CHALMERS

Evaluation of high rate real time GPS based tsunami warning system

Master of Science Thesis in Communication Engineering

JONAS SUNDSTRÖM

Supervisor: Jan Johansson

Department of Radio and Space Science

Space Geodesy and Geodynamics Research Group

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden, 2009

Evaluation of high rate real time GPS based tsunami warning system

JONAS SUNDSTRÖM

Department of Radio and Space Science
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2009

Evaluation of high rate real time GPS based tsunami warning system
JONAS SUNDSTRÖM

Technical report
Department of Radio and Space Science
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Department of Radio and Space Science, Göteborg, Sweden 2009

Evaluation of high rate real time GPS based tsunami warning system

JONAS SUNDSTRÖM
Department of Radio and Space Science
Chalmers University of Technology

Abstract

Jet Propulsion Laboratory (JPL) is transmitting real time stream of GPS measurement over Internet via a UDP socket. It is a high rate stream that sends 1 Hz data from several stations, today the total number is 70 – 80 stations.
This thesis is focusing on finding problems not only with the stream itself, also current problems with the actual measurement. There will be some improvement described however no focus is in network setup or geodetic detection of an actual earthquake.
The delay of the stream is gradually increasing from 5 to 15 seconds and drops down to 5 seconds again. This delay is end to end time and have a rate of about one to one and a half hour. Over this time period JPL receives data, process it and resend it to end user. At the receiver end not all data arrives. There are 10 missing measurement in a row at the time the delay drops. The statement is that anywhere in the processing, either at JPL or at receiving computer there is a overflow in a buffer which explains the drop.
There are other drops of data, which can be from seconds to hours in length. Whenever the dropped station starts to transmit again, sigma tends to be very large, which is a good thing. However it returns very quickly to normal, much faster than the actual GPS position. A sigma can take in the order of minutes to return while positions takes hours. This problem can be misleading and can create further problem for filters or other processing jobs where sigma is used. Different stations seem to be affected by the missing measure differently. This can explain the variation in result of displacement error.

The measurement is subject to a centimeter to decimeter displacement error for a ten-minute gap of data. The displacement error is larger when ever a loss of data occurs. During the next coming hours a large variation of the measurement occurs and a displacement error will be much greater. The solution for this is either to minimize any drop of data which can be a matter of investing in better connection for the receiver stations around the world to JPL or look into the models JPL have and give a better sigma estimate.
There are a significant autocorrelation in the signal, the peak can be up to 80% of top value and is close to a sidereal day. Because of this continues correlation a filter can be used to improve the time series. A five-day median filter gives an improvement up to 60%. This number will increase whenever the stated problems above are resolved.

Keywords: tsunami warning system, GPS, real time, high rate GPS

Acknowledgements

I would first thank my fiancé Medina, who let me take off for five month and do my project in Reno. She has been a very good support whenever a problem has occurred and I was very happy to take 2 weeks of vacation after finishing up the project.

I also would like to thank Geoffrey Blewitt, not only for being a great supervisor but also for taking care of all the administrative problems that has occurred during my stay in Reno. Both he and his wife Debra have given me a great welcome and have inspired me to continue with a PHD studdy.

Special thanks to Jan Johansson for supervising me back from Sweden and getting me started in the field of GPS and geodesy.

I would also thank all the other people I have had a chance to hang out with in Reno: Kurt, Amber, Toby, Stuart, Harold and a hole bunch of others…

1 Table of Contents

2 Introduction

Jet Propulsion Laboratory, from now on called JPL have set up or collaborates to receive measurement from several stations around the world. The result is a real time GPS measurement stream that gives a new measurement every second for all selected stations. The Cartesian coordinates are sent over Internet using a UDP (User Datagram Protocol) socket and presented in a single text line. The purpose of this stream is to use the measurement to detect movement and displacement which can be used in a model for faults close to the sea and in case of a large enough fault, warn for a tsunami. In order to not get false alarm a combination of stations close to each other will be used. For networks in that order several thousand stations is needed around the world. This thesis is not covering the network structure or models for detecting faults or earthquakes. The emphasis is to investigate the status of the stream and the quality of the measurement and give a hint on the accuracy the measurement have.

For this project a large number of data is needed, this is saved in hourly files and is converted to local coordinates, East, North and Vertical for better categorization of any GPS related problems. In some test the raw stream will still be used to test the quality of the stream.

3 Background

In this chapter the background of the project will be discussed. An introduction to Global Positioning System and tsunamis will be mentioned.

3.1 Global Positioning System

The basic idea is to calculate the range to the satellites in orbit around the earth and make a 3D triangulation. In order to do this calculation the receiver has to have some knowledge about the satellites. In order to triangulate, the distance is needed. This is calculated according to equation 1.

$$R = c \cdot (\tau - t) \tag{1}$$

Where R is the distance from satellite to receiver, c is the speed of light, τ is the receiver time and t is the satellite time. To get the time difference correct, both clocks have to be synchronized and no variation in clock speed present. In reality the clocks are not perfectly synchronized and an error is introduced.

$$\hat{R} = R + \Delta R = R + c \cdot \delta \tag{2}$$

R_hat is called the pseudorange, δ is the time bias from the clocks. The receiver clock in a hand held device is very poor and in that case the satellite error can be neglected. The satellites uses atomic clocks, hence the error is very small. In order to synchronize the signal in the receiver, a known code is integrated in the signal. The code will be repeated every millisecond. The code is unique for every satellite and the receiver can easily figure out which satellite that is sending. The message contains the current position of the satellite in X, Y and Z coordinates. Using this coordinates the position on the receiver can be calculated using.

$$P_n = \sqrt{(x_n - x)^2 + (y_n - y)^2 + (z_n - z)^2} + c \cdot \delta \tag{3}$$

Where x_n, y_n and z_n are the coordinates for a certain satellite and x, y and z is the receiver position. In this equation there are four unknown, hence four satellites is needed to calculate the current position. The equation is nonlinear and in order to solve a Tailor expansion is used. A Kalman filter is used to solve the linear expansion in a least square sense [1]. Whenever more than four satellites are used a solution with better accuracy is obtained.

All satellites orbiting the earth on a distance of 20,200 km, the orbits are placed with an angle difference of 60 degrees and an inclination angle of 55 from the equator. There are six orbital planes and for every plane at least 4 satellites are orbiting, at this point all planes have five satellites with at least four up and running. This gives 24 to 35 satellites totally. Because of the symmetry, at least six satellites are in line of sight for a receiver. The orbital period for a satellite is 11 hours and 58 minutes [1].

In order to obtain the best knowledge of satellite positions there are several ground stations monitoring and recalculating the positions of the satellites, the ground stations are transmitting updated positions to the satellites. The communication is done on a separate frequency from the normal use frequency. The satellites use two frequencies for sending data, L1 at 1,575.42 MHz and L2 at 1,227.60 MHz. In order to differentiate the different satellites a code modulation is used. There are two different codes, C/A code and P code. The C/A code is an open code that repeats itself every 3 ms, and consists of 1023 binary digits. The P code is closed code for military use only. The P code repeats itself every 266 days, the P code is encrypted with an unknown W code. The resulting code is called Y code. The C/A code is transmitted on L1 only and the Y code is transmitted on both L1 and L2.

Instead of using P and Y code to determine the distance to the satellite an estimation of how many cycles the wave have propagated. These unknown cycles are called phase ambiguity and will be converted to the distance using.

$$R = N \cdot \lambda + \varphi \cdot \lambda \tag{4}$$

Where N is the number of cycles, λ is the wavelength and ϕ is the fraction of the last wavelength. For any receiver, there is no problem maintaining and calculating change in number of cycles. The big problem is to know the initial number of cycles, fortunately for the rest of the tracking time. This number is the same and can be corrected to better and better guess. Whenever the signal is blocked by something, a house or a tree the receiver will miss the information and does not know how many cycles that have passed during the break. This effect is called cycle slip and can be avoided by combining several signals [1].

There are several error sources that make the uncertainty larger. These sources will be discussed below.

3.1.1 External error sources

In the distance between the satellite and the receiver the signal has to travel thru the ionosphere and troposphere. The signal will be delayed and phase shifted during this travel. The earth is not exactly elliptic and the earth model is not exact. In this chapter all these effects will be discussed.

3.1.1.1 Ionosphere

The ionosphere is the top layer of the earth atmosphere. The layer consists of free electrons and electric charge molecules. The charged molecules change the speed of the phase depending on the diffraction of air according to the equation below.

$$n = 1 + \frac{c_2}{f^2} + \frac{c_3}{f^3} + \ldots \tag{5}$$

This means that the measurement is delayed and the phase are in advanced. The delay and advance are different for L1 and L2 that can be used to remove the delay. This is done using equation (6)

$$L_3 \equiv 2.5464 \cdot L_1 - 1.5464 \cdot L_2 \tag{6}$$

Which removes all Ionospheric effect, however the noise and other error are amplified using this method [3].

3.1.1.2 Troposphere

The troposphere is the lower part of the atmosphere that contains of dry gases and water vapor. The water vapor is placed close to the earth crust and is not uniformly distributed. The distribution can change quickly and this is hard to estimate. The water vapor only represents 1/10 of the total delay [2]. The effect of all molecules is a delay of the signal and it affects the same way for everything below 15 GHz. The distance that the signal is propagating in the troposphere is depending on the angle between incoming signal and earth. In order to take this into account a mapping function is used. This function is mapping different angles into a zenith angle that is the shortest distance that the signal can travel thru the troposphere.

$$\Delta L = m_h(\alpha) \cdot ZWD + m_h(\alpha) \cdot ZHD \tag{7}$$

Where ZWD is the zenith wet delay, ZHD is the zenith hydrostatic delay, m_h is the mapping function and α is the angle of the signal.

