Bachelor's Thesis in Physics

Studies of Acoustic Position Calibration for the Neutrino Telescope KM3NeT

Submitted by

Isabella Probst

August 11, 2021

Erlangen Centre for Astroparticle Physics

Friedrich-Alexander Universität Erlangen-Nürnberg

Supervisor: PD Dr. Robert Lahmann

Abstract

The KM3NeT Neutrino telescopes are built to detect high-energy neutrinos of cosmic origin and determine the mass hierarchy of neutrinos. Therefore two large underwater Cherenkov detectors are currently under construction in the Mediterranean Sea. These detectors will form the ARCA and the ORCA telescope. The detectors will consist of vertical strings holding optical modules. As this modules can move a acoustic system is used for position calibration. Former investigation have shown, that the error of this position reconstruction is higher than expected and seems to be caused by digital effects. In this bachelor's thesis these effects are reproduced and then it will be investigated if it can be reduced with appropriate methods.

Contents

1 Introduction

KM3Net is an international collaboration for neutrino research and is constructing a network of neutrino telescopes in the Mediterranean Sea. Neutrinos are elementary particles, that only interact via the weak interaction. There are three flavors of neutrinos, the electron neutrino ν_e, the muon neutrino ν_μ and the tau neutrino ν_τ. Each of this particles also has an antiparticle.
Due to their little interaction with matter, neutrinos are very interesting for astronomy and astroparticle physics. Unlike other particles, they travel on a direct path from their source to the earth. This makes it possible to determine the position of their source. Neutrinos are amongst others produced at acceleration sites of cosmic rays or their propagation. High energetic charged particles from outer space can be detected on earth, but their origin can not be determined, as they get deflected by magnetic fields on their way to earth. By detecting neutrinos from such acceleration sites, their position can be determined. Possible sources of cosmic rays and neutrinos are supernova remnants, pulsars, binary systems and more. [10]
An unknown property of neutrinos is their mass hierarchy. Experiments have shown, that neutrinos oscillate between their flavors. This is only possible, if they have three non-zero mass eigenstates. Until now, it is only known, that $m_1 < m_2$. This leaves two possible mass hierarchies, either a normal hierarchy with $m_1 < m_2 < m_3$ or a inverted hierarchy with $m_3 < m_1 < m_2$. [10]

The little interaction of neutrinos with other particles, makes them interesting for astronomy, but also hard to detect. A possible way is, to detect high-energetic, charged particles, which originate from neutrino matter interactions, via Cherenkov detectors. Therefore detectors with a large volume are necessary. To realise detectors with volumes up to km^3, detection modules are placed in ice or water. The, currently under construction, KM3NeT telescopes are build in the Mediterranean See, and use the seawater as detector medium. [1, 10]

At the KM3NeT telescopes, acoustic signals are used to determine the position of the DOMs at a certain time. These signals are analysed via cross-correlation. The

position reconstruction showed a higher deviation than expected, which seems to be caused by digital effects rather than from random noise [7]. In this B.Sc. thesis, this digital effects are analysed by comparing the time of arrival of one acoustic signal at different fixed detection points. In section 2 an overview of the KM3NeT Neutrino telescopes and neutrino detection is given. Followed by a more detailed explanation of the position calibration in section 3. In section 4 the time of arrivals given at the KM3NeT data base are analysed and reconstructed from the raw data files. Whereas in section 5 upsampling is used to reduce the deviation of the time of arrivals.

2 The KM3Net Neutrino Telescopes

2.1 The Telescopes

The KM3NeT neutrino telescopes are split in two projects, ARCA and ORCA with different goals and locations in the Mediterranean Sea. But both telescopes are constructed similar.
The detectors consist of several detection units (DUs). A detection unit is a string, that is anchored at the ground of the Mediterranean Sea. Each detection unit holds 18 optical modules (DOMs), each of which host 31 photo-multiplier tubes. At the upper end of the detection unit buoys are attached. This leads to a vertical orientation. Three building blocks, each consisting of 115 detection units, are foreseen. Two of this blocks will form the ARCA telescope and the last one the ORCA telescope. As of August 2021, six detection units have been deployed for both ARCA and ORCA. A schematic overview of a building block can be seen in Figure 1. [1, 11]

ARCA (Astroparticle Research with Cosmics in the Abyss) is located about 100 km off the coast of Porto Palo di Capo Passero in Sicily (Italy), see Figure 2a. The Sea is about 3500 m deep at this location. Each detection unit is about 700 m long and the detection units will be in average placed 95 m apart. The whole ARCA telescope will cover a volume of about 1 km^3 [6]. The purpose of the ARCA telescope is to detect high-energy neutrinos of cosmic-origin. Its position on the

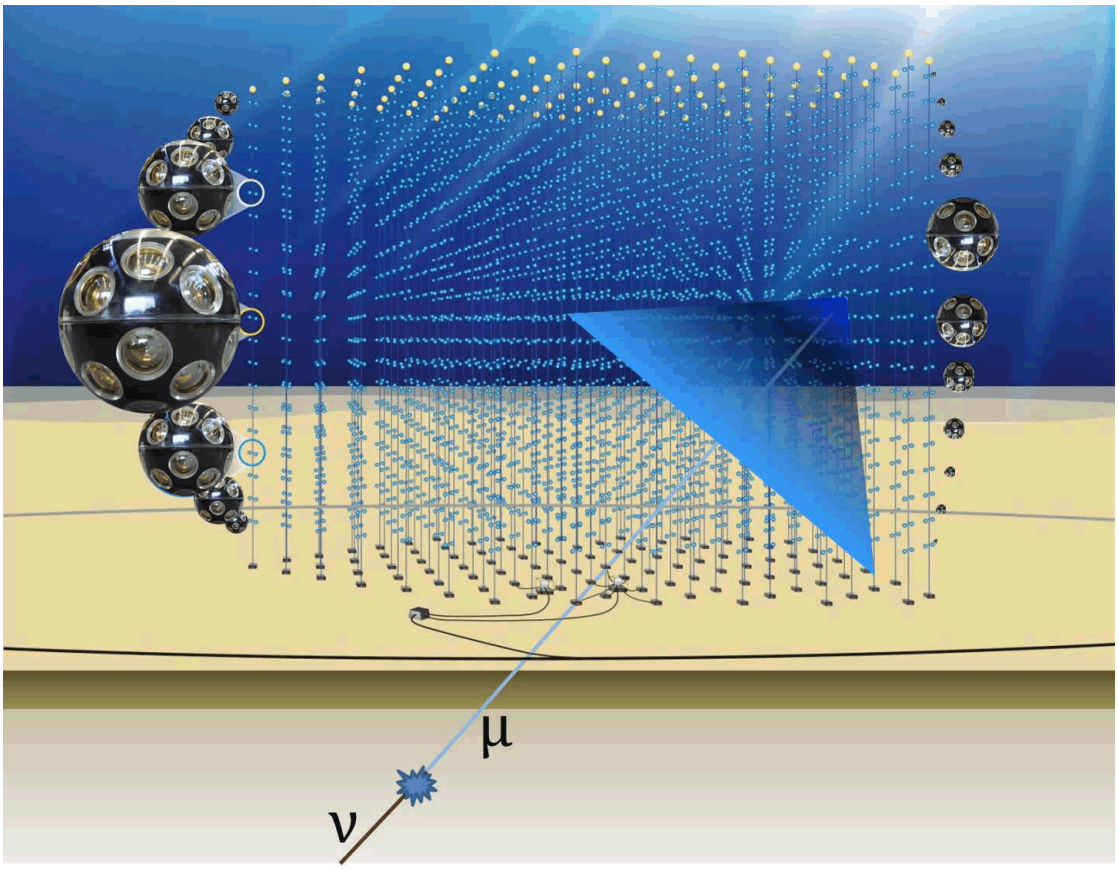


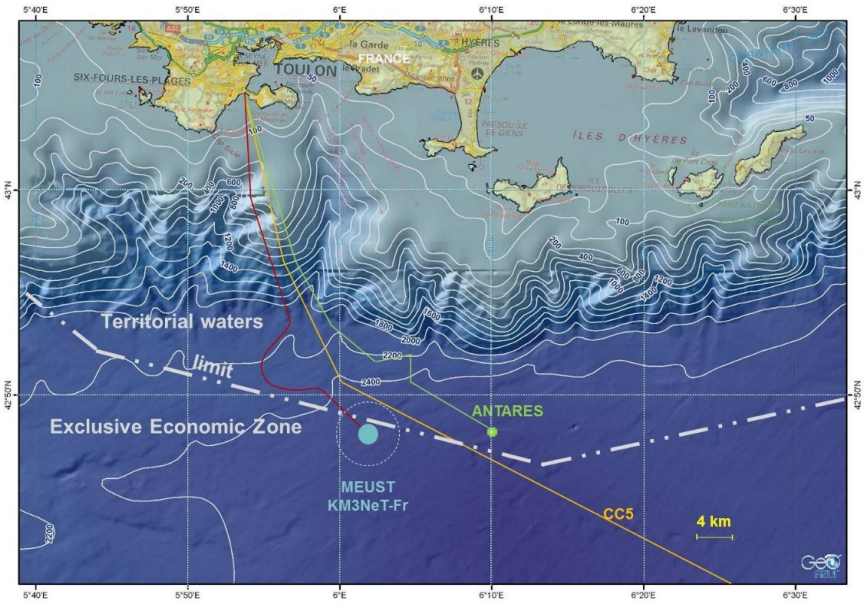
Figure 1: Schematic view of a KM3NeT building block. Shown are strings with DOMs and a example muon μ, originating from a neutrino ν interaction. Shown in a blue triangle is Cherenkov radiation, caused by the muon. Also shown are closeups of some DOMs. [3]

northern hemisphere makes it best suitable for observing the southern sky and a counterpart to the IceCub detector at the South Pole. [1]
The ORCA (Oscillation Research with Cosmics in the Abyss) detector is located about 40 km off-shore Toulon in France, see Figure 2b. The detection units are anchored in 2450 m depth and placed 20 m apart from each other. As the detection units are also only 200 m long the density of detection units is way higher than at the ARCA telescope. The goal of this detector is to determine the mass hierarchy of Neutrinos. Also shown in Figure 2b is the ANTARES telescope. This telescope is a smaller predecessor of the KM3NeT telescopes, for further information about ANTARES see [2]. [1]

2.2 Neutrino Detection

At KM3Net Neutrinos are not detected directly, but indirectly via particles from neutrino interactions. These secondary particles are detected via Cherenkov radiation. A charged particle moving faster through a medium, than the speed of light in this medium, causes the emission of Cherenkov radiation. This Cherenkov radiation is than detected by the photo-multipliers at the DOMs. For a high energetic neutrino telescope, one can differentiate between track-like events and cascade-like events.
Neutrino matter interaction can produce high-energetic muons. If these muons travel through the detector, they cause track-like events. For such a single particle, the Cherenkov radiation forms a cone around the particle, as shown in Figure 1. From the Cherenkov cone, the trajectory and the speed of the muon can be reconstructed. If an up-going muon track is detected, the muon has to originate from a neutrino interaction, since only neutrinos can cross the earth. Therefore such up-going muon tracks are the best way to determine the position of astrophysical neutrino sources. Neutrino interactions also produce hadronic and electromagnetic cascades. These cascades are typically shorter than a muon track and the Cherenkov light does not form a cone, as there are several particles moving in different directions. Therefore no track can be reconstructed. But this cascade-like events are well suitable to determine the energy of the neutrino. [10, 1]

(a) ARCA

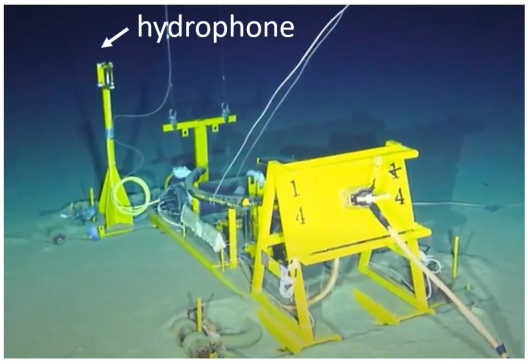


(b) ORCA

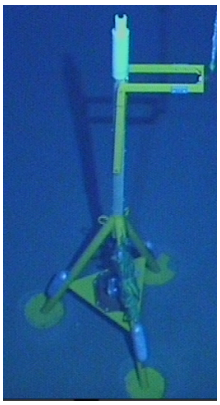
Figure 2: Locations of the ARCA (KM3NeT-It) and the ORCA (KM3NeT-Fr) telescopes in the Mediterranean Sea. Also shown in (b) is the ANTARES telescope. [1]

3 Position Calibration

As the DOMs are only attached to a string, which is anchored at the ground, they can move due to current. Therefore it is necessary to monitor their position. This is done with a acoustic system. An acoustic signal is emitted at a fixed point and detected at the DOM. As acoustic signal, chirps, also called sweeps, are used. Each chirp is 5 ms long and in this time interval the frequency of the signal is linear increased by 2 kHz. There are several emitters, called beacons, which each have unique chirp frequencies. At the DOMs, piezo-based acoustic sensors are installed. The recorded signal, does not only contain the chirps, but a lot of noise. With the help of cross-correlation the time of arrival of a chirp at the DOM is determined between the noise. As the time of emission of the chirp is also known, it is possible to calculate the time difference between emission and detection of a chirp. With help of this time and the acoustic velocity in water, the distance between emitter and DOM can be calculated. For the trilateration of an individual DOM, distances to at least three fixed points (beacons) need to be known.
Apart from the acoustic sensors at the DOMs, hydrophones are installed at the anchors of some detection units. These hydrophones have a fixed, but not exactly known position, but can be used to determine their own position and the positions of the beacons.



(a) Hydrophone at the anchor (Picture credit KM3NeT).



(b) Acoustic emitter (beacon) (Picture credit KM3NeT).

Figure 3: Fixed position calibration system at the ground.

3.1 Cross-correlation

To determine the time of arrival of the acoustic signal, cross-correlation is used. A cross-correlation compares the signal with a reference signal. In this case the signal is the recorded signal and the reference signal is waveform of the at the beacon emitted signal. For two continuous signals f and g the cross-correlation

$$(f \star g)(t) = \int_{-\infty}^{\infty} f^*(\tau) g(\tau + t) \mathrm{d}\tau \tag{1}$$

is the integral over the complex conjugated signal f^* and the signal g shifted by t [8]. Now the maximum of the function $(f \star g)(t)$, gives the value t, for which f and g are most correlated.
As in most practical cases, the acoustic signal is saved in samples and not in a continuous form. Therefore one needs a discrete form of the cross-correlation and the reference signal. In this case the integral becomes a sum

$$(f \star g)_i = \sum_k f_k^* g_{k+i} \tag{2}$$

[8]. The signal f is again complex conjugated and g is shifted by j. One can imagine, that g is shifted step by step over f and j is giving the number of steps. For $i = 0$ the first sample of the signal f is paired with the first sample of the reference signal g. For $i = 1$ the first sample of g gets paired with the second sample of f and so on. This is done until the last sample of g is paired with the last sample of f. Now the j, for which the sum is maximal, is the offset, for maximal correlation. For our acoustic signal, this means one can determine the time of arrival by searching for the maximum i_{max} in the discrete correlation function. The time of arrival toa is given by

$$toa = t_0 + \frac{1}{f_s} \cdot i_{max} \tag{3}$$

with t_0 the time of the first sample and $\frac{1}{f_s}$ the length of one sample given by the sampling frequency f_s. For both, the continuous and the discrete correlation, if $f = g$, the correlation is called auto-correlation.

