Analyses of detection efficiency biases in the KM3NeT research infrastructure

THESIS

submitted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE
in
PHYSICS

Author :	Mike Wang
Student ID :	s2538334
Supervisor :	Dorothea Samtleben
2^{nd} corrector :	Tjerk Oosterkamp

Leiden, The Netherlands, November 9, 2023

Analyses of detection efficiency biases in the KM3NeT research infrastructure

Mike Wang

Huygens-Kamerlingh Onnes Laboratory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

November 9, 2023

Abstract

The KM3NeT neutrino telescope, located presently at two different sites in the Mediterranean Sea, consists of two neutrino detecors As the telescope is still being built, the calibration is fully underway. This research focuses on four potential biases present in the astrophysical focused part of the detector, ARCA. Muon light, both detected and reconstructed from simulations and ^{40}K decay being used to determine the size of those biases. It is found that the PMTs shadowed by the titanium collar on the DOM show a different bias dependent on what hemisphere they are located. Furthermore it is found that the different gel transperacy in the DOMs causes different efficiencies, and that the new PMTs show lower efficiencies than the old PMTs.

Different colors, pale gray and green, purple, white, and gold, the play of light through the water,
Dumb swimmers there among the rocks, coral, gluten, grass, rushes, and the aliment of the swimmers,
Sluggish existences grazing there suspended, or slowly crawling close to the bottom

—Walt Whitman, *Leaves of Grass*

Contents

Chapter 1

Introduction

One of the great mysteries of life is the question of how the universe began.

Various hypotheses have been proposed: the ancient Greeks believed there was a nothingness, suddenly interrupted by the coming of the gods, which happened to create the universe as a sort of by-product of their feuds, while Christianity famously postulates that God created the world in seven days.

The current hypothesis holds that the universe exploded about 14 billion years ago from a singular point to expand ever (at least up until now) afterwards. Yet this theory also has it fair share of questions. What caused this bang exactly, and how does this relate into physics as we currently understand it? Would our current framework of physics be sufficient enough to explain exactly what happened?

It is also a difficult question to research, as it is difficult to trace particles (or anything for that matter) from such a long time ago. Most physical elements did simply not exist back then, and have been reshaped countless times since their creation. However, one avenue of research that may lead to any possible answers is the study of neutrinos. This particle has some very special properties which allow us to shed some light on this question.

This question, however interesting, is also very grand and may not be answered anytime soon. Luckily, there are also smaller problems to be answered. This research aims to help in understanding two of them; namely to improve the Standard Model of Particle Physics by testing to what extent the properties of neutrinos confirm to said model and improving said model where necessary, and to improve our understanding of cosmic accelerators.

1.1 Neutrinos

Neutrinos are elementary particles. They are fermions (so they have spin $\frac{1}{2}$), have no electrical charge, have a very small mass and come in three flavours: namely the electron neutrino ν_e, the muon neutrino ν_μ and the tau neutrino ν_τ and their anti-particles. A neutrino of one flavour could change into one of the other two via neutrino oscillation.

Known sources where neutrinos originate are from the Sun, supernovas and the Big Bang, but also from the inside of the Earth, the atmosphere of the Earth, nuclear reactors and even nuclear bombs.

Most importantly, neutrinos rarely interact with other particles and only do so via the gravitational force and the weak force interaction. They also move undisturbed through electromagnetic fields, as they have not electrical charge. This means they can travel very far, and conversely mean we could see into physical processes that would otherwise not be visible, which makes them invaluable for astronomical research.

1.2 Cherenkov radiation

From neutrino weak-matter interaction a hadron or a lepton results, such as a muon for example*. When travelling through matter at a speed greater than the speed of light in that matter (given by $v = c/n$), the charged particle polarises the medium and emits a wavefront according to following relation:

$$\cos\theta = \frac{1}{\beta n} \tag{1.1}$$

with θ being the angle along which the wavefront is emitted, $\beta = c/v$ with v being the speed of the particle, and n the index of refraction of the medium. This effect is often compared to the sonic boom experienced when an aircraft or a rocket crosses the speed of sound, since that also produces a wavefront, but is not entirely apt as there exists no metaphor for the polarisation aspect seen in Cherenkov radiation. Note that the required speed of the particle can be dramatically changed by the choice of matter: it is known that the speed of light in water for example is about three-quarters of that in air.

This effect also produces a faint ultraviolet light [2], which can be detected and therefore forms a method to identify neutrinos, thereby forming the core of the operation of a neutrino telescope.

*An overview of the relevant neutrino interactions can be found at [1].

1.3 Water/ice neutrino telescopes

First proposed by M.A. Markov in 1951, a water or ice neutrino telescope is a research infrastructure designed to detect (cosmic) neutrinos on a large scale. As neutrinos interact very rarely, a very large area of operation is needed to detect any significant amount of neutrinos. Combined with the operating principle of Cherenkov radiation light, this restricts the choice of matter in which to build the detector to only water or ice. Such a telescope typically consists of thousands of photomultiplier tubes (PMTs) with which to detect the Cherenkov radiation spread over a vast volume (in the newest generation detectors a cubic kilometer) located a few kilometers below sea level. These requirements present vast technical obstacles, and it is for this reason they could only be built from the 1980s onward.

The basic principle of operation relies on the photons of the Cherenkov radiation light, which is detected in several photomultiplier tubes, so that the heading of the light is known. From equation (1.1) one could calculate the angle of the emission of the wavefront, so that ultimately the trajectory of the original neutrino could be derived. In this research only the muon detection is considered, as this is the most relevant for the detector calibration [1].

Current generation neutrino telescopes are KM3NeT (Cubic-scale Neutrino Telescope) in the Mediterranean Sea, IceCube at Antartica, the Baikal Deep Underwater Neutrino Telescope, and the Super-Kamiokande in a former mine in Japan.

1.4 Other sources of light

In the case of an underwater neutrino telecope, one needs to consider other sources of light. Specifically in the case of KM3NeT, they are the following.

1.4.1 Bioluminescence

A neutrino telescope does not live in an isolated room unfortunately: it shares the space with other deep-sea lifeforms, some of which may emit light. The bioluminescence can be separated into two parts: periodic bursts with high rates connected to large lifeforms passing the detector and a contineous rate usually connected to bacteria nearby the detector. The bursts result in an increased hitrate of several seconds.

To account for this the output of any channel affected by this phenomenon is automatically blocked whenever the rate exceeds 20 kHz [3].

1.4.2 Potassium-40 decay

One of the naturally occurring isotopes in sea water is ^{40}K, which decays in most cases in the following matter.

$$^{40}K \rightarrow ^{40} Ca + e^{-} + \bar{v}_e \tag{1.2}$$

of which the electron and the can produce Cherenkov radiation. This decay occures at a rate of 10 kHz [4] produces most of the light seen in KM3NeT, but can be filtered out. In addition, this decay is also used to calibrate the PMTs in the telescope. This isotope is also present in the glass used in KM3NeT.

1.4.3 Dark current

Another source of light is the spontaneous emission of electrons within the photomultiplier tube. This happens at a rate of 1.5 kHz [5] and can also be filtered out.

1.4.4 Atmospheric muons

Cosmic rays are groups of high-energy particles that travel at nearly the speed of light. They interact with the atmosphere when hitting the Earth, decaying into, among other things, high energy muons. This muon flux is several orders of magnitude larger than the muon flux expected of cosmic neutrinos [1]. As the atmospheric muon flux can only come from the atmosphere, they can be filtered out by selecting this specific direction. However, multiple coincident atmospheric muons can also imitate a trajectory of a muon coming from a cosmic neutrino, so this filter is not fully accurate.

1.5 Goal of the research

The goal of this research is to investigate the effects of certain known and suspected biases upon the functioning of the KM3NeT. To this end primarily three quantities will be investigated of the photo-multiplier tubes: the hit rate, the total amount of hits, and the efficiency. The efficiency will be explained in next section.

Chapter 2

Methods

2.1 KM3NeT

The KM3NeT (Cubic-scale Neutrino Telescope) was conceived to detect neutrinos. It consists of two modules at present, both still under construction:

- Oscillation Research with Cosmics in the Abyss (ORCA), located about 40km south of Toulon, France at a sea depth of about 2450m. This site primarily focuses on the physical properties of neutrinos such as the weight of three known flavours of neutrinos.
- Astroparticle Research with Cosmics in the Abyss (ARCA), located about 100km south-east of Portopalo di Capo Passero at the south-eastern tip of Sicily, Italy at a sea depth of about 3500m. The research at this site primarily focuses on the astronomical part of neutrino research, which concerns e.g. the origin of the neutrinos or what dark matter really is consisted of.
- A third site is planned offshore off Pylos, Greece, but is slated for the very-far future.

Those three sites consist of the same building blocks: 31 photo-multiplier tubes (PMTs) located in a spherical Detection Optical Module (DOM), of which 18 are linked up vertically in strings anchored on the seabed in Detection Units (DUs).Multiple DUs make up a detection site. For both sites only about one-fifth of the DUs have been deployed so far; the final projected size for both sites is 115 DUs per site. The technology in both blocks is the same; the only difference is the geometrical layout. We will look more closely at those modules in the following sections.

2.1.1 Detection Units, Detection Optical Modules and the Photo-multipler Tubes

A DU, or string, consists of 18 DOMs spaced apart 36m (for ARCA), resulting in a total height of 700m, or 9m (for ORCA), resulting in a total height of 250m. The different vertical levels are called floors. The distance between the strings is 90m for ARCA and 20m for ORCA [3]. A DOM is about 43 cm in diameter and has two titanium collars, which are attached to two 4mm ropes supporting the DOMs vertically. The ropes are linked at the seafloor to an anchor, and at the sea level by a buoy to keep the whole structure vertical. Also attached to each DOM is an electro-optical cable, used for data transmission. The various strings are linked together on the seafloor by another datacable that sends the data to a shore computing centre for analysis.

Figure 2.1: *Left: image of a DU. In reality there are more DOMs located in a string. Right: a single DOM. Visible are the various PMTs, the electrical-optical cable on top, and the two titanium collars on the sides holding the rope supporting the DOM. Picture obtained from [6].*

Located in each DOM are 31 PMTs, seperated in five rings of six PMTs each and a sixth ”ring” consisting of one PMT facing straight down (towards the seafloor). A ring is defined along the azimuth of the DOM. The single PMT is labelled ring A, and progressing upwards are the rings B up to F. The succesive rings each are spaced by 30 degress, while the PMTs in each ring are spaced by 60 degrees [6]. Each PMT is also accorded a number from 1 to 6 in its ring, so that each PMT can be identified with a number and a letter. Rings E and F make up the upper hemisphere of the DOM, while rings A-D make up the lower half.

To ensure optical contact between the DOM and the PMT, the PMT is surrounded by an optical gel filling the space between the support of the PMT and the DOM.

The primary purpose of a PMT is to convert light to a digital signal, which it does so via the photoelectric effect. For KM3NeT two types of PMT are utilised: for the first batches of DOMs the Hamamatsu R12199-02 was used, while for the new batches this was replaced by the slightly improved Hamamatsu R14374 [7]. Another point of interest for this research was to investigate the potential biases of those two PMT types.

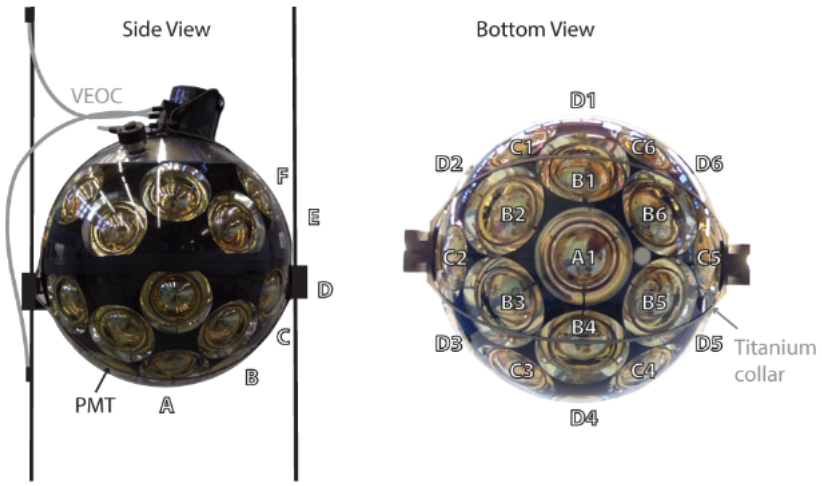


Figure 2.2: *The ring structure of the DOMs and the naming convention of the PMTs. Picture obtained from [3]*

2.2 Particle detection

Most of the light received by the KM3NeT telescope can be subdivided in two categories.

One are the level zero (L0) hits: those happen whenever a photon reaches a PMT. With each hit the unique PMT-identifier is recorded, along with the time at which the signal of a hit exceeds a certain threshold, and the time the signal of a hit stays above said certain threshold: the time over threshold (ToT). The ToT is an indication of the amplitude of the wave. All L0 hits are bunched together in time slices of 0.1s and sent to a data-centre on shore unprocessed.

The second are the level 1 (L1) coincidence: those occur whenever one DOM happens to have two or more L0 hits within a time window of 25 nanoseconds[3]. Those hits are most often the result of ^{40}K decay.

2.2.1 Efficiency calculation

The efficiency of a PMT is a measure of how many photons a PMT detects versus how many it should detect, and is therefore expressed in a ratio. The PMT efficiency in the KM3NeT neutrino telescope is a combination of the quantum efficiency of a PMT, the angle at which a photon excites the PMT, the absorption in the DOM gel and glass, and the collection efficiency of a PMT.

This ratio is calculated by taking all unique PMT pairs (465 in total), subtracting the random coincidence rate between uncorrelated hits, and then fitting coincidence distributions of all the pairs to a Gaussian each with area [5] *

$$f(\theta_{i,j}) \cdot \epsilon_i \cdot \epsilon_j \tag{2.1}$$

with θ_{i_j} denoting the space angle between both PMTs in a pair and ϵ denoting the efficiency of a given PMT.

As this approach yields 465 distributions with each five unknowns ($\theta_{i,j}$ can be calculated from the geometrical structure of the DOM, while in addition to the efficiency the PMT time offset and the transit time spread for each PMT in the pair are determined also) the efficiency of each PMT can be determined.

2.2.2 Muon simulation

To aid in the reconstruction of muon tracks, various simulation packages are used. These simulations also serve as comparison for the collected real-life data. We list here a few that are in use at KM3NeT.

MUPAGE (*Mu*on *Ge*nerator from *Pa*rametric Formulas) is a software package designed to simulate the flux of atmospheric muons in water or

*A full treatment can also be found at [5].

ice. These muons can be both single muons and muon bundles, and it takes as main inputs the zenith angle of the muons and the muon energy [8].

The Cherenkov photons and their paths are simulated with another software package, called JSirene. This package also simulates the detection of the photons by the DOMs. This data is then used along with the simulated muon flux to generate a data flow similar to the real data flow to the DOMs. Here various other effects are taken into account, such as bioluminesence, ^{40}K decay, and the behaviour of the electronics in the DOM.

To accurately determine the performance of the detector, detailled Monte Carlo (MC) simulations are used. These simulations rely on the estimated ^{40}K decay in the PMTs among other things, and the output can be then used to compare to the collected real-life data.

2.3 Used dataset

In this research data from exclusively ARCA is used. For the single rates the datasets used contain the data taken continuously starting 2023-01-12 at midnight and ending 2023-01-17 at 03h00, so in total 123 hours of data, and have been manually produced from the JRunAnalyser package. These sets contain the single rate hit distributions and can be accessed under the names `jra_133_14399.root` up to `jra_133_14440.root`. As this different datasets are very alike (standard deviation over time being about 0.01 to 0.02 percent of the mean of the values), the results have been condensed into effectively one dataset by taking the mean of the rates over time.

For the efficiency analysis the corresponding efficiency files.

`KM3NeT_00000133_00014399.v8.0_PMTeff_new.ToT.QE.PMTeff`

up to `..._00014400_...` were used spanning the same range of time. Of those files the mean over time was also taken.

For the total amount of hits per PMT the file `datav8.1.1.jchain.aashower.00013754.root` was used, with the corresponding Monte Carlos simulation files

`mcv8.1.mupage_tuned_100G.sirene.jterbr00013754.jchain.aashower.3xxx.root` with xxx $\in [092, 103]$.