3.1.1.3 Clock error

A very critical part of the GPS system is the time, because the signal propagates in the speed of light, the clocks have to be very accurate. In every satellite several atom clocks is used to calculate current time. Ground stations update this every hour in order to not have different drifts. The three different clocks used are Hydrogen masers, Cesium and Rubidium atomic clocks. The Hydrogen masers are very short term stable however they can have a constant drift over a longer period of time. The Cesium clock is not that short terms stable however over a longer period of time it is more stable. The Rubidium clock is used in older satellites and is phased out.

3.1.1.4 Multipath and shadowing

An antenna is exposed to not only the true signal, it is also exposed to reflections and diffracted replicas of the signal. Because the reflected and diffracted signals do not come from the direct path a delay occurs. The reflections and diffractions come from all objects in the surrounding area however the most significant is usually the closer objects due to spreading.

Another problem can occur when there is not a direct path between antenna and satellite, this can be a house or a tree that blocks the direct path, this phenomenon is called shadowing and combined with multipath the receiver can have problem finding the correct time interval, hence a miscalculation of one satellite will occur [7].

For a GPS station the satellite geometry will look the same every 23 hours 56 minutes and 4 seconds. This time period is called a sidereal day and come from the rotation of earth around its own axes. For a static GPS station on a remote location the surrounding environment will look the same and the multipath will be highly correlated over one sidereal day. According to [8] the cycle for a 1 Hz GPS measurement is approximately 9 seconds less than the mean sidereal day, 23 hours 55 minutes and 55 seconds.

3.1.2 Differential GPS

The concept of DGPS is to use two or more than receivers with at least one known position. The measurement of the known receiver is used to calculate the offset and subtract from the unknown position of the other receivers. In order to make this work, several satellites have to be shared between the receivers. To use this method around a larger area several reference stations can be used. Because the reference stations is far away the offset is not very accurate. One way to solve this is to use a virtual reference station that is calculated from the real reference stations. [3]

3.2 Tsunami

Whenever the sea surface makes an abrupt movement from an earthquake it pushes water upwards and creating circular waves. These waves can travel thousands of kilometers in the sea without losing much energy. The waves have a very long wavelength over deep water that makes the amplitude very small, about 1 meter. Whenever the wave hit shallow water, the wavelength become shorter and the amplitude rises. This makes the waves very hard to spot over deep water. The strength of the earthquake has to be over M 7 on Richter scale in order to create a tsunami. The estimation of greater magnitude for earthquake is hard, for the Sumatra earthquake (M 9.2) it was estimated to M 8 within 11 min, M 8.5 within 5 hours and several days later M 9.2. This miscalculation of magnitude can cause serious problem when calculating the impact of the earthquake.

3.3 Real time data

A large number of stations around the world are included in a network for reference receivers. The network of receivers uses GDGPS (Global Differential GPS) to process data, this gives the uncertainties in decimeter range. The frequency of measurement is 1 Hz, which gives a good time series. According to [10] the latency of the system is approximately 5 seconds from receiver to remodeled measurement. The transmission to client can be over Internet, VPN or satellite broadcast. For this project a UDP (User Datagram Protocol) stream over Internet is used. The reason UDP is used instead of TCP is the speed, as if there is a package loss for TCP, a retransmission is needed and the rest of the measurements will be further delayed in time.

3.4 Tsunami warning system

The project is aiming at contributing parts of a tsunami warning system that is multi agency cooperation. The system is going to be fully automatic and partly using GPS data in order to find slips and faults from earthquakes that potentially can create a tsunami. A model is built for known faults and movement is mapped with a magnitude earthquake that will be more accurate than real time seismic measurement.

4 Method

4.1 Xyz2files

The incoming x,y and z was converted to local coordinates and centered round a reference point. In order to convert the incoming x,y and z a conversion to polar angle coordinates where done using a GIPSY command xyz2gd, the angles where used in a transformation matrix that is calculated the first time a station occurs in the stream and saved into a file.

$$\begin{pmatrix} \Delta N \\ \Delta E \\ \Delta V \end{pmatrix} = \begin{pmatrix} -\sin(\varphi)\cdot\cos(\lambda) & -\sin(\varphi)\cdot\sin(\lambda) & \cos(\varphi) \\ -\sin(\lambda) & \cos(\lambda) & 0 \\ \cos(\varphi)\cdot\cos(\lambda) & \cos(\varphi)\cdot\sin(\lambda) & \sin(\varphi) \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{8}$$

Where λ is the longitude angle and ϕ is the latitude angle [4]. The transformation matrix G is fixed for any given station and can be pre calculated. The coordinates are written to a time tagged file using hours from first of January 1970. The last coordinates are also written into a separate file in order to make other programs not have to read a long file for an updated position.

In order to convert the x,y and z sigma more information is needed. As equation 9 show the covariance xy xz and yz is needed. Because the stream does not provide that at the moment they are assumed to be zero until the stream is updated.

$$C_{Local} = GC_{xyz}G^T \tag{9}$$

Where C_{Local} is the sigma matrix for local coordinates and C_{xyz} is the sigma matrix for x,y and z. A direct matrix computation converts the x,y and z sigma to local coordinates if assume zero covariance.

$$C_{Local} = \begin{pmatrix} \sigma_n^2 & \sigma_{ne} & \sigma_{nh} \\ \sigma_{en} & \sigma_e^2 & \sigma_{eh} \\ \sigma_{nh} & \sigma_{eh} & \sigma_h^2 \end{pmatrix} \tag{10}$$

$$C_{xyz} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix} \tag{11}$$

4.2 Statistic of GPS data stream

There are two major parts of statistic in this report, incoming data stream statistics and data series statistics. They are totally different from each other and have to be treated differently. For data stream, problems can occur in sender- receiver-program and UDP over Internet. Any problem can be narrowed down to these parts. The data series is more complicated, there can be problems in missing measurement, error in measurement, models or processing.

4.2.1 Stream measurement

The incoming data is ones every second, hence the difference in time tag for two next coming data should be one. The simple test of subtract the tag from next incoming tag shows any missing measurements. For any negative number the delay of the data is greater than one second and for larger number than one there is a gap in the stream. In order to remove uncertainties about the computer and buffer problems, several computers where used to collect data. These computers "gneiss", "response" and "promise" use different streams however same stations are received.

The delay from antenna receiver to end user computer program was also looked at. This delay should be as small as possible however some delay is acceptable.

4.2.2 Data series measurement

The quality and performance of the data series are to be investigated. This is done by a simple test of permanent displacement. An earthquake and the seismic wave's lasts for several minutes, during this time a lot of random movement will occur. In order to calculate the displacement, a mean value of the position before and after a hypothetical earthquake is done. The number of used values should be as few as possible for not delaying the test from real-time, however too small amount of measurement can create less accurate mean. For this test two minutes of data was used. The value of the gap, 10 minutes was based on the M 9.2 Sumatra earthquake (2004) [5].

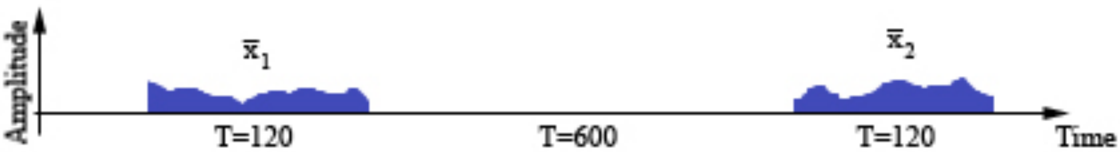

Figure 1: Time line for estimationg co-seismic displacement

For a non earthquake measurement, the displacement should be zero, in all series error and noise makes the test non zero. For a mean value the errors tend to cancel out and is the test is calculated according to.

$$dx = \bar{x}_2 - \bar{x}_1 \tag{12}$$

For a real measurement with noise, *dx* is a random variable and in order to understand the variation a cumulative distribution function can be calculated.

$$F_X(x) = \Pr(X \le x) \tag{13}$$

where X is a random variable and x is the distance from the mean value. This distribution can be interpreted as percent of the time that displacement accuracy is less than a certain distance [6].

4.3 Sidereal effects

The effect of a sidereal day is a slow variation of measurement based on time and space, in this case the position is fixed and the surrounding is almost the same for all time that gives only the variation in time. This time according to [8] is 23 hours 55 minutes and 55 seconds, which have to be confirmed. The only measurement available for the stations is the position and time the best way of measuring sidereal effect is to do an autocorrelation of the time series. The autocorrelation function will give a good estimate of how much effect the sidereal time have.

4.3.1 Sidereal filter

In order to build a filter to remove the sidereal effect the past days have to be used. If five days of data have high correlation for a sidereal day the next coming day will probably look similar. As all stations around the globe may not have the same sidereal day time the filter has to be updated every day in order to stay in good shape of incoming measurement.

The foundation of the filter is based upon the last five days measurement. However if there is a sigma problem the measurement will be taken from one day before. For this solution there will not be any spikes in the filter. Out of five days measurement the median will be used to create the filter value. The median gives a better value than the mean because the mean will be affected easier by even one large measurement.

$$M_n = M - F \tag{14}$$

Where M_n is the new measurement, M is the received measurement and F is the sidereal filter. For improvement purposes the same test performed for the data series will be performed on the filtered measurement.