3.2 Position Calculation

With the time of arrival of a signal obtained from the cross-correlation and the emission time of the signal the time of flight t is calculated. The distance d between the emitter and the hydrophone at the anchor is approximately given by

$$d = t \cdot c(z) \tag{4}$$

with $c(z)$ the speed of sound at the depth z, where the beacon and hydrophone are installed [11].
The speed of sound depends on the water temperature, salinity and pressure. Therefore approximations depending on the depth are used. At a depth of -2450 m where the ORCA detectors strings are anchored, the speed of sound is about 1545.65 $\frac{\mathrm{m}}{\mathrm{s}}$. For calculating the speed of sound at a given depth and in the following the distance corresponding to a time of flight, the script "timeDist.py" [14] is used. [9]

3.3 Observed problems with position calibration

Previous investigations showed, that the deviation of the position reconstruction is higher than expected. Furthermore it showed a pattern indicating an origin from digital effects rather than from noise, see [7]. In the following sections the effect will be reproduced and than investigated if it can be reduced by upsampling the recorded data and reference waveform.

4 Time of arrival

At the moment, the KM3NeT telescopes are still under construction, but parts of the detectors are already working and collecting data. For the following parts, data from the ORCA telescope will be used.

4.1 DUs and Beacons

Figure 4 shows the positions of the in the following used detection units (DUs) and beacons. A detection unit is a string with DOMs and the anchor at the ground. These units are already installed and part of the ORCA detector. The detector is located in the UTM zone 32N and the positions are given in meters. At DU2, DU3 and DU9 hydrophones are installed at the anchor of the strings. These hydrophones are not moving over time and therefore used for position calibration. In this work, the signal from two beacons will be used. Beacon 14 emits an acoustic chirp with a frequency of 30 kHz up to 32 kHz. The chirp of beacon 16 goes from 34 kHz up to 36 kHz. The chirps emitted from the beacons are referred as waveform 14 and waveform 16. All chirp signals are 5 ms long. Each beacon emits 11 chirps every ten minutes[1]. This 11 chirps are 5 s apart. [11]

To analyse the accuracy of the position calibration via cross-correlation, the time of arrivals of the chirps at different DUs are compared. For a well functioning position calibration, it is expected, that the time differences between two hydrophones is equal for each chirp. This can be expected, because of the fixed position of the hydrophones at the anchors.

4.2 Time of arrival database

At first the time of arrivals from the KM3NeT data base are analysed. For this purpose data from two runs is used. Run 8882 started on 09. November 2020 at 12:20:00 and was 1 h 40 min long. Run 10477 was taken on 19. July 2021 and

[1] Due to a malfunction, the beacon with waveform 12 emits only one chirp every 10 minutes and was not used, due to the resulting lower statistics. [11]

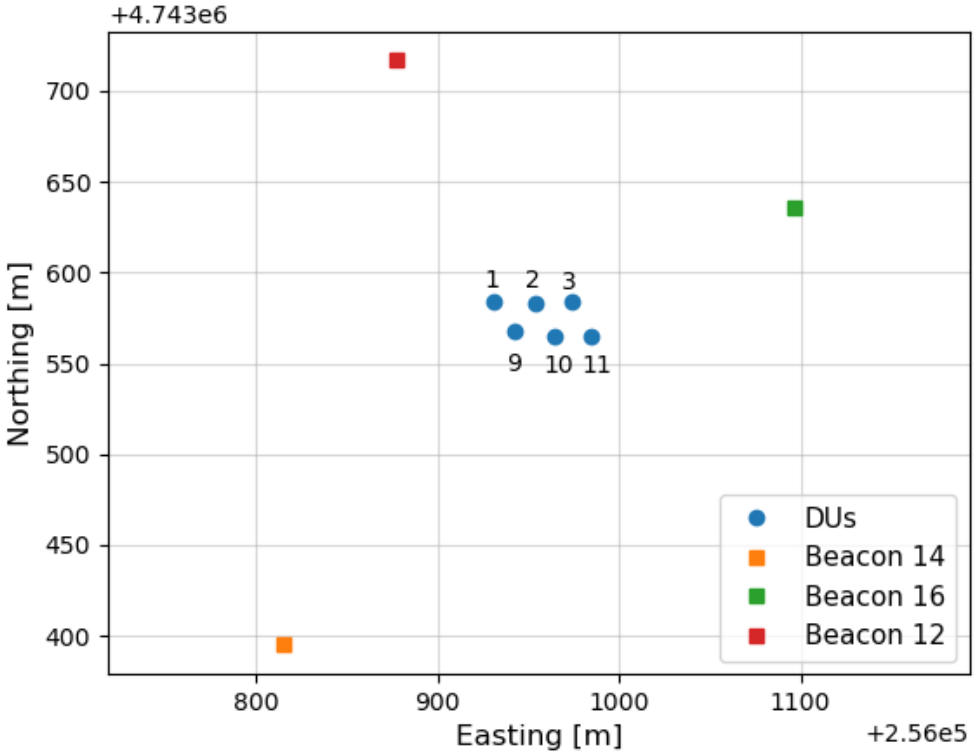


Figure 4: Positions of the used detection units (DUs) and beacons. The position is given in UTM standard, Zone 32N.

was only 9 min and 59 sec long. At the database the DUs are separated in floors and the hydrophones are referred to as floor 0. For reading the data the script "readDataDemo.py" [13] was used. This program also returns information about the date and length of the run.

4.2.1 Run 8882

In the first plot of Figure 5 the data obtained from the data base is shown for the signals from beacon 14. The y-axis shows the absolute value of the maximal achieved absolute cross-correlation value for each time of arrival. This maximum is called the quality factor, or q-factor.
One can separate times with a q-factor over $0.5 \cdot 10^6$ and way lower q-factors. For this data only times with a correlation over $0.5 \cdot 10^6$ are real detected signals. This data points show the expected structure of several data points in a short time interval, followed by a longer break. The times with a low q-factor are caused by noise or signals from other beacons. The maximal value of the q-factors vary from detection unit to detection unit, but for all a clear differentiation between signals and background can be made. The time of arrivals of run 8882 for beacon 16 show a similar pattern. At run 8882 both beacons emitted 110 chirps, but not every chirp was recorded by all three hydrophones.

The next step is to extract the time of arrivals with a high correlation and calculate the difference to the time of arrivals at the other detection units. A difficulty at this task is, that for waveform 14, some time of arrivals are missing. This is caused by some time intervals with no data. To be able to compare the time of arrival for one chirp, two arrays of same length are needed. Further the time of arrivals from one chirp has to have the same index. As there should be 11 signals with about 5 s difference, missing signals can be detected. For this missing signals a dummy variable is inserted in the time of arrival arrays. This dummy is than ignored, when calculating the time differences between the detection units.
The results for the time differences will be discussed in subsection 4.4 in comparison to time differences calculated from raw data files.

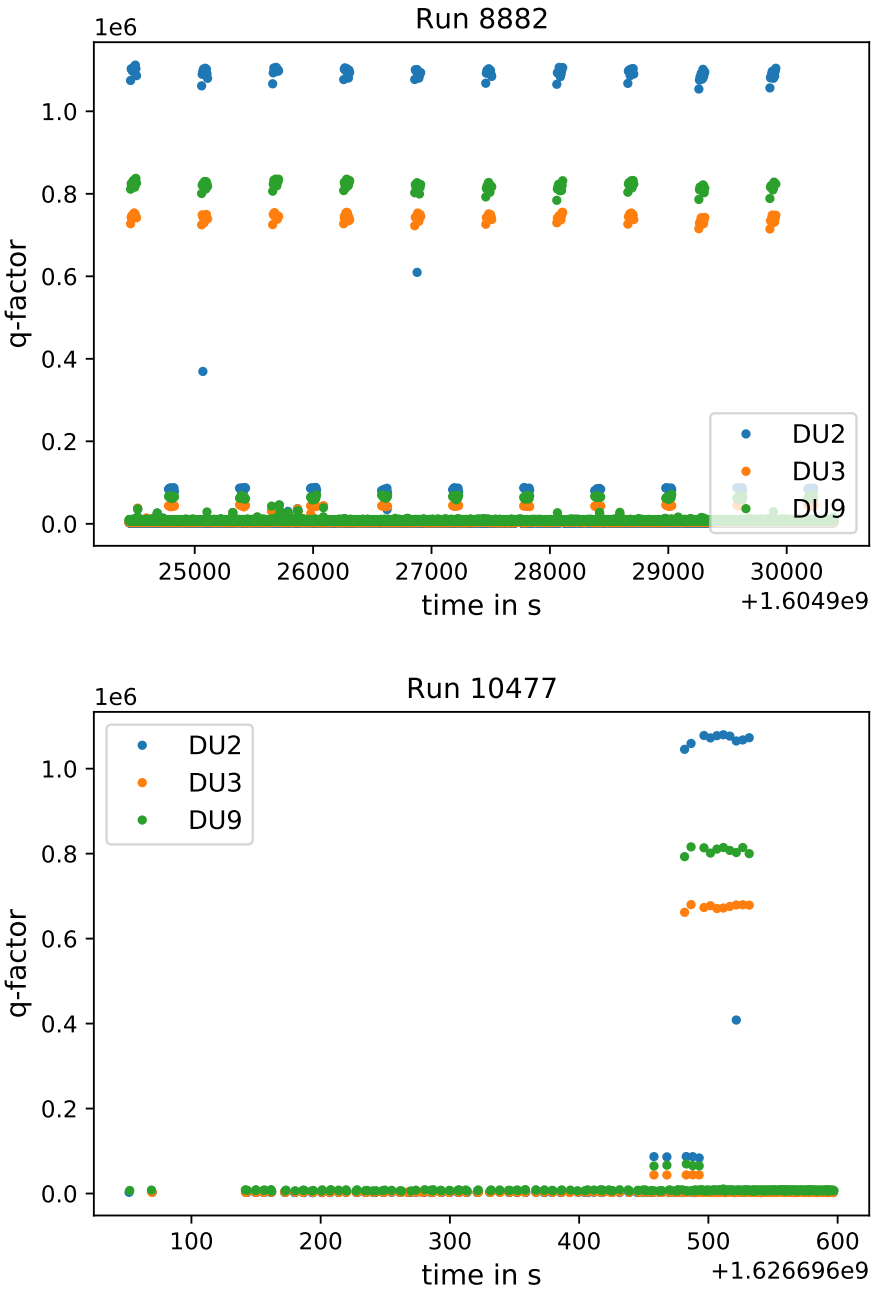


Figure 5: Time of arrivals for the at beacon 14 emitted signals, with data from the database. The y-axis shows the absolute value of the maximal cross-correlation for each time of arrival.

4.2.2 Run 10477

In run 10477 only 11 signals for waveform 14 and 8 signals for waveform 16 are detected, due to the short length of the run. The time of arrivals with q-factors for this run are shown in Figure 5. The maximal q-factors have similar values as for run 8882. For analysis, the same steps as for run 8882 were done, but in this run, no time of arrival was missing.

4.3 Analysis raw data

The next step was to reconstruct the time of arrivals from the database. Therefore a program for reading raw data and performing the cross-correlation was needed. This program is written in python and is an extension of the program "rawDataDemo.py" for reading the raw data [12]. The first part of the program reads the data from the raw data files and creates an array with the data and an array with the corresponding time. The raw data is saved in binary files, with each file containing data from a 6 s interval. The sampling frequency of the data is $f_s = 195312.5\,\mathrm{Hz}$, therefore one data sample is $5.12\,\mu\mathrm{s}$ long. The reference signal is saved as an array with the same sampling frequency.
For the cross-correlation the scipy function "scipy.signal.correlate" is used. This function calculates the discrete cross-correlation of two arrays as described in subsection 3.1. For further details see the documentation [4]. This function returns an array with the result of the correlation for each shift. The result of the cross-correlation shows a specific pattern as one can see in Figure 6. At this plot the correlation is plotted over the shift.
By using the option "valid", each shift in the correlation array equals one entry in the time array. The time of arrival is the maximal correlation of the signal and reference signal. The maximal correlation can be a positive or negative value. A positive value means that the signal and reference signal oscillate in the same direction. The signal and the reference signal being mirrored, results in a negative correlation value. Therefore the absolute values of the resulting array is searched for the maximum. The index of this maximum than can be used to find the corresponding time in the timearray. This time and the q-factor for this time

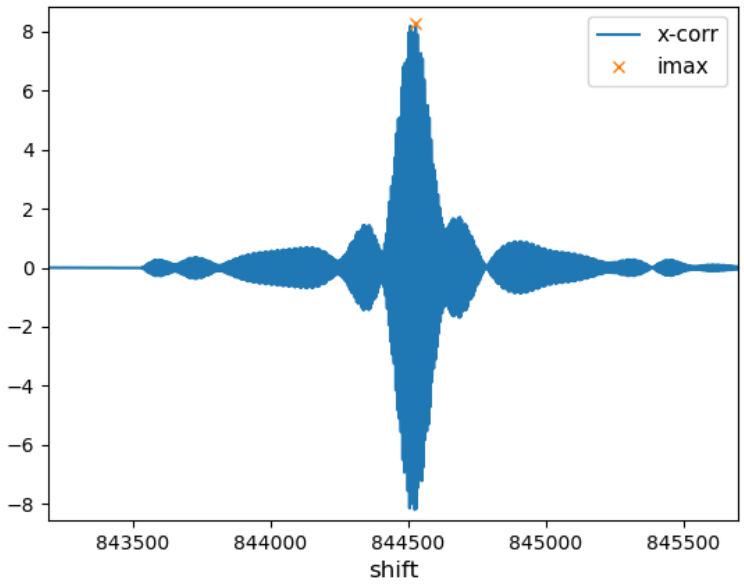


Figure 6: Result cross-correlation (x-cor) for a raw data file (DOM_808981515_CH1_1604927507.bin) from run 8882 and waveform 14, detected at the hydrophone from detection unit 2. The maximal correlation is marked as imax.

are saved in an array, to be analysed. One can already see in Figure 6, that the correlation does not have a smooth peak in the middle, but has some structure.
In Figure 7 the result of the cross-correlation for run 8882 and waveform 14 is shown. On the x-axis the time of arrivals are plotted with the q-factor on the y-axis. When regarding the absolute values of the q-factors, the plot shows the same pattern as for the time of arrivals from the database, where the q-factors are absolute values (see Figure 5). The values of the q-factors are about the same, apart from a factor 10^5, which is due to different normalisation. The plot of the raw data also shows less times with a small absolute q-factor, because only times with an absolute q-factor above 0.5 are saved after the cross-correlation.
For calculating of the time differences between the hydrophones, the absolute values of the q-factors are used to filter the time of arrivals for times with a high correlation. Then the time difference between the hydrophones is calculated as it was one with the data from the database.