2.4 Data-analysis

2.4.1 Kolmogorov-Smirnov test

A Kolmogorov-Smirnov (K-S) test is a test to determine the similarity of two distributions. The test returns two values, a test statistic and a p-value. The test statistic is a value of the maximum difference between the two distributions, while the p-value is a measure for how different the two distributions are. Specifically, the p-value is used to reject the null-hypothesis (usually that the two tested distributions are the same) when below a certain value (typically 0.05).[†]

The test has a few limitations, of which the most important is that the test has a bias towards the mean of the distribution (the others can be somewhat lessened with more data and/or more computing power). In this thesis, this test will be used a few times.

The test is used a few times to test whether a given distribution of the a given variable (efficiency, hit rate, amount of hits) over a selection of PMTs is similar to another distribution of the same given variable but over another selection of PMTs.

2.4.2 Biases

In the following chapter four items will be discussed, each affecting the efficiency evaluation of the telescope in their own manner. These are

- the titanium collar located on the DOM;
- sedimentation in the sea water;
- possible other gel transparency than assumed in some DOMs;
- differences in performance due to an improved version of PMT.

In their specific sections these items will be explained more.

2.4.3 Single rates and efficiency

For the single rates files the analysis process is as follows. The datasets `jra_133_14xxx.root` contain the distribution of all the hitrates for all PMTs. Over this distribution a Gaussian is fitted, and the mean of this

[†]For a comprehensive treatment, one could refer to [9] or to any undergraduate statistics course.

Gaussian is taken. This yields a single average rate of a PMT, expressed in units of kHz. Furthermore the corresponding efficiency for the same PMT is attached to the data of that PMT, yielding thus two numbers for each PMT. This process is repeated for all PMTs in all DOMs in all strings, resulting in data for 21 string over 18 floors for 31 PMTs, or about 22 000 different efficiencies and rates assuming one single run.

For the relation between single rates and efficiency a fit $y = a \cdot x$. This automatically fixes it at point $y = 0$; this seemed fit from a physical standpoint: if a PMT has efficiency zero it should not be able to detect any photons whatsoever. To improve the accuracy of the fit, all outliers located on a distance of at least three standard deviations were excluded from the plots and the fits. This is because a very low/high efficiency or hit rate are often caused by PMT-specific defects and not a bias which is shared among multiple PMTs.

Then this data was sorted on string number and floor number and plotted to see if there were any systematic deviations were visible, which can be indicative of a bias. In the case of a bias a hypothesis was prepared and the bias was investigated further.

2.4.4 Total number of hits and Monte Carlo simulation

For the total number of atmospheric muon hits and the MC simulation of the atmospheric muon hits a script kindly provided by Aart Heijboer was provided, in which the number of hits per PMT for all PMTs in ARCA was counted. This script considers the hits which are a consequence of reconstructed muon events, which are all the hits that match the expected time of arrival of a reconstructed muon trajectory in a certain time window.

This script was applied on both the real (measured) data and the corresponding MC simulation of that data. Then, for each DOM, the mean number of hits per PMT-ring was taken and sorted per floor and per string. This resulted in a heatmap with the following axes detailing the following data: on the x-axis the string was plotted, on the y-axis the floor, on the z-axis (the different colours) the amount of hits. The plots itself were of a certain theme, such as every individual PMT but usually those were the rings in the DOM, so that plot 1 was only ring A, plot 2 only ring B, etc. Again, any biases were further investigated. Non-operative DOMs were indicated by either a NaN (Not a Number) value or a zero; the split being an artefact from the debugging of the data-analysis script and being not especially relevant as only the operatative DOMs were analysed.

From this heatmap 1D distributions were also plotted by counting how

many times a specific entry occurs. Here most times the data from each individual PMT was taken, so that the distribution could be more accurate. This is useful to check for possible discrepancies in a selection of data: if say the distribution of total muon hits is shifted for e.g. string 27 compared to all the other strings, this could be indicative of some quirk in the operation of the detector and therefore possibly a bias in the efficiency evaluation.

As the rates, efficiency, real hits and Monte Carlo hits were often impacted by the same factors, some biases may have influence on multiple of those groups. For example, the Monte Carlo simulation relies partly on the efficiency, so that both the Monte Carlo hit rate and the efficiency could be influenced by the same bias.

Chapter 3

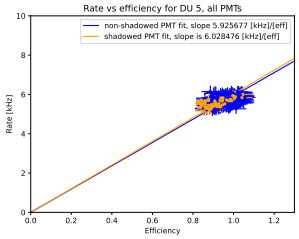
Results

3.1 Shadowing effects

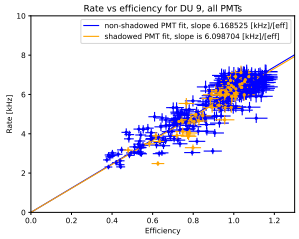
The first of the biases to be treated in this thesis is caused by the geometrical construction of the DOM. Four of the PMTs (C2, C5, E2, E5) are shadowed by the titanium collar, as can be seen in fig. (2.2), and get less hits as a result. Note that PMTs C2, C5 are shadowed from light from below and E2, E5 are shadowed from light from above.

The efficiencies for the PMTs should be impacted by this as there are less photons detected than there should be otherwise due to the shadowing. The single rates should likewise be affected, as they are drawn from the ^{40}K decay from far-away distances. Therefore the hypothesis here is that those two quantities should be changed when we compare non-shadowed PMTs to shadowed PMTs.

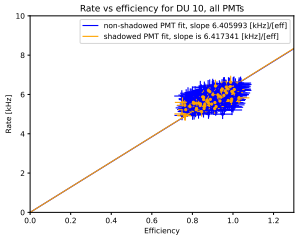
The data/MC ratio is expected to change, as the shadowed PMTs are expected to receive less photons from certain angles, leading to less hits. This means that the efficiency is a function of the direction at which a photon hits the PMT, and therefore that the used average efficiency is not entirely representative. To check this, we compare the the ratio data/MC of shadowed PMTs to non-shadowed PMTs of the same ring. By ring-wise comparison the other effects (especially sedimenation) would be minimised.



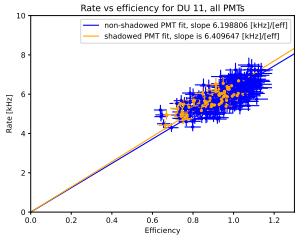
(a) *String 5*



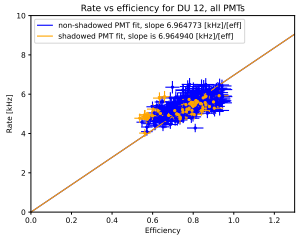
(b) *String 9*



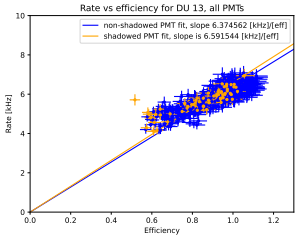
(c) *String 10*



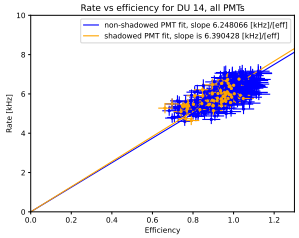
(d) *String 11*



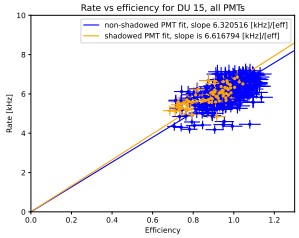
(e) *String 12*



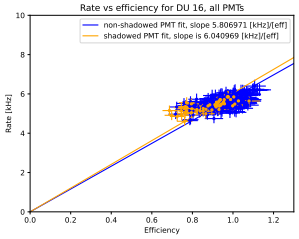
(f) *String 13*



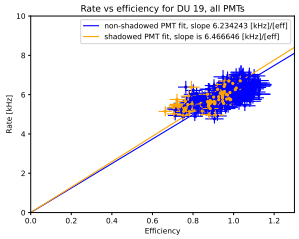
(g) *String 14*



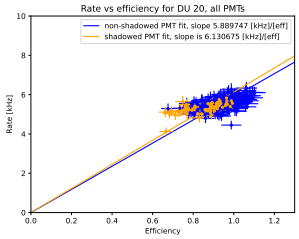
(h) *String 15*



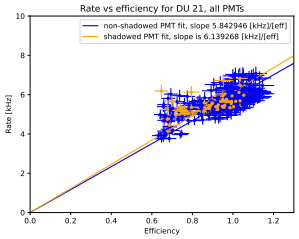
(i) *String 16*



(j) *String 19*

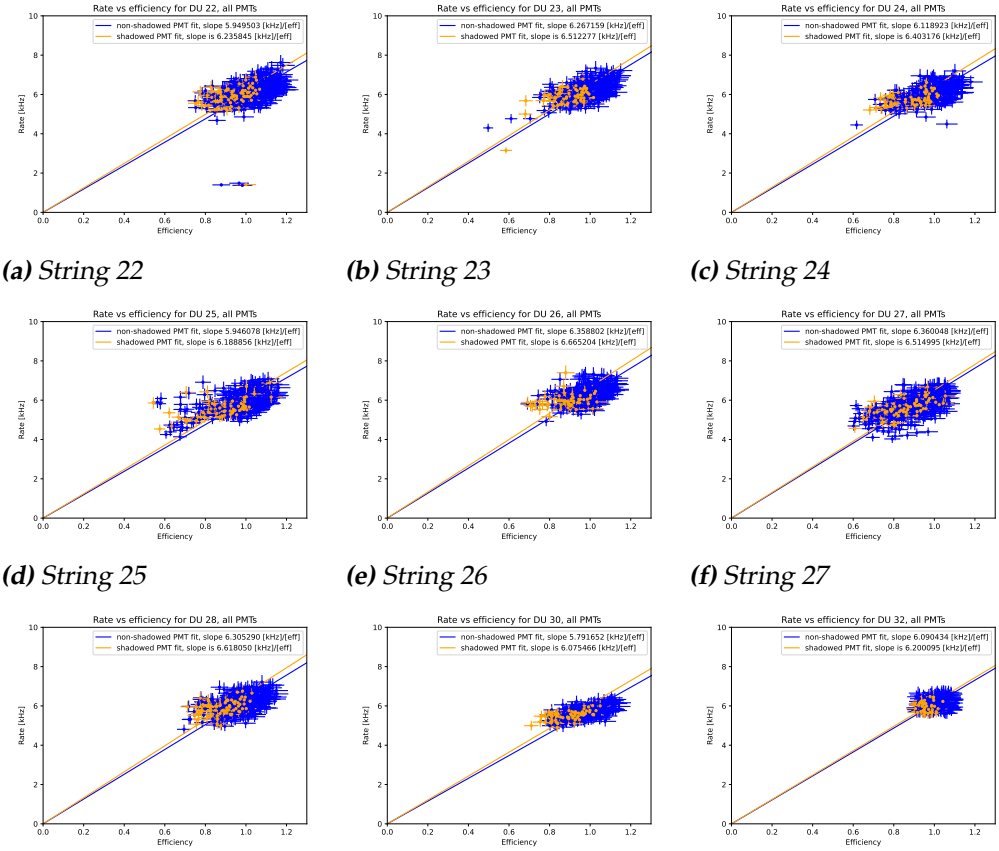


(k) *String 20*



(l) *String 21*

Figure 3.1: *Plots of the rates vs efficiency of all PMTs in one string split on the shadowed and non-shadowed PMTs. Continued on the next page.*



(a) *String 22* ***(b)*** *String 23* ***(c)*** *String 24*

(d) *String 25* ***(e)*** *String 26* ***(f)*** *String 27*

(g) *String 28* ***(h)*** *String 30* ***(i)*** *String 32*

Figure 3.2: *Plots of the rates vs efficiency of all PMTs in one string split on the shadowed and non-shadowed PMTs. Continuation of the previous page.*

As can be seen, the fits for the shadowed and the non-shadowed PMTs are not very different. To quantify this effect, the means of the fit coefficient have been calculated: for the non-shadowed PMTs 6.17 kHz/[efficiency] (recall that efficiency has no real unit as it is a fraction) and for the shadowed PMTs 6.36 kHz/[efficiency]. This is an increase of 3.1 percent. As the assumed error in the rates and efficiencies is 5 percent, this falls in the error range and is therefore not significant. Thus there is no real difference in the rates vs efficiency relation between shadowed and non-shadowed PMTs.

The assumption of a fit $y = a \cdot x$ is not always entirely apt, as can be seen for e.g. string 27. This might be because there are also dark currents – the spontaneous emissions of a photon, and therefore a hit in the PMT, without the presence of any muons. This impacts the efficiency evaluation

and therefore might cause another relation between rates and efficiency.

There might be for the sole efficiencies and the data/MC ratio, so we now look at those.

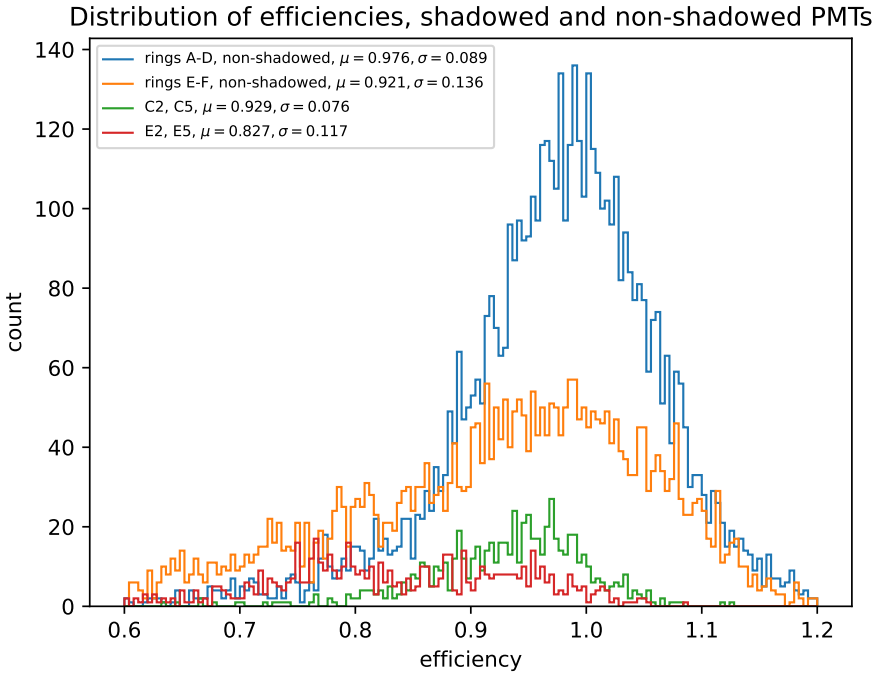
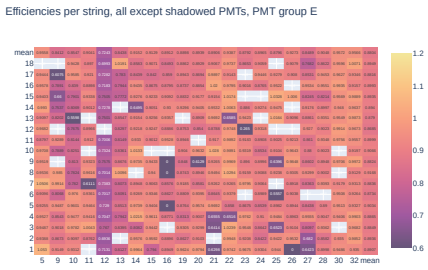


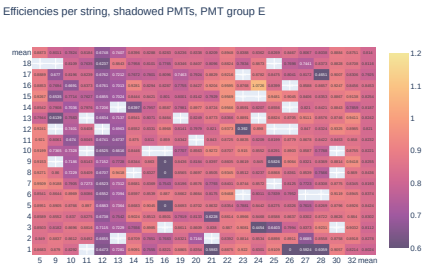
Figure 3.3: *Comparison of the distributions of shadowed and non-shadowed PMTs. The shadowed component has been splitted into both component rings, as ring E is known to suffer from sediment, which causes a lower efficiency, while ring C is clear from this issue. Summing the distribution for C2,C5 and E2, E5 yields the distribution for all shadowed PMTs. All values outside of [0.6; 1.2] were excluded.*

The efficiency in PMTs E2, E5 are somewhat lower than those in PMTs C2, C5. In group C2, C5 this is reflected by the main trend in rings A-D (the bottom PMTs) among the non-shadowed PMTs, with the difference between their means being 5.1 percent. This is significant and therefore seems to support the hypothesis that the efficiency would be impacted by the shadowing.

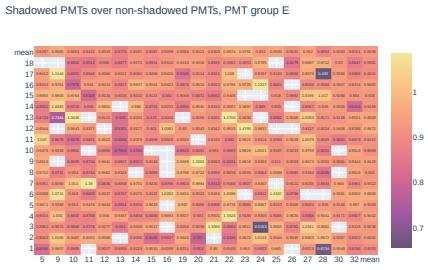
It is not the case with group E-F and E2 and E5, as the mean difference is there 11.3 percent. The distribution for rings E-F seems to be wider at the peak, while E2, E5 seems to be wide in general but barely peaks, with the difference between both their spreads being 16.2 percent.



