deviate significantly from the actual polynomial function.

Based on the results of table 1, a third- degree polynomial on 6 points will be used. This results in the following matrix:

$$\mathbf{A}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 10 & 100 & 1{,}000 \\ 1 & 20 & 400 & 8000 \\ 0 & 0 & 0 & 0 \\ 1 & 40 & 1600 & 64{,}000 \\ 1 & 50 & 2500 & 125{,}000 \\ 1 & 60 & 3600 & 216{,}000 \end{pmatrix},$$

The polynomial coefficients can now be determined by the formula:

$$\hat{\boldsymbol{x}} = (\mathbf{A}^{\mathrm{T}} * \boldsymbol{A})^{-1} * \boldsymbol{A}^{T} * \boldsymbol{d} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

where vector **d** contains the LRI range-rate observations and a_0 to a_3 are the polynomial coefficients. Note that this estimator $\hat{\boldsymbol{x}}$ implies white noise but these noise correlations are beyond the scope of this project.

With the formula $\hat{\boldsymbol{y}} = \hat{\boldsymbol{x}} * \boldsymbol{A}$, the original range-rates can now be estimated. Using design matrix A however will only result in the range-rate estimations of the neighboring epochs of the suspected outlier. Let us introduce matrix **B** which is the same as matrix **A**, but the middle row is now in line with the other rows :

$$\mathbf{B}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 10 & 100 & 1{,}000 \\ 1 & 20 & 400 & 8000 \\ 1 & 30 & 900 & 27{,}000 \\ 1 & 40 & 1600 & 64{,}000 \\ 1 & 50 & 2500 & 125{,}000 \\ 1 & 60 & 3600 & 216{,}000 \end{pmatrix},$$

Multiplying matrix **B** with the polynomial parameters in $\hat{\boldsymbol{x}}$ gives an estimation of the range-rates, including the estimated range-rate of the suspected outlier. This value found at the outlier epoch is the expected range-rate when there would be no outlier present, and therefore considered to be an estimation of the true value. One can now compute the difference between this estimation of the true value and the actual value in the range-rate dataset. This value is denoted as C.

When this estimation of the true value deviates too much from the original range-rate found in the dataset, one can conclude that this epoch contains an outlier. Whether this difference is too large is specified by the following equation:

$$\frac{|C|}{\widehat{\sigma_d}} > \sigma,$$

where $|C|$ is the absolute difference between the estimated true value and the original range-rate and $\widehat{\sigma_d}$ is standard deviation of the noise in C. This noise is defined as how accurate C can be computed on a time interval that contains no outlier. σ is a certain threshold. In order for a signal to be distinguishable it should be around five times the standard deviation of the noise, according to the Rose Criterion *(A. Rose, 1973).* Based on this statement σ is initially set to 5.

In order to find $\widehat{\sigma_c}$, the following formula is used:

$$\widehat{\sigma_c} = \sqrt{\sum_{i=1}^{m} \frac{r_i^2}{m-1}},$$

where r denotes the residuals computed by the formula:

$$\boldsymbol{r} = \boldsymbol{d} - \boldsymbol{B} * \hat{\boldsymbol{x}}$$

In order to find $\widehat{\sigma_c}$, 10,000 polynomials have been computed, starting from t=0 of the dataset. It is assumed that the first 10,000 polynomials of the dataset do not contain a lot of outliers and are representative for the whole dataset. Using design matrix A, 10,000 $\hat{\boldsymbol{x}}$ vectors are computed, which are then used in the equation above to calculate **r**. The residuals in **r** denote the difference between the true and the estimated range-rates where the error in the estimation of C is of interest. This means that from **r** only the middle value of each polynomial is of interest, giving us 10,000 residual values in total. This results in a $\widehat{\sigma_c}$ of:

$$\widehat{\sigma_c} = \sqrt{\sum_{i=1}^{10,000} \frac{r_i^2}{10,000 - 1}} = 1.61 * 10^{-6}\ m/s$$

In order to determine how many datapoints need to be investigated, it is smart to make a plot of the largest outlier values and plot them in descending order.

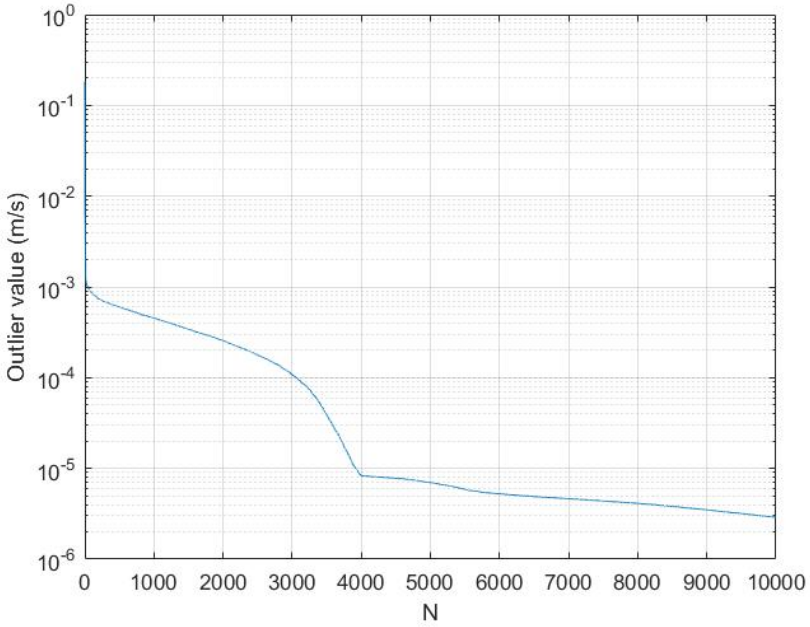


Figure 10: 10,000 largest outliers plotted in descending order

Figure 10 shows the 10,000 largest outliers and their corresponding value. It can be seen that the values decrease rapidly from N=0 to N=4000, after which the outlier values starts decreasing much more slowly

and shows a linear trend. Based on this plot, it has been decided to begin with an analysis of the 5000 largest KBR-LRI range-rate differences.

Let's take a look at the results when the $|C|/\sigma_C$ ratios are computed for these epochs:

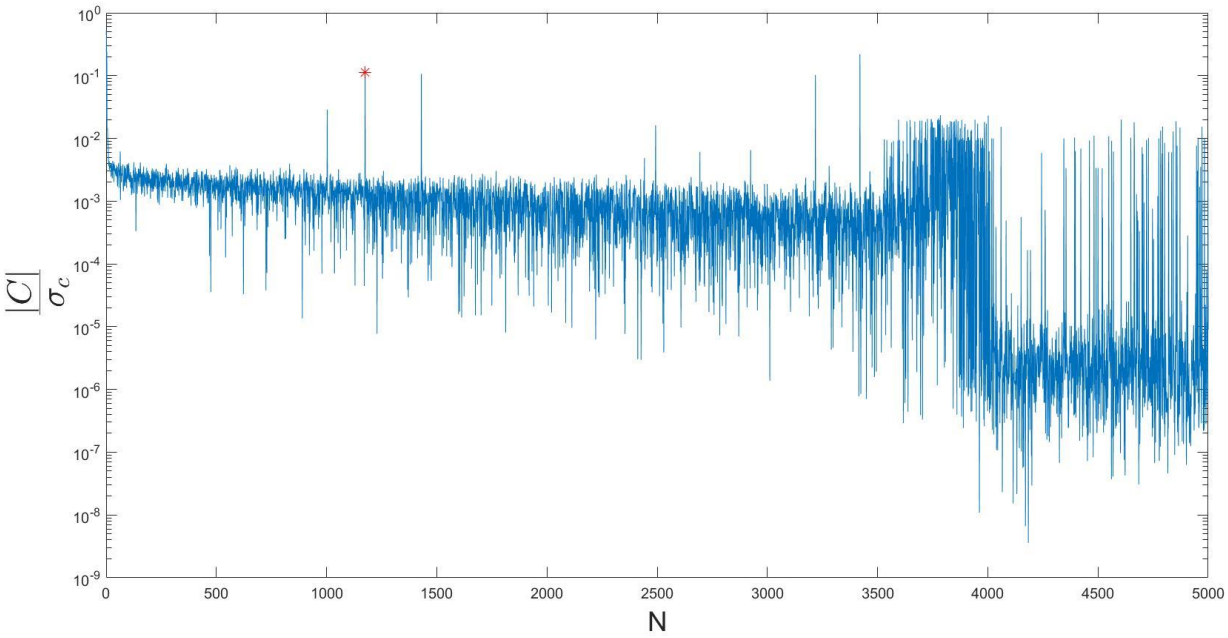


Figure 11: $\frac{|C|}{\widehat{\sigma_C}}$ ratio for the 5000 largest LRI-KBR range-rate differences (high to low)

Figure 11 shows the results of the largest 5000 LRI-KBR range-rate differences, ranked from high to low. It is clear that there is a correlation between the magnitude of the LRI-KBR range-rate differences and the $|C|/\sigma_C$ ratio but the values are highly varying and there are some large peaks that need further exploration. A sudden drop around N=3800 is still clearly visible however. Let me first discuss the unexpected peaks.