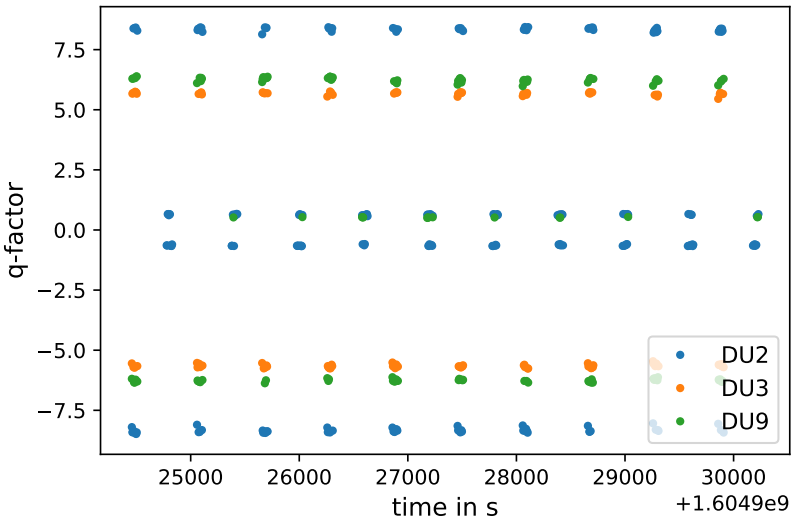


Figure 7: Time of arrivals for the at beacon 14 emitted signals, for the raw data from run 8882. The y-axis shows the maximal received correlation for each time of arrival.

4.4 Results

4.4.1 Run 8882

Figure 8 shows the time difference for acoustic signals from beacon 14 and Figure 9 from beacon 16. Shown in blue are the time differences calculated from the database data and in orange the differences calculated from the raw data. Overlapping of an orange and a blue bar results in a brownish color, for entries, that occur in both, the data base data and the raw data. One bar in the histograms is 2.56 μs wide, which equals half the length of one data sample.

Comparing the data from the database and the raw data in the plots, the distributions are nearly the same. There are deviations of one or two bins, but this is most likely caused by numeric effects in combination with the binning of the data. The similarity between the database and the raw databases also shows in the mean and the standard deviation of the distributions given in Table 1 for the database data and Table 2 for the raw data. Also given is the difference between the minimal and maximal time difference for each distribution.

waveform 14

	mean in μs	std in μs	spread in μs
DU2 and DU3	-8282.84	16.98	114.20
DU2 and DU9	13835.98	35.62	128.98
DU3 and DU9	22117.93	37.76	128.75

waveform 16

	mean in μs	std in μs	spread in μs
DU2 and DU3	12367.92	33.08	73.91
DU2 and DU9	-11779.21	12.21	73.19
DU3 and DU9	-24147.13	33.56	73.67

Table 1: Time differences for run 8882 with data from the KM3NeT database.

waveform 14

	mean in μs	std in μs	spread in μs
DU2 and DU3	-8284.16	17.23	116.35
DU2 and DU9	13836.47	36.05	130.89
DU3 and DU9	22119.74	38.00	130.65

waveform 16

	mean in μs	std in μs	spread in μs
DU2 and DU3	12365.40	33.09	76.77
DU2 and DU9	-11778.30	12.84	77.25
DU3 and DU9	-24143.52	33.49	76.77

Table 2: Time differences for run 8882 with times obtained via cross-correlation from raw data.

The standard deviations for the distributions are different for the two waveforms and also for the regarded time differences. For waveform 14 and raw data, the distribution for DU2 and DU3 has a standard deviation of 17.23 μs. This is less than half the standard deviation of the other two distributions with over 36 μs. This leads to the conclusion, that for waveform 14, the time of arrival at DU9

varies the most. For waveform 16, the standard deviation of the time difference of DU2 and DU9 is way smaller than for the other distributions. Here the standard deviation is for the raw data is 12.84 μs compared to over 33 μs. Here leads DU3 to a high variation. The distributions from the database show the same pattern only with small variations in the standard deviations.
The histograms, both for the database data and the raw data, do not have one single peak as expected, but the differences vary. Noticeably is, that for waveform 14, the histograms all have a three peak pattern. There is a higher peak in the middle with two smaller peaks on each side. Most of these peaks are two to three bins wide and the gap between is three to four bins. The difference between the centers of this three peaks is about 15.4 μs, which equals about half a cycle duration of the chirp. The starting frequency of waveform 14 is 30 kHz, which equals a cycle duration of about 33.3 μs. Therefore the distance between the peaks can be explained with the chirp been detected half a period earlier or later. Apart from this three peaks, there are also time differences, that differ even more from the middle of the three peaks structure. Regarding the raw data and the difference between DU2 and DU3, there is a difference of 100 μs between the highest bar and the bar most far apart. For the difference between DU2 and DU9 this is 115 μs and for DU3 and DU9 112 μs. In distance a deviation of 100 μs is approximately 0.15 m [2].

The distributions for waveform 16 do not show a clear three peaks pattern, as the distributions for waveform 14, but the time differences also vary over more than 70 μs. At the distributions of the time difference between DU2 and DU3 and the difference between DU3 and DU9, there are two peaks with a smaller distance, than a slightly bigger gap and two peaks with a smaller gap in between again. The two peaks with the smaller gap, are in average five bins and therefore about 13 μs apart. The frequency of waveform 16 is at the beginning 34 kHz, this equals a cycle period of roughly 29 μs. Therefore the two peaks are also resulting from the signal being detected half a cycle earlier or later, as was the three peak structure at waveform 14. Calculating the difference between the first and the third peak, or the second and the fourth, leads to a time difference of about 60 μs.
The distribution for the difference between DU2 and DU9 has one big peak with a

smaller peak on one side. The difference between the first and the second peak is about 16 μs, so again half a cycle of the signal. The difference between the middle and the right peak is again about 60 μs. A time difference of 60 μs means a 0.09 m difference in distance[2].
As there are for both waveforms gaps occurring in all the distributions, that are particular to the waveform. These bigger gaps must be dependent on the frequency of the acoustic signal.

[2]calculated with the script "timeDist.py" [14]

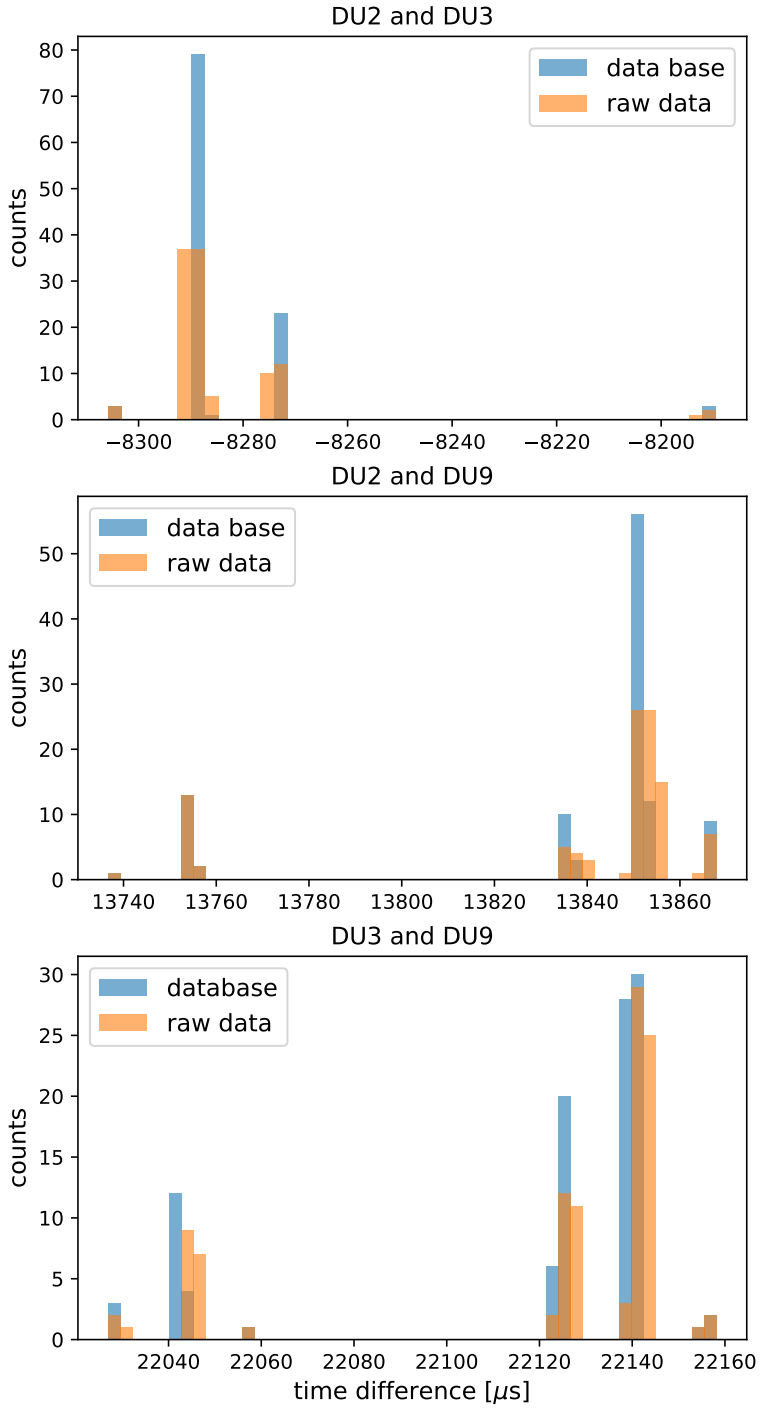


Figure 8: Time difference between the DUs for run 8882 and waveform 14. Shown in blue are the time differences from the database data, in orange the time differences from the raw data analysis.

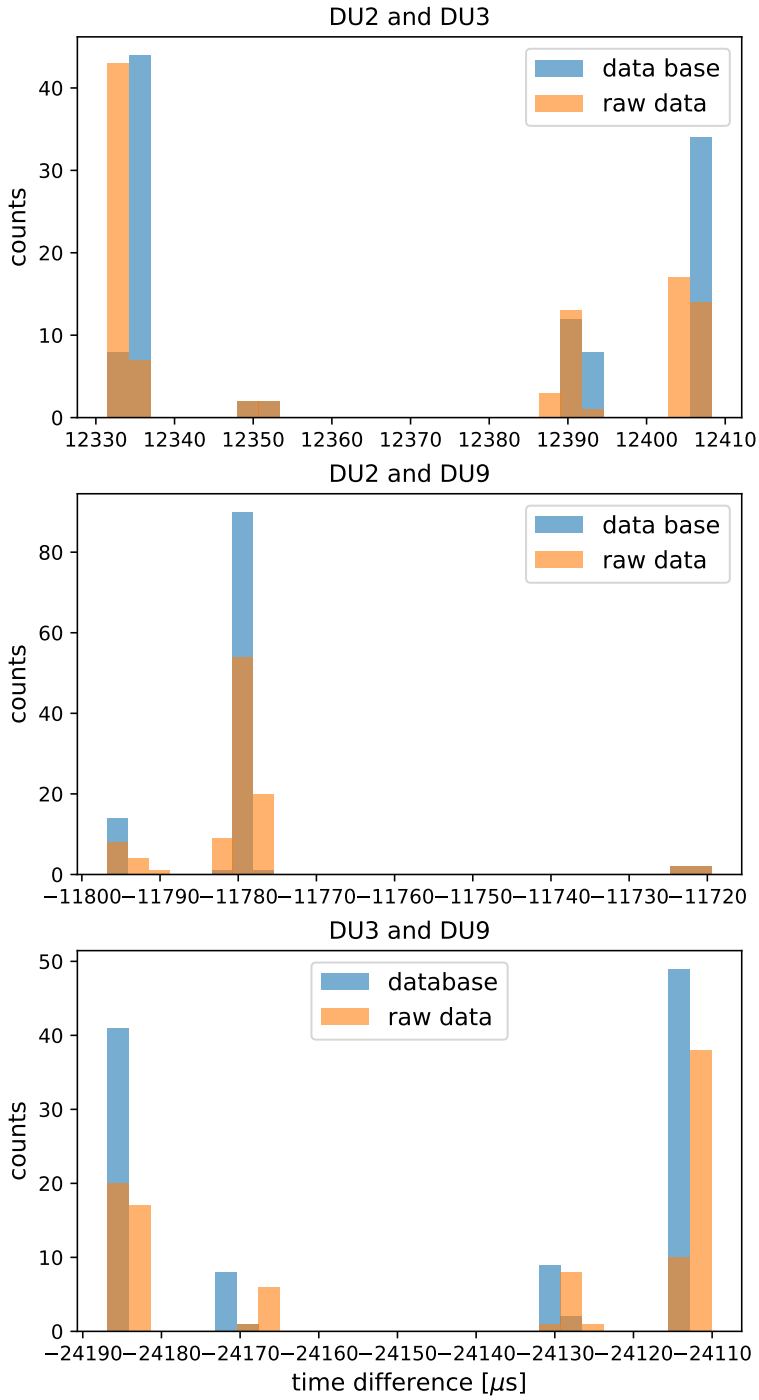


Figure 9: Time difference between the DUs for run 8882 and waveform 16. Shown in blue are the time differences from the database data, in orange the time differences from the raw data analysis.

4.4.2 Run 10477

For run 10477 results of the time differences are also plotted as histograms. Figure 10 shows the distributions for waveform 14 and Figure 11 the distributions for waveform 16. The distributions obtained from the database data are again very similar to the distributions from the raw data. Small deviations of a bin occur, but no larger deviations. The width of one bin is again half the length of one raw data sample. As run 10477 is way shorter than run 8882, the histograms consist of less data points. This leads to fewer bars with entries and fewer entries in those bars. The distributions for waveform 14 do not show the three peak pattern as seen in run 8882.
By comparing the distributions from run 8882 and 10477, one also recognises, that the histograms change for both waveforms. The times with peaks change and also the height ratio of the bars differ. The change of height ratio is caused by the small number of data points of run 10477. This also explains the fewer peaks in the histograms. But for the distribution for the time difference between DU2 and DU9 and between DU3 and DU9 the values of the time differences also change. This can also be seen at the mean values of the time differences given in Table 3 for the database data and in Table 4 for the raw data.

waveform 14

	mean in μs	std in μs	spread in μs
DU2 and DU3	-8281.37	53.52	113.01
DU2 and DU9	13823.99	38.47	98.71
DU3 and DU9	22105.36	44.69	112.77

waveform 16

	mean in μs	std in μs	spread in μs
DU2 and DU3	12369.50	24.32	70.10
DU2 and DU9	-11761.95	22.52	56.98
DU3 and DU9	-24131.44	6.41	17.17

Table 3: Time differences for run 10477 with data from the KM3NeT database.

waveform 14			
	mean in μs	std in μs	spread in μs
DU2 and DU3	-8287.03	54.59	118.02
DU2 and DU9	13816.33	30.22	97.27
DU3 and DU9	22105.88	45.91	116.59
waveform 16			
	mean in μs	std in μs	spread in μs
DU2 and DU3	12368.11	25.53	73.43
DU2 and DU9	-11761.67	22.53	56.51
DU3 and DU9	-24129.77	6.57	16.93

Table 4: Time differences for run 10422 with times obtained via cross-correlation from raw data.

Calculating the difference between the mean values of the two runs gives following results:

	waveform 14	waveform 16
DU2 and DU3	2.87	-2.70
DU2 and DU9	20.14	-16.64
DU3 and DU9	13.85	-13.74

The mean values for the time difference between DU2 and DU3 is less than 3 μs and therefore smaller than one data sample with 5.12 μs. This can be considered as noise and no significant difference. But for the other two time difference distributions, the difference between the two runs is between 13.9 μs and 20.1 μs. This between two and four data samples and is more than noise. A deviation of 20 μs equals 3.1 cm deviation of the position[3].
The standard deviations also differ to run 8882, but this can be expected, as there are way less data points.