4.4 Two hour high pass filter

There are almost always small earthquakes just before a large one. They can appear several hours before and may move the position of the station. In order to remove any old displacement a high pass filter can be implemented. In this particular case a weighted mean value is calculated from coordinates and sigma according to equation 15.

$$X_f = X - \hat{X} \tag{15}$$

Where X_f is the filtered position, X is the received position and $\hat{X}$ is the weighted mean value calculated:

$$\hat{X} = \sum_{i=1}^{N} X_i \cdot w_i \tag{16}$$

Where N is the total number of measurement for the filter and ω is the weight calculated from equation X using sigma.

$$w_i = \frac{\frac{1}{\sigma^3}}{\sum_{j=1}^{N} \frac{1}{\sigma^3}} \tag{17}$$

Where sigma is weighted more than least square because it is not completely random.

4.5 Presenting the data

In order to visualize the data in real time a plotting subroutine [11] need to be used. Because it is in real time no more than last 30 min is needed. For the presentation, the three coordinates are plotted with current time (local and UTC). The data used is unfiltered, sidereal filter, two-hour high pass filter, and a combination of sidereal and two-hour high pass filter.

In order to get an overview of all stations movement, another overview plot will be used. The plotting subroutine [11] have a world map that will be used and stations will be plotted at correct position with the current movement. Stations that are not sending any data will be viewed, as another color and direction and movement will be plotted with vectors.

5 Results

The first thing to do when using a data stream is to look at the quality and properties. This is done using programs written in C and FORTRAN 95 and plotted using Matlab.

5.1 Statistics of data stream

Almost all data was delivered however there were some problems for almost all stations. The problem with missing measurements is consistent for every station. As seen in figure (2) this happens several times during the period of 8 days.

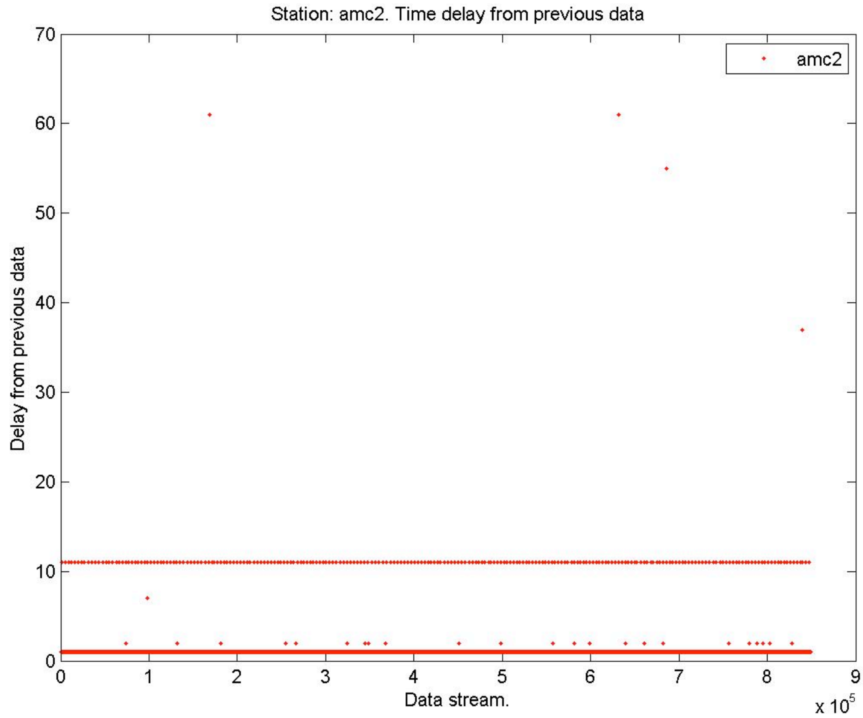

Figure 2: Missing tags for station AMC2. Notice the ~10 missing measurements consistent over 8 days

There is some random missing measurement with different amplitude usually 1, 2 or larger around 40 seconds, up to a minute. There are consistently ~10 missing measurements during this time.

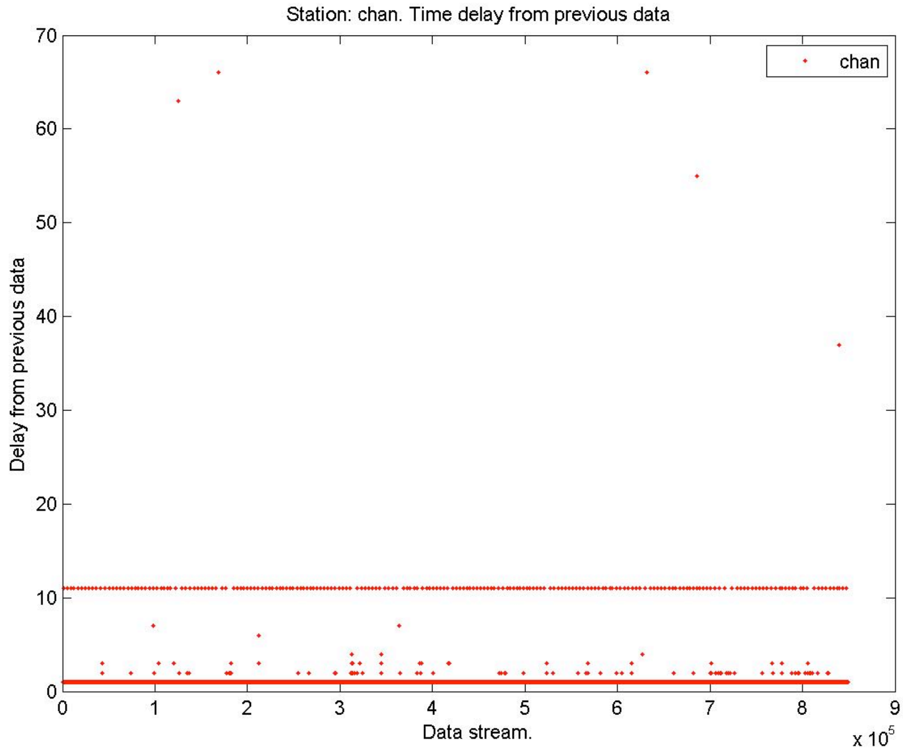

Figure 3: Missing measurement for station CHAN, notice the missing ~10 measurement.

As seen in figure (3) the different stations have different missing measurements. When looking into the missing measurements the repeat of ~10 missing measurement almost every hour seems to be consistent for almost all stations, Further investigation of one hour gives a good look at when the loss of measurement is happening. As seen in figure (4) there is no correlation between the different streams or station to station.

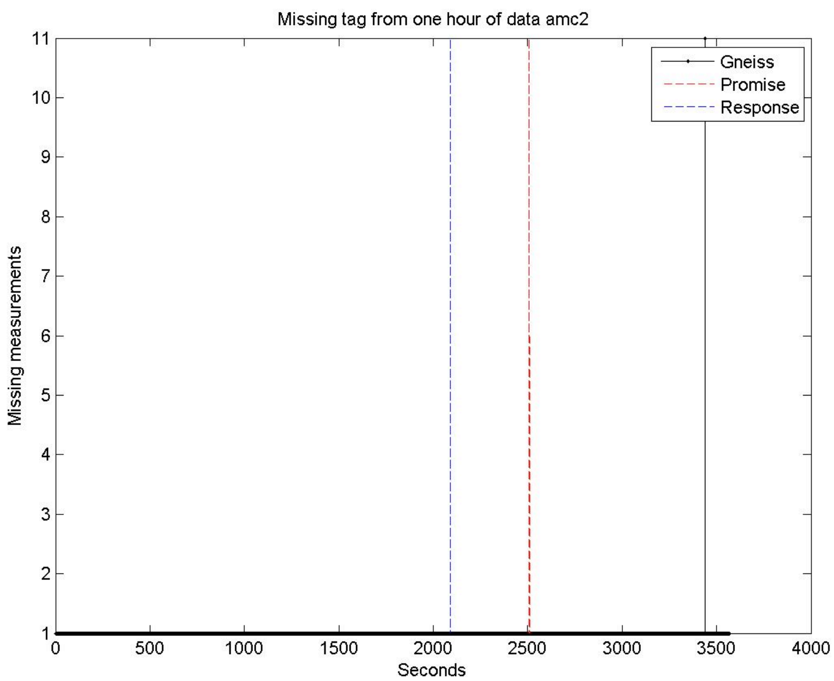

Figure 4: Further investigation for station AMC2, notice the time when ~10 missing measurement occur

The other problem that seems to occur ones in a while is that a station might not get any new measurement for an unknown period of time. This effect can be applied for seconds up to several hours or days. Figure (2) shows the different kind of lost measurements that can occur to a station. It seems that this happens to all stations in different ways and time periods. Some stations are more exposed to this than others. All streams have the same missing tags which indicate that this is a different problem than the ~10 missing measurements.

The delay of the measurement seems to be varying from six to fifteen seconds, and increases linearly in periods. This can be viewed in figure (5) and the time period is about one hour.

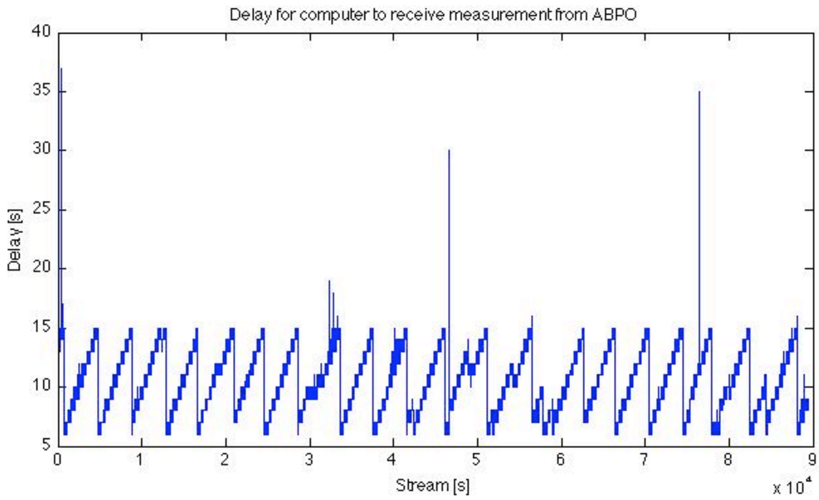

Figure 5: Delay plot for station ABPO.