The LRI range-rates of the neighboring points of the suspected outlier are mathematically described by a polynomial fit. An estimation of the true value at the epoch of the suspected outlier is then computed based on this polynomial parameters. In order to find a reliable estimation of the true value, it is necessary for the neighboring range-rates to be correct and don't contain any outliers as well.
Let's illustrate this statement with an example.

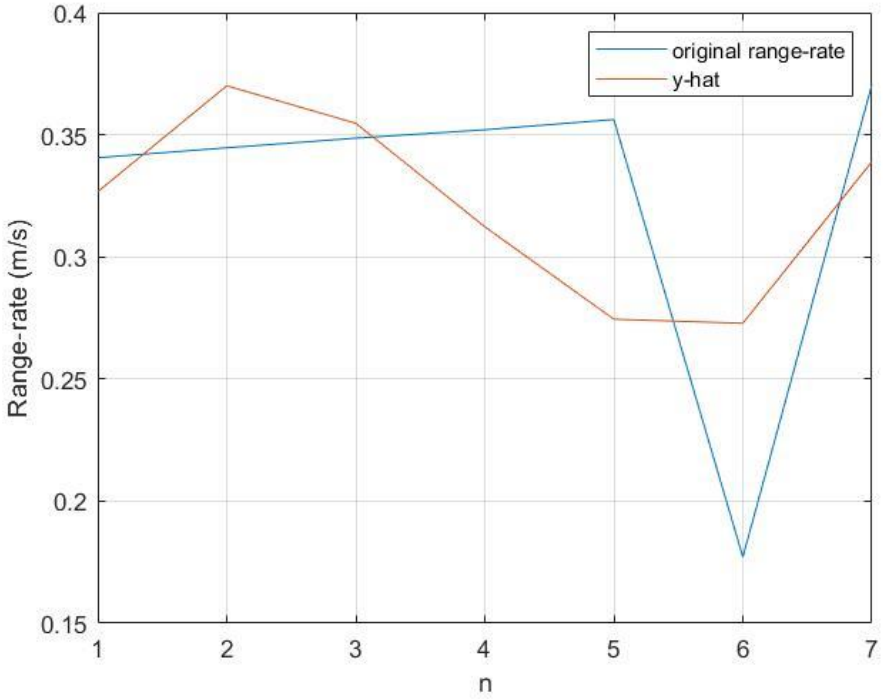


Figure 12: Original range-rates vs estimated range-rates y-hat at noise id N=1175

As can be seen in figure 11, there is a sudden peak at N=1175 (indicated with a red star). Figure 12 shows the original LRI range-rates and the estimated range-rates at this particular point. The suspected outlier is located at n=4. The other 6 epochs are used as input for matrix A to estimate $\hat{\boldsymbol{x}}$**.** The original range-rates show a clear outlier at n=6, which is causing an incorrect trend and results therefore in a wrong value for C at n=4.

So in order to determine if the range-rate at n=4 is an outlier, the method is only fully reliable when there are no other outliers presents in the data range from n=1 to n=7. When the $|C|/\sigma_C$ ratio gives a rather large value, this doesn't mean that the outlier at n=4 is present but could be anywhere from n=1 to n=7.

A possible solution would be that once a big outlier is found, the value is replaced with the estimated true value. This method is not so convenient for the current dataset however, as can be seen in the following figure.

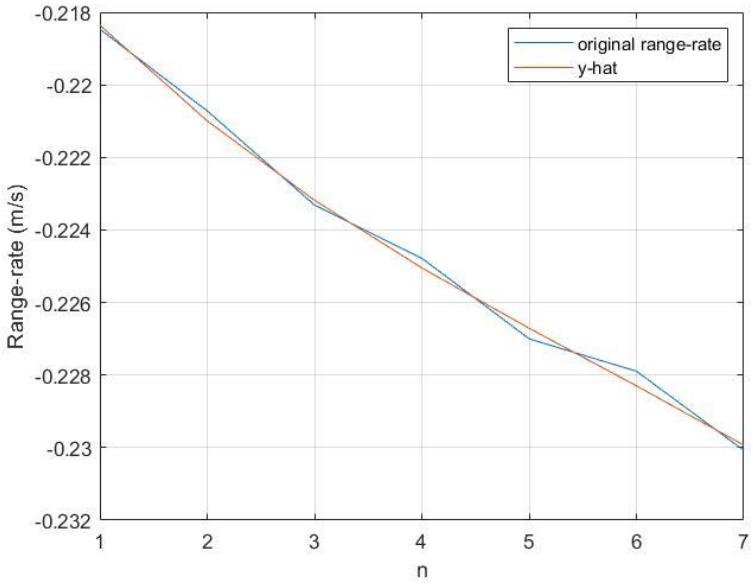


Figure 13: Original range-rates vs estimated range-rates y-hat at Noise id N=3300

Figure 13 shows the original range-rates vs the estimated range-rates at N=3300. The original range-rates show strong deviations over the entire time interval. One can now not simply replace a single outlier by an estimated true value since the outliers seems to be everywhere. This leaves us with the result that in the end some range-rate data will be labelled as an outlier while it is actually correct.

Let's take a look at the $|C|/\sigma_C$ ratios but now sorted from large to small.

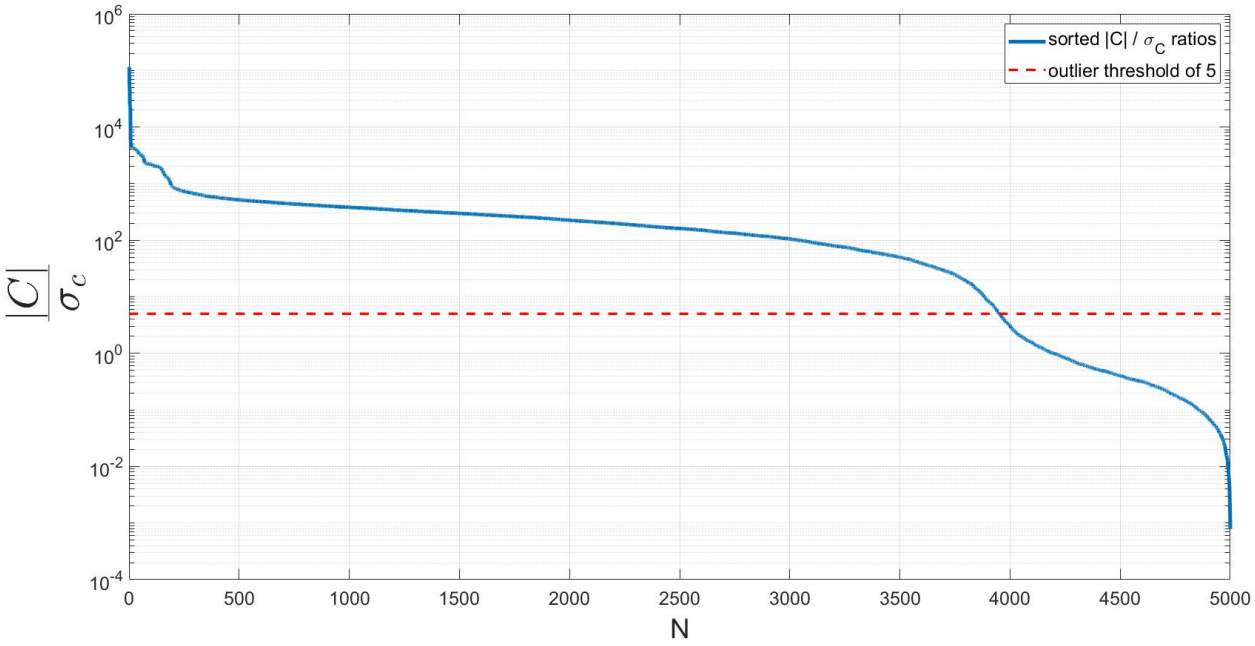


Figure 14: sorted $\frac{|C|}{\widehat{\sigma_C}}$ ratios of the 5000 largest noise values

Figure 14 shows the sorted 5000 largest $|C|/\sigma_C$ ratios and a red dashed line indicates the discussed outlier threshold of $\sigma = 5$. The intersection of the two lines is at N=3950. In order to check if this is a good threshold, we can look at the sorted |C| values themselves.