[3]calculated with the script "timeDist.py" [14]

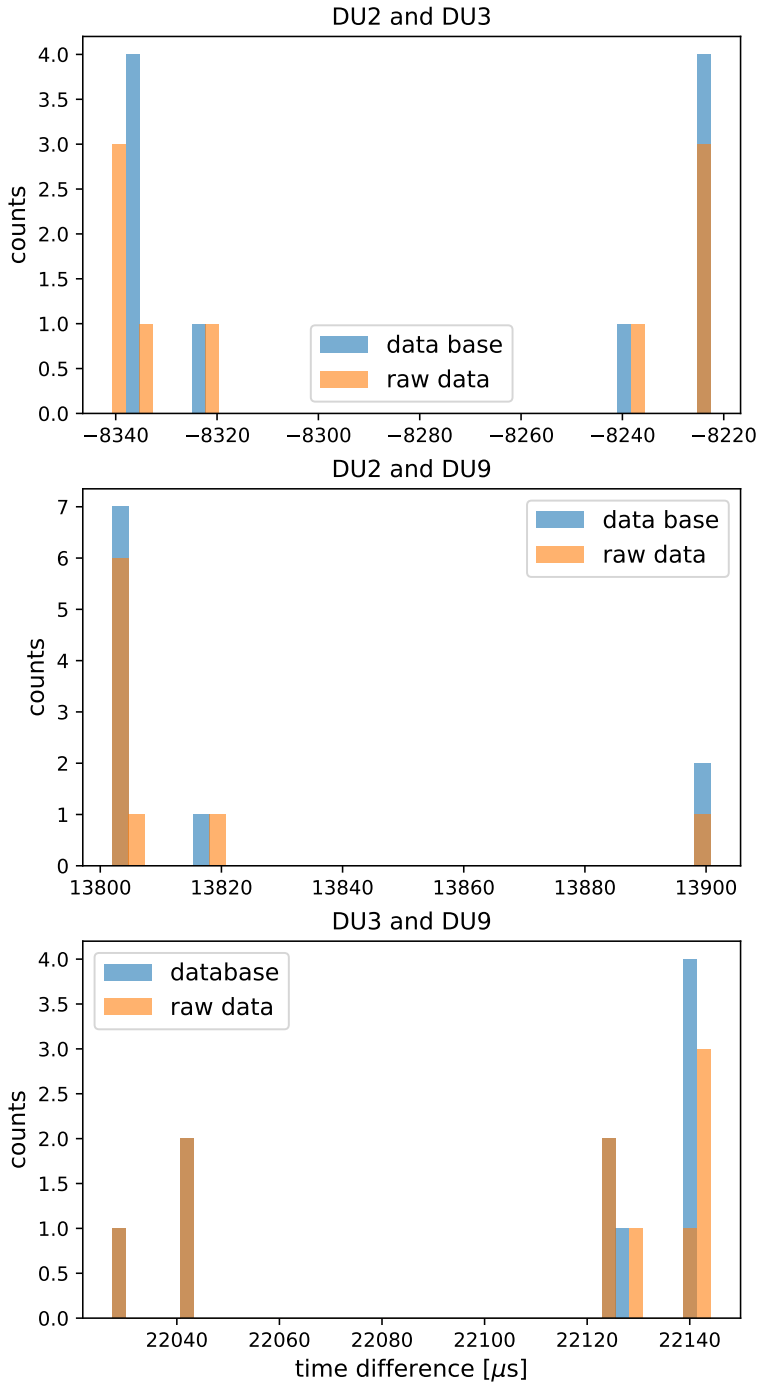


Figure 10: Time difference between the DUs for run 10477 and waveform 14. Shown in blue are the time differences from the database data, in orange the time differences from the raw data analysis.

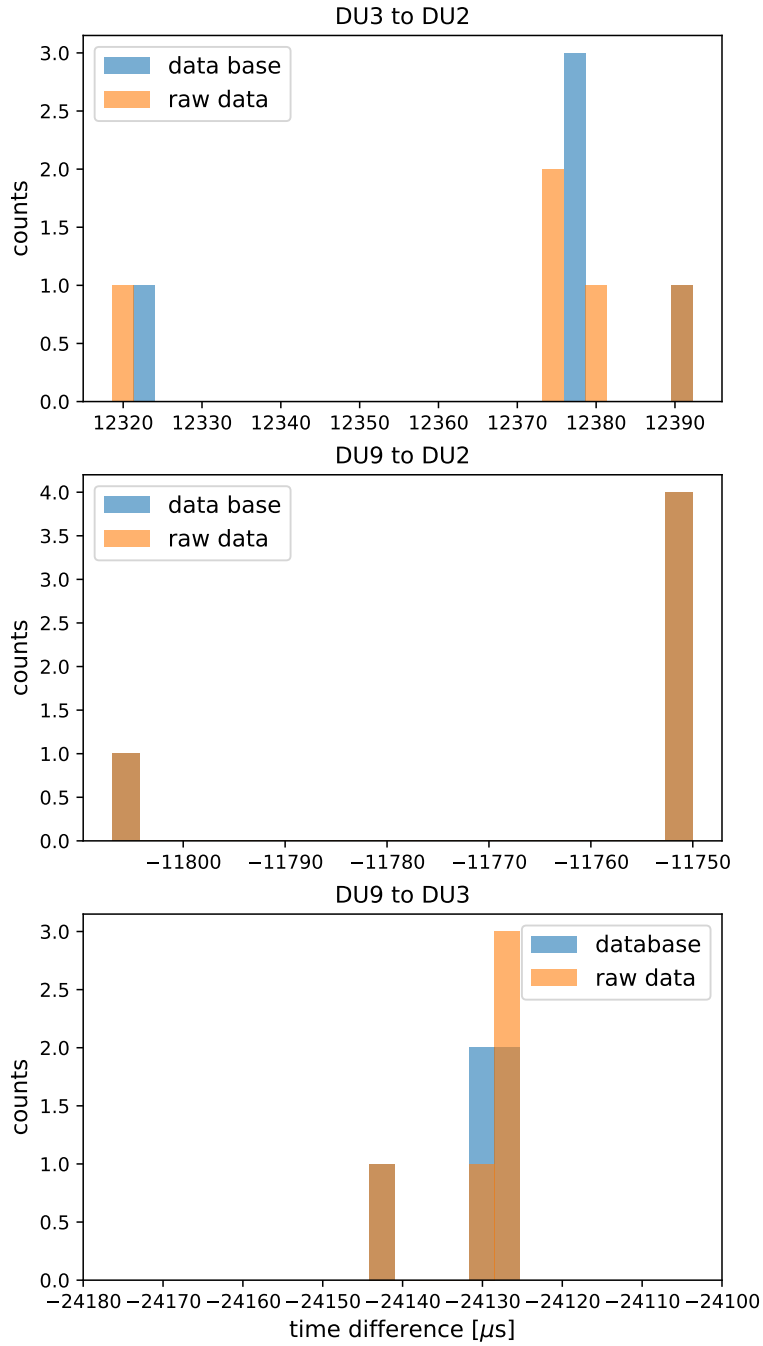


Figure 11: Time difference between the DUs for run 10477 and waveform 16. Shown in blue are the time differences from the database data, in orange the time differences from the raw data analysis.

4.5 Reasons for deviations of the time of arrival

To investigate why the distributions for the time differences, differ so much, a closer look at the result of the cross-correlation is needed. Shown in Figure 12 is an example of a from the hydrophone at DU2 detected signal from a raw data file of run 8882. The upper plot shows the beginning of the detected signal in blue and in orange the reference signal. While the lower plot shows the end of the signal. The reference signal starts at sample with the highest absolute q-factor from the cross-correlation. But for this data file, this is not the start of the signal, but already the fifth cycle. With a frequency of 30 kHz at the beginning of waveform 14, five cycles equal about 167 μs. This is even more, than the maximal deviation of the time differences calculated before. By looking at several signals with obtained time of arrival, multiple to late times, but only very few to early times were found. Further the amplitude of the recorded signal is comparatively small at the beginning caused by transient oscillation. This leads to the emitted signal not looking exactly as the reference signal and also makes it difficult to determine the start of the signal.

Shown in Figure 13 is a close-up off the upper part of the middle peak of the cross-correlation shown in Figure 6. Apart from the cross-correlation as obtained from the raw data, also the inverse of the cross-correlation is plotted. This means, both together show the behaviour of the absolute values of the cross-correlation, which are used to find the maximal correlation. The maximal correlation is also shown in the plot as imax. When regarding the envelope of the cross-correlation, it does not rise and fall monotonic, but oscillates. This leads to the maximum not always being in the middle of the peak, but depending on the oscillation of the envelope. There can either be a maxima of the oscillation in the middle of the peak, or a minima of the oscillation. Both cases are shown in Figure 6. The maximal q-factor is most likely a maxima of the oscillation and therefore not always at the middle of the correlation peak, which leads to a deviation of the time of arrival. The cycle of such an oscillation is about 20 correlation shifts. As one shift equals one sample of 5.12 μs, two peaks of the oscillation are over 100 μs apart. This is is also about the time difference recognised in subsubsection 4.4.1 for the difference between the highest peak and the peaks the furthest apart.

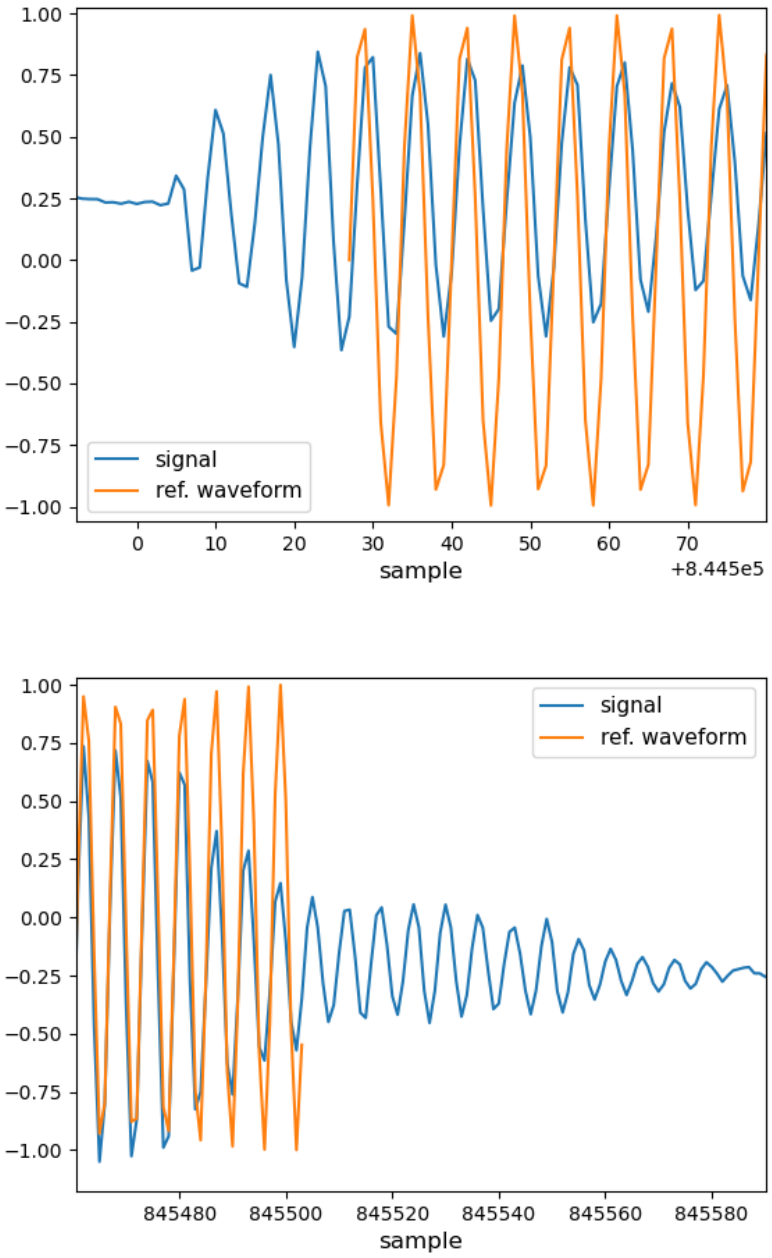


Figure 12: Signal from beacon 14 with reference signal starting at the time of arrival obtained from cross-correlation. The upper plot shows the beginning and the lower plot the end of the chirp. This signal is from raw data file DOM_808981515_CH1_1604927507.bin from run 8882. The signal and waveform are normalized to one.

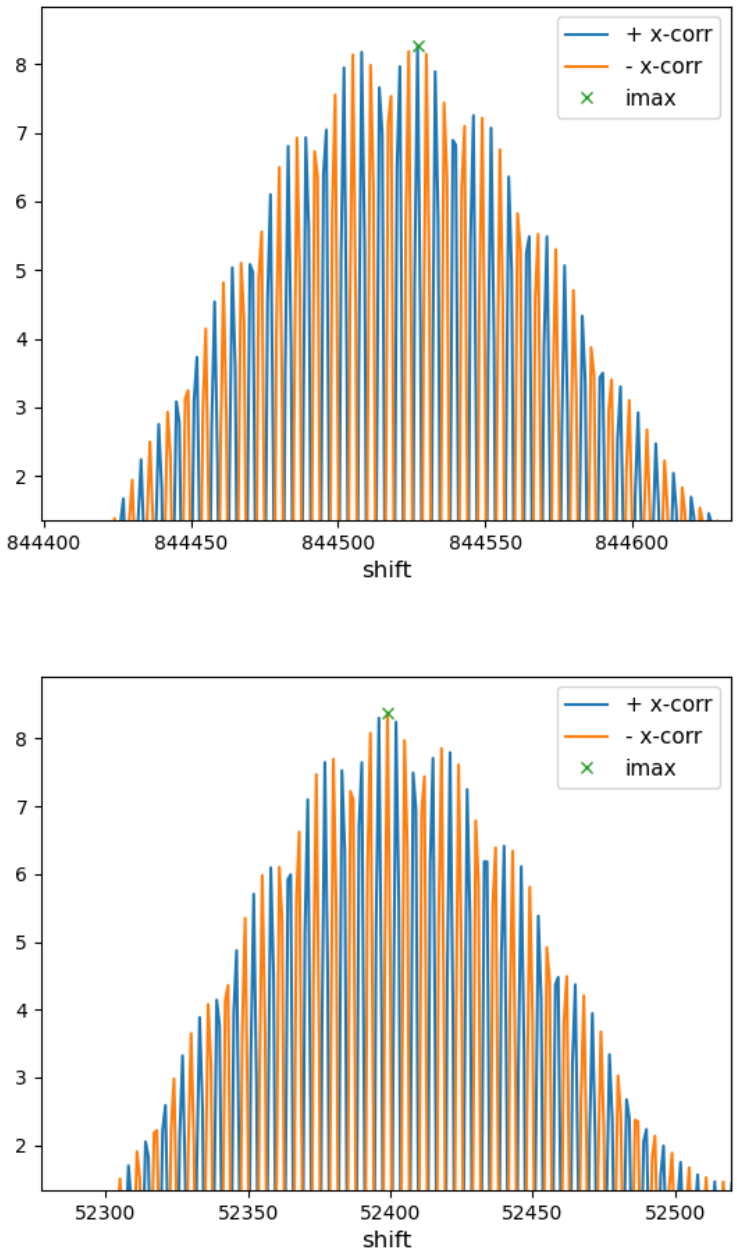


Figure 13: Closeup at the top of middle peak cross-correlation of the signal in the raw data file DOM_808981515_CH1_1604927507.bin (upper plot) and file DOM_808981515_CH1_1604927501.bin (lower plot) from run 8882 and waveform 14. Plotted in blue is the cross-correlation as calculated, in orange minus the cross-correlation and imax is the shift for the maximal correlation.