If missing 10 measurements and the delay are shown in the same figure the correlation between missing measurement and drop in delay is very clear. It seems to happen at the same time as the delay drops 10. In the plot delay is moved down to zero for better view together with the delay.

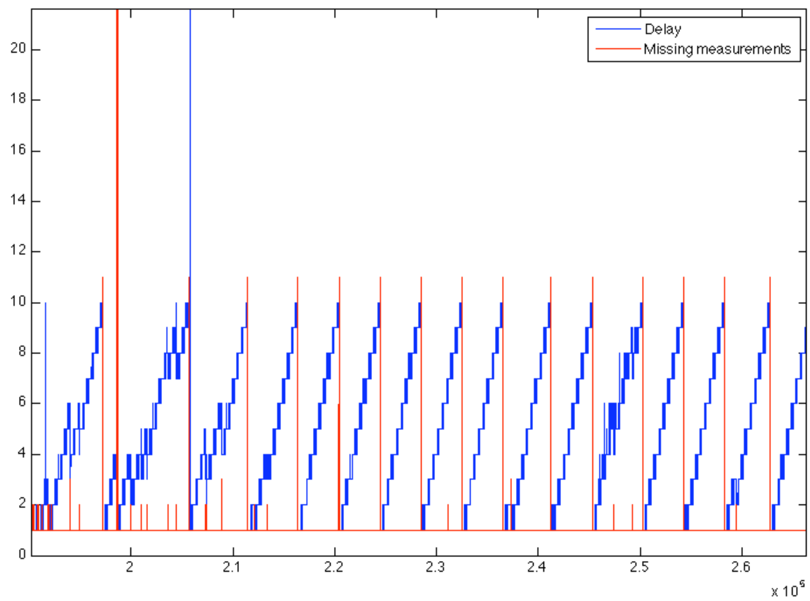

Figure 6: Delay and missing measurement plot for station ABPO

5.2 Statistics of measurement

Every station seems to behave differently and need to be looked at independent. When looking at the earth coordinates some of the stations have very large sigma often. This large sigma (over 0.1 m) has to be treated as non-received measurement because it cannot be trusted. As seen in figure (7) and (8) and in table (1) the number of times sigma is large is very different.

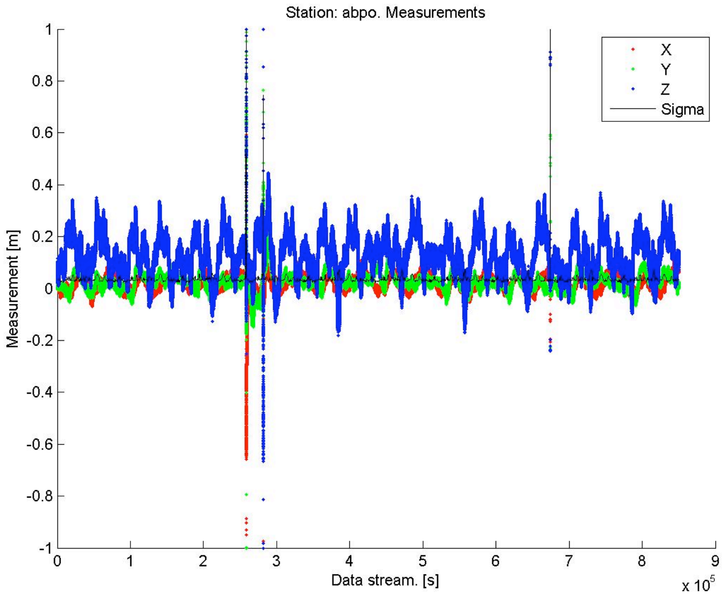

Figure 7: Signal plot for station ABPO

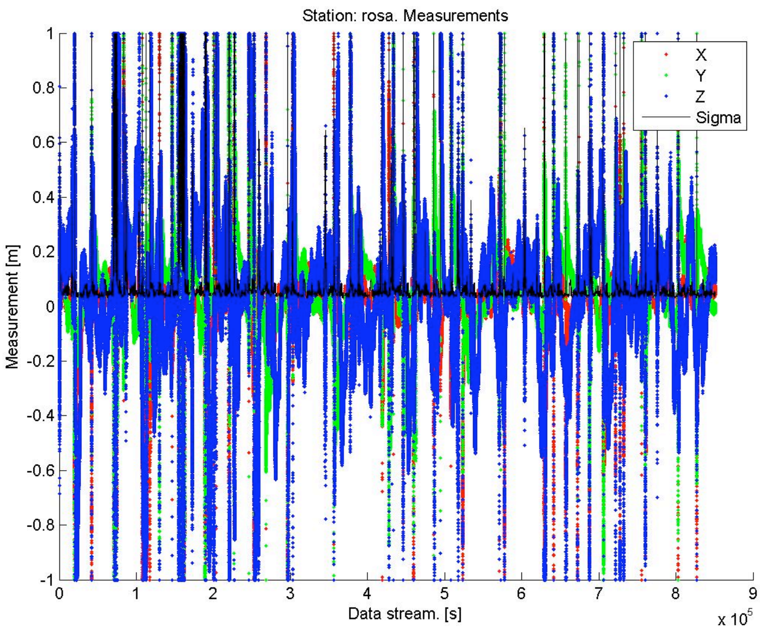

Figure 8: Signal plot for station ROSA

Station	#Large Sigma >0.3 m	# Measurements	Percentage
gold	1778859	1778859	100.000
toky	26087	26087	100.000
suth	791743	1009907	78.398
atjn	351459	1556721	22.577
kely	40850	882097	4.630
nisa	53580	1769150	3.030
falk	48060	1694754	2.840
brew	43350	1672256	2.590
xian	42414	1760103	2.410
buru	30885	1778232	1.740
areq	23666	1645811	1.440
rosa	25128	1780303	1.410
gode	19153	1778528	1.080
pol2	7634	1168132	0.654
kwj2	9203	1779536	0.517
chan	8736	1764752	0.495
usud	2702	567911	0.476
raro	7181	1780363	0.403
port	6826	1702206	0.401
okc2	6761	1779529	0.380
kgn1	6089	1779419	0.342
sydn	3966	1193101	0.332
cata	5825	1775272	0.328
inve	5060	1774856	0.285
kokb	4565	1779491	0.257
pie1	4421	1779735	0.248
bris	4381	1773021	0.247
farg	4266	1778880	0.240
joha	3766	1778255	0.212
wesl	3770	1781009	0.212
mkea	3750	1779352	0.211
cro1	3424	1778122	0.193
perm	3196	1778701	0.180
zwei	3018	1781066	0.169
guaz	2364	1777159	0.133
jplz	2326	1779160	0.131
perc	2291	1771976	0.129
tidb	1634	1778785	0.092
madr	1369	1756006	0.078
ispa	511	732245	0.070
mbar	984	1720665	0.057
fair	752	1777728	0.042
ykro	454	1183051	0.038
nssp	476	1741839	0.027
bogt	284	1502757	0.019
pimo	200	1164213	0.017
hraz	171	1133896	0.015
jplt	196	1415348	0.014
iisc	150	1760453	0.009

gala	140	1772975	0.008
quin	84	1184873	0.007
natl	96	1422012	0.007
abpo	94	1774408	0.005
usn3	89	1736616	0.005
vzpr	72	1780874	0.004
sant	67	1747841	0.004
mcm4	58	1778200	0.003
dlf3	48	1778265	0.003
amc2	33	1777906	0.002
dgar	26	1778539	0.002
palm	24	1779957	0.001

Table 1: List of large sigma for all stations

The displacement is affected by the large sigma measurement very hard and the stations performing badly have higher rate of large sigma. As seen in figure (9) the displacement plots for a non-earthquake situation is fluctuating with a small variation.

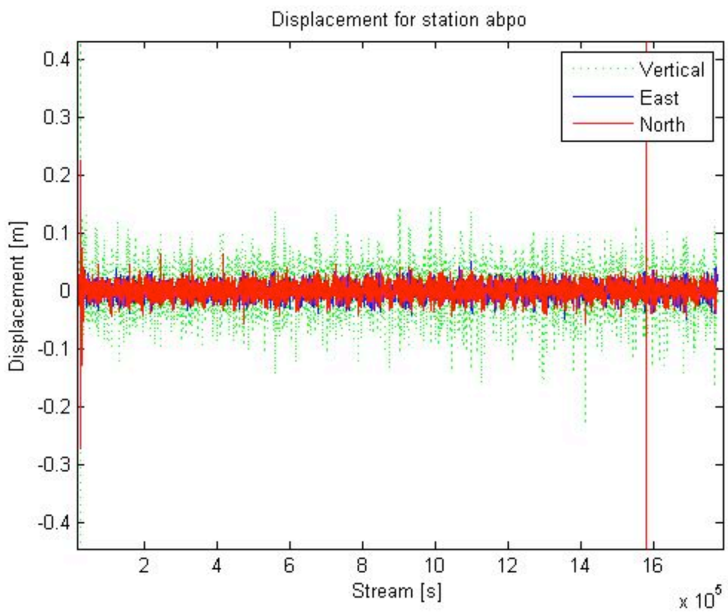

Figure 9: Displacement plot for station ABPO, notice the large sigma in the beginning and end which makes the displacement larger