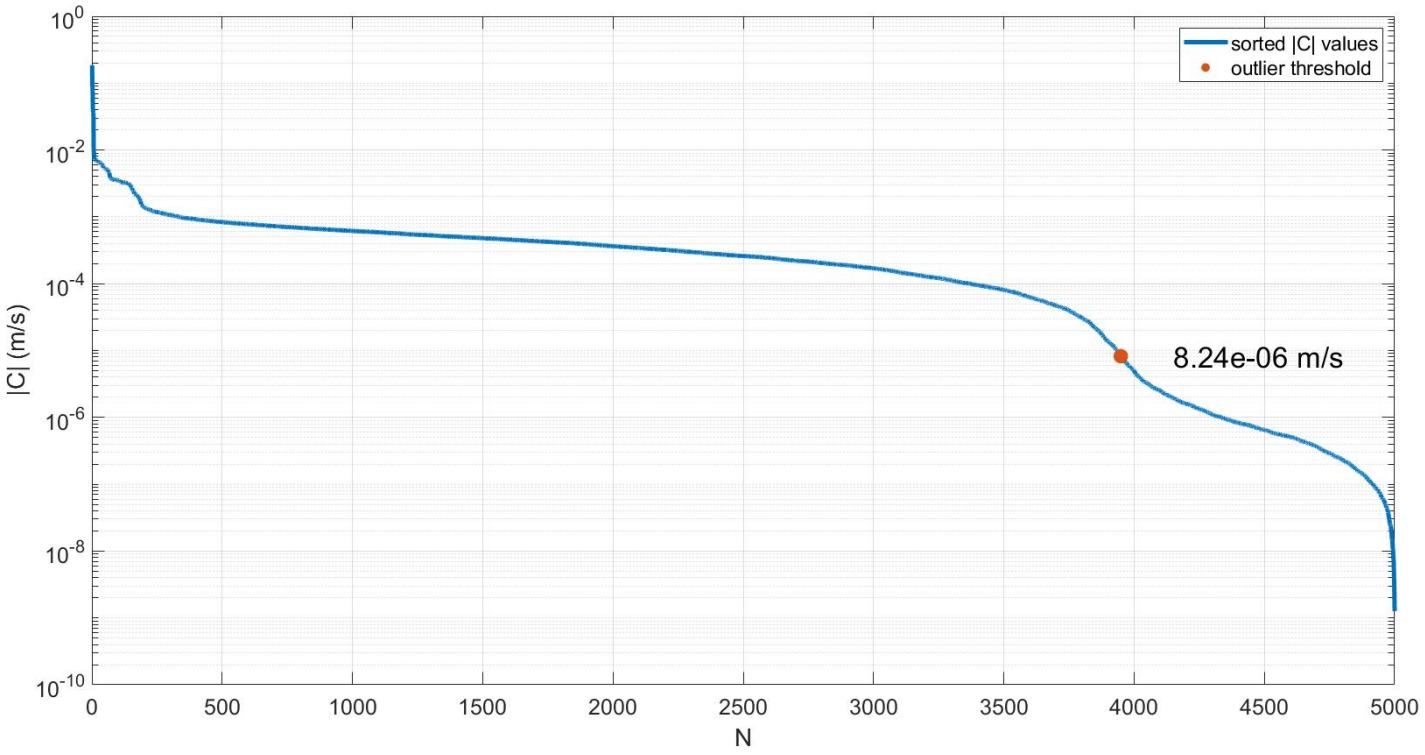


Figure 15: Sorted |C| values of the 5000 largest noise values

Figure 15 shows the sorted |C| values with the threshold location and corresponding |C| value pointed out. So based on this graph it can be concluded that whenever the LRI range-rate at an epoch deviates more than $8.24*10^{-6}$ m/s from the estimated true value, it is labelled as an outlier.
Based on the fact that LRI range-rates are of order 10^{-9} m/s accuracy, this value of $8.24*10^{-6}$ m/s is too high and more epochs should be labelled as an outlier.

It is possible to apply an improvement on the model. This involves the use of the original non-overlapping LRI range-rates. In order to process the noise, the KBR and LRI timeseries needed to be aligned. For the LRI dataset this meant that the data at t=2,4,6, 8 had to be removed.
The remaining LRI data was not only used for the processing of the noise but was also input for the outlier detection described above. It is also possible to use the original LRI data as input for the design matrix A, which will then result in a 2 second time-step instead of 10.

6.2: Use of original LRI range-rates

With the original non-overlapping LRI range-rates, design matrix A now looks as follows:

$$A=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \\ 0 & 0 & 0 & 0 \\ 1 & 8 & 64 & 512 \\ 1 & 10 & 100 & 1000 \\ 1 & 12 & 144 & 1728 \end{pmatrix},$$

The same procedure is followed as before. $\widehat{\sigma_c}$ is now found to be 3.35*10^{-9} m/s.
Compared to the originally found value of $1.61 * 10^{-6}$ m/s, this is a significant improvement.
Let's take a look at the final results:

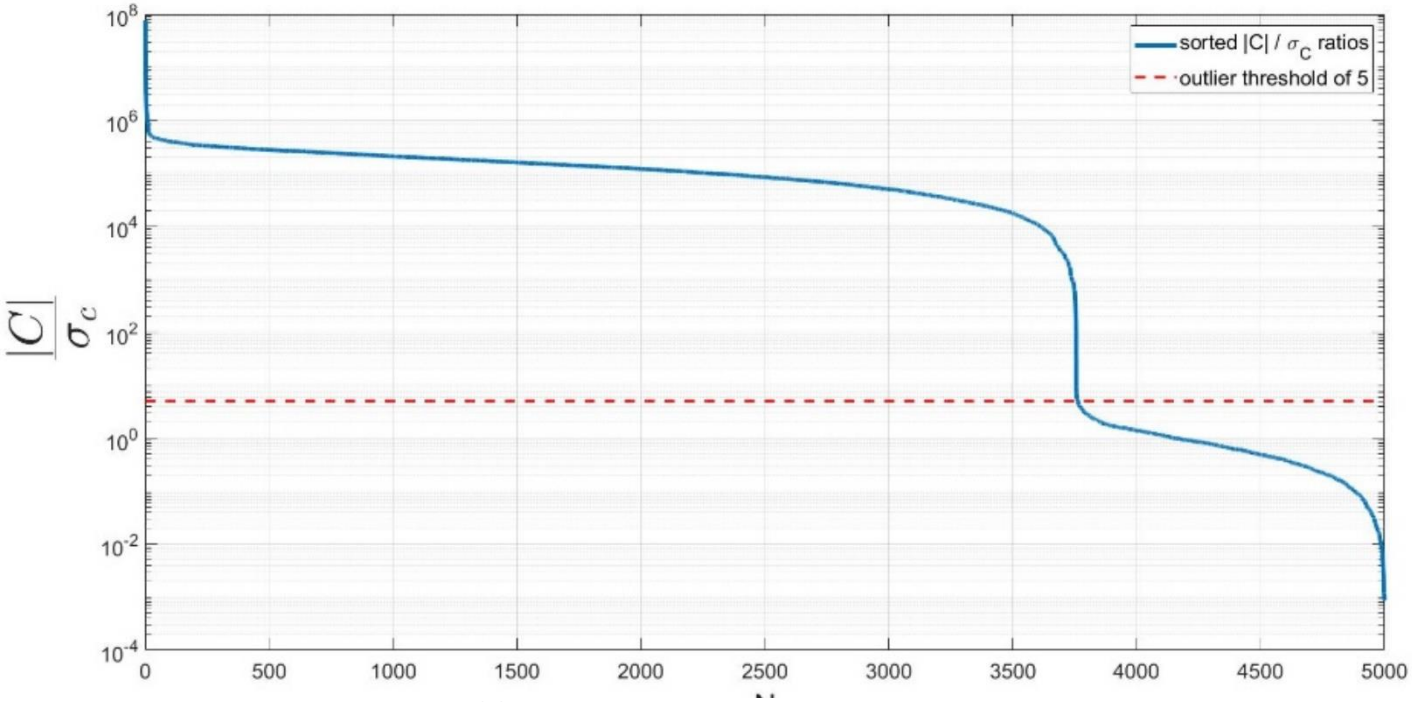


Figure 16: Sorted $\frac{|C|}{\widehat{\sigma_c}}$ ratios of the 5000 largest noise values using original LRI range-rates as input