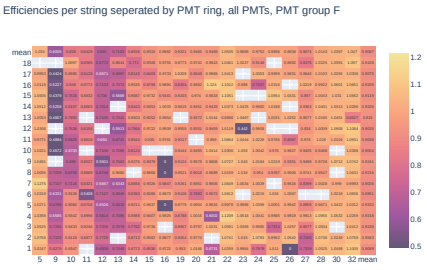
(a) *Ring E, non-shadowed*



(b) *Ring E, E2 and E5*



(c) *Ring E, ratio shadowed/non shadowed*

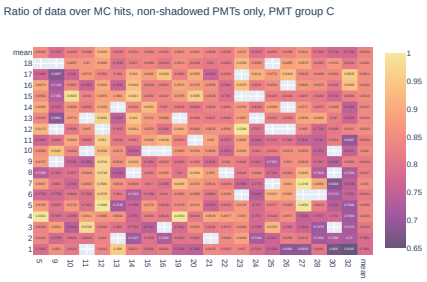


(d) *Ring F, in full*

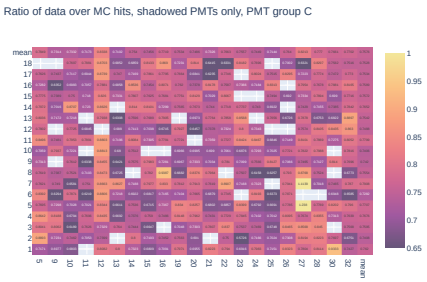
Figure 3.4: *Comparison of the efficiencies of ring E. The shadowed and non-shadowed PMTs have been included, as a ratio between those two and the ring F for comparison. The ratio was calculated by dividing heatmap (a) by (b), and therefore has not been weighted on the number of PMTs. The mean for E2, E5 differ slightly as this time there were no values excluded.*

The ratios for the PMTs seem to be relatively evenly distributed, with all string and floor means in the interval [0.89; 0.94]. This discounts any theory that the affected efficiencies would be dependent on location in the detector, and therefore seems to be inherent to the physical location in the DOM. Compared to ring F the non-shadowed PMTs show comparable efficiencies, with their means only differing 2.2 percent.

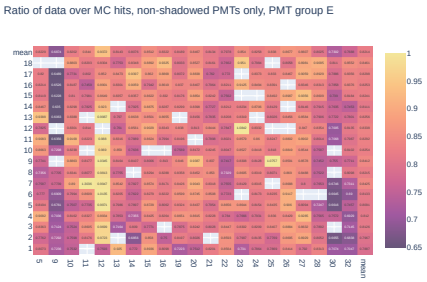
When we compare the data/MC ratio for shadowed and non-shadowed PMTs, we find the following.



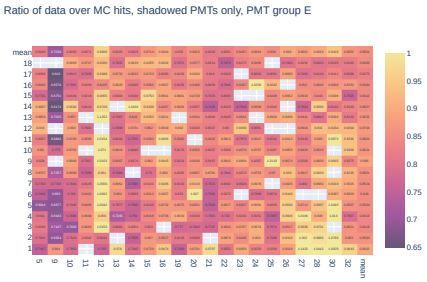
(a) *Ring C, non-shadowed*



(b) *Ring C, C2 and C5*

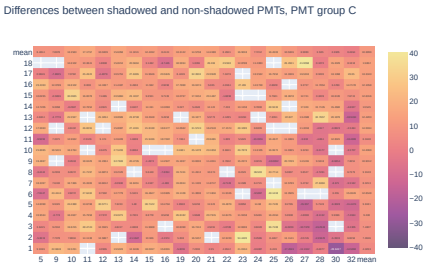


(c) *Ring E, non-shadowed*

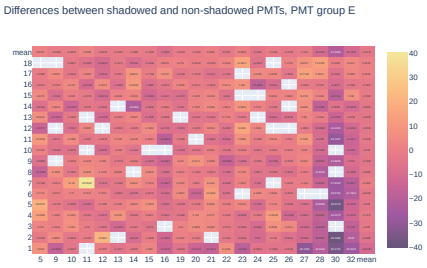


(d) *Ring E, E2 and E5*

Figure 3.5: *Comparison of the data over MC of rings C and E: in the left column, the non shadowed PMTs; in the right, the shadowed PMTs; upper row, ring C; lower row, ring E. Scale is uniform across those four plots.*



(a) *Ring C, differences between non-shadowed and shadowed*



(b) *Ring E, differences between non-shadowed and shadowed*

Figure 3.6: *Comparison of the the differences between the non-shadowed and the shadowed PMTs of ring C and E, in percentage. For the differences the results in fig. (3.5) have been taken, viz. subplot (b) has been compared to subplot (a) and the same with subplots (d) and (c).*

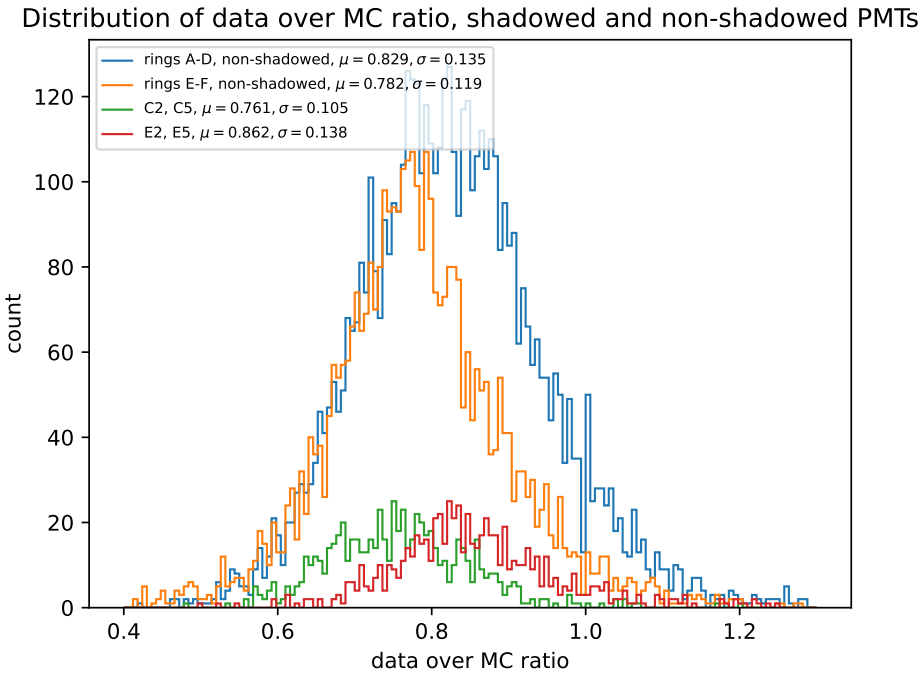


Figure 3.7: *Comparison of the distributions of data/MC ratio of shadowed and non-shadowed PMTs. The split in upper and lower rings is due to the same reason as fig. (3.3). Summing the distribution for C2,C5 and E2, E5 yields the distribution for all shadowed PMTs. All values outside of [0.6; 1.2] were excluded.*

The differences in fig. (3.6) have been calculated by taking the non-shadowed PMTs as a baseline: i.e. a negative percentage shows that the ratio for shadowed PMTs is higher than that for non-shadowed PMTs, and vice versa.

Both non-shadowed ring C and non-shadowed ring E show approximately the same mean value on a floor level. On a string level, this changes for some strings quite drastically (string 9, 19, 28) but more-or-less stays the same for other strings. The fact that the data/MC ratio is below one can be interpreted as that the MC simulation is overestimating the amount of hits a PMT sees.

String 12 shows in fig. (3.5) in all four subplots a different behaviour than the other strings but we suspect that that is caused by another gel inside the DOM (this will be treated in a later section). For this reason we do not analyse or interpret it here.

The shadowed PMTs of group C show a drastically lower ratio than the non-shadowed PMTs of group C, the difference between the means of

their heatmap being 9.8 percent. What furthermore springs to mind is that the actuall DOM-to-DOM differences in fig. (3.6) differ wildly and there seems not to be much of a pattern to be spotted there.

This is less so the case for the differences in percentage in ring E shadowed and non-shadowed, where it can be seen clearly that the shadowed PMTs have up to a 30 percent higher data/MC ratio than the non-shadowed PMTs. This effect can also be seen in the lower two-thirds of DOMs in string 28 and 30, where the effect is so pronounced that, the average ratio of non-shadowed PMTs in those strings is 10 percent lower than the average ratio of shadowed PMTs (the average ratio up to floor 12 in those three DOMs is 19.9 percent less than non-shadowed PMTs). This actually means that, when one also refers to the pure ratio plot of shadowed ring E, that the data and the Monte Carlo actually agree very nicely in those DOMs. Nonetheless, this might still be indicative of a bias as the Monte Carlo directly draws from the calculated efficiency, which is different for the shadowed PMTs than for the non-shadowed ones.

This different behaviour of strings 28, 29 and 30 can also be seen with some difficulty in ring C (those three appear to agree on the data/MC ratio between shadowed and non-shadowed, while every other string has the non-shadowed with a higher ratio).

For the data/MC ratio too the shadowed PMTs in the C-ring show a lower mean location than the mean location in non-shadowed rings A-D; this value being 14.7 percent lower. On the contrary, E2, E5 have a mean location 10.2 percent higher than the non-shadowed rings E-F. This discrepancy is curious, as one might expect that the shadowing would have a similar effect on the PMTs: the data/MC ratio either would be more for both or less for both C2, C5 and E2, E5, but might be explained due the effects of sedimentation.

It can be concluded that the both the data/MC ratio and efficiencies for C2, C5 is lower than the mean in rings A-D. For E2, E5 the efficiency is lower compared to rings E-F but the data/MC ratio is actually higher.

The fact that the data/MC ratio for E2, E5 are higher than the non-shadowed PMTs of rings E-F might be indicative of an efficiency bias in those two PMTs. The calculated efficiency of E2, E5 are lower than the rest of rings E-F, so the Monte Carlo hit rates should also be lower than the real data, since this draws from the calculated efficiency. The real data draws from a real efficiency, which is higher than the calculated efficiency as the calculated efficiency is an average of the current and neighbouring PMTs whereas the real efficiency is drawn from a single (current) PMT. For the shadowed PMTs the calculated efficiency is higher than the real one as the real one is impacted by the shadowing from above whereas for

the calculated one this effect is mitigated by virtue of being an average. Therefore there should be more MC hits seen than real ones. Combining all this yields that E2, E5 should have a higher data/MC ratio than the rest of rings E-F, but the opposite result is seen: namely that E2, E5 have a lower data/MC ratio than the two rings.

The same argument applies to C2, C5 but with the inverse values as they are shadowed from below instead of from above: they should have a lower data/MC ratio than the rest of rings A-D, but the opposite again is seen: namely that C2, C5 have a higher data/MC ratio than the other four rings. This could point to a bias in the calculated efficiencies of the shadowed PMTs: the Monte Carlo simulations return too many (in the case of E2, E5) or too few (in the case of C2, C5) photons, and as the MC simulations are a function of the calculated efficiency, this could be a bias in the calculated efficiencies.

3.2 Sedimentation

One of the facts of life in KM3NeT is sedimentation: the accumulation of particles on barriers in the water. Over time, the number of those particles increases and blocks out light coming into the telescope. This only occurs to the upper half of the DOMs (rings E-F); the lower half (rings A-D) do not suffer from this effect. As the efficiency calculations are reliant on ^{40}K decay from the sea water, these efficiencies might not be accurate anymore for sediment-affected PMTs, and as the Monte Carlo hit rates are also reliant on the efficiencies, these would also be changed.

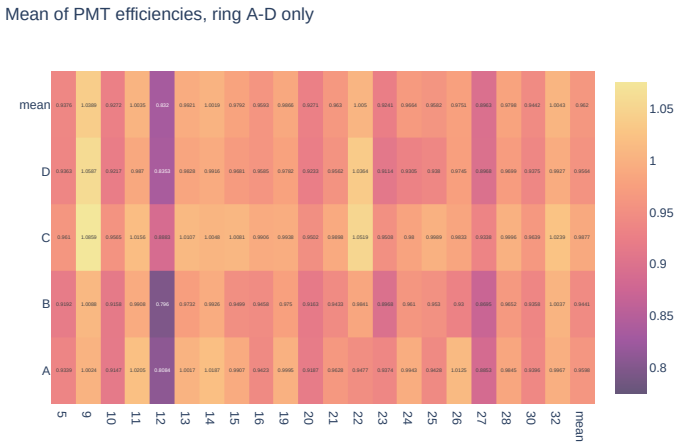
Specifically in the efficiency evaluations the effect of glass radioactivity is neglected, because the effect is small relative to the ^{40}K sea water decay. In the glass of the DOMs the ^{40}K isotope is also present and decays also according to eq. (1.2), but the more a DOM gets sedimented, the greater the effects of this glass radioactivity relative to the ^{40}K sea water decay. This will then lead to a bias, as the efficiency is no longer properly evaluated because the glass radioactivity, while not accounted for in the efficiency evaluations, does have an impact on said evaluations.

The sedimentation is also a function of time. This is logical: the longer a barrier spends in the water, the longer particles would be able to accumulate on it. String 9, for example, has been in deployed since late 2015 and it will be shown that this string is frequently an outlier in the following data.

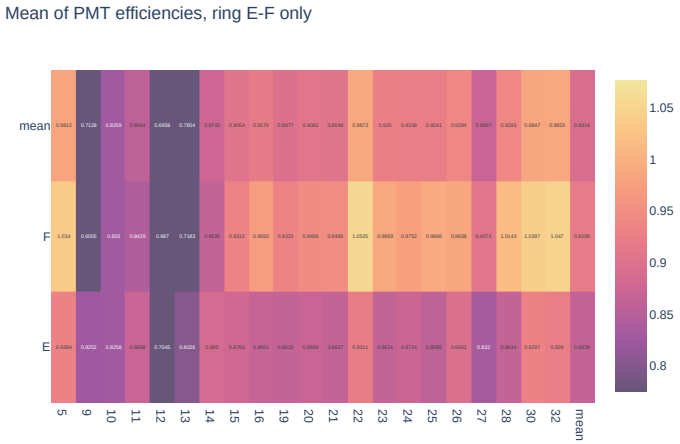
The ratio of data versus MC simulation hits should be lower for the PMTs effected by sediment, since the MC simulation hits are drawn from

the efficiency evaluation, among other things, and therefore have a bias due to the glass radioactivity. We compare the efficiencies, the data/MC ratio, and the hit rates of the upper two rings versus the lower four rings.

The effects for the efficiencies summarised by string are as follows.