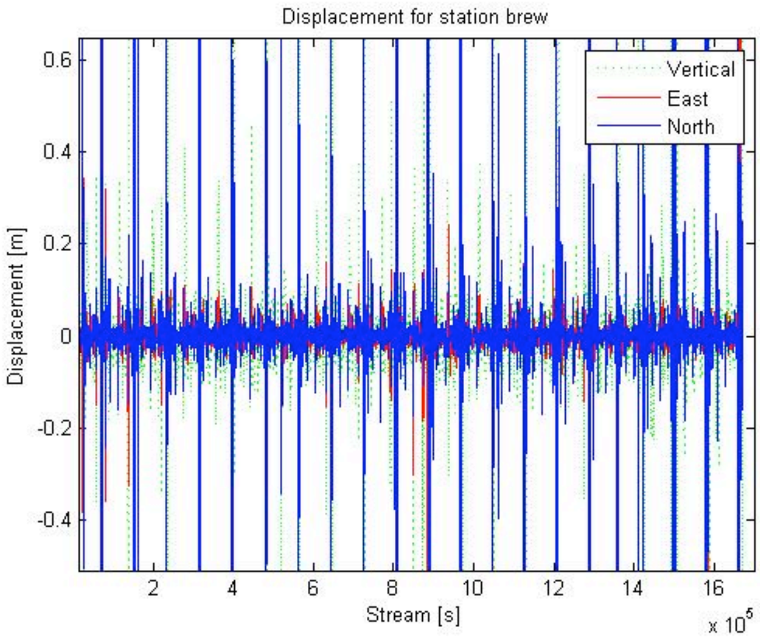

Figure 10: Displacement plot for station BREW, notice the large spikes from large sigma

When looking at a displacement an important thing to know is how often the displacement is under a certain threshold. This is a measurement of error and noise level for a station and as seen in table (2) the difference between stations is large. The best performing station have only 3 cm horizontal displacement error in 99% of the time while the poor performing station have over a meter in error.

Horizontal threshold [meter]				Vertical threshold[meter]			
Station	**50%**	**99%**	**99.9%**	**Station**	**50%**	**99%**	**99.9%**
mcm4	0.0075	**0.0330**	0.0516	mcm4	0.0210	0.0871	0.1073
dlf3	0.0080	**0.0434**	0.0882	dlf3	0.0119	0.0752	0.1168
nisa	0.0088	**0.0453**	0.0656	nisa	0.0169	0.0889	0.1433
gala	0.0087	**0.0491**	0.0759	gala	0.0168	0.1166	0.2482
usn3	0.0107	**0.0497**	0.0749	usn3	0.0186	0.0911	0.1205
abpo	0.0094	**0.0530**	0.2977	abpo	0.0212	0.0942	0.2560
palm	0.0072	**0.0554**	0.0746	palm	0.0155	0.0989	0.1396
dgar	0.0094	**0.0582**	0.1941	dgar	0.0204	0.1132	0.1919
gold	0.0001	**0.0587**	0.5478	gold	0.0001	0.0797	0.5567
amc2	0.0108	**0.0601**	0.0998	amc2	0.0192	0.1169	0.2207
natl	0.0115	**0.0661**	0.1998	natl	0.0247	0.1324	0.2336

quin	0.0083	**0.0691**	0.5099	quin	0.0163	0.0959	0.3265
madr	0.0127	**0.0715**	0.3284	madr	0.0216	0.1147	0.2715
pie1	0.0139	**0.0735**	0.1998	pie1	0.0235	0.1257	0.1727
bogt	0.0108	**0.0743**	0.3720	bogt	0.0208	0.1907	0.3715
hraz	0.0105	**0.0825**	0.7874	hraz	0.0183	0.1035	0.2084
iisc	0.0104	**0.0827**	0.2970	iisc	0.0230	0.1771	0.6459
jplt	0.0115	**0.0840**	0.8003	jplt	0.0215	0.1601	0.6674
kokb	0.0123	**0.0888**	0.2769	kokb	0.0224	0.1791	0.3933
sant	0.0131	**0.0908**	0.2571	sant	0.0220	0.1406	0.4919
ispa	0.0141	**0.0939**	0.1507	ispa	0.0273	0.1612	0.2536
ykro	0.0101	**0.1026**	0.5156	ykro	0.0217	0.1952	0.5923
kgn1	0.0124	**0.1057**	0.4020	kgn1	0.0219	0.1949	0.8297
jplz	0.0132	**0.1249**	0.2015	jplz	0.0245	0.1907	0.2946
pimo	0.0115	**0.1350**	0.5250	pimo	0.0212	0.1580	0.3465
pol2	0.0107	**0.1479**	0.6253	pol2	0.0233	0.2398	0.8473
fair	0.0113	**0.1560**	0.5124	fair	0.0213	0.1336	0.6664
okc2	0.0132	**0.1713**	0.6293	okc2	0.0234	0.2736	0.7237
tidb	0.0114	**0.2049**	0.4923	tidb	0.0212	0.2557	0.6228
chan	0.0121	**0.2096**	0.5244	chan	0.0185	0.2898	0.7985
guaz	0.0133	**0.2207**	0.9047	guaz	0.0305	0.4247	0.8451
nssp	0.0121	**0.2304**	0.7757	nssp	0.0224	0.4629	0.7188
cro1	0.0114	**0.2383**	0.7365	cro1	0.0244	0.2939	0.8337
kely	0.0098	**0.2533**	0.7516	kely	0.0196	0.4297	0.8961
gode	0.0137	**0.2612**	0.8347	gode	0.0213	0.2860	0.8492
areq	0.0135	**0.2782**	0.7389	areq	0.0231	0.3795	0.7483
mkea	0.0158	**0.3020**	1.0276	mkea	0.0276	0.3192	0.8586
usud	0.0212	**0.3588**	0.6783	usud	0.0347	0.5303	0.9797
xian	0.0149	**0.3620**	1.0671	xian	0.0228	0.5587	0.9182
mbar	0.0100	**0.4241**	1.0844	mbar	0.0213	0.3772	0.8668
brew	0.0162	**0.4331**	1.0192	brew	0.0248	0.5971	0.95

vzpr	0.0164	**0.5049**	1.1495	vzpr	0.0308	0.5060	0.9366
sydn	0.0155	**0.5109**	0.8916	sydn	0.0287	0.8238	0.9750
bris	0.0190	**0.5554**	1.1276	bris	0.0334	0.6494	0.9459
falk	0.0165	**0.5970**	1.2252	falk	0.0287	0.5725	0.9380
joha	0.0162	**0.6412**	1.1019	joha	0.0241	0.6942	0.9522
port	0.0172	**0.6612**	1.2210	port	0.0274	0.6627	0.9532
inve	0.0171	**0.6699**	1.1893	inve	0.0261	0.7482	0.9696
zwei	0.0160	**0.7504**	1.2832	zwei	0.0229	0.6655	0.9503
wesl	0.0171	**0.7524**	1.2034	wesl	0.0289	0.7526	0.9691
raro	0.0280	**0.7822**	1.2935	raro	0.0439	0.8538	0.9853
perm	0.0157	**0.8101**	1.2869	perm	0.0243	0.6900	0.9615
buru	0.0302	**0.8334**	1.3093	buru	0.0440	0.8176	0.9820
farg	0.0204	**0.8498**	1.3093	farg	0.0288	0.7518	0.9703
rosa	0.0272	**0.9134**	1.3450	rosa	0.0431	0.8019	0.9787
perc	0.0188	**0.9143**	1.3301	perc	0.0275	0.7448	0.9642
kwj2	0.0243	**0.9214**	1.3582	kwj2	0.0372	0.7504	0.9667
cata	0.0221	**0.9312**	1.3211	cata	0.0321	0.7842	0.9759
atjn	0.0601	**1.1781**	1.3864	atjn	0.0967	0.9226	0.9929
toky	0.0636	**1.3845**	1.4112	toky	0.1162	0.9888	0.9994

Table 2: Horizontal and vertical error for all stations.

As table (2) shows there is a wide spread of performance and the difference is over ten times between the best and worst performing stations. The cumulative distribution function of all stations show the difference and for some stations the large sigma can be looked at visually. This is shown in figure (11) and (12), the right part of figure (12) have a slow bow which indicate that many measurement is one meter and above.