The result of the cross-correlation with waveform 16, also has a oscillating envelope as one can see in Figure 14. Here the frequency of the oscillation is higher than for waveform 14 which leads to a smaller deviation for the time of arrivals. For this waveform, the cycle of one oscillation is between 13 and 14 shifts, this equals about 70 μs. The maximal difference between the time differences for waveform 16 is between 73 μs and 77 μs as given in Table 1 and Table 2. So this spread also can be explained with this oscillation.

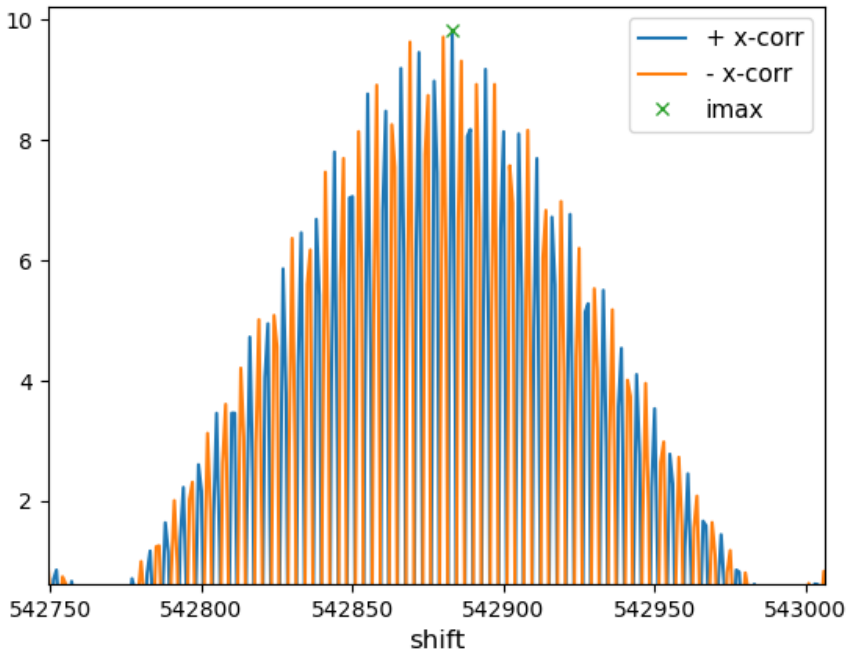


Figure 14: Closeup at the top of middle peak cross-correlation of the signal in the raw data file DOM_816978571_CH1_1604928379.bin from run 8882 and waveform 16. Plotted in blue is the cross-correlation as calculated, in orange minus the cross-correlation and imax is the shift for the maximal correlation.

The oscillation of the envelope is caused by the low sampling rate of the signal and reference signal. The signal and reference signal get scanned every 5.12 μs and the for waveform 14 with a frequency of 30-32 kHz, the length of one period is between

31.25 μs and 33.33 μs. This is not a whole multiple of 5.12 μs, therefore the signal does not get scanned at the same position every cycle, as one can see in Figure 15 for an oscillation with a frequency of 36 kHz. This leads to a beat of the signals' amplitude, which than also occurs at the result of the cross-correlation. The same is true for waveform 16 with a frequency of 34-36 kHz, which is also never a whole multiple of 5.12 μs.

The position of the maxima of the envelope depends on the sampling points of the signal and the reference signal. If both signal get sampled at the same points, the envelope has a maxima at the middle of the correlation peak. Sampling at different points leads to a shift of the maxima away from the middle of the correlation peak and therefore a deviation of the time of arrival. For the 11 chirps, emitted at a short distance of time, two emitted chirps are 5 s apart. Dividing 5 s by the length of one sample gives 5 s/5.12 μs = 976562.5. Therefore every second chirp gets sampled at the same points. This means if for one chirp, there is a maxima in the middle, for the next one it is a minima. This can be seen in Figure 13, where the result of the cross-correlation for two consecutive chirps is shown. For a correlation peak with a minima in the middle, the maximum of the correlation can be either on the maxima before the minima or after the minima. This leads either to a to early or to late time of arrival. At run 8882 every 10 min 11 chirps are emitted. Dividing 10 min by the length of one sample gives 600 s/5.12 μs = 117187500, which is a whole number. Therefore the chirps get sampled at the same points as 10 min before. This leads to the observed structure. But there is a shift of the mean time differences between run 8882 and run 10477. This can be explained by the clocks of the beacons not being synchronised with the sampling at hydrophones. Therefore the sampling points of the signal change slightly over time, which leads to a shift of the position of the maximas, as the sampling points of the reference signal stay the same. [11]

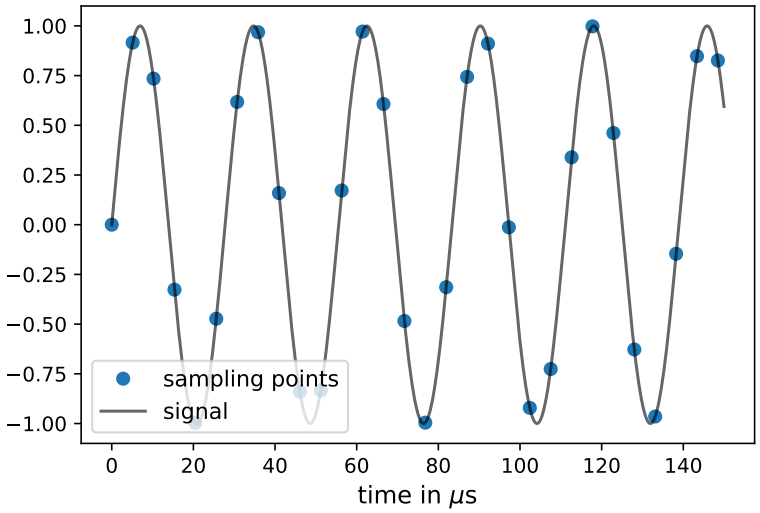


Figure 15: A 36 kHz oscillation with sampling points being 5.12 μs apart, which equals a sampling frequency of $f_s = 195312.5$ Hz.

5 Upsampling

5.1 Program

For reaching a better result at the cross-correlation, upsampling of the signal and the reference signal is tried. The upsampling process is done via the scipy function "scipy.signal.resample". This function uses Fourier transform to resample the input array in a array with a given length [5]. With the higher sampling frequency, the sampling points differ only marginal for each cycle of the oscillation, as one can see in Figure 16 for a signal with a frequency of 36 kHz and a ten times higher sampling frequency than original.

As upsampling by a factor x, returns x times the original amount of data, the cross-correlation also has to be calculated for a x times larger amount of data. This lead to the program for the correlation taking multiple hours for all files from run 8882. Therefore the raw data files are limited to data files, where a signal was found at the cross-correlation without upsampling, for the cross-correlation with upsampling. One other change to the cross-correlation program without

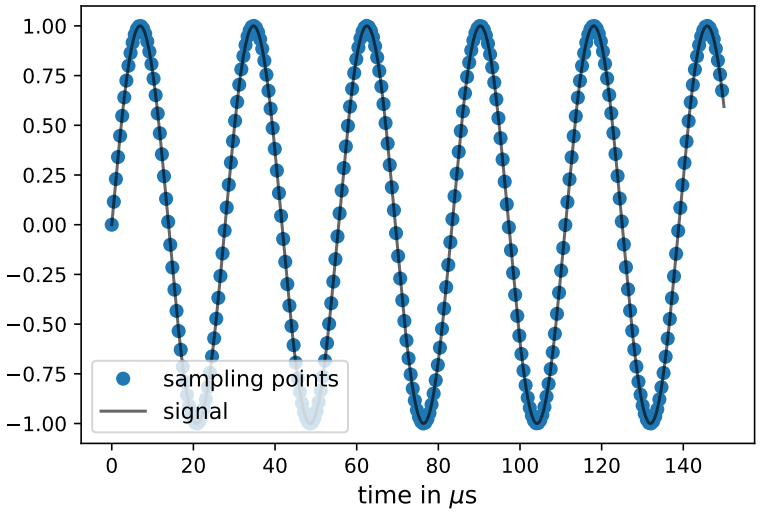


Figure 16: A 36 kHz oscillation with sampling points being 0.512 μs apart, which equals a sampling frequency being 10 times higher than the original frequency.

upsampling, is the calculation of the time of arrival from the index of the maximal correlation. Instead of upsampling the timearray, the time of arrivals is calculated from the original timearray. There are time intervals in the raw data files, where no data was collected, this leads to jumps in the timearray. Therefore following algorithm was used for calculating the time of arrival:

$$toa = \text{timearray}[b] + \frac{1}{f_s} \cdot c \tag{5}$$

with

$$\begin{aligned} a &= \frac{i_{max}}{x} \\ b &= \lfloor a \rfloor \\ c &= a - b\,. \end{aligned} \tag{6}$$

Where f_s is the original sampling frequency and i_{max} the index of the maximal correlation. $\lfloor a \rfloor$ returns the floor of a, which is the largest integer smaller or equal to a. The time of arrivals with corresponding q-factors are again saved as arrays.

5.2 Results

The from the cross-correlation obtained time of arrivals with q-factors are analysed the same way as the data without upsampling. For the upsampled data, the q-factors increase by the upsampling factor x, as seen in Figure 17 for upsampling factor $x = 10$. This plot is for run 8882 and waveform 14 and shows the absolute value of the q-factors plotted over the time of arrivals.

Taking a look at the plotted cross-correlations result at the upper plot of Figure 18, the curve around the maximum looks smoother, than without upsampling, shown in Figure 6 for the same raw data file. The second plot of Figure 18 shows the top of the middle peak of the plot above. This plot has again plotted the negative cross-correlation in addition to the result of the cross-correlation. After the upsampling, no more oscillation of the envelope occurs. Upsampling the signals, means scanning the signal more often. Therefore the amplitude of the signals stays constant. Without the oscillating envelop, a smaller variation of the time differences can be expected.
Figure 19 shows the same signal as Figure 12, but after the upsampling. The signal and the reference signal are smother than before. The reference signal is again at the point of maximal correlation. This point is after the upsampling one cycle earlier, but still not at the real start of the signal. This happens not only for this signal, but is typical for detection unit 2. Therefore there is a systematic offset for the time differences. This is most likely caused by the amplitude of the chirp being lower for the first few cycles than the rest.

5.2.1 Run 8882 with upsampling by factor $x = 10$

Comparing the time differences between the hydrophones leads to distributions with a lower spread than without upsampling. Shown in Figure 20 and Figure 21 are the time difference distributions for waveform 14 and 16 from run 8882 with and without upsampling. By comparing the distributions obtained with upsampling to the distributions from the raw data without upsampling, one sees clearly, that the variation of the time differences is smaller for the upsampled data. The time

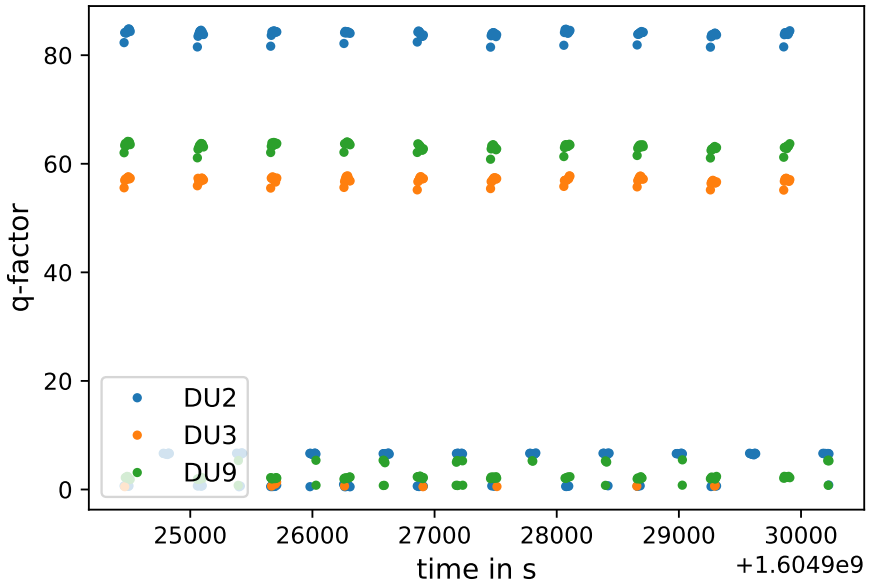


Figure 17: Absolute q-factors from the cross-correlation with upsampled data from run 8882 and waveform against the time of arrival. Upsampling factor $x = 10$.

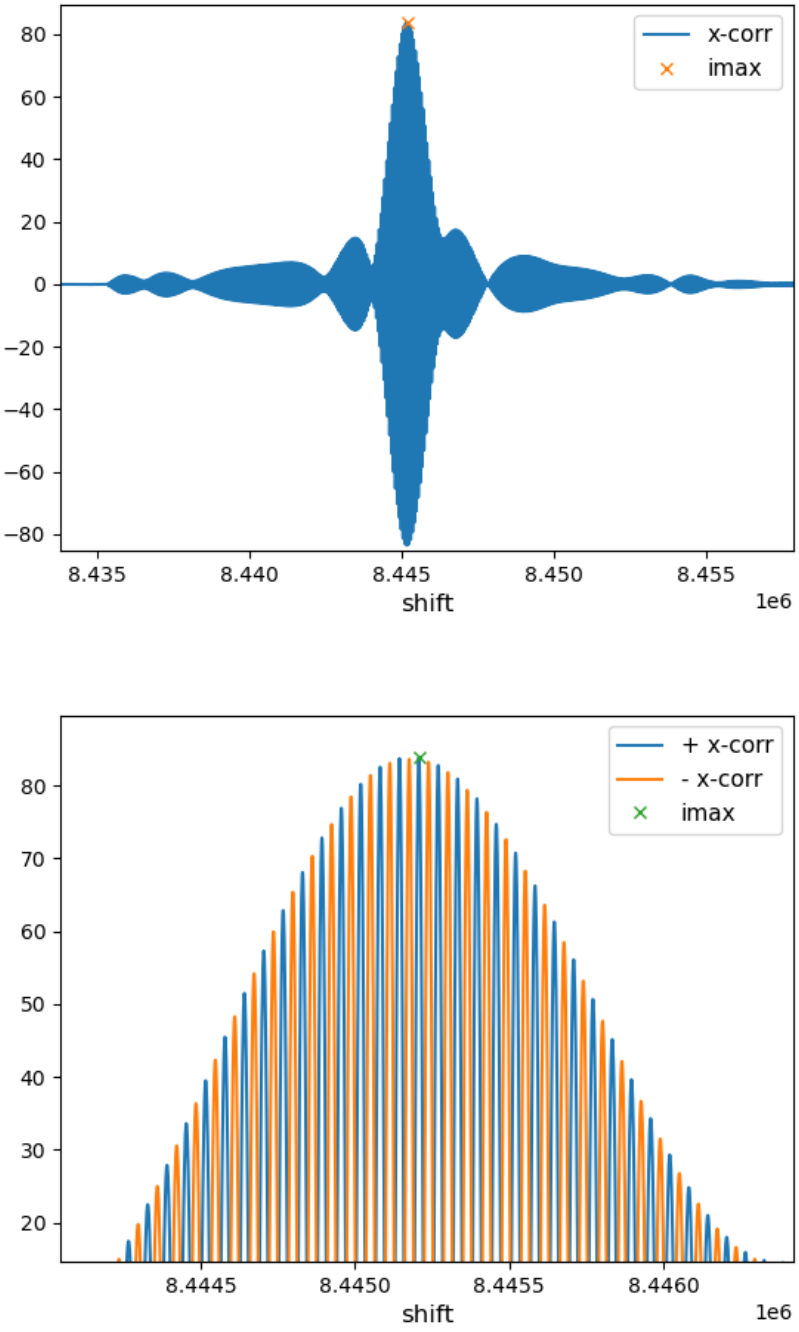


Figure 18: Results of the cross-correlation after upsampling the by factor $x = 10$ for the raw data file DOM_808981515_CH1_1604927507.bin from run 8882 and waveform 14 detected at DU2. Marked as imax is the maximum of the absolute correlation, and plotted at the lower plot is also the negative result of cross-correlation as -x-corr.