(a) *Mean efficiencies per DOM of all DOMs, lower rings only. Scale is the same across both subplots.*



(b) *Mean efficiencies per DOM of all DOMs, upper rings only. The choice for two heatmaps here is to better visualise the differences between the upper and lower rings.*

Figure 3.8: *Mean efficiencies per DOM seperated by rings.*

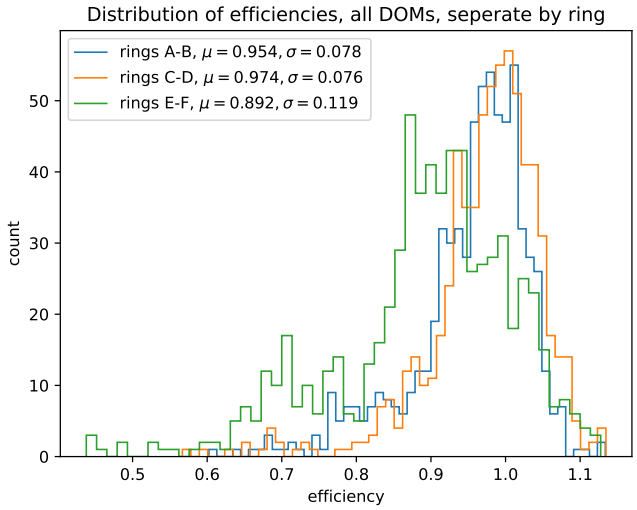


Figure 3.9: *Distribution of all the efficiencies in fig.(3.8). The distribution for ring A-D have been split in two so that they not totally eclipse the distribution for ring E-F. All points below efficiency 0.4 have been left out as those efficiencies can probably be ascribed to other issues.*

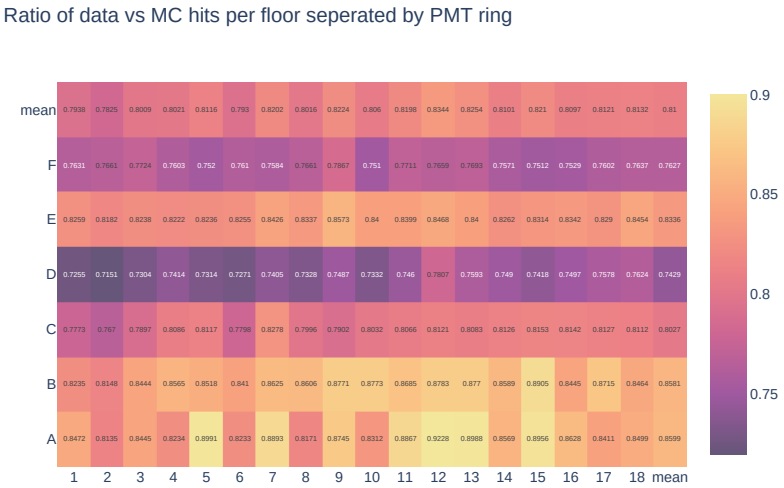


Figure 3.10: *Ratio of real data hits versus the MC simulation summarised per floor. Here the average for each floor over every string is taken, resulting in a average value for each floor per ring.*

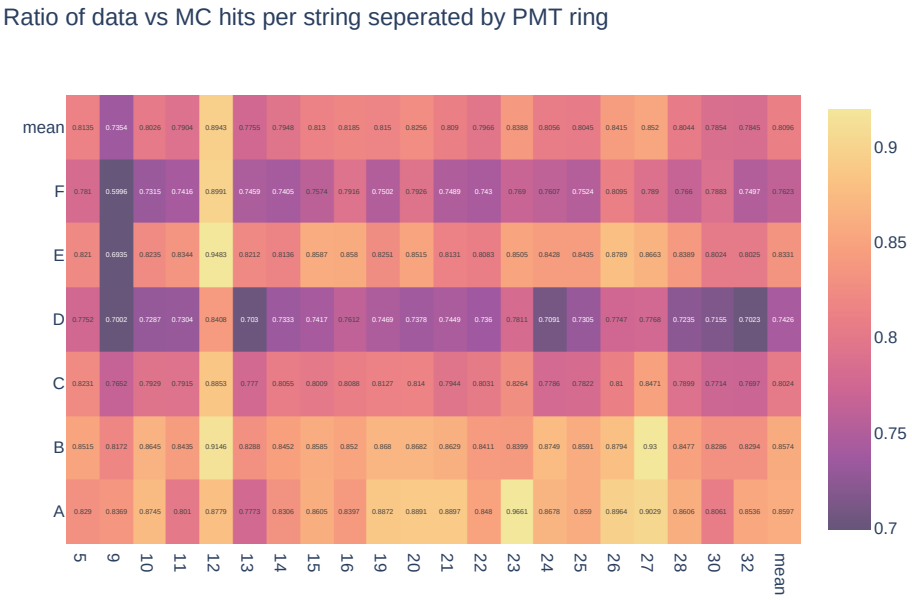


Figure 3.11: *Ratio of real data hits versus the MC simulation summarised per string. Here the average for each string over every floor is taken, resulting in a average value for each string per ring. Note that scales are slightly different between this figure and the one above.*

A number of trends can be identified:

- The effects for ring A vary significantly. This is due to the fact that ring A only consists of a single downward facing PMT, and therefore has in the first place six times as less data as the other rings do, making it more suspectable to statistical fluctuations.
- Ring E and F have a significantly lower efficiency than the rest of the rings. This effect appllies somewhat less to ring E but it is suspected that this ring suffers less from this issue as two of the six PMTs (E2, E6) are shadowed by the titanium collar, keeping them somewhat more clear from sediment. The titanium collar shadowing effect is accounted for in the efficiency evaluations, so the lower efficiency in ring E is already partly accounted for. When one corrects for the shadowing the efficiency for ring C is improved by about 3 percent, but decreases very slightly for ring E.
- The data/MC ratio differ significantly for rings E and F. This is unexpected as both should be exposed to sedimentation and therefore

according to the hypothesis both should have a lower ratio than the rest of the rings. A possible cause for this is the shadowing effect seen by two of the six PMTs in ring E: it is already shown in fig. (3.7) that the shadowed PMTs have a significantly higher data/MC ratio than the non-shadowed PMTs in that ring. We shall therefore exlcude those PMTs in the rest of this section, as that would lower the data/MC ratio.

- On a similar note, the data/MC ratio for ring D seems also te be strange. There is something in this ring that is not well-reflected in either the MC simulation or the photon intake, but it is not clear yet what this exactly might be.

- String 12 seems to be have a significantly higher data vs MC ratio (it is a lot closer to 1 than the other strings), with a mean of 10.5 percent higher than the mean over the total heatmap. This could be caused due to the suspected gel transparency bias, as the MC might not sufficiently adjusted for the other gel properties or there is a difference between received photons with other strings.

- String 9 however seems to be veering to the other side, sitting about 9.2 percent below the mean of the heatmap. As this string suffers the most from sedimentation, we will take this string as a baseline for further investigations.

A next step would be to compare the strings in different catagories of performance to the performance of string 9. One indicator could be the raw number of hits in the upper and lower hemisphere of the DOM; as the lower hemisphere should not be impacted by sedimentation, the number of hits should there be higher.

In fig (3.12) string 9 has a significantly lower ratio than the rest of the strings, 39.4 percent lower than the mean over the entire heatmap. When one only considers floors 10-17, this increases to 50 percent lower. OTher strings do not even come close to this difference in ratio.

The high spot at string 23 floor 12 should be disregarded, as this DOM suffers from an absymally low efficiency and therefore is not representative of any systematics in ARCA.

Furthermore, strings 12 and 13 seem to have a ratio slightly less than the norm. Here the suspected gel issue for string 12, which will be treated later, is considered less relevant as this heatmap purely considers intra-DOM performance; in other words, it is a ratio in which both parts suffer

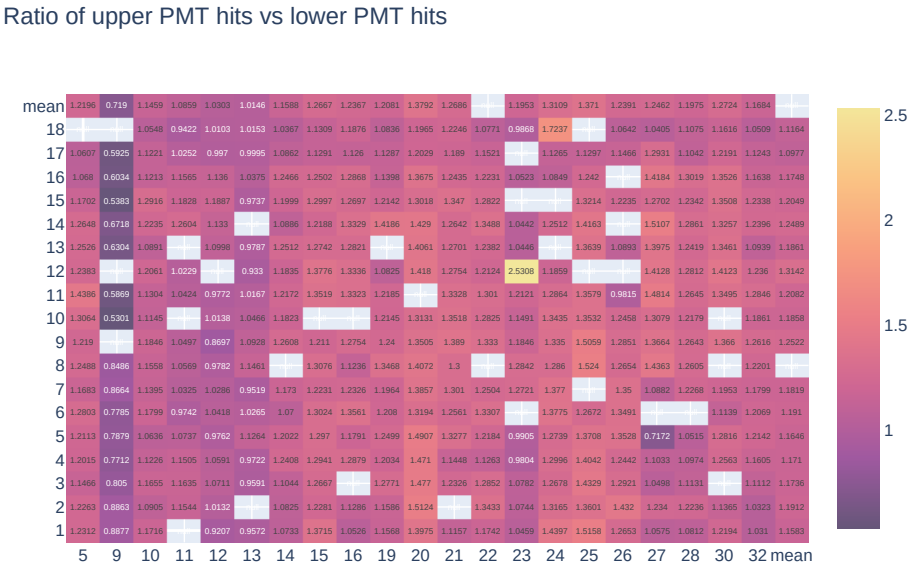
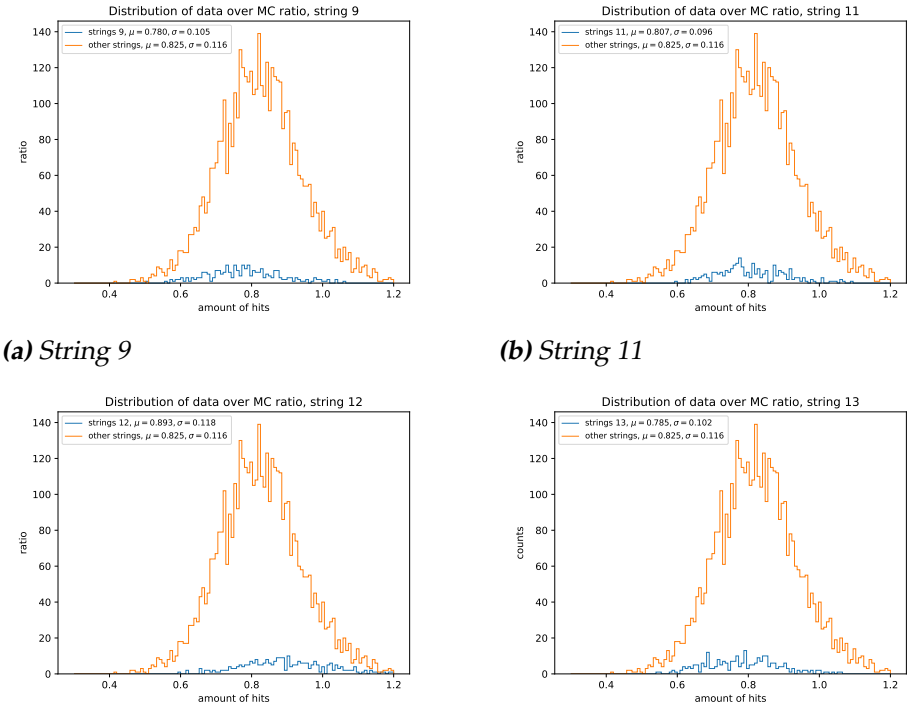


Figure 3.12: *Heatmap of the ratio of upper (rings E,F) PMTs divided by lower PMTs (rings A-D). The ratio has been weighted on the number of PMTs to correct for the fact that the lower PMT ring has about 75 percent more PMTs than the upper ring. Average ratio is 1.1867.*

from the same bias, thereby cancelling out the bias. Nevertheless, the figure are that string 12 has a ratio of 13.2 percent less and string 13 has a ratio of 14.5 percent less. Compared to string 9 string 12 has a ratio of 30.2 percent more and string 13 has a ratio of 29.1 percent more. Both strings are thus located slightly more closely to the mean of the total heatmap than to the ratio of string 9.

It should be noted that string 9 has been in the water since 2015, and strings 11 through 13 inclusive have been deployed since mid-2021, all being among the longest deployed strings in ARCA. They are expected to suffer from sedimentation the most for this reason.

The next lowest string, string 11, has a ratio of 8.4 percent lower than the heatmap mean, while this percentage is roughly halved for the next lowest, string 10. The four strings with the lowest mean (9, 11, 12, 13) considerably drag the mean down as almost every other string has a mean above the current heatmap mean. When one excludes those four strings, the heatmap mean would be a ratio of 1.2403, or about 4.5 percent higher.



(a) *String 9* ***(b)*** *String 11*

(c) *String 12* ***(d)*** *String 13*

Figure 3.13: *Distribution of the suspect sediment strings against the distribution of all strings except those four suspect strings. Here only the upper rings are plotted (ring E-F), so that all PMTs in this figure in theory could suffer from sedimentation.*

As can be seen in figure (3.13) string 11 and 12 are possibly less hard hit by sedimentation than string 9 and 13. String 12 does appear to have an slight overestimation relative to the other strings of the data/MC ratio: possibly this is due to the gel transmission properties bias as it should be expected that sedimentation would cause the same effect as can be seen in string 9: an underestimation of the data/MC ratio.

String 13 has a mean ratio of 0.005 higher than string 9, and a similar spread to string 9. The mean ratio is 5.1 percent lower than the mean over all non-suspect strings. For this reason it is likely that it also suffers from sedimentation, in combination with the lower efficiency seen in fig. (3.8).

It should be noted that string 13 is somewhat of an oddity: the sedimentation features were already present at the deployment of the string. As the effects of sedimentation do not occur directly after deployment, the effects now seen are probably caused by another issue and warrant further investigation.

What would also be interesting to see is the distributions of the real/MC data hits split over upper (ring E-F) and lower (rings A-D) group PMTs. As we know that sedimentation mostly happens on the upper group, and that the MC does not reflect the sedimentation very well, it may be interesting to see how the distributions compare.

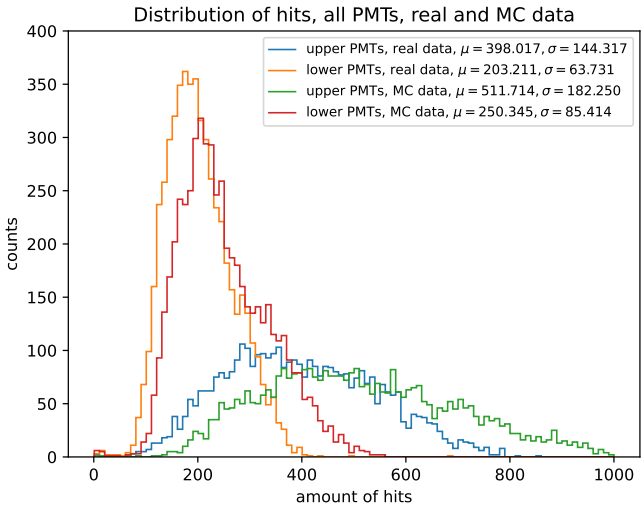
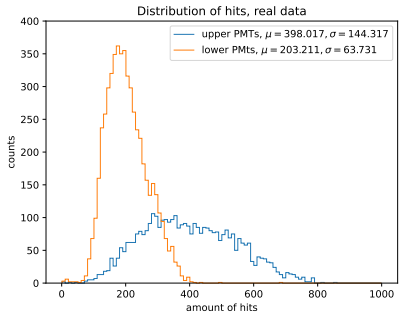
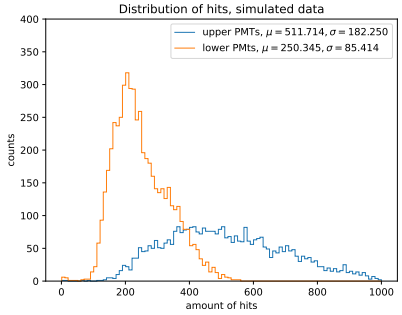


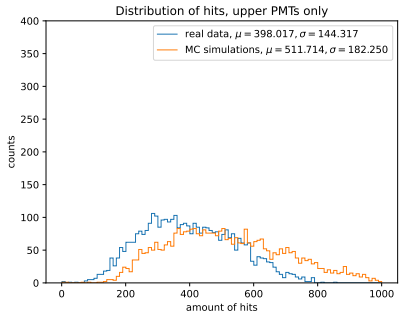
Figure 3.14: *Distribution of real and simulated data, split in the upper and lower group PMTs. This figure is essentially fig. (3.15) condensed in one plot.*



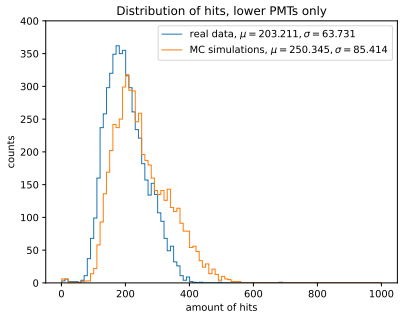
(a) *Real data*



(b) *MC data*



(c) *Upper PMTs*



(d) *Lower PMTs*

Figure 3.15: *Collection of various distributions over the upper and lower group PMTs, and real and simulated (MC) data. Hits refers to the amount of hits one PMT has counted, and counts refers the amount of PMTs that have a given amount of hits.*

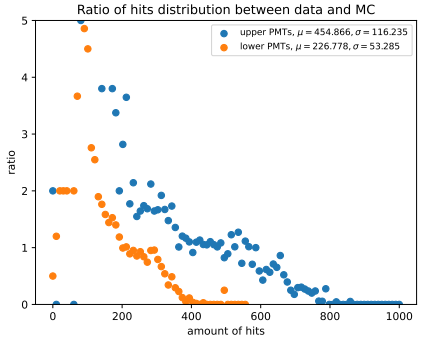


Figure 3.16: *Ratio of the distribution seen in fig. (3.14). It is calculated by dividing, the real data by the corresponding Monte Carlo simulated data. This is done for upper and lower group PMTs seperately. The ratio actually peaks at a ratio of 80 but above the limits of this plot it is clear that the real data distribution eclipses the MC simulation.*

The data and MC mostly agree on both upper and lower PMTs, although this difference is greater for upper PMTs than for lower; the difference in mean is greater while the difference in spread is about the same relatively. It is clear that the MC tends to underestimate the counts at low hits, while it overestimates the counts at high hits.