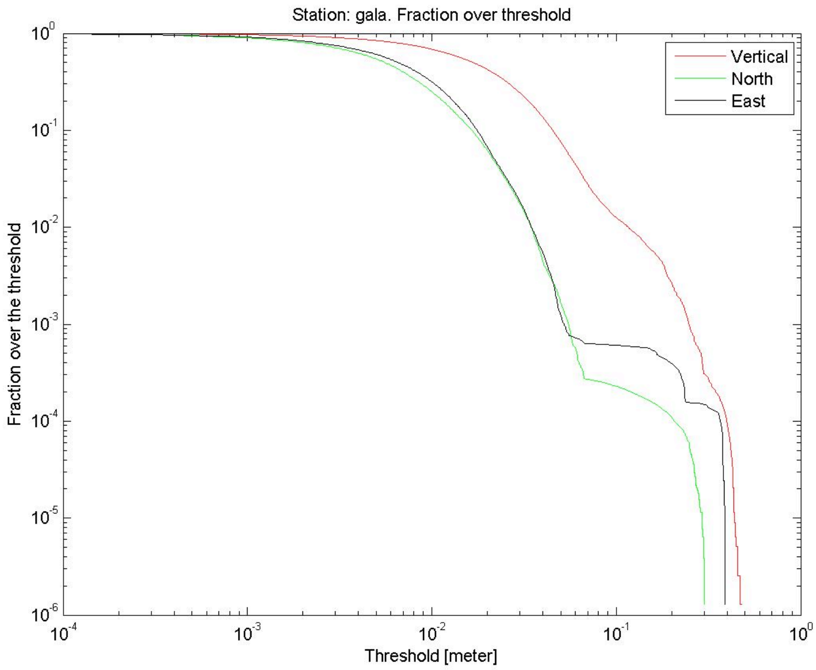

Figure 11: Cumulative distribution function for station GALA

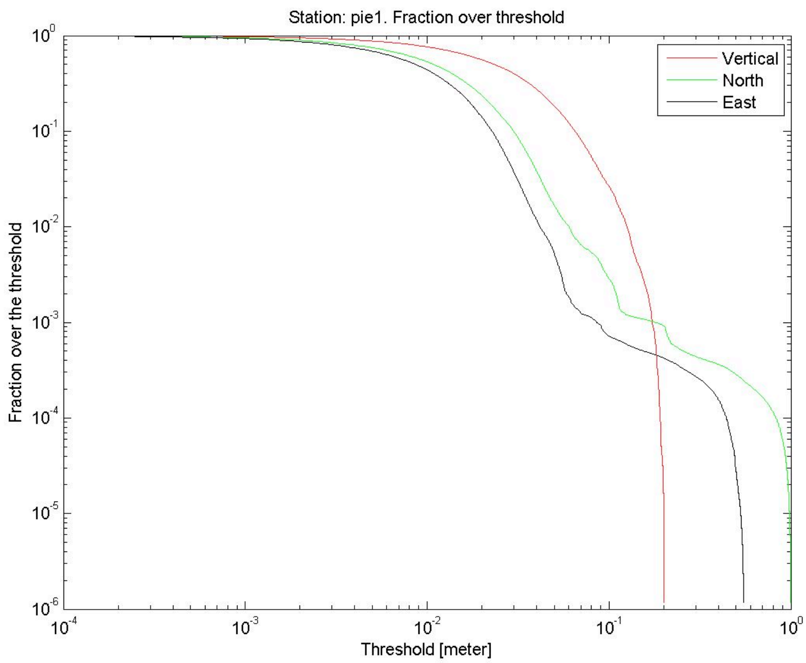

Figure 12: Cumulative distribution function for station PIE1

Figure 11 and 12 show a knee shaped part and that can be interpreted as the border between normal and large sigma measurement. If the problem with sigma would be ignored the performance would be a drop down curve instead of the knee shaped curve which would make the threshold significantly lower for a 99% case.

Whenever a large sigma occurs the measurement usually moves over a meter, however when sigma returns to normal the measurement have not. The sigma takes several minutes to return to normal however the measurement will take several hours return back to normal. The effect gives a false displacement position for several minutes and can trigger a false displacement alarm. As seen in figure (13) the sigma is returning to normal much faster than the measurement.

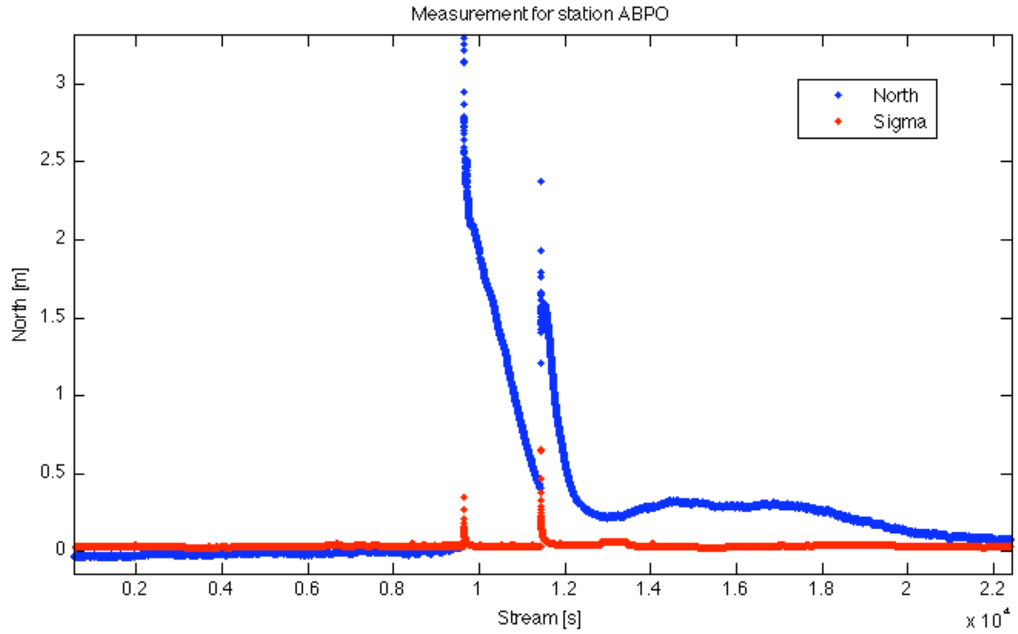

Figure 13: The problem with fast returning sigma. The measurement take longer time to return to normal

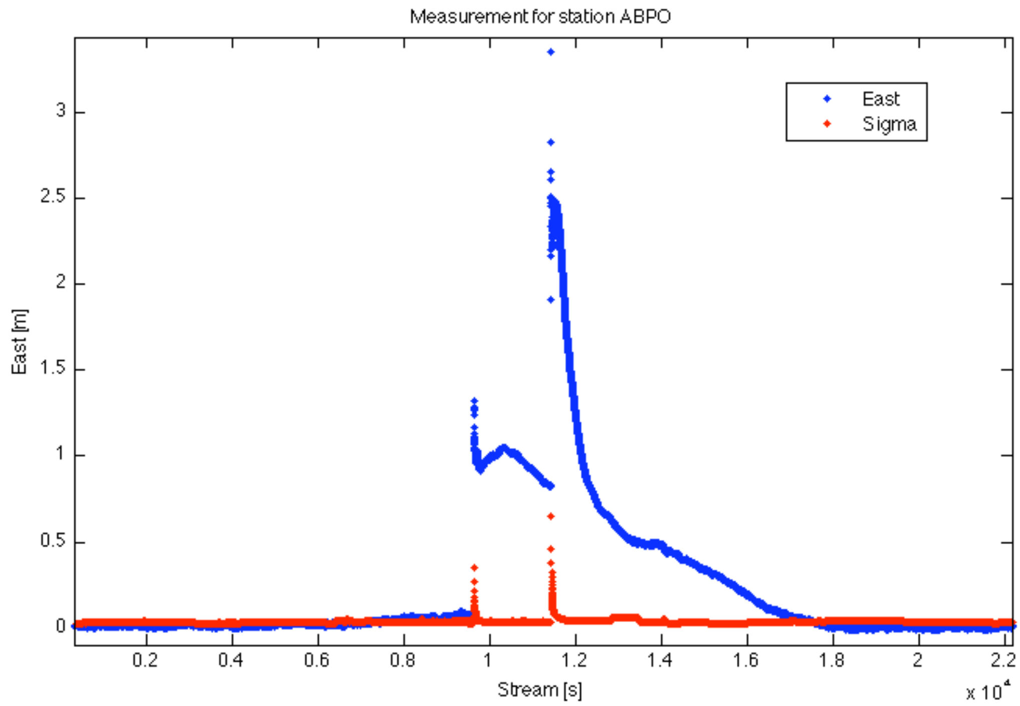

Figure 14: The fast returning sigma, East component have larger values however it returns to normal faster.

The east component takes shorter time to return to normal, however the displacement error is larger. Whenever a large number of measurements are dropt and missing the sigma tend to get large. This was believed to be the only time when large sigma occur, however for figure (13) and (14) this is not true. There where no missing measurement before the large sigma.

On the current project homepage [9] a weird displacement of measurement occurs several times. As seen in figure (15) there is a big jump for north and up component.

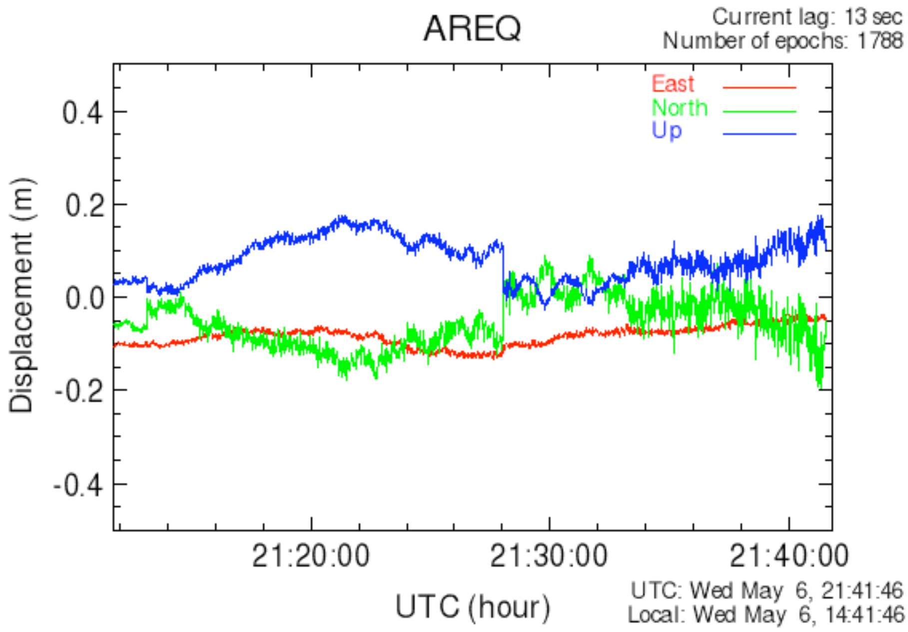

Figure 15: Displacement for station AREQ

A further investigation shown in figure (16) for the core stream and is done for Z and Y component only. This affects the transformation to earth coordinates. This movement looks like a displacement and can be confused as an earthquake.