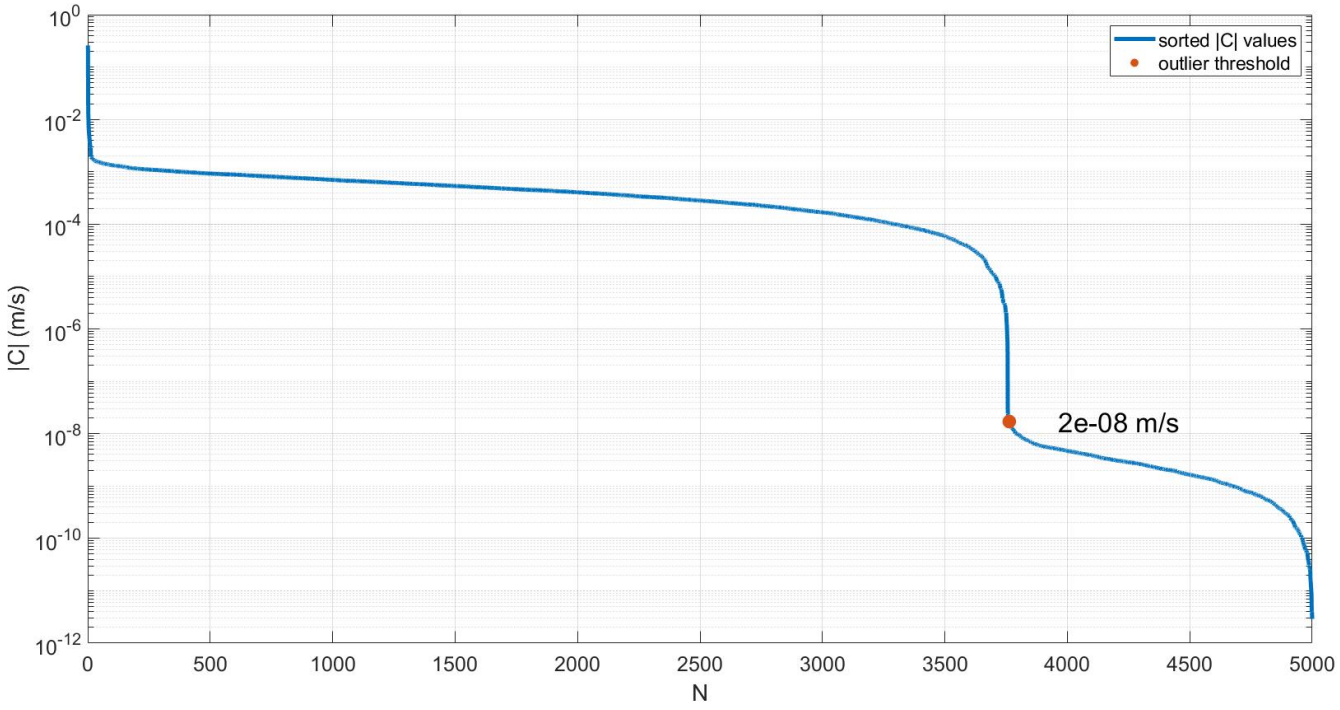


Figure 17:Sorted |C| values of the 5000 largest noise values using original LRI range-rates as input

Figure 16 and figure 17 show the sorted $|C|/\sigma_C$ ratios and sorted |C| values, respectively. When the outlier threshold of σ is set to 5, the intersection in figure 16 is found at N=3764. This value is located at the lower part of a big drop. It can then be derived from figure 17 that all epochs with a |C| value larger than $2*10^{-8}$ m/s will be labelled as an outlier. The large drop takes place just before the intersection from N=3756 to 3757. Here the $|C|/\sigma_C$ ratio drops from 144.5 to 7.2 in just a single step. Then the ratio gradually decreases to below 5 at N=3764. It can be concluded that using the original non-overlapping LRI range-rates improves the model. Using a 2-second time step in **A** results in a much lower $\widehat{\sigma_c}$ and therefore in a more accurate estimation of the true value C. When figure 15 and figure 17 are compared it can also be seen that in figure 16 there is a much steeper slope around the location of the intersection.

6.3: KBR range-rate data

The same Procedure can be used for removing the KBR range-rate outliers.
Since it was shown that using the original LRI range-rates gave more accurate results, the same will be done for the KBR range-rates. This gives the following design matrix A

$$\mathbf{A}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 5 & 25 & 125 \\ 1 & 10 & 100 & 1000 \\ 0 & 0 & 0 & 0 \\ 1 & 20 & 400 & 8000 \\ 1 & 25 & 625 & 15{,}625 \\ 1 & 30 & 900 & 27{,}000 \end{pmatrix},$$

$\widehat{\sigma_c}$ is now found to be 1.53*10^{-7} m/s. Let's plot the results:

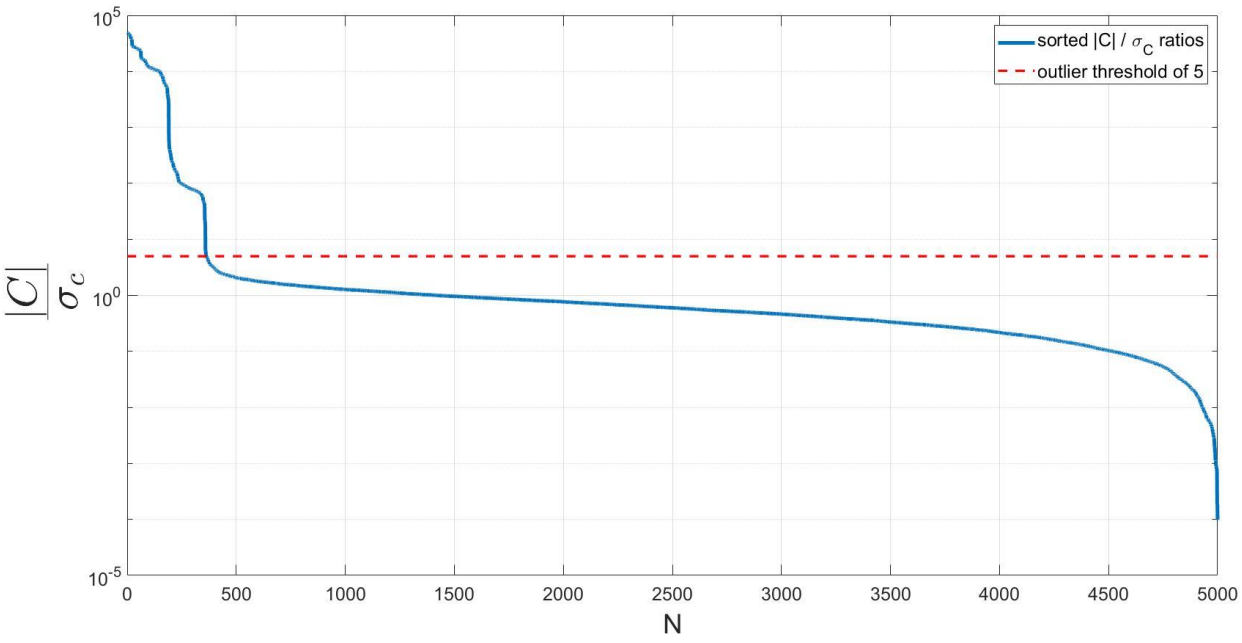


Figure 18: Sorted $\frac{|C|}{\widehat{\sigma_c}}$ ratios of the 5000 largest noise values using original KBR range-rates as input

Figure 18 shows the sorted $|C|/\sigma_C$ ratios when the original non-overlapping KBR range-rates are used as input. The intersection with the σ=5 line is found at the lower part of a steep slope at N=362.
The smallest C that will be identified as an outlier has a value 7.7*10^{-7} m/s. From N=358 to N=359 there is a large $|C|/\sigma_C$ drop from 18.7 to 6.2, from which it then gradually decreases. This is a similar feature that was discussed in the outlier detection of the LRI range-rate dataset.

6.4: Visualization of outlier epochs

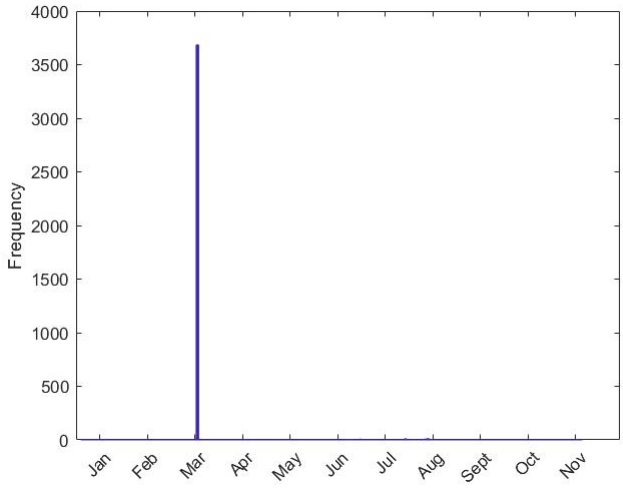


Figure 19: Histogram of LRI outlier epochs (200 bins, roughly 1.5 day per bin)

Figure 19 shows a histogram of the epochs of the LRI outliers. 3683 out of 3764 outliers are part of one day, March 19. This is just two days after the system was rebooted on March 17. Therefore, the outliers are likely to be correlated to this event, an indication of startup problems. It would be a recommendation for future analysis to remove the LRI data of March 19.

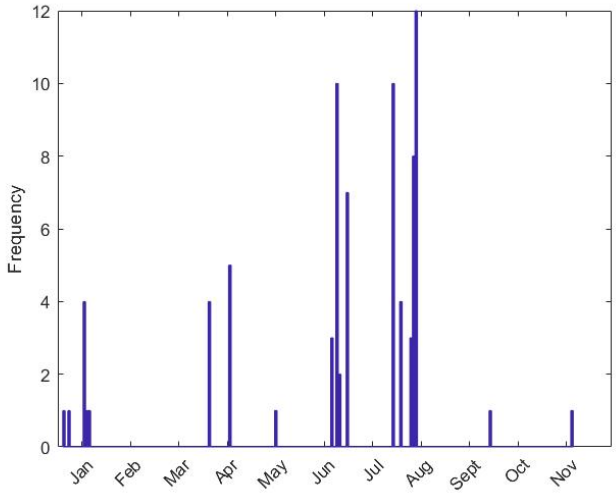


Figure 20: Histogram of LRI outlier epochs with outliers of March 19 removed (200 bins)

Figure 20 shows a histogram of the epochs of LRI outliers but with the outliers of March 19 removed. Most of the remaining outliers occur from June to August but they are also clearly present in the other months.