It is also interesting to see that in regards to the upper and lower PMTs that the lower PMTs receive more hits, some 44.9 percent more. This is not a surprise, as the lower PMT group consists of 18 PMTs whereas the upper group consists of just 12, but the lower PMTs receive slightly less than the ratio should strictly indicate (namely 50 percent). This is barely significant, but a cause outside of statistical fluctuations could be that more muons from the sky reach the detector than through the earth.

Another interesting fact is that the data and MC simulations almost agree on the amount of total hits over all PMTs: 3780 vs 3764 for real data and MC respectively for the upper group and 5476 vs 5474 for real data and MC respectively for the lower group. This is a sign that the simulation works well, just that it needs to be ironed out a bit.

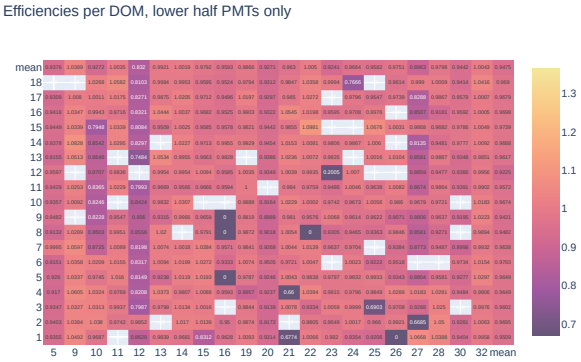
3.3 Gel transparency effects

It is suspected that the gel properties in some of the DOMs may be different. The efficiency evaluation is dependent on ^{40}K decay of nearby seawater as noted earlier, while the transmission through the gel might also be different. This evaluation is therefore dependent on a given (by the

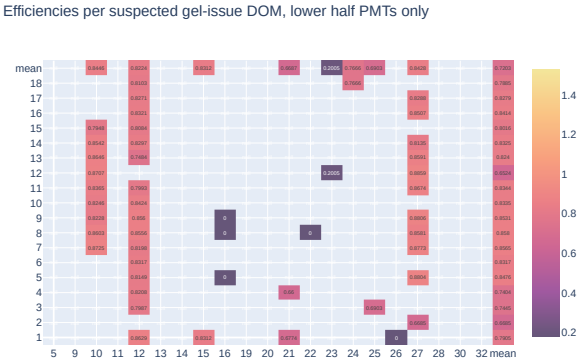
nearby ^{40}K decay light) wavelength spectrum, and can be translated to the wavelength spectrum of far-away light. However, when the transparacy is different, this translation of the wavelength spectrum would be incorrect. This concretely implies that the efficiencies of DOMs with a different gel transparency would also be different.

Another effect is caused by the ^{40}K isothope present in the glass of the DOM. This ^{40}K decays too, but is normally accounted for in the efficiency evaluations. When sedimentation occurs on the DOMs, the light gets increasingly blocked out, and thus the ^{40}K decay of the nearby seawater is also less visible. The effect of this is that the ratio of ^{40}K decay in the DOM increases relative to the ^{40}K decay of nearby seawater, leading to a flawed efficiency rating.

We suspect from [10] that in any case ARCA.0012 in full and ARCA.0010 floors 7-15 have this gel transparency bias. Those groups have therefore been taken as a baseline for analysis in the following section. Filtering out all efficiencies lower than the mean of those strings yields the second plot in the figure below. Hereby only the lower four rings (A-D) have been selected so to filter out any sedimentation bias.



(a) *Mean efficiencies per string of all DOMs with only rings A-D selected.*



(b) *Mean efficiencies per string of DOMs with only rings A-D selected.*

Figure 3.17: *Comparison of the full efficiency DOM map versus those filtered by the average efficiency of suspected affected DOMs. Note that scales are not the same. We refer to the appendix for the full overview of the efficiency maps.*

For plot (b) the DOMs were selected according to following formula:

$$\mu < (\mu_{12} + \mu_{10})/2 + 2 * \sqrt{0.25 * \sigma_{10}^2 + 0.25 * \sigma_{12}^2} \tag{3.1}$$

with 10 and 12 referring to string 10 floors 7-15 (for brevity's sake: in the rest of this section if string 10 is mentioned only floors 7-15 are meant) and string 12 all floors. We have therefore set here as upper bound for the averages the mean of the mean of both groups plus twice the propagated standard deviations of both groups, and selected for all efficiencies below

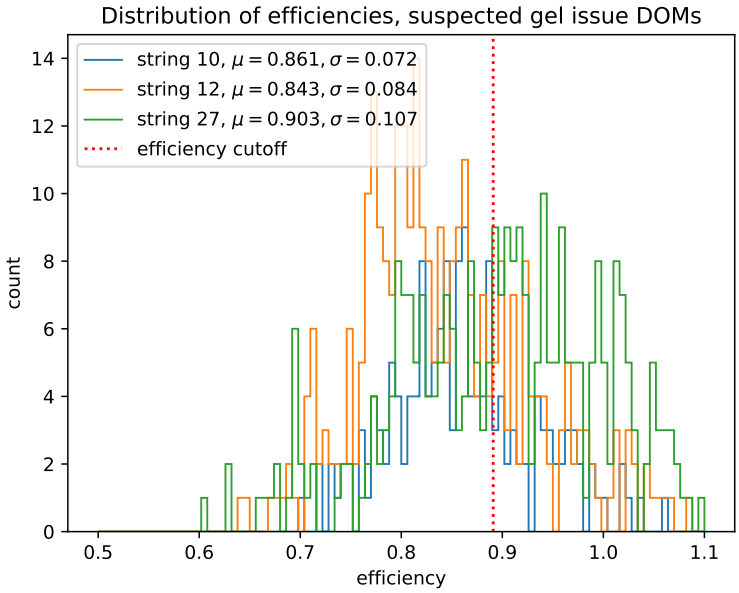


Figure 3.18: *Distribution of the efficiencies of the three suspected gel-issue strings. Note that they all have a similar spread but a shifted mean point. Here the efficiencies for all the DOMs in string 27 have been plotted.The efficiency cut-off indicates the point whereto the suspected DOMs were selected as per eq. (3.1).*

that value (0.891 in this case). A factor two was chosen so that the filter was somewhat less restrictive; otherwise only scattered DOMs were selected.

This filter also selects a few other scattered DOMs but those were not inlcuded for further analysis as other nearby DOMs seem to perform with at least the average efficiency. The low efficiency might here be caused due to other issues. They might still be worth looking in to if they turn out to be in the same constructed batch or if there is some other common denominator among them, but the constructed batch denominator would be unlikely as multiple of the same batch tend to end up in the same string.

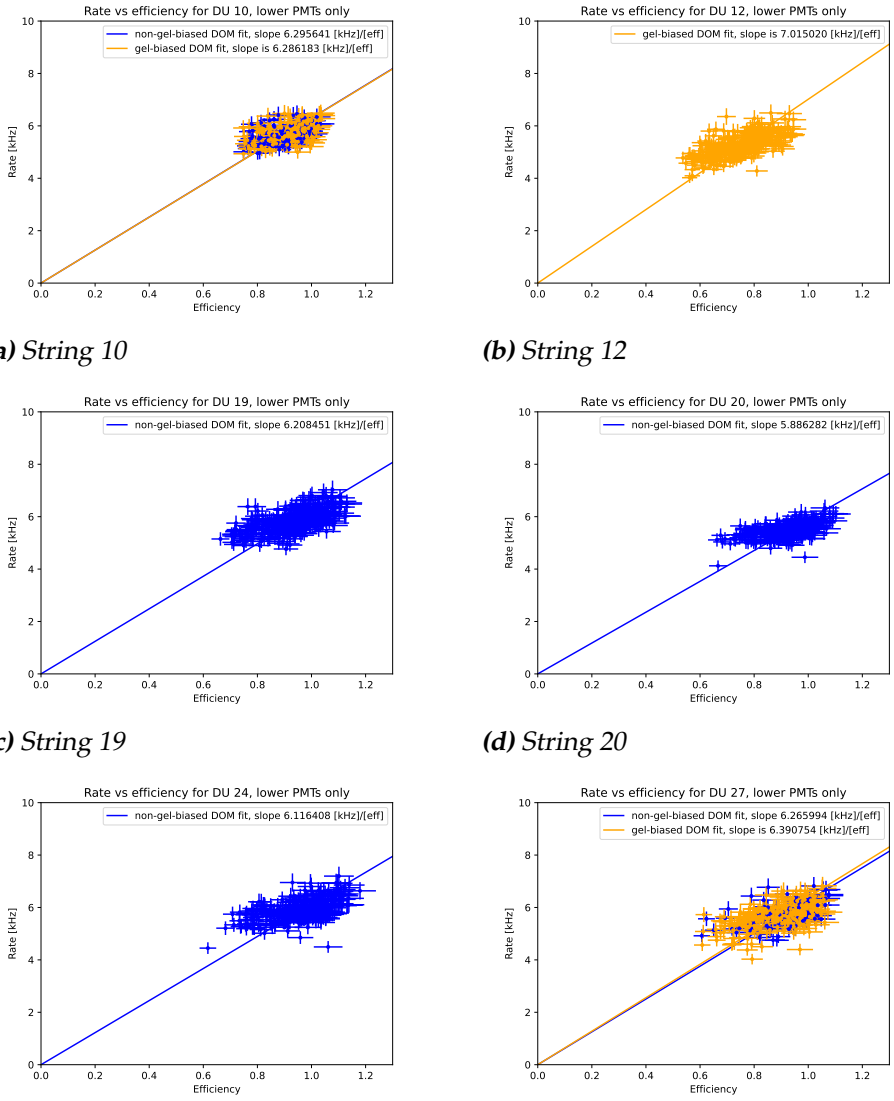
We can see a clearly different colour for the aforementioned subgroups. From figure (3.17) it can also be seen that string 27 might also be affected; 11 of its 18 DOMs have a lower efficiency value than was set by the filter.

The selected DOMs of string 27 have a mean efficiency of 11 percent lower than all the DOMS of all the strings, irrspective of they have been selected by eq. (3.1) or not. This is significant and warrants further investigation, which we could do by looking at the distribution of the efficiencies.

As can be seen in fig. (3.18), most efficiencies in string 27 appear to be above the cutoff line. This may indicate that most DOMs might be good

and not be suffering from this issue. Still, it is possible that only a few DOMs ahve this issue, as is the case with string 10. Another way to test the hypothesis if string 27 would be affected by this gel transparency bias is to check how the single rates behave.

The single rates are the hits from far-away photons. As the calibration for the gel transparency bias affected DOMs would be different than the other DOMs, the single rates would also be different. Therefore there should be a change in relation between single rates and efficiency in the gel transparency bias affected DOMs.



(a) *String 10* ***(b)*** *String 12*

(c) *String 19* ***(d)*** *String 20*

(e) *String 24* ***(f)*** *String 27*

Figure 3.19: *Plots of the rates vs efficiency of some strings. Blue indicates no suspected gel-bias, whereas orange does indicate that bias. The blue plots were selected because they were the three plots closest to the average fit value. Here again only the results for rings A-D were plotted.*

The higher coefficient indicates that the rate increases more sharply as function of efficiency, and therefore that the affected DOMs might indeed be calibrated otherwise than those not suffering from this specific bias. It might also indicate a deficit in the amount of photons received, which may also be caused by the different gel transmission properties.

As mentioned earlier, the rates vs efficiency was fitted to a function $y = ax$. This yields as average for the rings A-D non-gel transparency biased DOMs an average of $6.10 + -0.20$ kHz/[efficiency] and for the same rings but biased DOMs an average of $6.56 + -0.33$ kHz/[efficiency], which is a difference of about 7.6 percent. The average over both biased and non biased DOMs is $6.16 + -0.27$ kHz/[efficiency].

This result also supports the hypothesis that string 27 might suffer in part from this gel transparency bias, as this string has a coefficient of 6.39 kHz/[efficiency]. This is in itself not excessively high (two other strings without this suspected bias have a higher coefficient) but combined with the low efficiency rate this string is suspect. There is therefore a case that string 27 suffers from the gel transparency bias, but we would definitely recommend more research to verify whether this is the case.

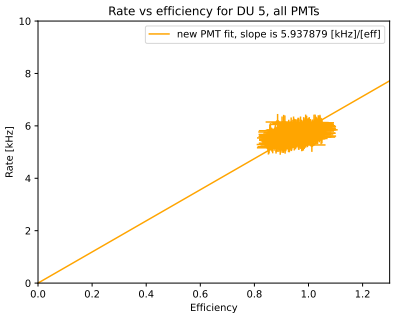
As for the other strings: we have not found a reason to suspect them, as their efficiencies are in line with other results and their rate vs efficiencies plots also show no oddities.

3.4 PMT version effects

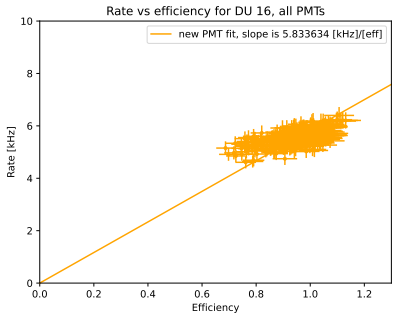
Another possible source of bias is the difference in PMT model. The most recent PMTs are of model Hamamatsu R14374 (the new model) [7], while the rest are of model Hamamatsu R12199. The main differences are that the new model PMTs have quite a lot less afterpulses (about 3- percent as opposed to 5-15), less delayed pulses (+- 2 percent instead of 3-4), but quantum efficiency stays the same [11].

The model update in PMT may however be a cause for another bias, as there is quite an integral part of the DOM changed, and so the calibration of that part must also be possibly changed. As the new model PMTs have a lot less afterpulses, this also changed the single rates detected: these should logically be lower for the new PMTs.

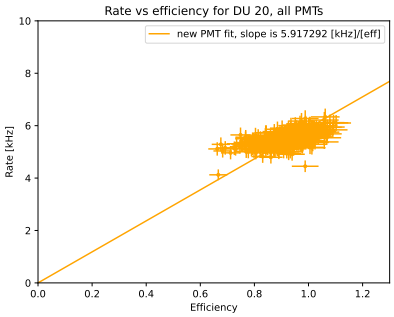
To this end we have plotted the single rates vs efficiency plots per string, but splitted them into a fit for the new PMTs and the old PMTs. The new PMTs have lower rates; therefore we expect a lower fit coefficient for them. The distributions of the efficiencies will also be plotted.