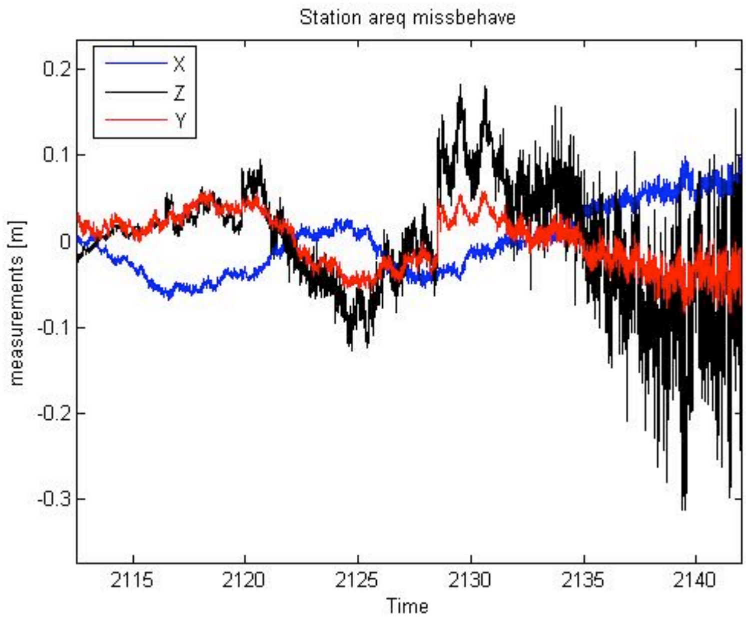

Figure 16: Core stream measurement for station AREQ

5.3 Sidereal filter

The autocorrelation of several days' measurement have the largest peak at 86156 seconds for the stations with a higher peak than 0.4. The time is 23 hours, 55 minutes and 56 seconds, which confirms the theoretical values of a sidereal day. As seen in figure (17) the correlation is high while figure (18) show low correlation.

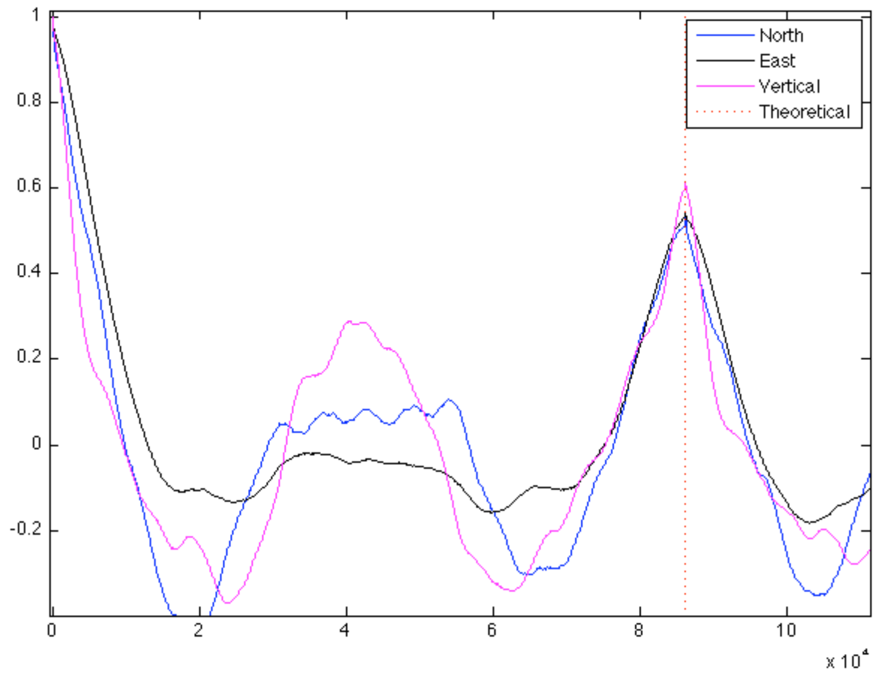

Figure 17: Autocorrelation function for station ABPO, high correlation for a sidereal day

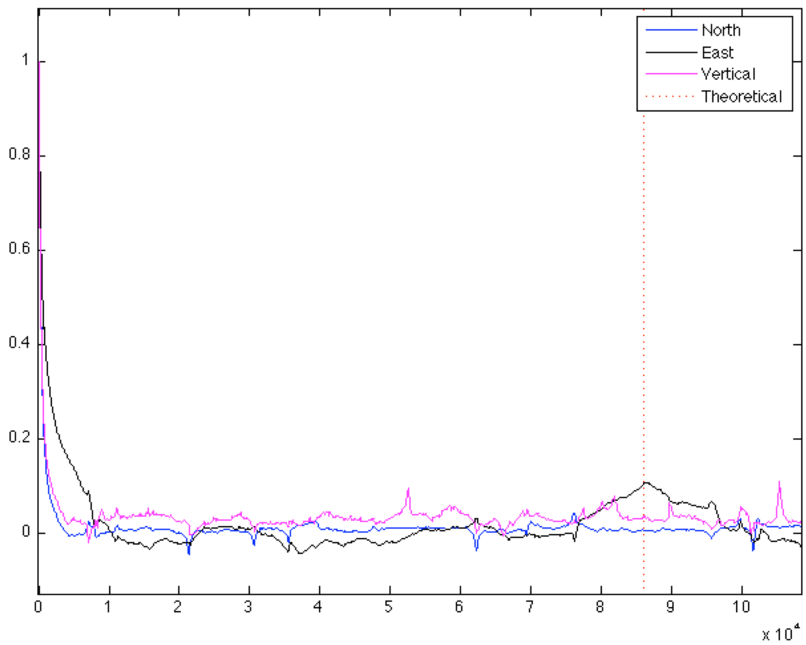

Figure 18: Autocorrelation function for station CATA, almost no correlation

After computing the filter a comparison of the next coming day was done. As figure (19) shows the match is very accurate. The drops down to zero are not received measurement, which has to be spaced out in order to make the filter and measurement the same length.

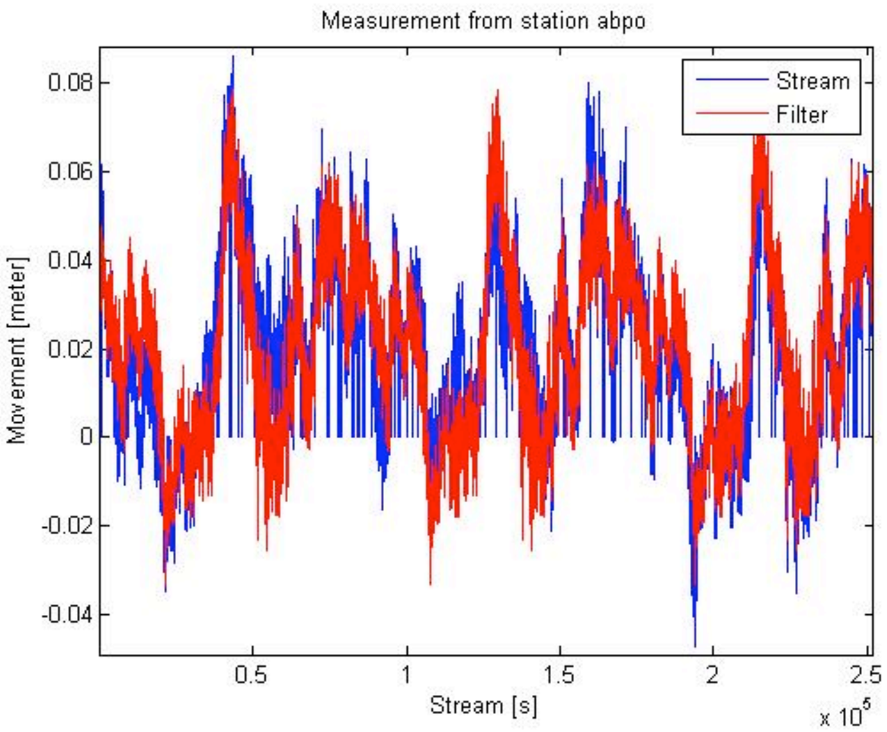

Figure 19: The filter and next day plotted together, notice the similarities in movement.

When implementing the filter, clear improvements is obtained and figure (20) show the result. The higher spikes are missing for the filtered measurement and the variation is smaller.

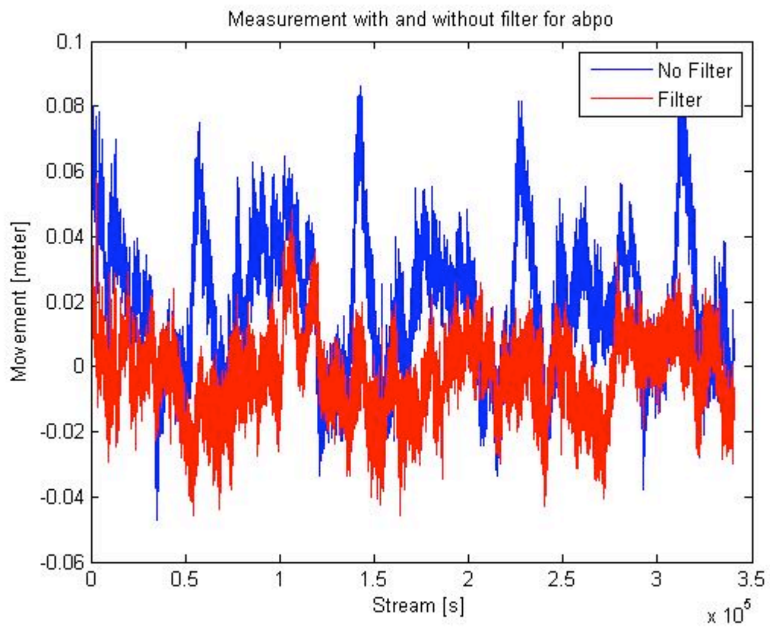

Figure 20: Filtered and non-filtered stream for station ABPO, the filter decreases the high spikes.