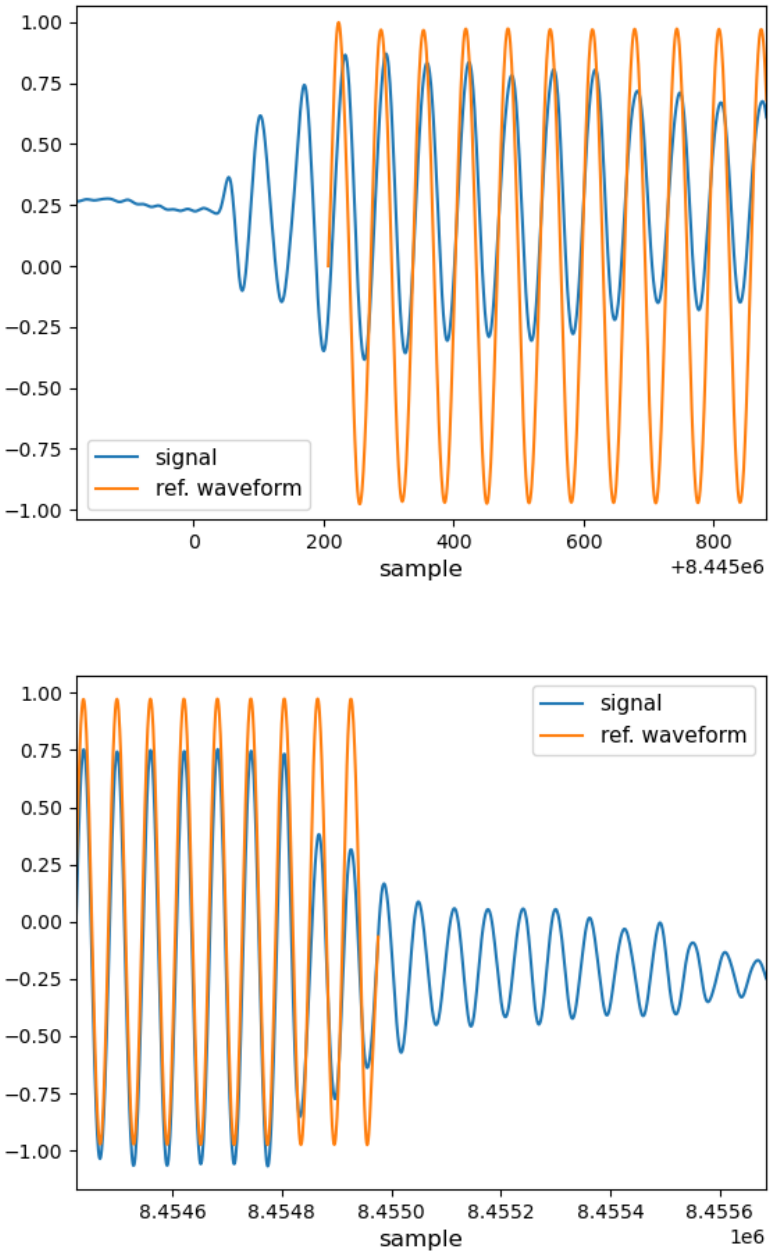


Figure 19: Signal from beacon 14 with reference signal, both upsampled by factor $x = 10$. The reference signal starts at the time of arrival obtained via cross-correlation. The signal is from raw data file DOM_808981515_CH1_1604927507.bin from run 8882. The signal and waveform are normalized to one.

differences, that differ largely from the mean do not occur with the upsampled data. Whereas the three peak pattern around the mean still appears, but at different time differences than for the raw data without upsampling. This also shows at the mean values and standard deviations of the distributions, shown in Table 5. The standard deviations go down to values between 6.7 μs and 11.0 μs, compared to values between 12 μs and 38 μs. In distance a deviation of 10 μs are 1.5 cm. This is consistent with a decrease of the maximal differences between the time differences. Only at the time difference between DU3 and DU9 for waveform 16 occur data points, which do not belong to the three peak pattern. At this distribution, there are two more bars, one at each side of the three peak pattern. But with the same distance to the next peak, as between the three peaks in the middle. Therefore this equals again half a cycle of the signal. Also this bars consist only of one or three entries. This the leads to a maximal spread of the time differences of 60 μs, which is slightly higher, than for the other distributions of this run with a spread between 31 μs and 37 μs. But the standard deviation is not higher, than for the other distributions.

waveform 14

	mean in μs	std in μs	spread in μs
DU2 and DU3	-8278.026	9.45	36.24
DU2 and DU9	13831.43	10.96	37.43
DU3 and DU9	22109.59	9.30	36.72

waveform 16

	mean in μs	std in μs	spread in μs
DU2 and DU3	-11776.60	8.41	32.42
DU2 and DU9	12360.86	6.73	31.47
DU3 and DU9	-24137.42	9.81	56.98

Table 5: Time differences for run 8882 obtained from raw data with upsampling by factor $x = 10$.

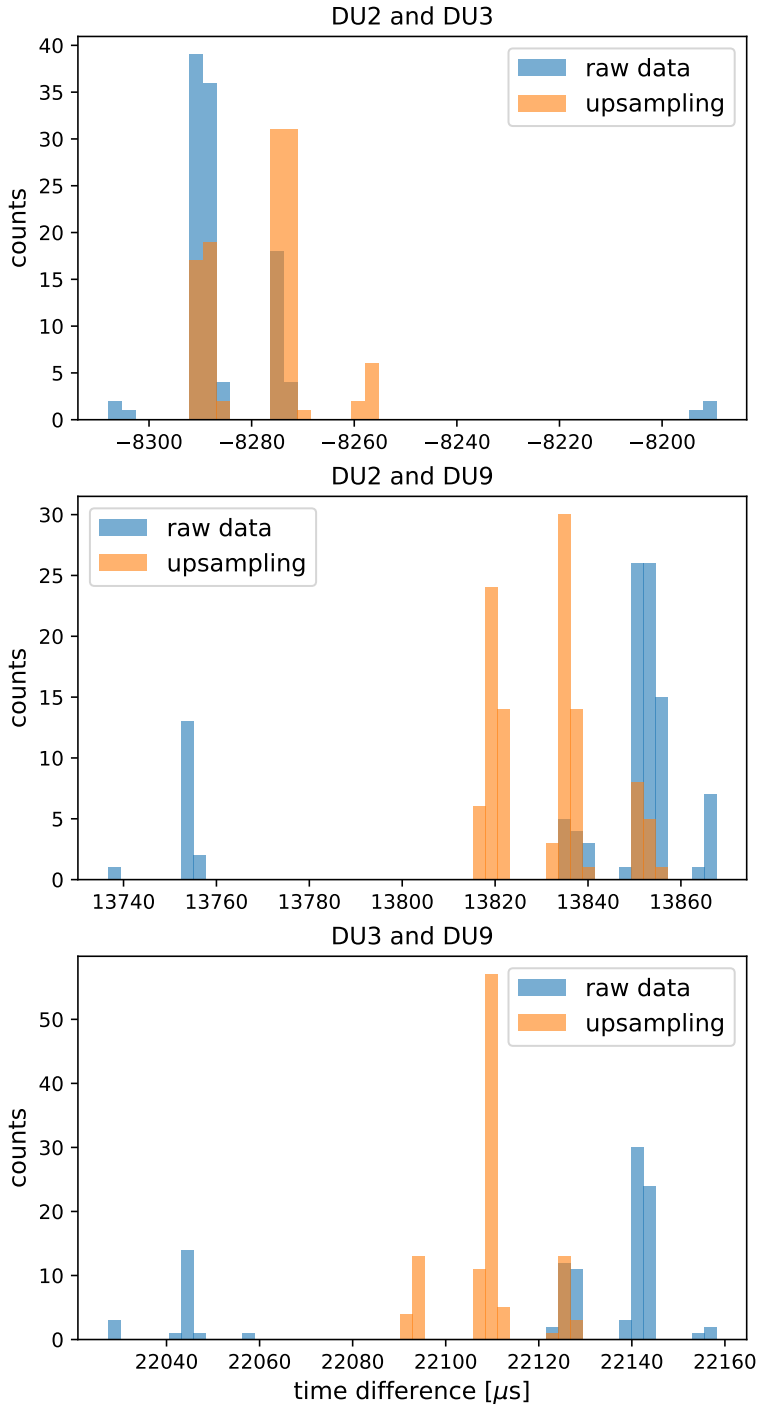


Figure 20: Time difference between the DUs for run 8882 and beacon 14. Shown in blue are the time differences from the raw data, in orange the time differences with upsampled data.

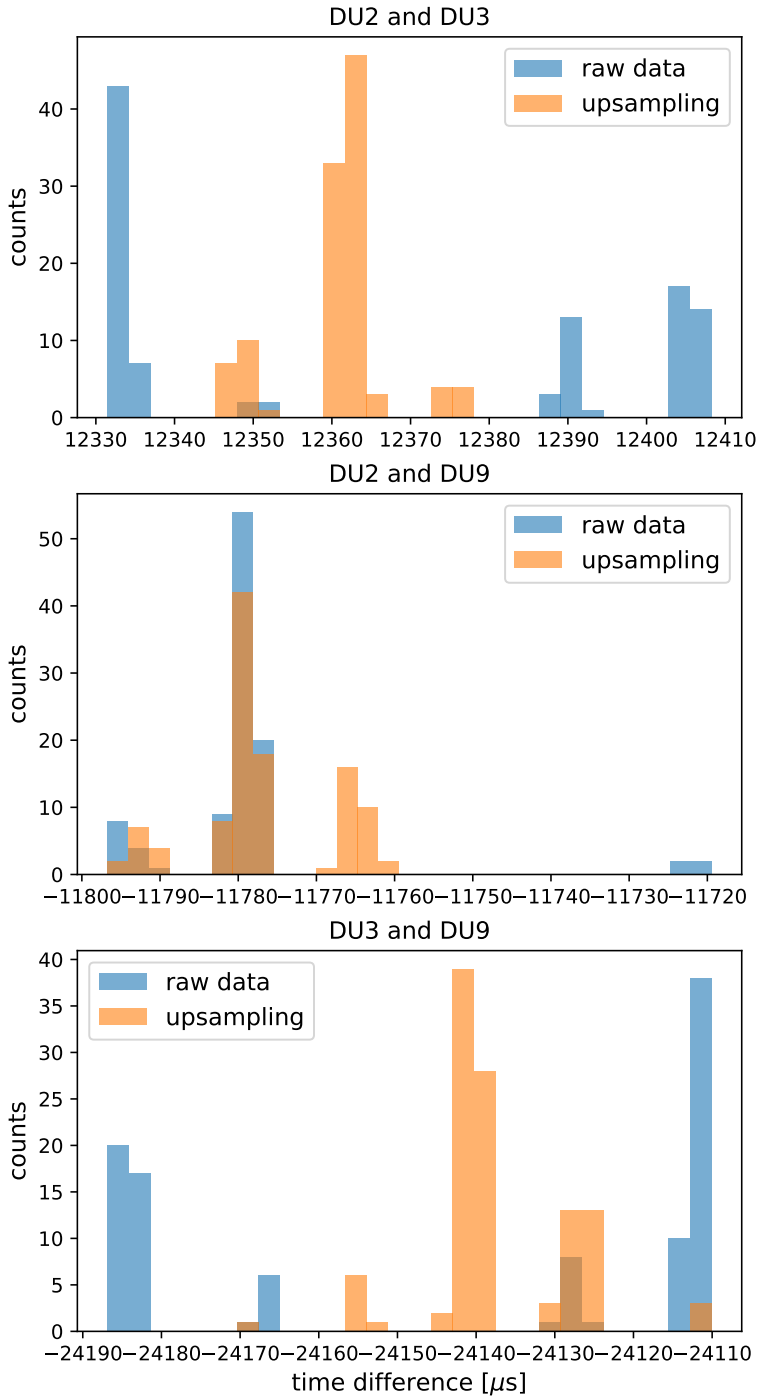


Figure 21: Time difference between the DUs for run 8882 and beacon 16. Shown in blue are the time differences from the raw data, in orange the time differences with upsampled data.

5.2.2 Run 10477 with upsampling by factor $x = 10$

For this run, the spread of the time differences distributions also get smaller. As one can see in Table 6 the maximal difference between time differences from this run goes down to 31 μs. The standard deviation also decreases and is for the upsampled data between 7 μs and 13 μs and is therefore similar to run 8882. The distributions are shown in Figure 22 for waveform 14 and Figure 23 for waveform 16. The three peak pattern can still not be seen for most distributions, but this is due to the small amount of signals. Apart from the lower spread, for this run there is also a shift of the mean time differences. But for the upsampled data, the mean values are nearly the same for both runs. The differences between the runs are:

	waveform 14	waveform 16
DU2 and DU3	-0.30 μs	-1.62 μs
DU2 and DU9	1.82 μs	2.04 μs
DU3 and DU9	2.25 μs	3.70 μs

for the upsampled data. This are values smaller than the length of one original sample with 5.12 μs. Therefore this can be considered as noise and there is no significant difference between the runs after the upsampling process.

waveform 14

	mean in μs	std in μs	spread in μs
DU2 and DU3	-8277.73	8.28	21.93
DU2 and DU9	13829.61	7.66	17.88
DU3 and DU9	22107.34	8.22	30.99

waveform 16

	mean in μs	std in μs	spread in μs
DU2 and DU3	-11778.64	12.74	28.85
DU2 and DU9	12362.48	8.81	27.18
DU3 and DU9	-24141.12	7.35	22.89

Table 6: Time differences for run 10477 obtained from raw data with upsampling by factor $x = 10$.