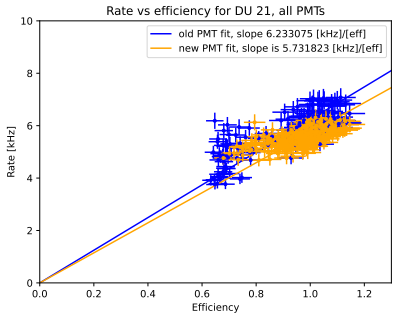
(a) *String 5*



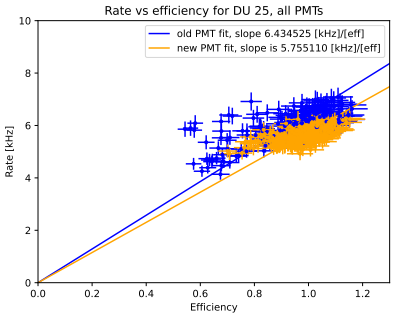
(b) *String 16*



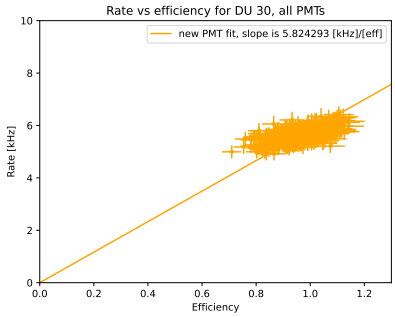
(c) *String 20*



(d) *String 21*

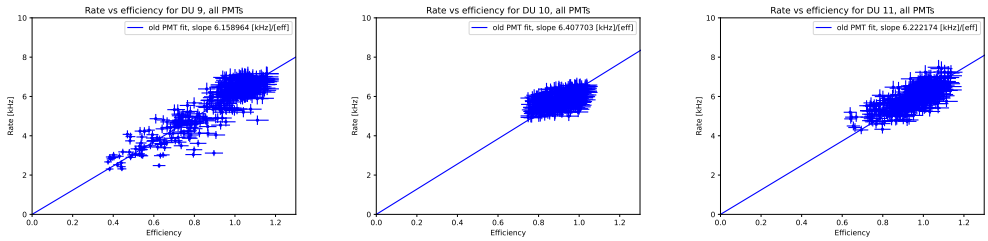


(e) *String 25*



(f) *String 30*

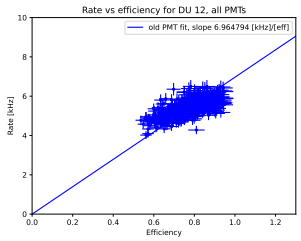
Figure 3.20: *Plots of the rates vs efficiency of all strings with new PMTs. Blue indicates the old PMT version, whereas orange indicates the new version. Some strings are equipped with DOMs with only new PMTs and DOMs with only old PMTs (there is no DOM with both PMT versions).*



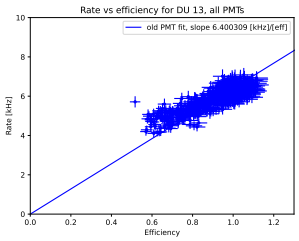
(a) *String 9*

(b) *String 10*

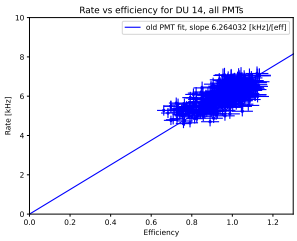
(c) *String 11*



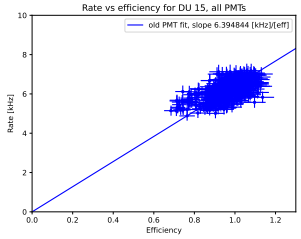
(d) *String 12*



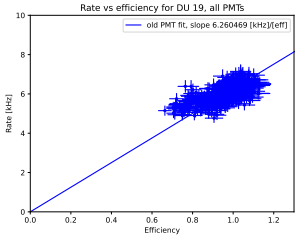
(e) *String 13*



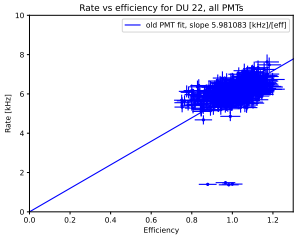
(f) *String 14*



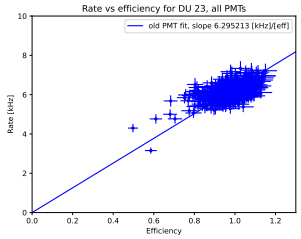
(g) *String 15*



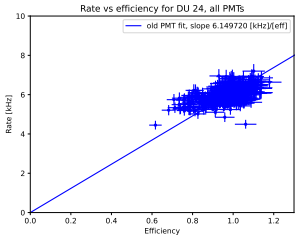
(h) *String 19*



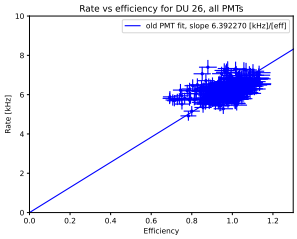
(i) *String 22*



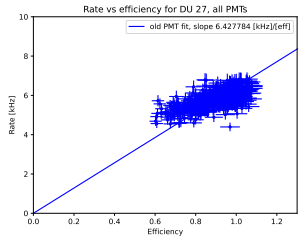
(j) *String 23*



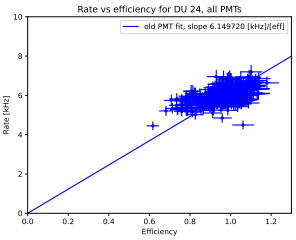
(k) *String 24*



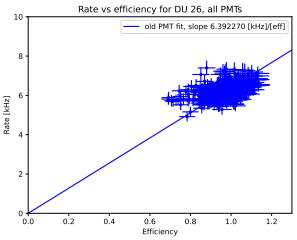
(l) *String 26*



(m) *String 27*



(n) *String 28*



(o) *String 32*

Figure 3.21: *Plots of the rates vs efficiency of all strings with old PMTs.*

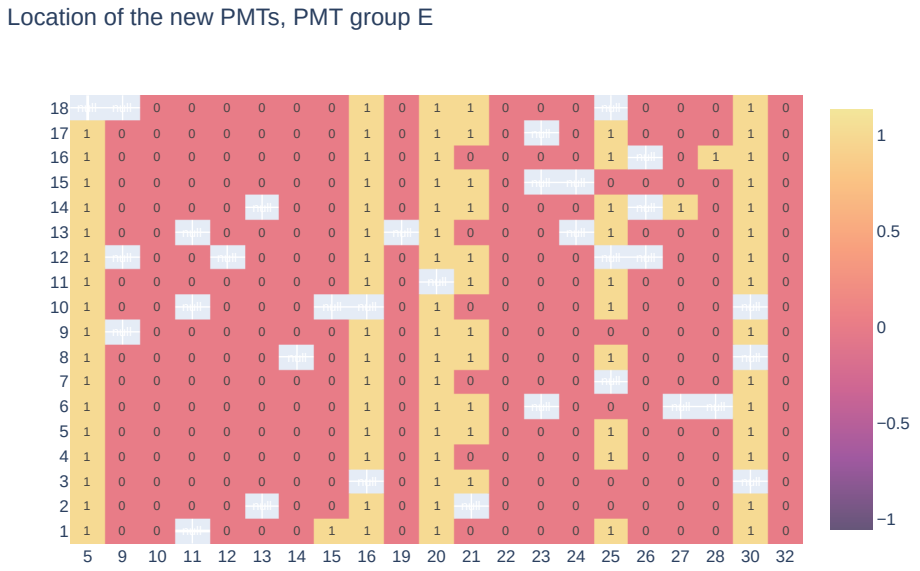


Figure 3.22: *Overview of which DOMs consist of the new PMTs. A "1" indicates they have the new version PMT, whereas "0" indicates they have the old version PMT. The choice for this ring is completely arbitrary as results across the other five rings are exactly the same. There are therefore no DOMs with a mix of new and old version PMTs. With the exception of string 15, all these strings have been deployed in or after 2022.*

For the plots that contain both the old and new versions one can see that the new PMTs have a somewhat lower slope. To further quantify the slope, the means of the slope coefficient of the new and the old PMTs have been calculated.

	old version	new version	mean
lower PMTs	6.28 +- 0.24	5.80 +- 0.07	6.26 +- 0.13
upper PMTs	6.37 +- 0.18	5.87 +- 0.08	6.12 +- 0.10
all PMTs	6.32 +- 0.20	5.83 +- 0.07	6.08 +- 0.11

Table 3.1: *Means of the slope of the fits of the rate vs efficiency plots. Unit is kHz/[efficiency]. Old version denotes PMT version R12199, while new version denotes R14374. Lower PMTs include rings A-D, while upper PMTs include rings E and F. Standard deviations have been included here. The individual lower PMT fits and upper PMT fits have been included in the appendix.*

There is some difference between the lower and upper group of PMTs, but this appears to be relatively the same for both old and new PMTs. The spread in data points in the new PMTs appear to be lower, which might be an indicator of the absence of sedimentation, as string 9, which suffers the most from sedimentation, has the most spread in its data set.

It is more probable that it is caused caused due to the improved properties of the new PMTs: less afterpulses means that the data is grouped closer, which leads to a reduced spread.

The slope coefficient is consistently lower for the new version PMTs, which indicates either lower single rates for a given efficiency or lower efficiencies for a given rate. As the efficiencies should be somewhat improved, the latter option seems unlikely.

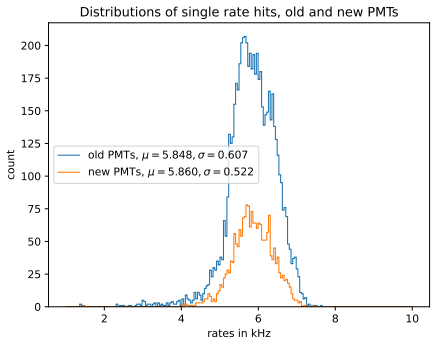


Figure 3.23: *Distribution of single rates hit for both old and new version PMTs, over all DOMs. The distribution per string is included in the appendix.*

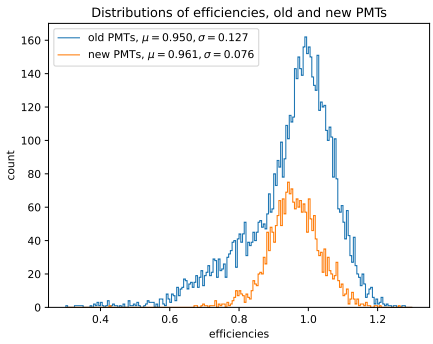


Figure 3.24: *Distribution of efficiencies hit for both old and new version PMTs.*

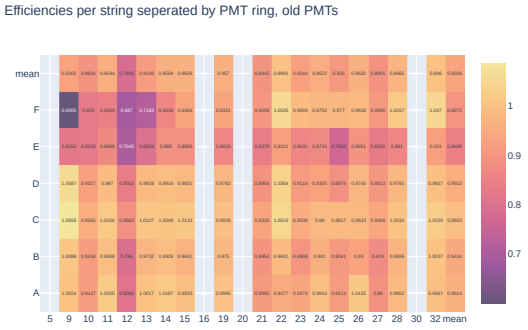


Figure 3.25: *Heatmap of mean PMTs per ring per string, old PMTs only.*

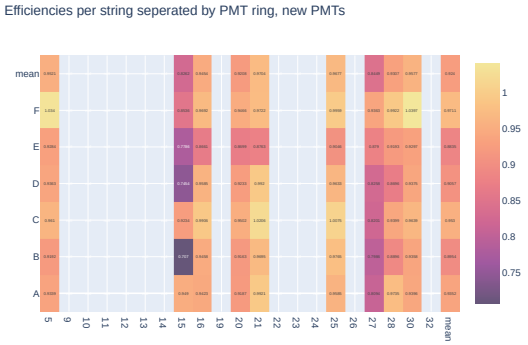


Figure 3.26: *Heatmap of mean PMTs per ring per string, new PMTs only.*

The mean efficiencies per hemisphere can be calculated and compared. For the lower hemisphere (rings A-D) the new PMTs have a mean efficiency of 4.8 percent lower than the old PMTs whereas for the upper hemisphere (rings E-F) the new PMTs have a mean efficiency of 6.2 percent higher than the old PMTs. Combined this yields for the new PMTs an efficiency of 1.1 percent higher than the old PMTs.

The distributions for the new version PMTs for the single rate seems to be approximately the same as that for the old version PMTs, albeit with a slightly higher mean and a lower spread.

For the efficiencies this is not the case; especially the tail at the left end for the old PMTs is a lot longer, ranging up to an efficiency of 0.6, than for the new PMTs. The mean of the new PMTs is slightly higher, but this is possibly caused due to the distribution being slightly right-skewered.

This can be confirmed by using a Kolmogorov-Smirnov (K-S) test, which is used to determine the similarity of two distributions. Testing the distribution of the new PMTs against that of the old PMTs yield a test statistic of 0.355 and a p-value of 1.38e-11. This confirms that the distribution for the efficiencies of the new PMTs are different than those for the old PMTs.

Chapter 4

Conclusion

The goal of this research was to identify potential biases in the efficiency evaluations of KM3NeT, using data from the ARCA of KM3NeT. To that end, this research was effectively split into four parts, each of which will be shortly treated below.

With regards to the shadowing bias we have found that of the shadowed PMTs (C2, C5, E2, E5) C2 and C5 show a an efficiency of 5.9 percent lower than the non-shadowed PMTs in their hemisphere (rings A-D), whereas E2 and E5 show a difference of 11.3 percent lower than their hemisphere (rings E-F). Relative to their respective hemispheres C2 and C5 have together a mean of the data/MC ratio that is 14.7 percent lower than the mean in the non-shadowed rings and for E2, E5 this result is actually 10.2 percent higher. This result is not as expected as C2, C5 are shadowed from below and therefore should have a higher data/MC ratio than the other PMTs in their hemisphere and E2, E5 are shadowed from above and therefore should have a lower data/MC ratio than the other PMTs in their hemisphere, and is further indicative in a bias in the efficiency evaluation.

For the sedimentation we found that string 11, 12 and 13 possibly also suffer from an improper efficiency evaluation caused by sedimentation next to string 9. We will first list here the results for the efficiencies per string, and then the results for the data/MC ratio. For the efficiencies we found that the upper rings of string 13 have a mean efficiency of 14.6 percent lower than all upper rings of all strings, whereas for string 11 we found that the mean efficiency of the upper rings is 4.1 percent lower than than the mean efficiency of upper rings of all strings. For the data/MC ratio we found that string 13 upper rings have a mean ratio of 4.7 percent lower than the mean of all non-sediment strings (that are all strings except 9, and 11 through 13 inclusive), and string 11 has a data/MC ratio of 2.2

percent lower than the data/MC ratio of all upper rings of all strings.

For the efficiencies we found that the lower rings of string 13 have a mean efficiency of 3.1 percent higher than all lower rings of all non-sediment strings, whereas for string 11 we found that the mean efficiency of the lower rings is 4.3 percent higher than than the mean efficiency of lower rings of all non-sediment strings For the data/MC ratio we found that string 13 lower rings have a mean ratio of 5.8 percent lower than the mean of all non-sediment strings (that are all strings except 9, and 11 through 13 inclusive), and string 11 has a data/MC ratio of 2.7 percent lower than the data/MC ratio of all upper rings of all strings.

The results for string 12 are similar but should be held in more suspect as string 12 is also known to be possibly suspect to the gel transperacy bias. It should be noted that the effects of sedimentation in string 13 were seen directly after the deployment of this string, and that it is therefore probable that the effects seen are caused by something else than sedimentation. We recommend this specific string for further research. The results for string 11 are indicative of sedimentation, as the figures for the upper hemisphere of that string are worse than for the lower hemisphere.

With regards to the gel transparency bias, it is known that 9 DOMs in string 10 and all of string 12 is suspect. Furthermore it is found that string 27 is possibly also suspect, having a similar low efficiency across 11 of its DOMs like string 10 and string 12. The mean efficiency of this group is 0.8428, or about 12.4 percent lower than the mean over all analysed strings. The single hit rate-efficiency coefficient is here 6.39 kHz/[efficiency], sitting about midway to both the suspect (6.65 kHz/[efficiency]) and the non-suspect coefficients (6.16 kHz/[efficiency]). As the decrease in efficiency is higher than the decrease in single hit rate-efficiency coefficient, it is likely that string 27 has in general a different gel transparency than the other strings.

For the PMT version bias we have compared the new version PMT (Hamamatsu R14374)versus the old version PMT (Hamamatsu R12199-02). For the lower hemisphere the new version PMTs have an mean efficiency of 4.6 percent lower than the old ones whereas for the upper hemisphere the new PMTs have an mean efficiency of 6.1 percent higher than the old version PMTs. The results for the lower hemisphere is more indicative of the real situation as the upper hemisphere is impacted by sedimentation, from which the older PMTs suffer the effects more. The single hit rate appears to be mostly the same across both old and new version PMTs, whereas the relation between the single hit rates and the efficiency being changed: the fit coefficient between those two is lowered by 8.2 percent for the lower PMTs, which is indicative of a slightly worse efficiency.

Chapter 5

Acknowledgements

I would like to thank my supervisor Dorothea Sambtleben for her guidance and extensive insights during this project; without her help I would not be able to produce this research. I also want to thank Aart Heijboer for supplying part of the code necessary for my project. Additionally I want to thank the people at Nikhef and especially the KM3NeT research group for their support also.

Bibliography

[1] J. Carr, G. Hallewell, A. Holford, U. Katz, P. Kooijman, L. Moscoso, P. Piattelli, and P. Rapidis, *Conceptual design report for KM3NeT*, Technical report, KM3NeT Collaboration, 2008.

[2] L. Bergstrom and A. Goobar, *Cosmology and particle astrophysics*, Springer Praxis Books / Astronomy and Planetary Sciences, Springer, 2nd edition, 2006.

[3] KM3NeT Collaboration et al., *Dependence of atmospheric muon flux on seawater depth measured with the first KM3NeT detection units*, arXiv (2019).

[4] P. Bagley, J. Craig, A. Holford, A. Jamieson, T. Niedzielski, I. G. Priede, M. de Bell, J. Koopstra, G. Lim, and E. De Wolf, *KM3NeT:*, Technical report, 2015, [Online; accessed 26. Jul. 2023].

[5] K. Melis, *In-Situ Calibration of KM3NeT*, ResearchGate , 1059 (2017).

[6] S. Adrián-Martínez et al., *Letter of intent for KM3NeT 2.0*, J. Phys. G: Nucl. Part. Phys. **43**, 084001 (2016).

[7] S. Aiello et al., *The KM3NeT multi-PMT optical module*, J. Instrum. **17**, P07038 (2022).