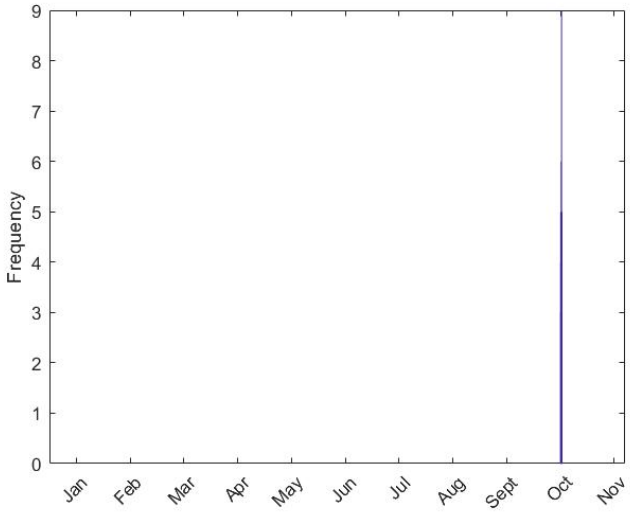


Figure 21: Histogram of KBR outlier epochs

Figure 21 displays the epochs of the KBR outliers. All 362 epochs occur on two consecutive days, October 16 and October 17. These epochs are not close to a large system shutdown like the LRI outliers. Investigation of the cause of these outliers is out of the scope of this project but would be interesting to examine in another project.

7. KBR noise visualization after outlier-removal and further analysis

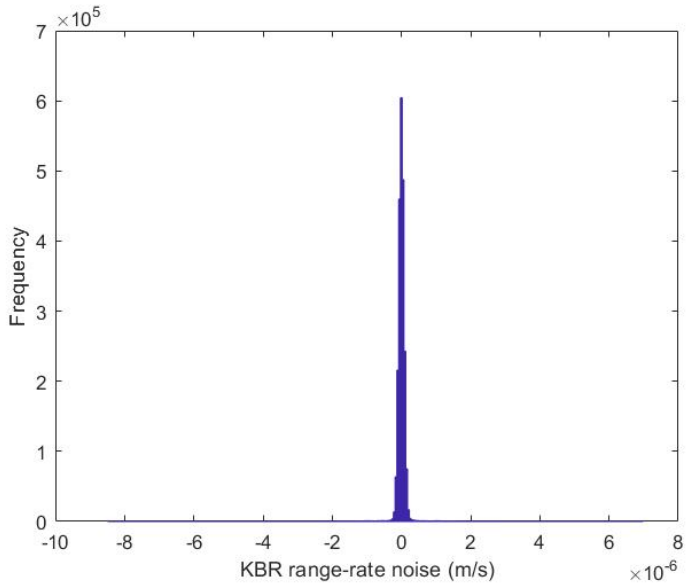


Figure 22: KBR range-rate noise histogram after LRI and KBR outlier removal (300 bins)

Figure 22 displays the KBR range-rate noise after both the LRI and KBR outliers are removed. 3764 epochs due to LRI outliers and 362 epochs due to KBR outliers are removed from the dataset. The KBR range-rate noise does not show a normal distribution that one would expect. In order to get that normal distribution the noise values should fall in the interval $<-3*10^{-7}$ m/s $3*10^{-7}$ m/s>. This is not the case. In fact there are around 59,000 points out of approximately 2,250,000 that still fall out of the expected range. The outlier detection method doesn't capture these points. Let us take a closer look at the epochs of these points.

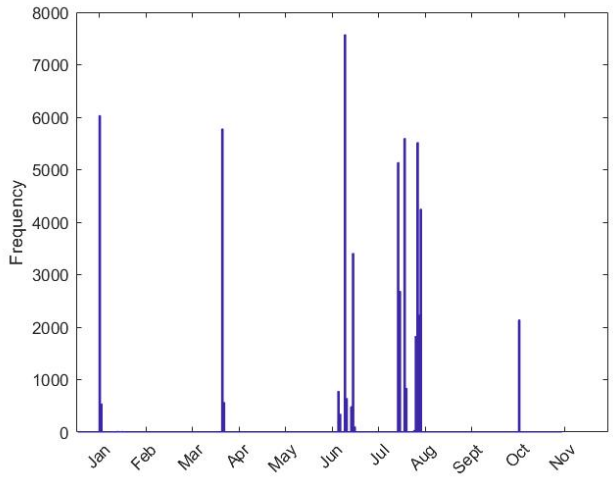


Figure 23: Histogram of epochs of the remaining KBR range rate noises that fall out of the interval $<-3*10^{-7}$ m/s, $3*10^{-7}$ m/s> (300 bins)

Figure 23 shows a histogram of the epochs of the 59,000 noise values that fall out of the expected noise interval. As can immediately be seen, these epochs are not randomly distributed but occur in clusters. Let's take a closer look at these clusters. Two epochs are defined to be in the same cluster if they are within half an hour time span, this is an arbitrary value.
Using this criteria, there are 11 major clusters that together form $\frac{56,878}{59,171}$*100=96,1% of the investigated noise values.

Table 3: Date and original GPS time of 11 clusters

Cluster	Date	Original GPS time
1	January 16*	600868800- 600955090
2	April 6*	607780800- 607867190
3	June 25 (6)	614692810- 614779180
4	June 30	615133800- 615160290
5	June 30	615162150- 615182980
6	June 30	615190530- 615211190
7	July 30*	617716800- 617803190
8	August 3	618062470- 618148790
9	August 10- August 11	618750000- 618839990
10	August 13*	618926400- 619012790
11	October 16-October 17	624505340- 624575420

Table 3 shows the date and original GPS times of the 11 clusters. A date with an asterisk means that the beginning of the cluster starts exactly at the beginning of the day, so at 0:00. What is interesting for these clusters is not only that they start exactly at 0:00, but also that the duration is more than 23 hours and 58 minutes, so the 4 clusters with an asterisk span almost a full day.

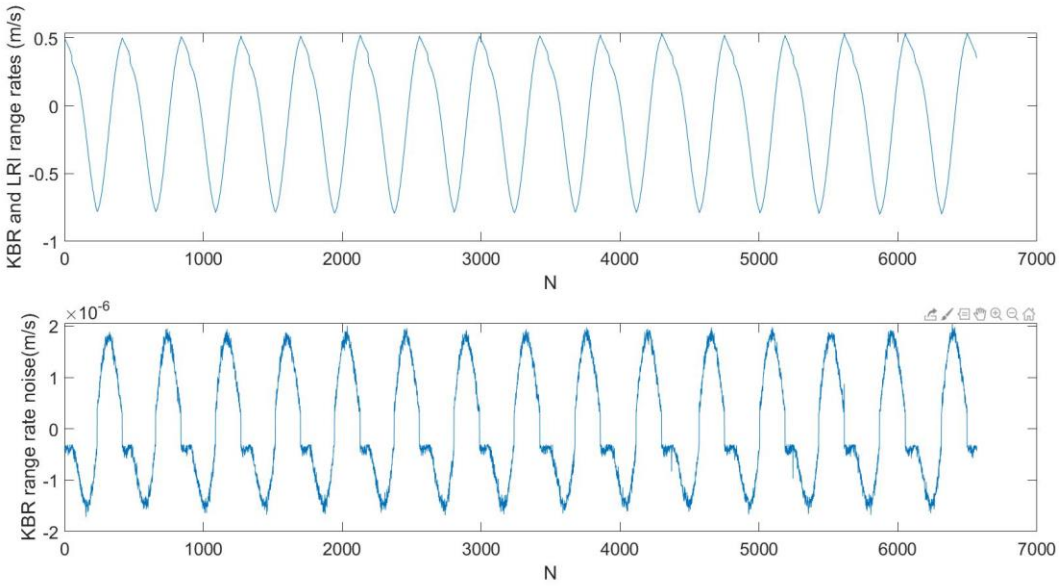


Figure 24: Subplots corresponding to cluster 1. Top: KBR and LRI range-rates, bottom: KBR range-rate noise