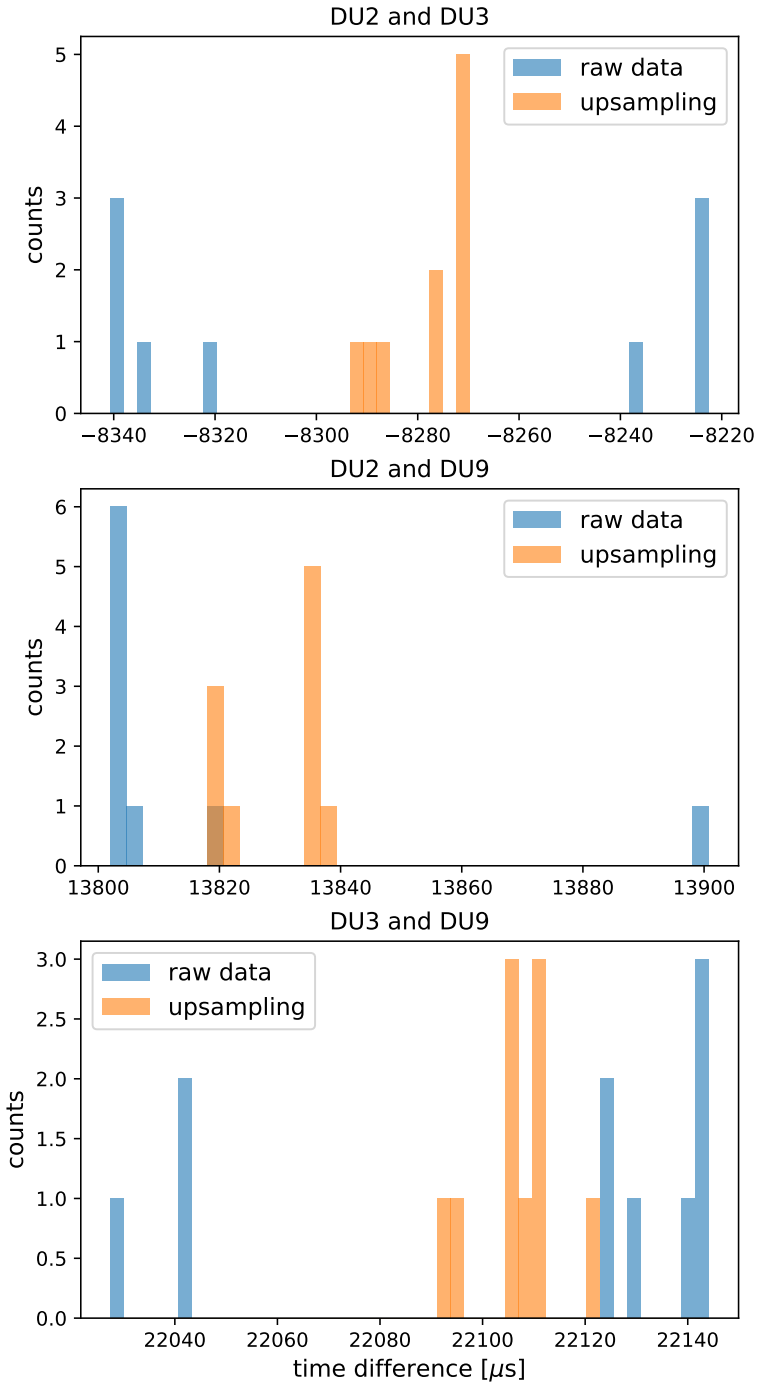


Figure 22: Time difference between the DUs for run 10477 and waveform 14. Shown in blue are the time differences from the raw data, in orange the time differences with upsampled data.

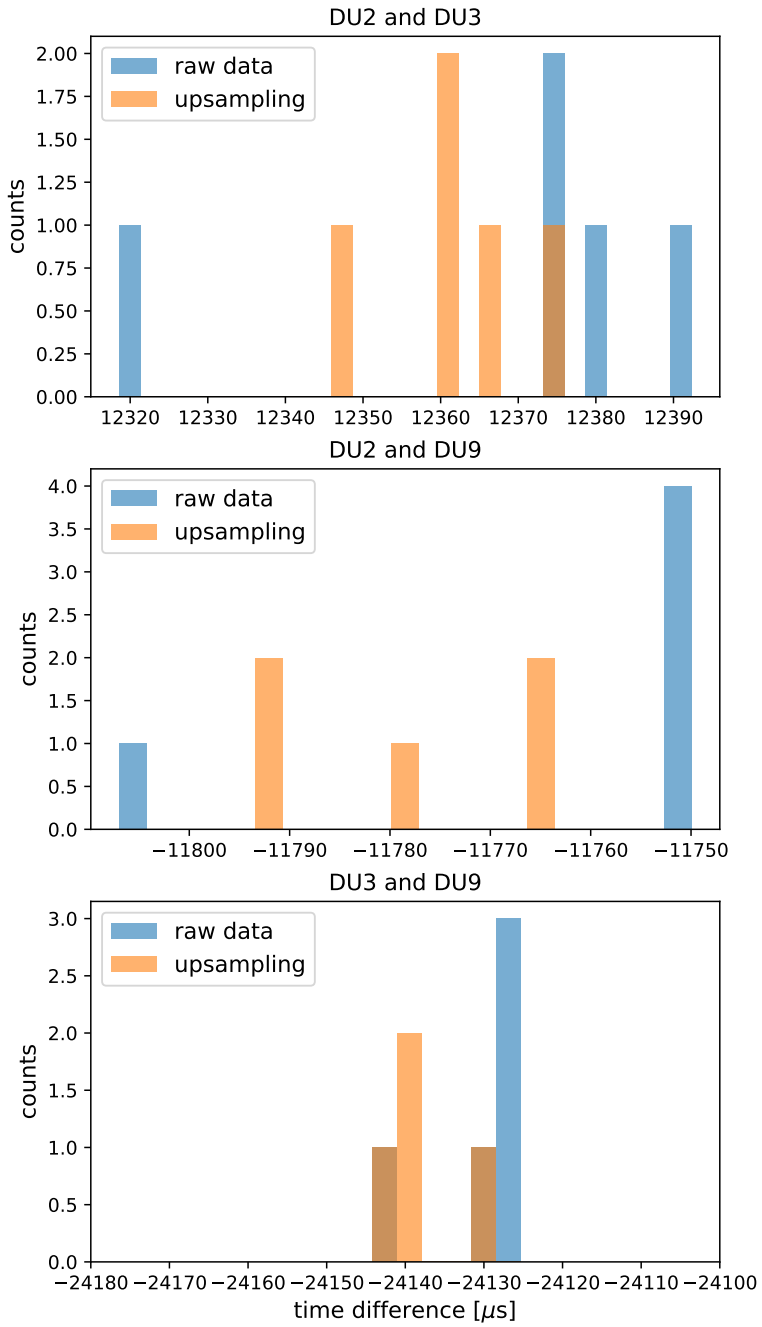


Figure 23: Time difference between the DUs for run 10477 and waveform 16. Shown in blue are the time differences from the raw data, in orange the time differences with upsampled data.

5.3 Comparison time of arrivals with and without upsampling

To investigate, how much the time of arrival of one chirp changes with upsampling, the differences between the times with and without upsampling are calculated and plotted. The result for both runs of waveform 14 can be seen in Figure 24 and those for waveform 16 in Figure 25. One bar in the histograms is again 2.56 μs wide. Noticeable is, that the time of arrivals change as expected by discrete values. Also not every time of arrival changes with upsampling.
For waveform 14 the difference between the peaks is 15.8 μs for both runs, which equals about half a cycle of the 30 kHz signal at the beginning of the waveform. Furthermore the maximal change is -64.6 μs and +64.6 μs, this equals two cycles of the signal.
The distance between the peaks for waveform 16 is 13.4 μs, which is smaller than for waveform 14, but the frequency at the beginning of waveform 16 is with 34 kHz higher. Therefore 13.4 μs is about half a cycle of waveform 16. The maximal change is with $\pm$42.9 μs, also smaller and equals only 1.5 cycles of waveform 16. Overall the time of arrival changes about similar often to higher and to lower values, with small differences between the DUs and waveforms. This is consistent, with the deviations of the time of arrival being caused by the oscillation of the envelope. Where a maxima of the envelope can be before or after the middle of the correlation peak.

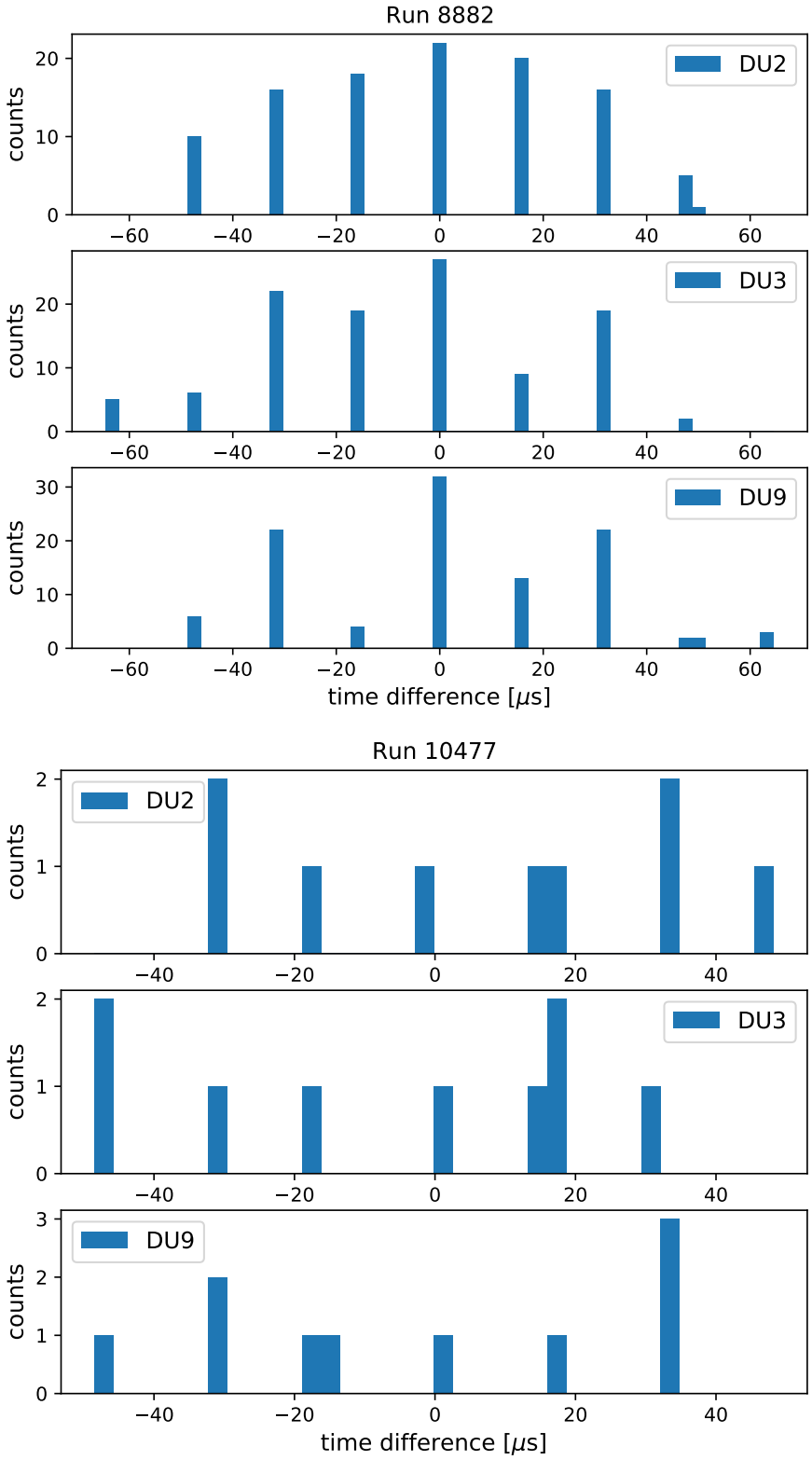


Figure 24: Difference between time of arrivals with and without upsampling for waveform 14.

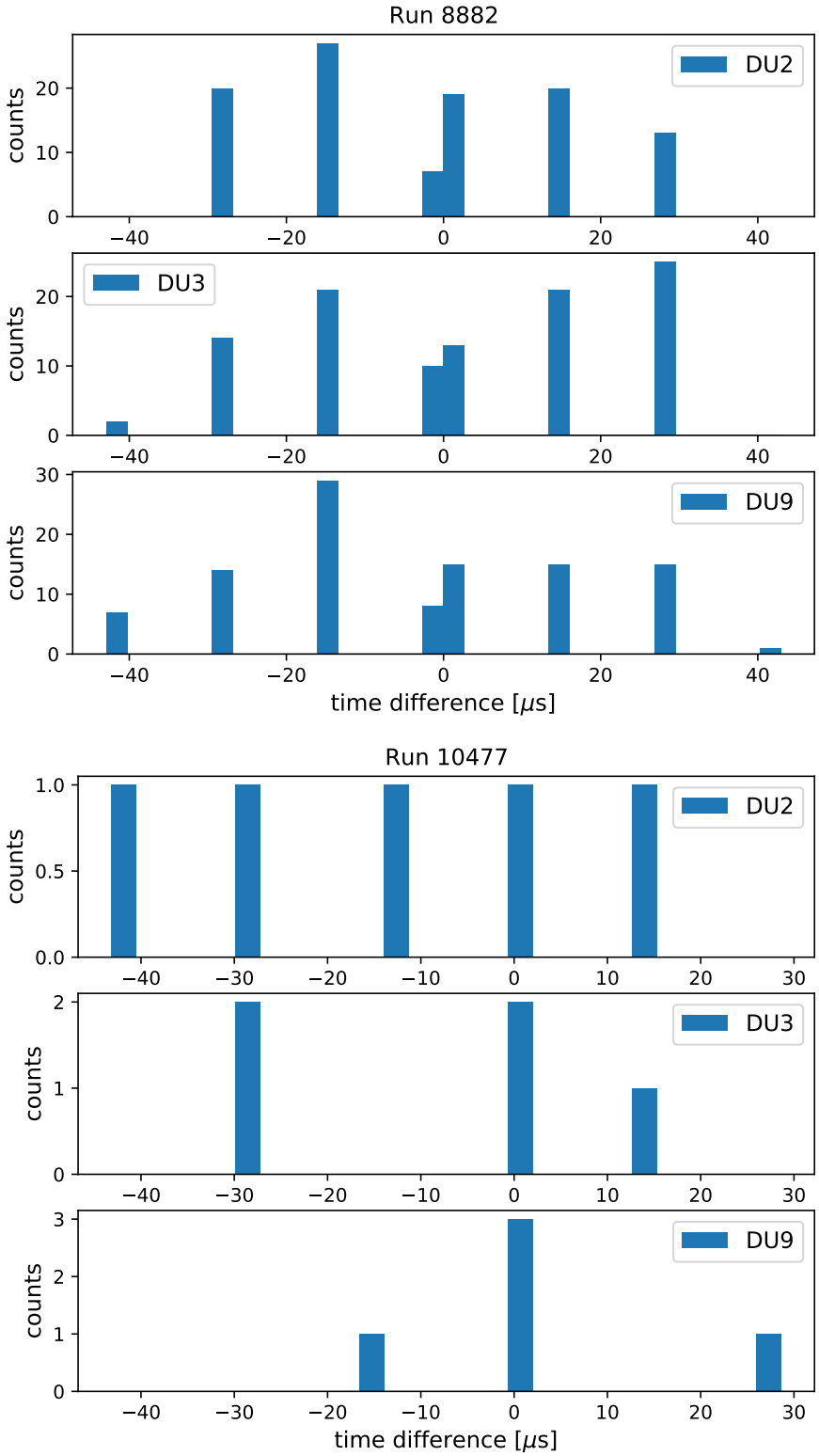


Figure 25: Difference between time of arrivals with and without upsampling for waveform 16.