[8] G. Carminati, M. Bazzotti, A. Margiotta, and M. Spurio, *Atmospheric MUons from PArametric formulas: a fast GEnerator for neutrino telescopes (MUPAGE)*, Comput. Phys. Commun. **179**, 915 (2008).

[9] *NIST/SEMATECH e-Handbook of Statistical Methods*, 2022, [Online; accessed 20. Aug. 2023].

[10] A. Konings, *Evaluation of potential biases in the detection efficiencies of the KM3NeT ARCA and ORCA detectors*, 2022, [Online; accessed 19. Aug. 2023].

[11] A. Simonelli, C. M. Mollo, P. Migliozzi, and on behalf of the KM3NeT collaboration, *Characterisation of an improved 3" Hamamatsu photomultiplier for the KM3NeT Neutrino Telescope*, J. Phys. Conf. Ser. **2429**, 012031 (2023).

Chapter 6

Appendix

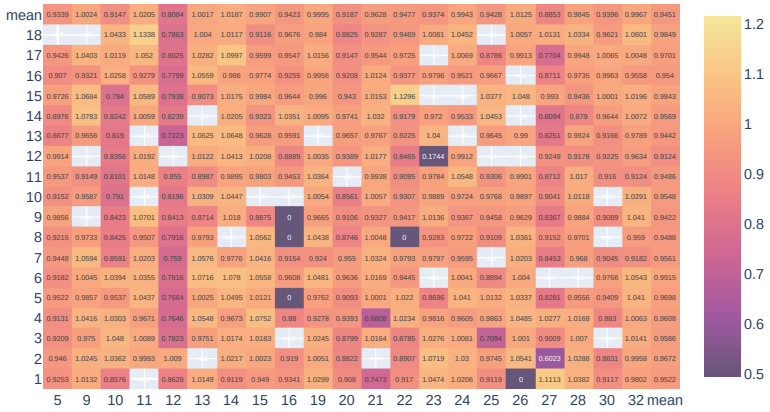
6.1 Appendix A: Plots

In this section some of the raw plots are included. These were not included in the main report as they tend to take up a very large amount of space, but might still be of interest to the interested reader.

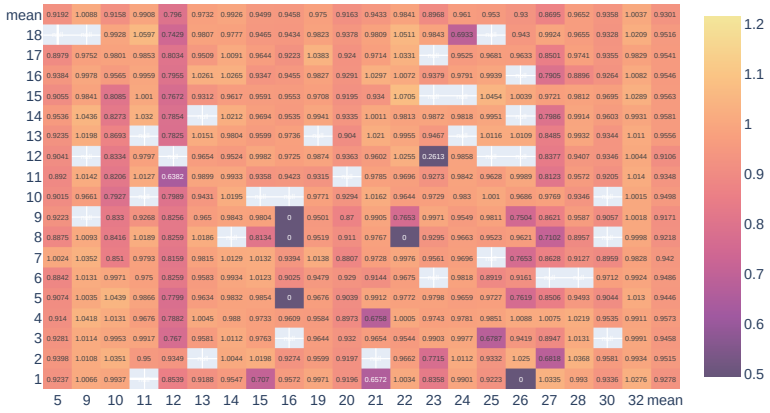
6.1.1 Efficiency plots

Printed below are the efficiencies per DOM sorted by ring. Fig. (3.8) is effectively a summary of the first four plots.

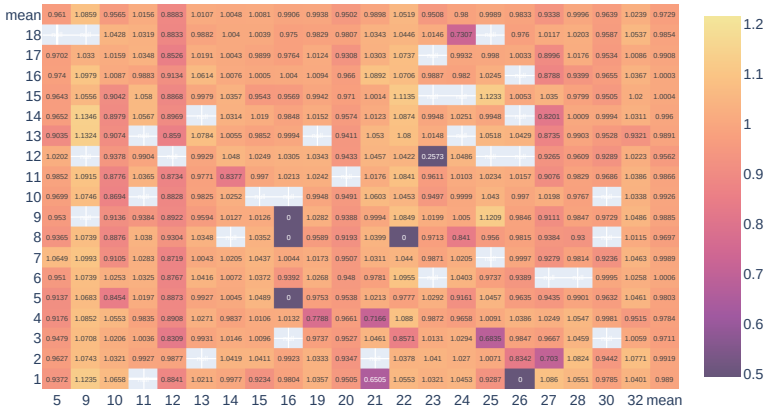
Efficiencies per string seperated by PMT ring, all PMTs, PMT group A



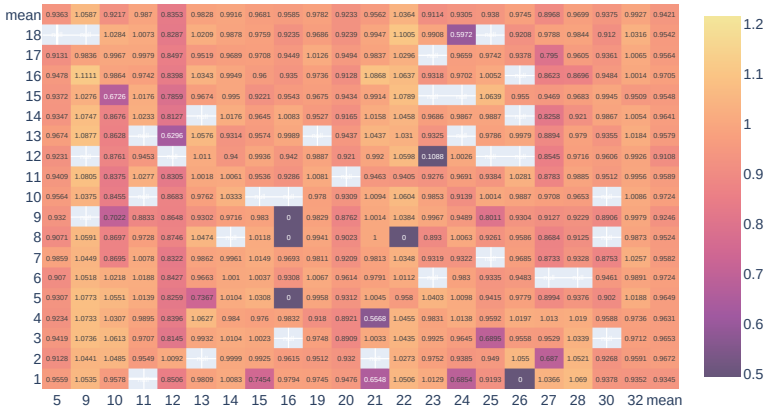
Efficiencies per string seperated by PMT ring, all PMTs, PMT group B



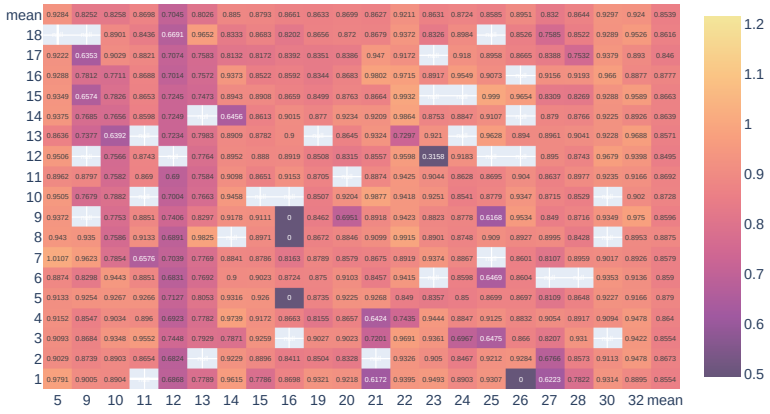
Efficiencies per string seperated by PMT ring, all PMTs, PMT group C



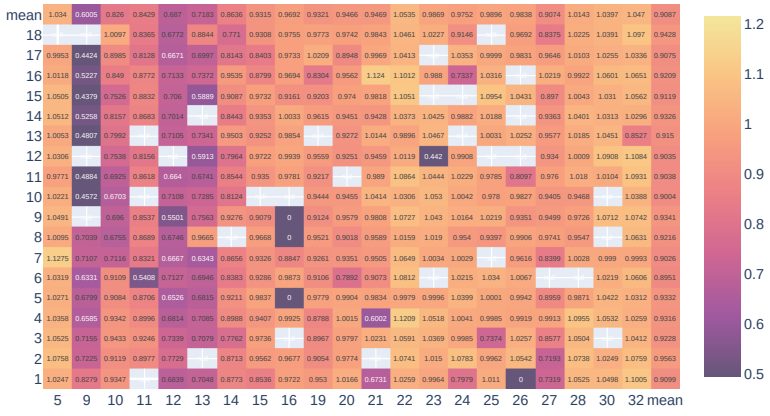
Efficiencies per string seperated by PMT ring, all PMTs, PMT group D



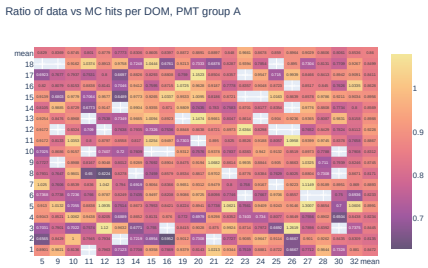
Efficiencies per string seperated by PMT ring, all PMTs, PMT group E



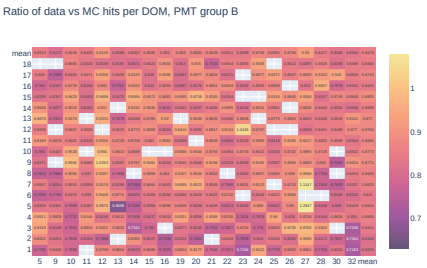
Efficiencies per string seperated by PMT ring, all PMTs, PMT group F



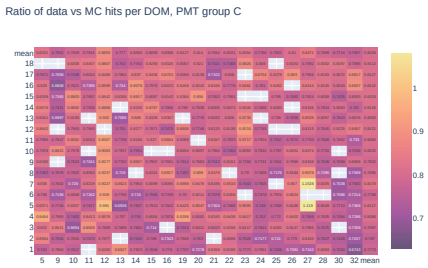
6.1.2 Data/MC plots



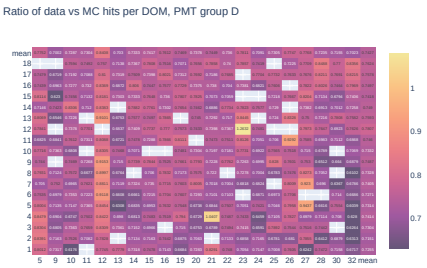
(a) *Ring A*



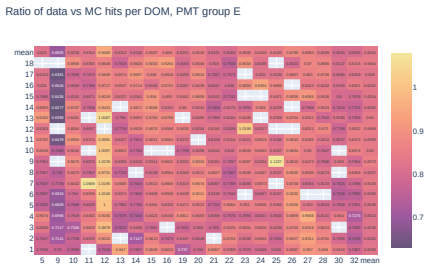
(b) *Ring B*



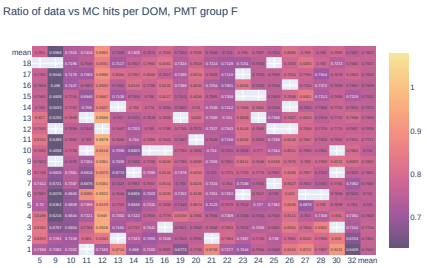
(c) *Ring C*



(d) *Ring D*



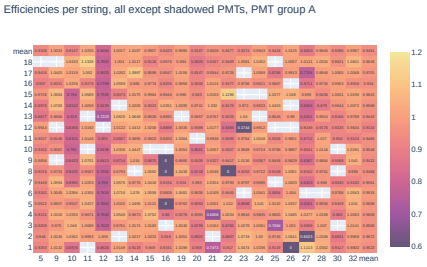
(e) *Ring E*



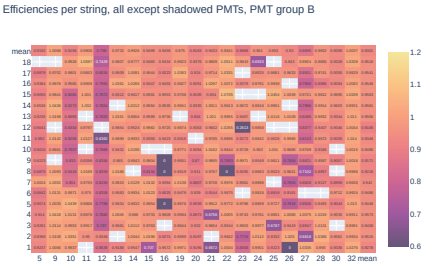
(f) *Ring F*

Figure 6.1: *Plots of the data-MC ratio sorted by ring.*

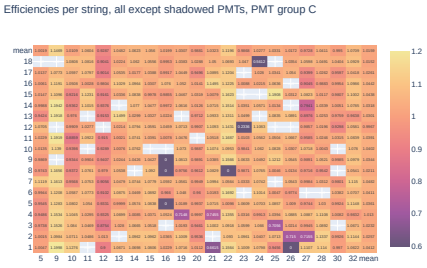
6.1.3 Efficiencies



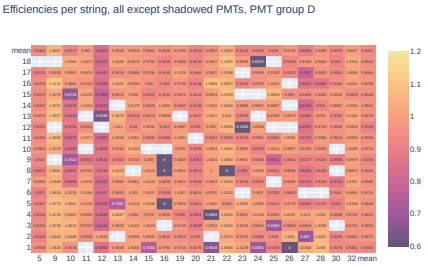
(a) *Ring A*



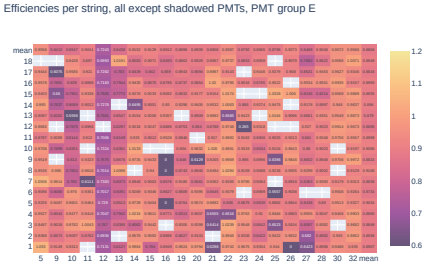
(b) *Ring B*



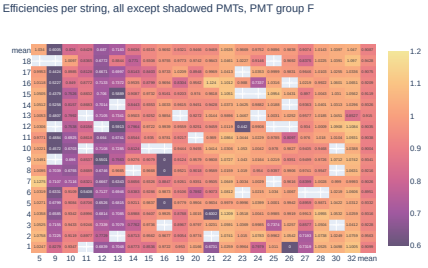
(c) *Ring C*



(d) *Ring D*



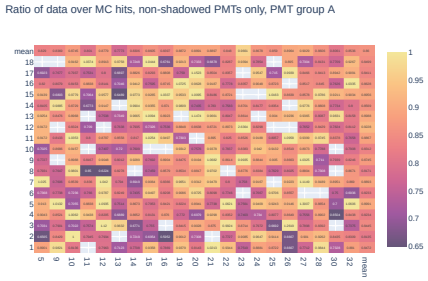
(e) *Ring E*



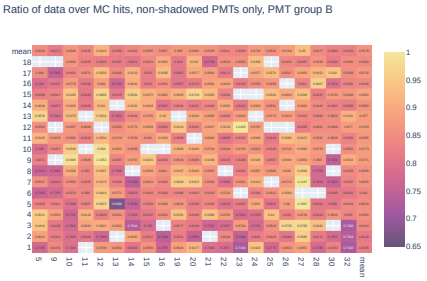
(f) *Ring F*

Figure 6.2: *Plots of the efficiencies sorted without any shadowed PMTs.*

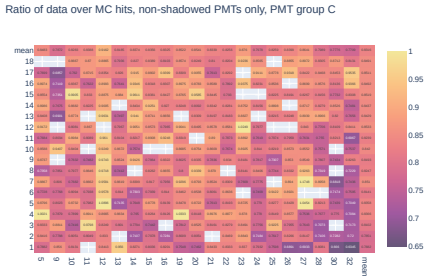
6.1.4 Data over MC ratio



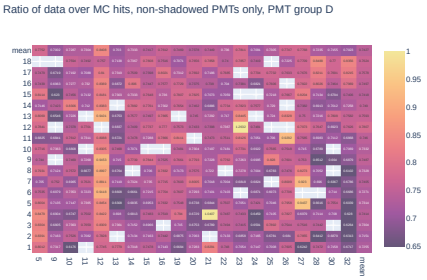
(a) *Ring A*



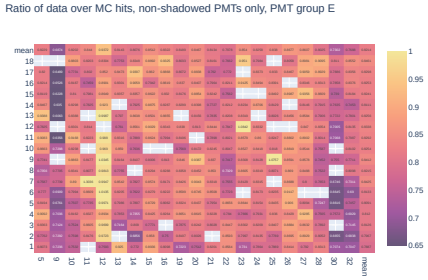
(b) *Ring B*



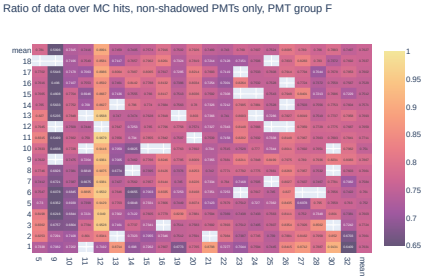
(c) *Ring C*



(d) *Ring D*

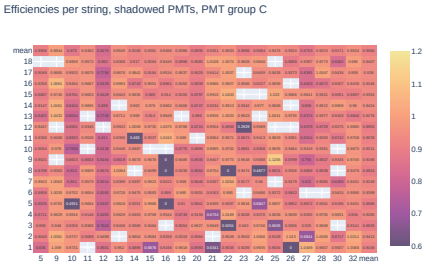


(e) *Ring E*

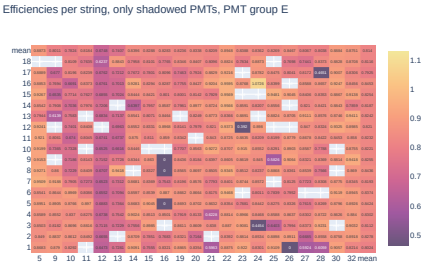


(f) *Ring F*

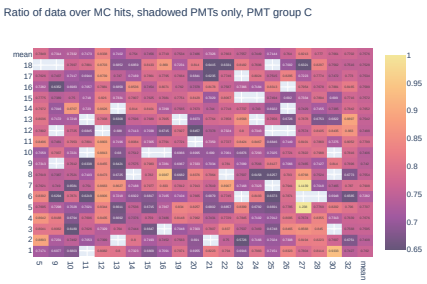
Figure 6.3: *Plots of the data/MC sorted without any shadowed PMTs.*