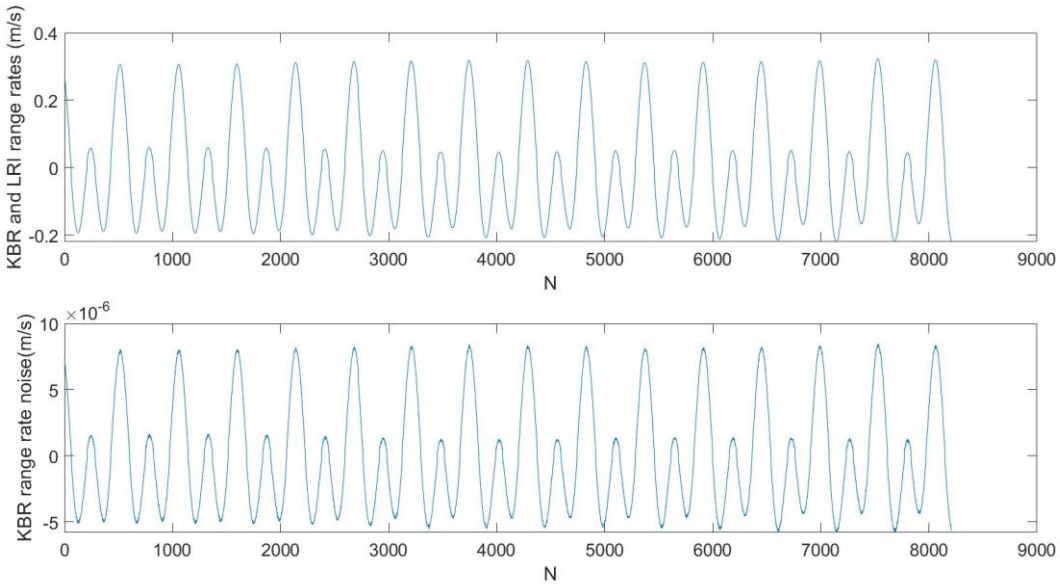


Figure 25: Subplots corresponding to cluster 3. Top: KBR and LRI-range rates, bottom: KBR range-rate noise

Figure 24 and figure 25 contain two subplots each. At the top the KBR and LRI range-rates, at the bottom the corresponding KBR range-rate noise. What can be seen is that both cluster 1 and cluster 3 show noise values that follow the same pattern as the KBR and LRI range-rate data. This is an indication that there is a structural error in these clusters that causes this strange noise pattern. The most plausible explanation is a clocking error, meaning that for these clusters the KBR and LRI measurements were not timed at the same moment but with a small offset. This would then result in shifted data, and when one takes the difference between two points that are labelled as the same epoch, the same pattern as in the range-rates will show up in the noise values. This phenomenon can be found in clusters 1 to 10.

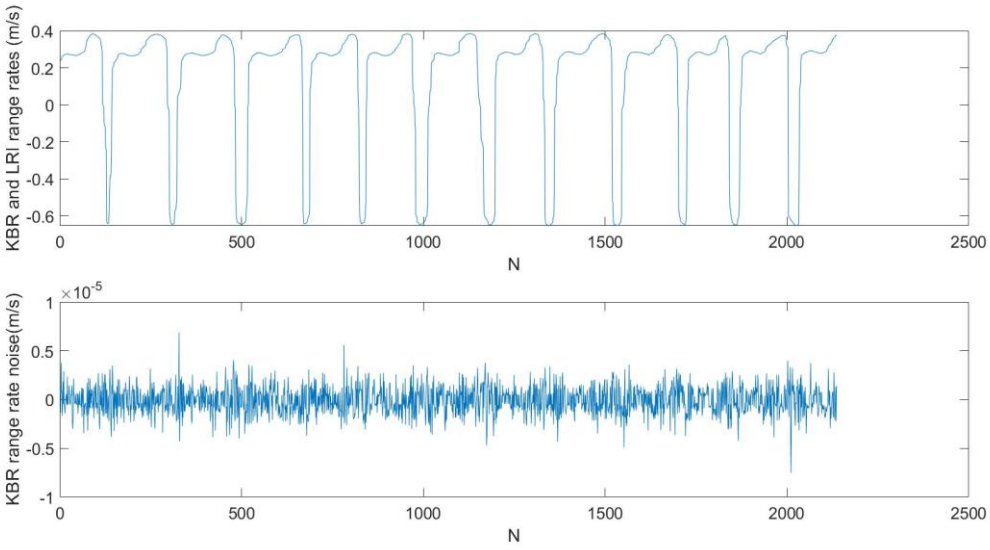


Figure 26: Subplots corresponding to cluster 11. Top: KBR and LRI rang- rates, bottom: KBR range-rate noise

Only cluster 11, displayed in figure 26, shows noise values that are more randomly distributed. These are only 2000 values however. The other 54,000 values show the same pattern in their noise as in the original range-rate data. In the script used to generate the figures above, the clusters are automatically found. What needs to be done manually however, is to determine whether the noise values of a particular cluster show the unexpected pattern as in figure 21 and 22. In this case cluster 1 to 10 are labelled to be incorrect and these epochs are therefore removed from the dataset.

In the KBR range-rate noise there are now only 6791 values that fall out of the <-3*10^{-7} m/s, 3*10^{-7} m/s> interval. A final and rather simple outlier removal method that is applied, is removing all the data that falls out of the 3σ range. The standard deviation σ is calculated with the formula

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|A_i - \mu|^2}$$

Where A is a random range-rate and μ is the mean of the range-rates. This gives a σ of 1.34 *10^{-7} m/s and a 3σ threshold value of 4.03*10^{-7} m/s. using this threshold will remove an additional 4972 noise values.

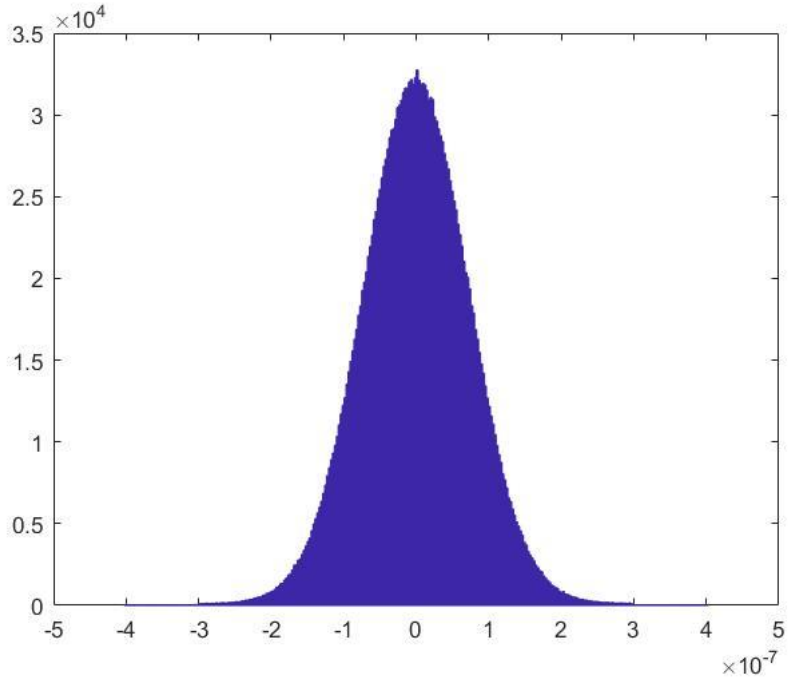


Figure 27: Histogram of KBR range-rate noise (300 bins)

A histogram displaying the final KBR range-rate noise is displayed in figure 27. It can be seen that there is a normal distribution of noise achieved which will be used as input for an estimation of the power spectral density.

8. Power spectral Density estimation of KBR range-rate noise

Now that there is a final KBR range-rate noise dataset, it is convenient to make a periodogram to get an estimation of the spectral density. There are different ways to do this, in this project the Welch's method has been used. This method divides a signal into different segments where each periodogram is calculated by the discrete Fourier transform. Then the periodograms are averaged to get one final periodogram. The input parameter that is of particular interest, is the length of the segments.