As seen in figure (21) and (22) the improvement is highest for north component where all stations with higher correlation than 40% while east have several stations with negative improvement.

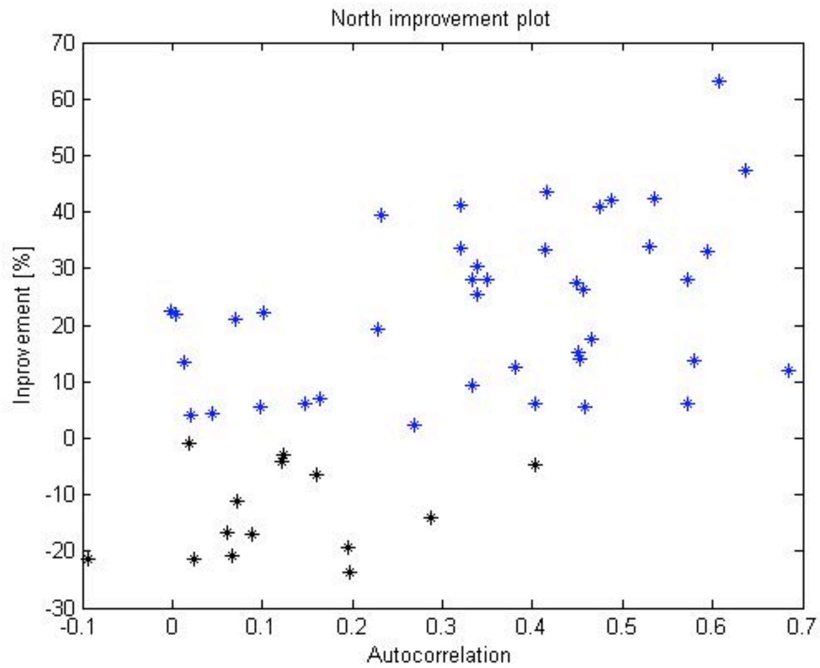

Figure 21: Improvement plot for north component, blue is positive improvement and black negative.

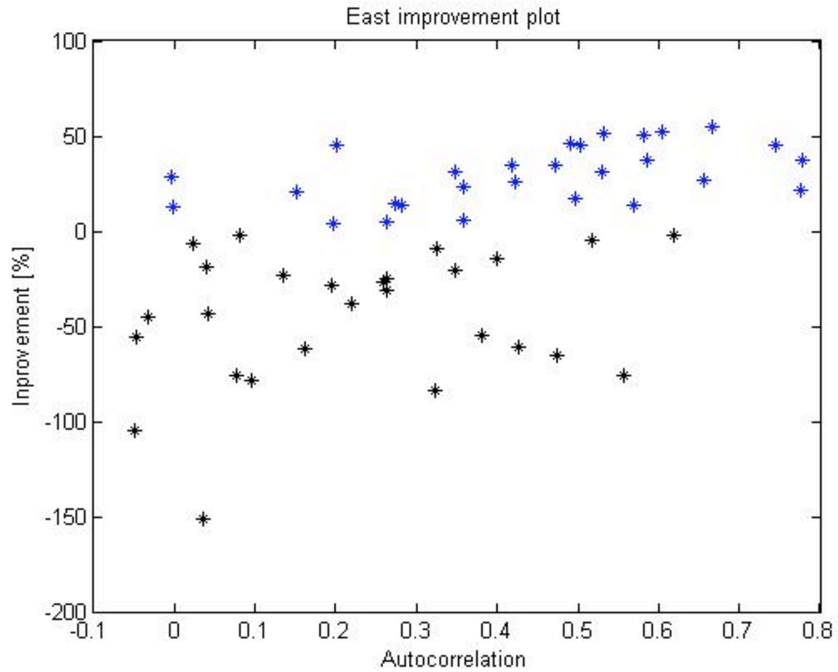

Figure 22: Improvement plot for east component, blue is positive improvement and black negative

6 Discussion

6.1 Statistics of stream

The missing 10 measurements are different for every stream. They occur at the same time as the drop in delay, this indicates a buffer problem. Because different stations have different delays at any given moment, which means that, the drop of data is done differently for different stations, the receiving part of the program could not have the buffer problem. The conclusion is that the problem is on the sender side. Which is like a black box for this project, however as stated in a memorandum to JPL about the problem, they have the chance to correct the problem.

Every stream has a different receiver however the same sender program, the guess is that it is probably in that program rather than in any calculation or model where the overflow is occurring. The suggestion for a solution is to look at the sender side if there are still 10 measurements missing before sending or even before pipe it into the sending program.

All other missing measurements are the same for all streams and this concludes that the problem is not the same as the 10 missing measurement. For any one or two missing measurement, there might be UDP drop in the network. Many stations are placed on very remote places and might have problem deliver data without loss. This might be the explanation for the longer periods of loss data. This cannot be solved for more than invest in better hardware for transmitting data from the remote stations to the core network, which is not up to the tsunami project.

6.2 Statistics of measurement

The non-earthquake displacement shows various errors for different stations. The best stations have a displacement error in cm range while the worst performing stations have up to a meter in displacement error. The sigma spikes can explain the difference in performance, which makes the slow declining measurement a large factor for displacement error. The worst performing stations have a large number of spikes while the best performing stations have very few spikes. If the measurement is not in the state of large sigma or declining from a large sigma spike, the performance looks very similar. The displacement error is in the range of less than one decimeter.

For the purpose of detecting earthquake the measurement have to be more reliable. As the result show, there are many stations with lots of lost data. This can be overcome by having several stations close to each other and the probability that several stations have drops at the same time is significantly less. Some stations are performing almost without large drops and could probably be used without any improvement.

Whenever a large drop of data occurs the sigma spikes which is a good thing. The return of sigma to normal is however very optimistic and gives a false statement of the measurement. The recommendation is to look into the sigma calculation and models in order to get rid of the optimistic sigma after any loss data. Some of the large sigma behavior is not caused by loss of data and have to be further investigated.

The false displacement for z and y component is a serious problem and need to be addressed. This can trigger a false alarm for earthquake, which is a serious problem.

6.3 Sidereal filter

For stations which have few sigma spikes have a high correlation within a sidereal day. The theoretical value and the calculated spike value are within a few seconds. This daily variation can be subtracted and the performance is improved with up to 60%. As the problem stated in the beginning of the discussion will be solved. The improvement will increase further while drop of data will increase and if sigma is more realistic the easier it will be to create a better filter that is weighted by a true sigma instead of only a median.

7 Conclusion

The current status of the stream is not enough to build a tsunami warning system and the current number of stations is not enough. As the project evolves JPL will be able to get more and more stations along the cost and problems such as missing measurement will not be as fatal as it is with the current set up. If more than one station is placed on one fault then it does not matter as much if there is a problem with one of them.

The problem with optimistic sigma will make it hard to build any filter which weights the measurement using sigma. It makes it harder to detect whenever a station is in bad shape or actually moves which can cause false displacement or not detect any real displacement. The slow decline in measurement should be addressed, if a drop of data occur, several hours of data should not be lost.

Whenever a station is performing as normal (no large sigma or missing measurement) the quality is in cm to dm range. This can be improved using sidereal filtering and the improvement when station behaves without sigma problem or missing measurement very good.

8 Further work

The next step in this project is to investigate in the problems stated in this thesis. The drop of data every time the delay is above 15 seconds is a buffer overflow and is only a small programming problem that can be worked out. The missing measurement may not be a solvable problem. The remote stations may need better hardware or the position is to remote to get all data. This is a money issue and can be very hard to solve. However the slow decline can be addressed as a separate problem.

Whenever these problems are fixed, this thesis can be a base in order to calculate improvement and can be used as a reference of problem. The same tests should be done and comparison with current results hopefully shows a better performance in future.

9 References

[1] El-Rabbany, Ahmed (2002) Introduction to GPS: The Global Positioning System. USA: Artech House

[2] Parkinson, W. Bradford (1996) Global Positioning System: Theory and Applications, Volume I. USA: American Institute of Aeronautics and Astronautics, Inc.

[3] Hofmann-Wellenhof, B., Lichtenegger, H. & Collins, J. (2001) GPS Theory and practice. Austria: Springer-Verlag

[4] Jonsson, Bo. Gävle (1997) Geodetic Applications of GPS. Sweden: Båstad.

[5] Blewitt, Geoffrey. Hammond, William C. Kreemer Corné. Plag, Hans-Peter. Stein, Seth. Okal, Emile (2008) GPS for real-time earthquake source determination and tsunami warning systems. Springer-Verlag

[6] Miller, Scott L. Childers, Donald G. (2004) Probability and Random processes, Burlington: Elsevier Academic Press

[7] Kaplan, Elliot D, Hegarty Christoffer J. (2006) Understanding GPS: Principles and Applications, Second Edition. Bedford, Massachusetts

[8] Choi, K. Bilich, A. Larson, K, M. Axelrad, P. (2004) Modified sidereal filtering: Implications for high-rate GPS positioning, Department of Aerospace Engineering Sciences, University of Colorado, Boulder, Colorado, USA

[9] Plag, Hans-Peter (2009) GREAT http://cheryl.nbmg.unr.edu/GREAT/ [Viewed 06-05-09]

[10] Bar-Sever, Yoaz (2009) The NASA Global Differential GPS System – System Description http://www.gdgps.net/system-desc/index.html [Viewed 15-07-09]

[11] Plplot developer community (2008) Plplot homepage http://plplot.sourceforge.net/ [Viewed 20-06-09]