5.4 Comparison of different upsampling factors

The mean and standard deviation of the time difference distributions change for different upsampling factors x. Figure 26 shows the mean time difference between the DUs for run 8882 and waveform 14 plotted over the upsampling factor. For waveform 16 the mean values are shown in Figure 27. For waveform 14, the mean values stay nearly the same for upsampling factor 7 and higher. For waveform 16, higher upsampling factors are necessary to reach a constant value. But again the highest changes are before $x = 10$.
The standard deviations, shown in Figure 28, behave similar. For waveform 14, nearly constant values are reached for $x \geq 7$. Where they vary more for waveform 16. The highest standard deviation for high upsampling values occur for the difference between DU3 and DU9 for waveform 16. For this difference, the standard deviation even increases again from $x = 10$ to $x = 15$. The standard deviations for the other two distributions of waveform 16 stay nearly constant for $x \geq 10$.
As increasing the upsampling factor, also increases the computing time and needed memory a upsampling factor as small as possible should be chosen. Considering the behaviour of the means and standard deviations, no upsampling factor above 10 is needed.

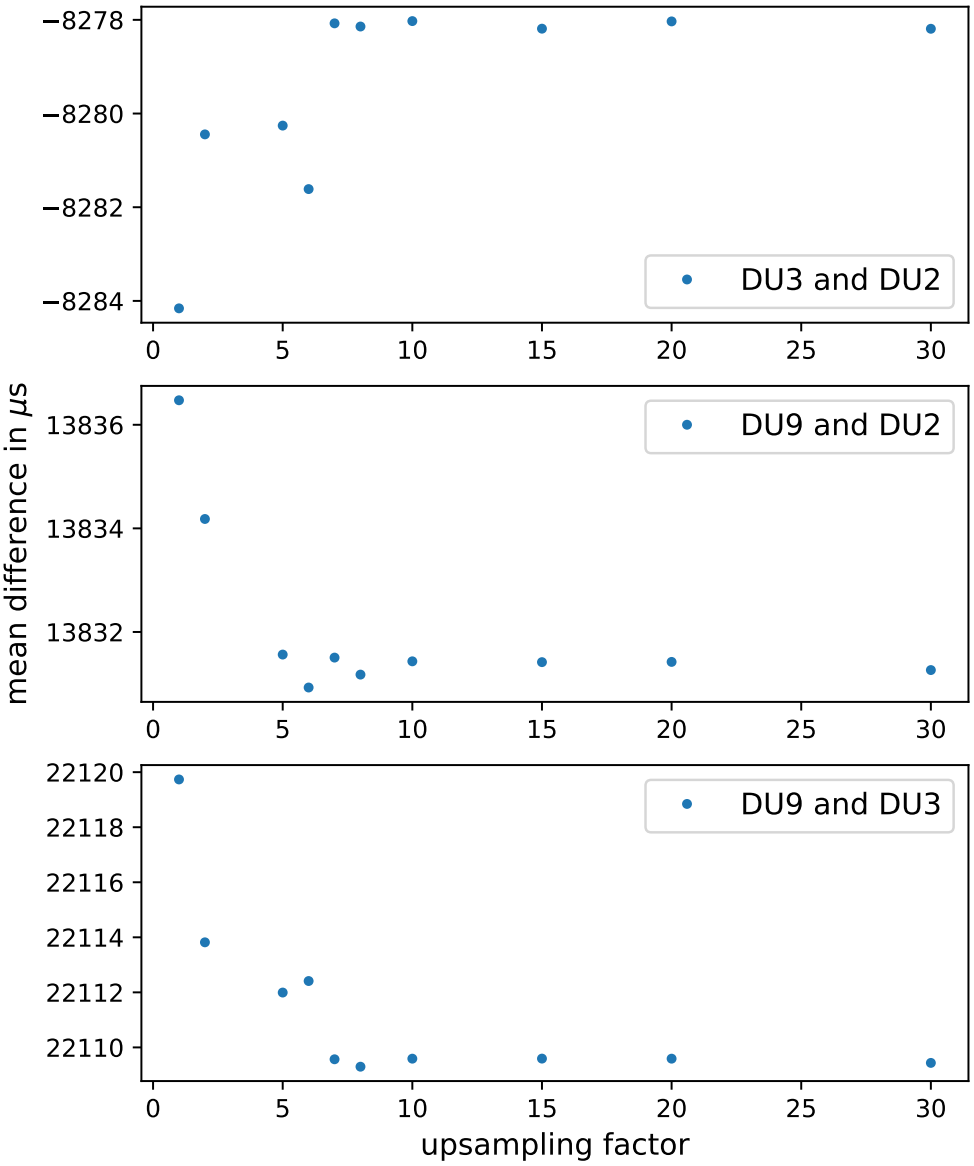


Figure 26: Mean time differences per upsampling factor between the DUs for run 8882 and beacon 14.

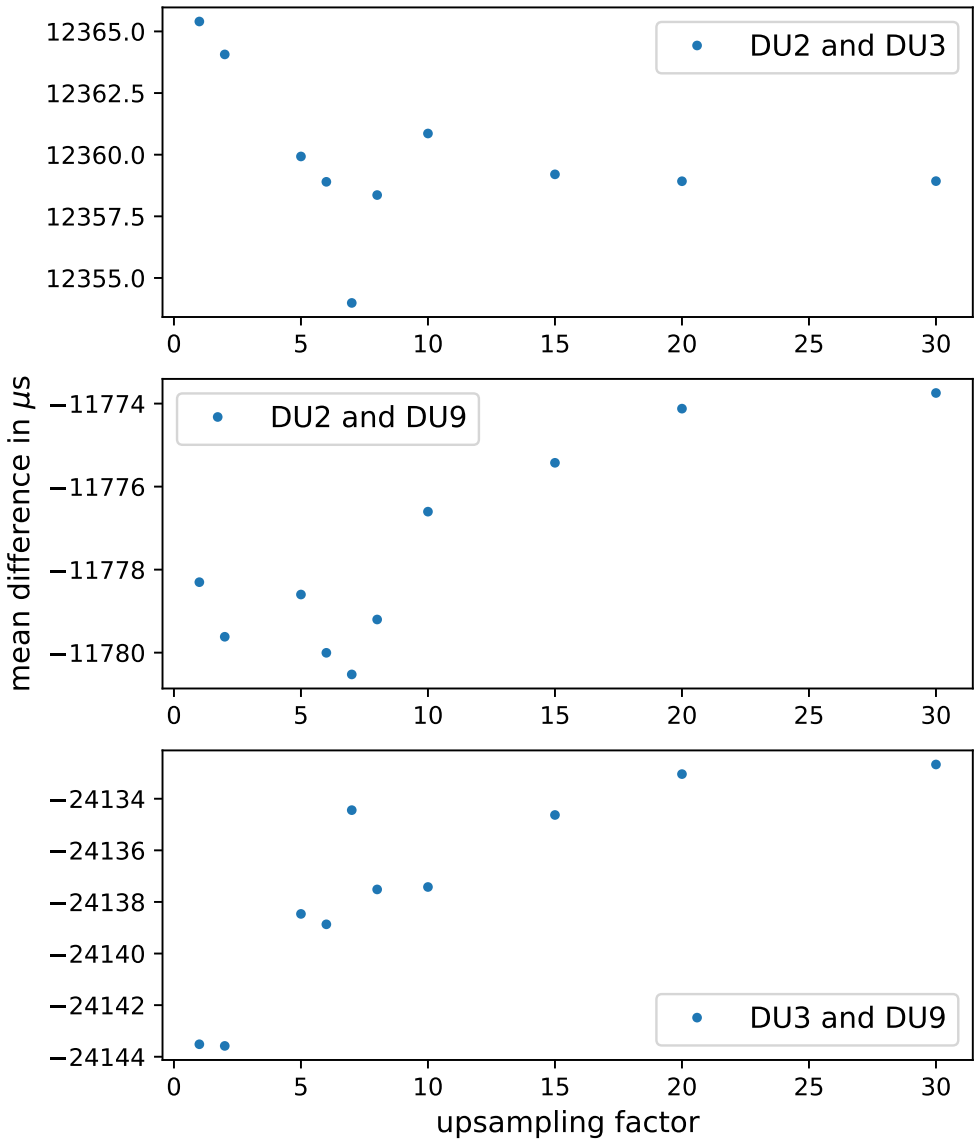


Figure 27: Mean time differences per upsampling factor between the DUs for run 8882 and beacon 16.

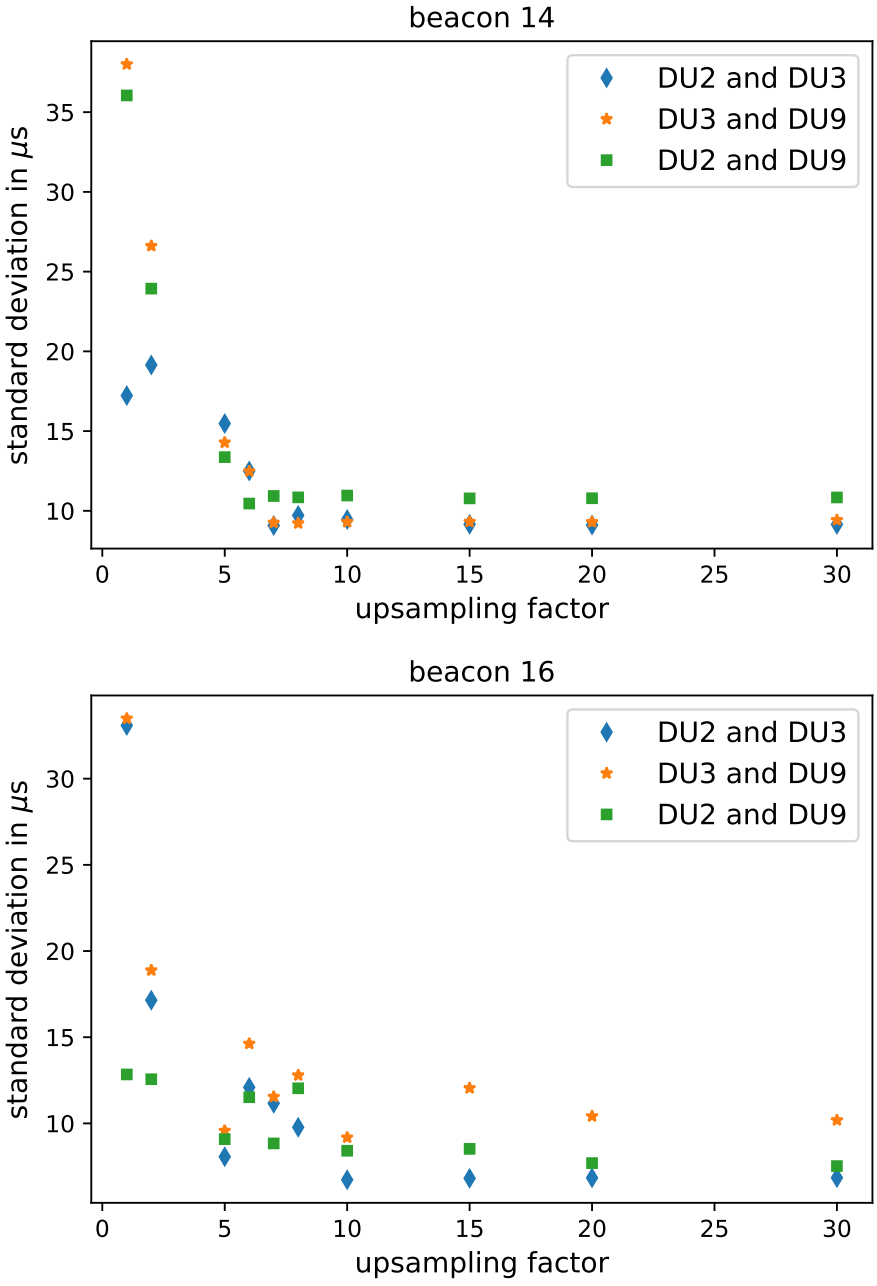


Figure 28: Standard deviation per upsampling factor for run 8882.

6 Conclusion

Overall the precision of the time of arrival determination for the acoustic signals can be improved with upsampling. Therefore also the statistical error of the DOMs positions is decreased. Decreasing the average standard deviation of the time difference distributions from 28.4 μs to 9.1 μs at run 8882 with upsampling by factor 10, decreases the error of the position from 44 mm down to 14 mm [4]. For run 10477 the standard deviation goes down from 31.7 μs to 8.8 μs by upsampling with $x = 10$. For the distance, this equals a decrease from 49 mm to 14 mm [4]. Apart from the statistical error, there is still the possibility of systematic errors occurring. Therefore further investigations may be needed. The time of arrivals at the hydrophones can be used to determine the position of the beacons and the hydrophones. Regarding the upsampling, it seems like factor $x = 10$ is the best compromise between result and computing time.

[4] calculated with script "timeDist.py"[14]

References

[1] S Adrián-Martínez et al. "Letter of intent for KM3NeT 2.0". In: *Journal of Physics G: Nuclear and Particle Physics* 43.8 (June 2016), p. 084001. ISSN: 1361-6471. DOI: 10.1088/0954-3899/43/8/084001. URL: http://dx.doi.org/10.1088/0954-3899/43/8/084001.

[2] Michel Ageron et al. "ANTARES: the first undersea neutrino telescope". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (2011), pp. 11–38.

[3] The KM3NeT Collaboration. *KM3NeT*. URL: https://www.km3net.org (visited on 07/09/2021).

[4] The SciPy community. *scipy.signal.correlate*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html (visited on 07/20/2021).

[5] The SciPy community. *scipy.signal.resample*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.resample.html (visited on 07/22/2021).

[6] R. Coniglione. *High-energy neutrino astronomy with KM3NeT-ARCA*. 2017. arXiv: 1701.05849 [astro-ph.IM].

[7] M. deJong. *ELOGKM3NET: Position Calibration, Time-of-arrival data*. Sept. 23, 2020. URL: https://elog.km3net.de/Calibration/298 (visited on 08/07/2021).

[8] Stefan Gerlach. "Daten- und Signalanalyse". In: *Computerphysik: Einführung, Beispiele und Anwendungen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, pp. 245–268. DOI: 10.1007/978-3-662-59246-5_18. URL: https://doi.org/10.1007/978-3-662-59246-5_18.

[9] *Internal KM3NeT document*. URL: https://git.km3net.de/calibration/input_tables# (visited on 08/02/2021).

[10] U.F. Katz and Ch. Spiering. "High-energy neutrino astrophysics: Status and perspectives". In: *Progress in Particle and Nuclear Physics* 67.3 (July 2012), pp. 651–704. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2011.12.001. URL: http://dx.doi.org/10.1016/j.ppnp.2011.12.001.

[11] Robert Lahmann. personal communications.

[12] Robert Lahmann. *Script rawDataDemo.py*. URL: https://git.km3net.de/calibration/positioning/-/tree/rl-positioning.

[13] Robert Lahmann. *Script readDataDemo.py*. URL: https://git.km3net.de/calibration/positioning/-/tree/rl-positioning.

[14] Robert Lahmann. *Script timeDist.py*. URL: https://git.km3net.de/calibration/positioning/-/tree/rl-positioning.

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich diese Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Die Arbeit stimmt weder vollständig noch in wesentlichen Teilen mit einer Arbeit überein, die bereits einer anderen Prüfungsbehörde vorgelegt wurde.

Erlangen, den 11.08.2021

Isabella Probst