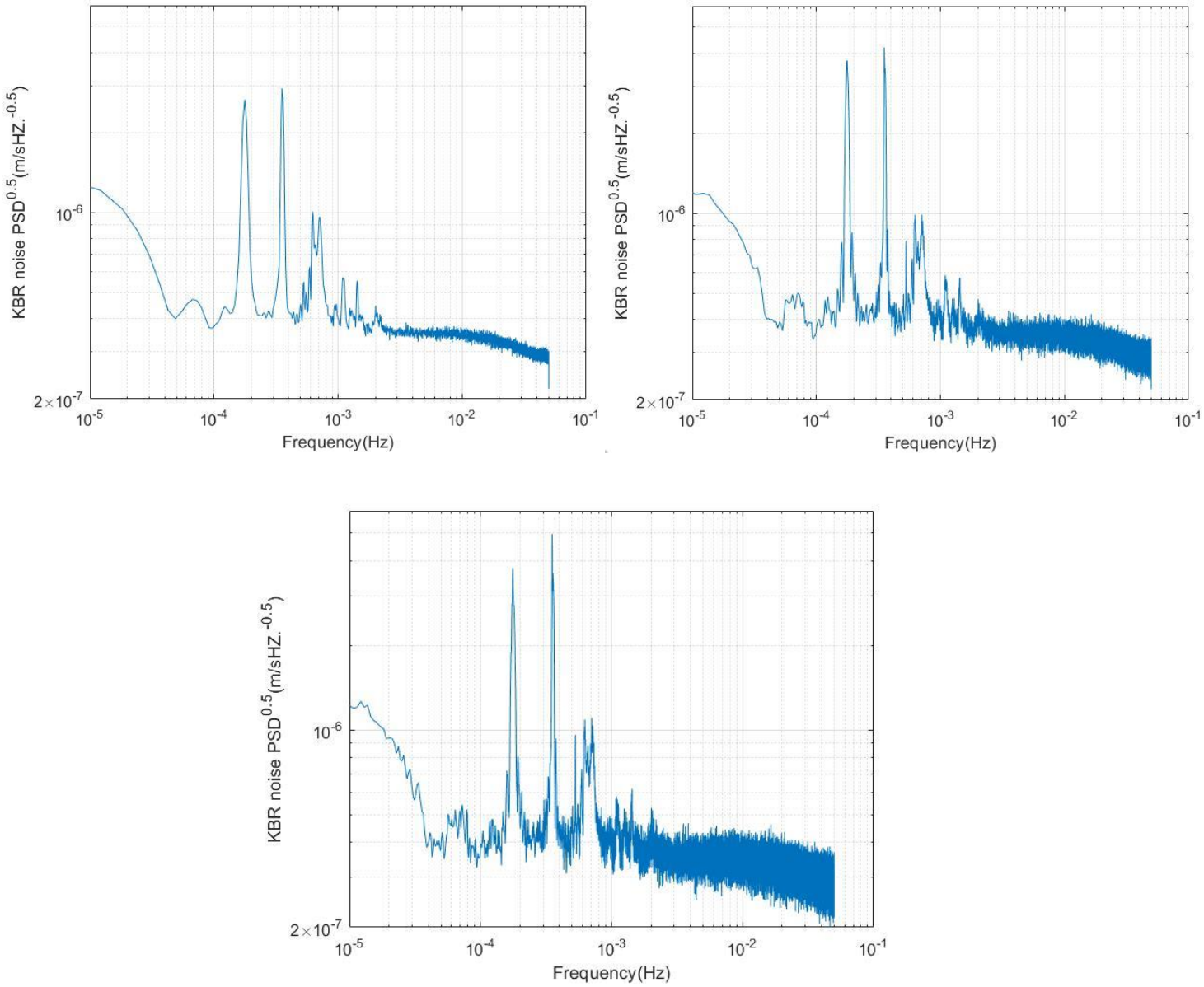


Figure 28-30: PSD of KBR range-rate noise dividing the signal into segments 10,000 samples in length(left top), 50,000 samples in length(right top) and 100,000 samples in length(bottom)

Figures 28 to 30 show the PSD of the KBR range-rate noise with three different segment lengths of 10,000, 50,000 and 100,000 respectively. It can be seen that when the segment length increases the plot becomes more detailed but the signal is less smooth and it becomes a bit harder to distinguish characteristic features. To find a suitable segment length value it is 'trial and error'. What can be seen in the figures is that there two main peaks around $2*10^{-4}$ Hz and $3.5*10^{-4}$ Hz. What is interesting to see is that a larger segment length results in higher peaks. For a segment length of 10,000 the PSD peak values are both of order $3*10^{-6}$ m s^{-1} $Hz^{-0.5}$ and with a segment length of 100,000 samples the PSD peak values are of order $4*10^{-6}$ m s^{-1} $Hz^{-0.5}$ and order $5*10^{-6}$ m s^{-1} $Hz^{-0.5}$ respectively. I don't have an explanation for these two distinctive peaks, it would be a recommendation to do an investigation on this in future research. After these two peaks there is one smaller peak at around $7*10^{-4}$ Hz with values tipping the order of 10^{-6} m s^{-1} $Hz^{-0.5}$. After the third peak the PSD values of the KBR range-rate noise decrease to around $3*10^{-7}$ m s^{-1} $Hz^{-0.5}$ at the higher frequency range of about $5*10^{-2}$ Hz.

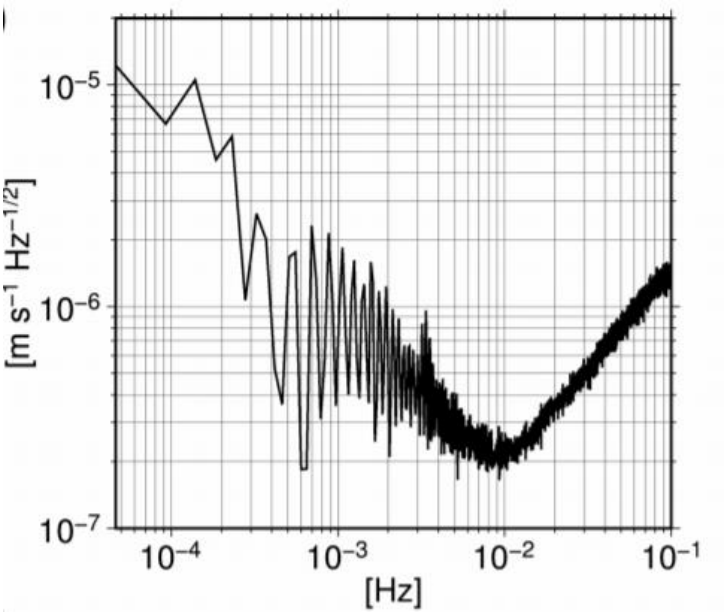


Figure 31: PSD of range-rate residuals of the GRACE mission of December 2008. (Behzadpour et al.,2019)

Figure 31 shows the PSD of the range-rate residuals of December 2008 of the GRACE mission. What can be seen is that the PSD values here start at a frequency of $5*10^{-5}$ Hz with values around 10^{-5} m s^{-1} $Hz^{-0.5}$ and gradually decrease to values of order $2*10^{-7}$ m s^{-1} $Hz^{-0.5}$ at around 10^{-2} Hz. Then the values gradually increase again to around 10^{-6} m s^{-1} $Hz^{-0.5}$ at 10^{-1} Hz. This plot shows a different pattern than the previous three plots. There are no sudden peaks but the PSD values are larger at most frequency ranges. It is interesting to compare the three PSD plots of the GRACE-FO (fig.28-30) with the one of the GRACE mission (fig.31). The lower frequency range is defined as the frequency range before the first peak of the GRACE-FO PSD plots ($5*10^{-5}$ Hz to $1.05*10^{-4}$ Hz), the mid- frequency range as the part containing the three peaks ($1.05*10^{-4}$ Hz to $8*10^{-4}$ Hz) and the higher frequency range as the frequencies that are followed after the three peaks ($8*10^{-4}$ Hz to $5*10^{-2}$ Hz). In the lower frequency range the PSD values of the KBR range-rate noise of the GRACE-FO are between 10^{-6} m s^{-1} $Hz^{-0.5}$ and $4*10^{-7}$ m s^{-1} $Hz^{-0.5}$ while the KBR range-rate noise for the GRACE mission is of order 10^{-5} m s^{-1} $Hz^{-0.5}$, which is almost two orders of magnitude higher than the ones found for the GRACE-FO. At the mid-frequency

range for the KBR range-rate noise with LRI it was shown that the magnitude of the peak depends on the segment length. For the comparison the second plot with segment length of 50,000 samples will be used. The first peak of the PSD of the KBR range-rate noise of the GRACE-FO is of order $4*10^{-6}$ m s^{-1} $Hz^{-0.5}$ and the PSD values of the KBR range-rate noise for the GRACE mission are around 10^{-5} m s^{-1} $Hz^{-0.5}$ s^{-1} $Hz^{-0.5}$. The second peak of the PSD of the KBR range-rate noise of the GRACE-FO is also of order $4*10^{-6}$ m s^{-1} $Hz^{-0.5}$ and the PSD values for the GRACE mission are around $2*10^{-6}$ m s^{-1} $Hz^{-0.5}$ .The third peak is around 10^{-6} m s^{-1} $Hz^{-0.5}$ and the GRACE mission PSD value is of order $7*10^{-7}$ m s^{-1} $Hz^{-0.5}$. So the first peak of the GRACE-FO is lower than the GRACE mission but the other two peaks are higher. In the higher frequency range the PSD values of the KBR range-rate noise of the GRACE-FO are going from order $4*10^{-7}$ m s^{-1} $Hz^{-0.5}$ to $3*10^{-7}$ m s^{-1} $Hz^{-0.5}$ and the PSD values for the KBR range-rate noise for the GRACE mission first drop to the order order $2*10^{-7}$ m s^{-1} $Hz^{-0.5}$and then the values gradually increase again to around 10^{-6} m s^{-1} $Hz^{-0.5}$. At this lowest point of $2*10^{-7}$ m s^{-1} $Hz^{-0.5}$ the PSD values of the KBR range-rate noise of the GRACE mission are a bit lower than the PSD values of the GRACE-FO but after this point the GRACE-mission PSD values rise again to around 10^{-6} m s^{-1} $Hz^{-0.5}$, while the GRACE-FO PSD values stay around $3*10^{-7}$ m s^{-1} $Hz^{-0.5}$.

Conclusion

The goal of this project was to do an investigation of noise in KBR ranging data of NASA's and GFZ's GRACE Follow-On mission. The two main objectives were to find a suitable outlier detection method and to compare the PSD values of the KBR range-rate noise with results found in literature.

I decided to investigate the KBR range-rate noise. Since LRI ranging data is two orders of magnitude more accurate compared to the classical K-Band ranging data, noise was defined as the difference between KBR and LRI range-rates.

Initially it was aimed to perform a KBR noise analysis by using the difference between KBR range-rates and the interpolated LRI range-rates. The aimed interpolation error of 10^{-10} m/s was not achieved. Therefore only the overlapping epochs were used as input for outlier analysis. It is a recommendation for future studies to investigate if it is still feasible to use the interpolated LRI range-rates for the KBR noise analysis.

In order to get a realization of KBR range-rate noise it is a prerequisite that the KBR and LRI range-rates are correct and that therefore outliers are removed. The method for outlier detection is based on comparing a particular range-rate with an estimation of the true range-rate. This estimation of the true range rate is based on a third degree polynomial fit where the polynomial coefficients are found by using the range-rates of the six neighboring epochs. With this method it is possible to label LRI epochs as outliers where the range-rate deviates by more than $2*10^{-8}$ m/s from the estimation of the true range-rate. For the KBR range-rates all the epochs where the range-rate deviates by more than $7.7*10^{-7}$ m/s were labeled as outliers. 3764 LRI and 362 KBR epochs out of 2,250,000 epochs in total were found to be outliers. 3683 out of 3764 LRI outliers are part of one day, March 19 and this is likely related to the fact that the system was rebooted on March 17. It is a recommendation to remove the LRI data of March 19 for future analysis. All 362 KBR outlier epochs occur on two consecutive days, October 16 and October 17. The reason behind this is not found in this project and it a recommendation to investigate this in future studies.

After the outlier detection a histogram was made of the remaining KBR and LRI range-rate differences. To get a normal distribution of the data, the noise interval had to be around $<-3*10^{-7}$ m/s, $3*10^{-7}$ m/s. This was not the case and around 59,000 points still fell out of this interval. The exact time intervals are given in table 3. The time intervals were investigated and it was found that 54,000 of these points formed clusters where subtraction of the KBR range-rate from the LRI range-rate did not give a realization of noise but a pattern that was similar to the original signal. The most plausible explanation for this is a clocking error, so KBR and LRI measurements were not timed at exactly the same moment. It is a recommendation to do further analysis on these intervals and investigate where these errors exactly come from.

Using the Welch method it was then possible to create three different PSD plots with different segment lengths (10,000 , 50,000 and 100,000 samples) of the KBR range-rate noise and this has been compared to a PSD plot of the KBR range-rate noise of the GRACE mission of December 2008. In this project there were two large peaks and one smaller peak between 10^{-4} Hz and 10^{-3} Hz. These peaks were not present in the PSD plot of the GRACE mission. Except from the second and the third peak, the PSD values of the

KBR range-rate noise of this project were lower, up to two orders of magnitude in the lower frequency range (around 10^{-4} Hz) and one order of magnitude in the higher frequency range (around 10^{-1} Hz). It is a recommendation to investigate the three peaks between 10^{-4} Hz and 10^{-3} Hz in future studies.

To finalize I think that the first objective to detect LRI and KBR range-rate outliers has been achieved. The method proved capable of detecting LRI range rates that deviate by more than $2*10^{-8}$ m/s from the estimation of the true range-rate. For the KBR range-rates all the epochs were the range-rate deviates by more than $7.7*10^{-7}$ m/s were removed from the dataset. This was not enough however to get a nice distribution of KBR range-rate noise, there turned out to be clusters were taking the difference between the KBR and LRI range-rates did not give a realization of noise. These clusters should be investigated in future studies.
The second objective to compare the PSD values of the KBR range-rate noise with results found in literature has also been achieved although it would have been better if there were more PSD plots from the literature to compare it with, especially since it turned out that the PSD plots show different patterns.

References

Abich K, Christina B, Claus B, Karsten D, Marina D, Martin G, Alexander G, Gerhard H,Mark H, Christoph M, Vitali M, Kolja N, Josep S, Daniel S, Benjamin S, Gunnar S, Kai V. (2015). *GRACE-Follow on Laser Ranging Interferometer: German contribution*. Journal of Physics: Conference Series. 610. 012010. 10.1088/1742-6596/610/1/012010.

Behzadpour S, Mayer-Gürr T, Flury J, Klinger B, Goswami S. (2019). *Multiresolution wavelet analysis applied to GRACE range-rate residuals*. Geoscientific Instrumentation, Methods and Data Systems. 8. 197-207. 10.5194/gi-8-197-2019.

Dahl C, Baatzsch A, Dehne M, Gilles F, Hager P, Herding M, Nicklaus K, Voss K, Abich K, Braxmaier C, Gohlke M, Guenther B, Sanjuan J, Zender B, Barranco G.F., Görth A, Mahrdt C, Müller V, Schütze D, Stede G, Heinzel G. (2017). *Laser ranging interferometer on Grace follow-on.* Proc. SPIE 10562, International Conference on Space Optics — ICSO 2016, 105623V

GRACE-FO. (2019) . *GRACE-FO Level-1B Release version 4.0 from JPL in ASCII*. Ver. 4. PO.DAAC, CA, USA. Dataset accessed [2019-12] at https://doi.org/10.5067/GFL1B-ASJ04

Kim J & Lee S.W. (2009). *Flight performance analysis of GRACE K-band ranging instrument with simulation data*. Acta Astronautica, Vol. 65, 2009, pp. 1571-1581

Kramer H.J. (n.d.). *GRACE-FO (Gravity Recovery And Climate Experiment - Follow-on)*. Retrieved from https://earth.esa.int/web/eoportal/satellite-missions/g/grace-fo#foot14%29

Renishaw plc. (n.d.). *Interferometry explained*. Retrieved from https://www.renishaw.com/en/interferometry-explained--7854

Webb F, Flechtner F, Landerer F, Watkins M, Save H, Dahle C. (2019*). GRACE Follow-On Science Data System Newsletter.* Report: Jan-Mar 2019 (No. 2)

Wen H.Y., Kruizinga G, Paik M, Landerer F, Bertiger W, Sakumura C, Bandikova T, Mcculough C. (2019). *Gravity Recovery and Climate Experiment Follow-On (GRACE-FO) Level-1 Data Product User Handbook.* NASA Jet Propulsion Laboratory California Institute of Technology.

n.d. (2017). *GRACE-FO: Tracking Earth's Mass in Motion*. Retrieved from https://www.jpl.nasa.gov/news/brochure/gracefo_brochure.pdf

Appendix

KBR1B dataset full description

Column	Description	unit
1	Continuous seconds past 01-Jan-2000 11:59:47 UTC	s
2	CRN-filtered biased inter-satellite range with ionospheric correction	m
3	First time derivative of biased-range	m/s
4	Second time derivative of biased-range	m/s^2
5	biased ionospheric correction for biased-range, for Ka frequency	
6	Light time correction for biased-range	m
7	Light time correction for range-rate	m/s
8	Light time correction for range-acceleration	m/s^2
9	Antenna phase center offset correction for biased-range	m
10	Antenna phase center offset correction for range-rate	m/s
11	Antenna phase center offset correction for range-accleration	m/s^2
12	SNR of K band for GRACE-FO C satellite	0.1 db-Hz
13	SNR of Ka band for GRACE-FO C satellite	0.1 db-Hz
14	SNR of K band for GRACE-FO D satellite	0.1 db-Hz
15	SNR of Ka band for GRACE-FO D satellite	0.1 db-Hz

LRI1B dataset full description

Column	Description	unit
1	Continuous seconds past 01-Jan-2000 11:59:47 UTC	s
2	CRN-filtered biased inter-satellite range with ionospheric correction	m
3	First time derivative of biased-range	m/s
4	Second time derivative of biased-range	m/s^2
5	estimated scale correction for biased range, range rate, and range accl due to unknown onboard LRI frequency (assumed nominal frequency is 281614.8 GHz)	
6	Light time correction for biased-range	m
7	Light time correction for range-rate	m/s
8	Light time correction for range-acceleration	m/s^2
9	Not defined	
10	Not defined	
11	Not defined	
12	CNR of laser ranging for GRACE-FO C satellite	db-Hz
13	Not defined	
14	CNR of laser ranging for GRACE-FO D satellite	db-Hz
15	Not defined	

GNV1B dataset full description

Column	Description	unit
1	Continuous seconds past 01-Jan-2000 11:59:47 UTC	s
2	GRACE-FO satellite identifier	
3	Coordinate reference frame	
4	Position, X value	m
5	Position, Y value	m
6	Position, Z value	m
7	Formal error of xpos	m
8	Formal error of ypos	m
9	Formal error of zpos	m
10	Velocity along x-axis	m/s
11	Velocity along y-axis	m/s
12	Velocity along z-axis	m/s
13	Formal error of xvel	m/s
14	Formal error of yvel	m/s
15	Formal error of zvel	m/s