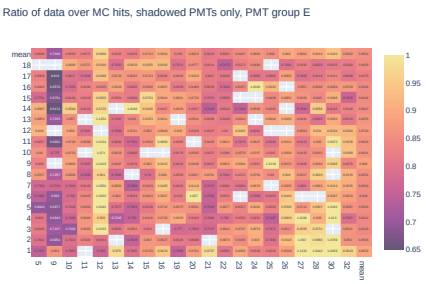
(a) *Ring C, shadowed*



(b) *Ring E, shadowed*



(c) *Ring C, shadowed*

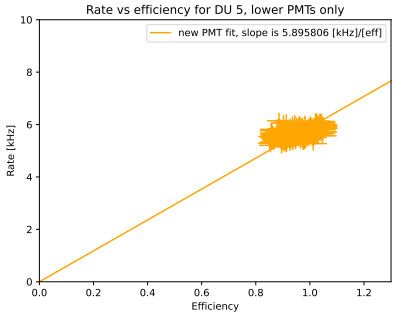


(d) *Ring E, shadowed*

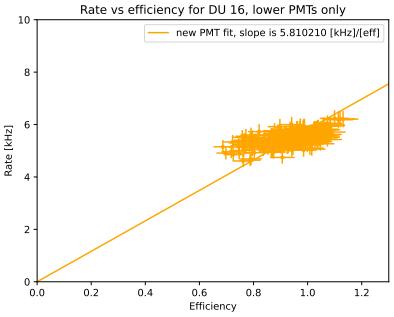
Figure 6.4: *Plots of the efficiencies and data/MC of only shadowed PMTs.*

6.1.5 PMT version plots

Rates vs efficiencies

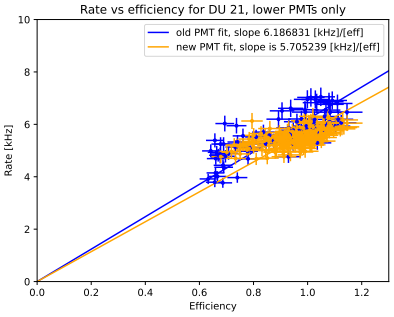


(a) *String 5*

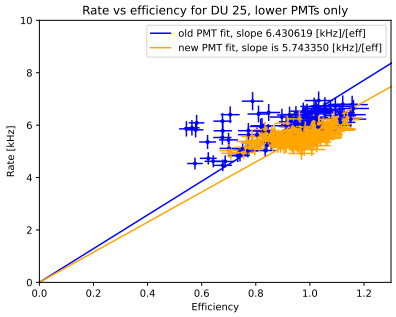


(b) *String 16*

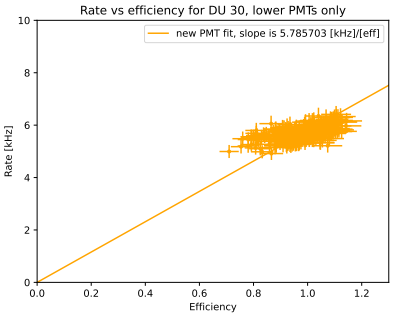
(c) *String 20*



(d) *String 21*

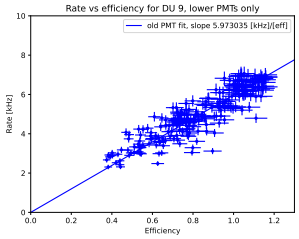


(e) *String 25*

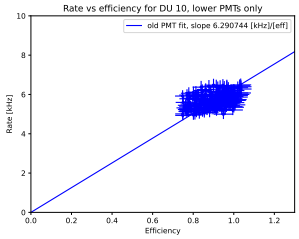


(f) *String 30*

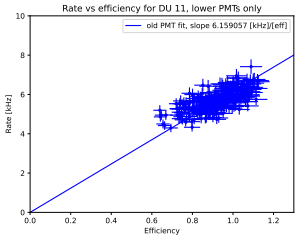
Figure 6.5: *Plots of the rates vs efficiency of all strings with new PMTs. Blue indicates the old PMT version, whereas orange indicates the new version. Lower index only (PMT ring A-D, DAQ-index 0-18 inclusive).*



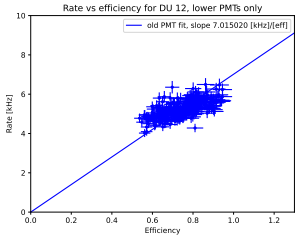
(a) *String 9*



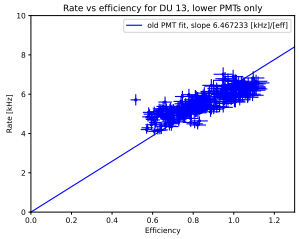
(b) *String 10*



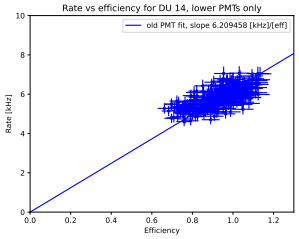
(c) *String 11*



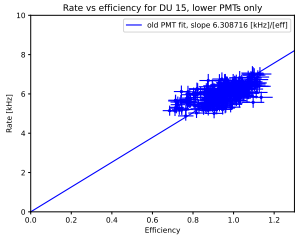
(d) *String 12*



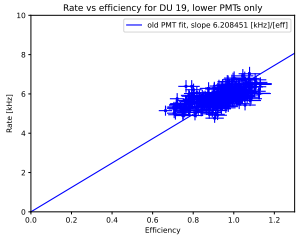
(e) *String 13*



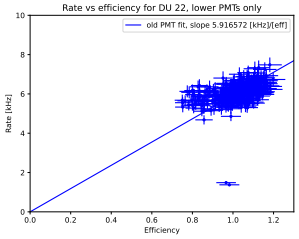
(f) *String 14*



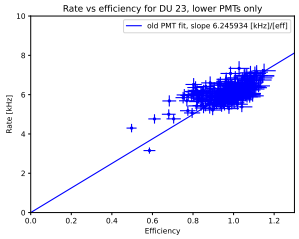
(g) *String 15*



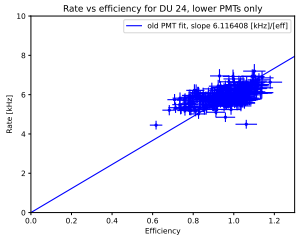
(h) *String 19*



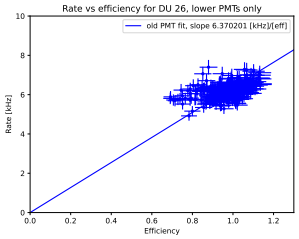
(i) *String 22*



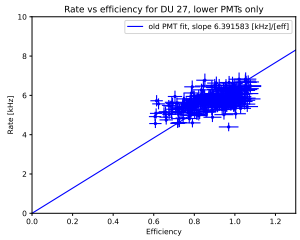
(j) *String 23*



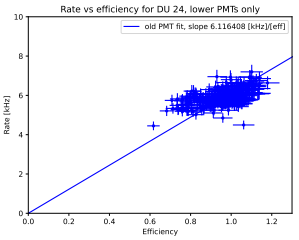
(k) *String 24*



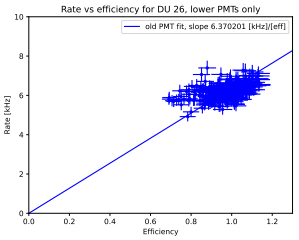
(l) *String 26*



(m) *String 27*

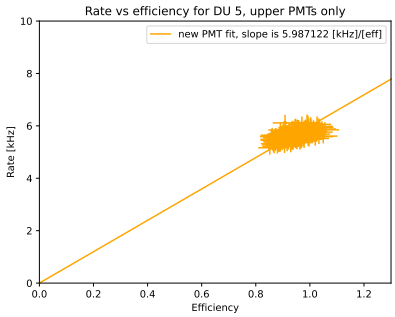


(n) *String 28*

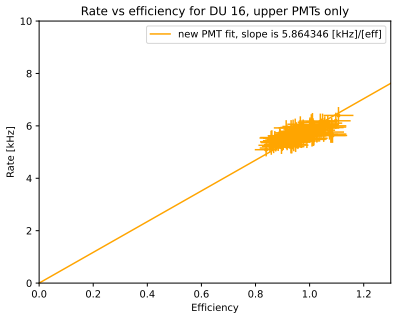


(o) *String 32*

Figure 6.6: *Plots of the rates vs efficiency of all strings with lower and old PMTs.*

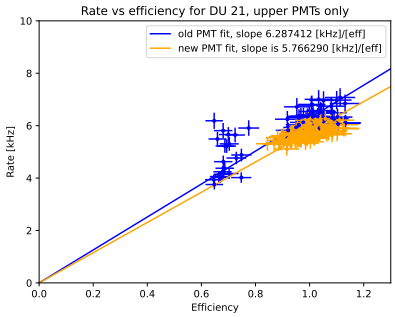


(a) *String 5*

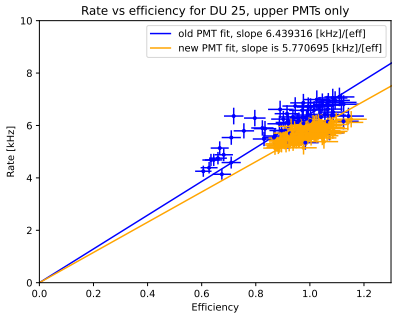


(b) *String 16*

(c) *String 20*



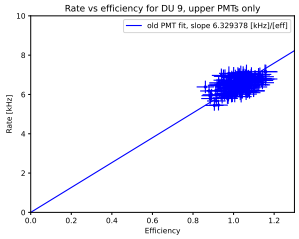
(d) *String 21*



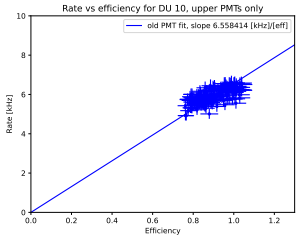
(e) *String 25*

(f) *String 30*

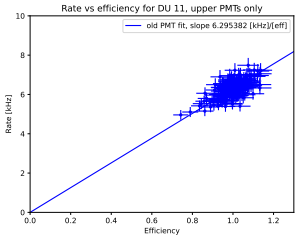
Figure 6.7: *Plots of the rates vs efficiency of all strings with new PMTs. Blue indicates the old PMT version, whereas orange indicates the new version. Upper index only (PMT ring A-D, DAQ-index 0-18 inclusive).*



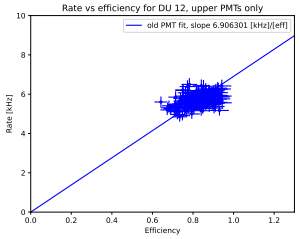
(a) *String 9*



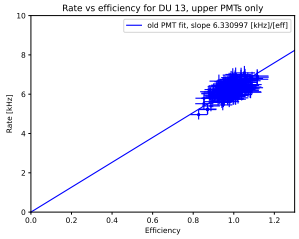
(b) *String 10*



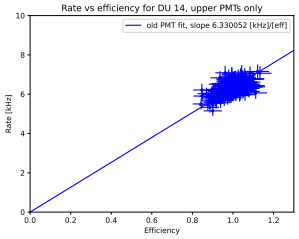
(c) *String 11*



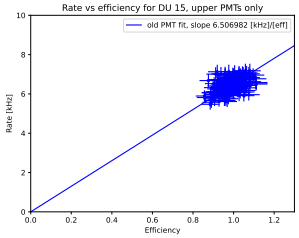
(d) *String 12*



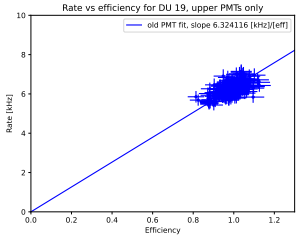
(e) *String 13*



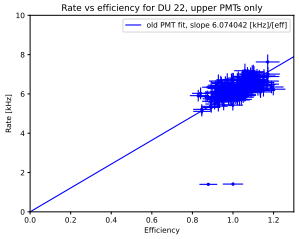
(f) *String 14*



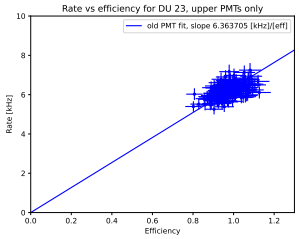
(g) *String 15*



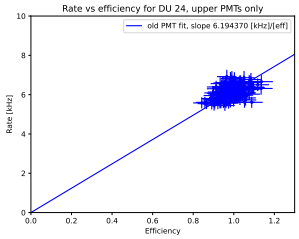
(h) *String 19*



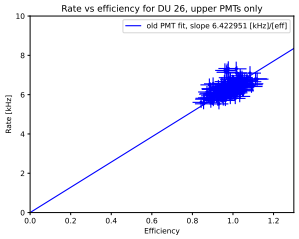
(i) *String 22*



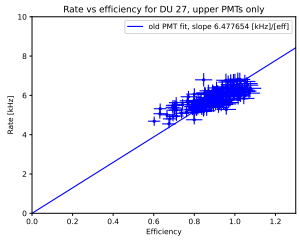
(j) *String 23*



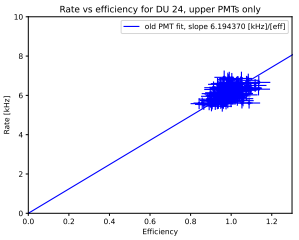
(k) *String 24*



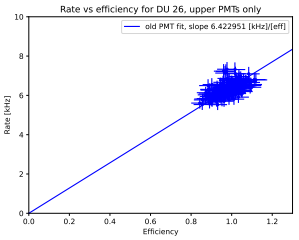
(l) *String 26*



(m) *String 27*



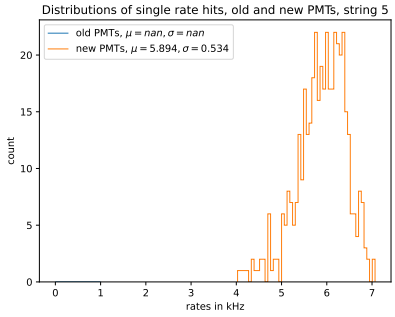
(n) *String 28*



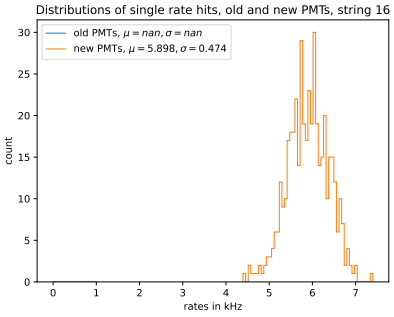
(o) *String 32*

Figure 6.8: *Plots of the rates vs efficiency of all strings with upper and old PMTs.*

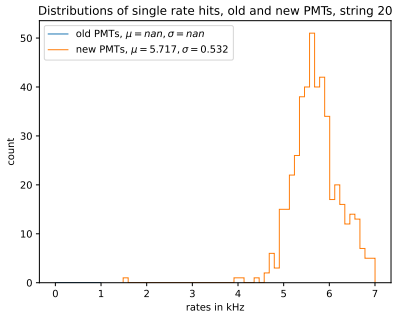
6.1.6 Distribution of rates



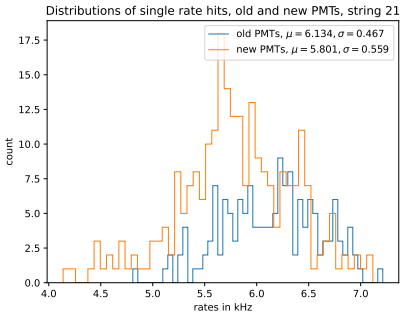
(a) String 5



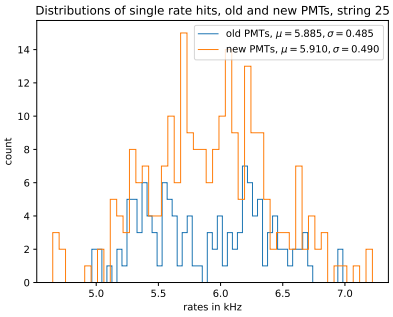
(b) String 16



(c) String 20

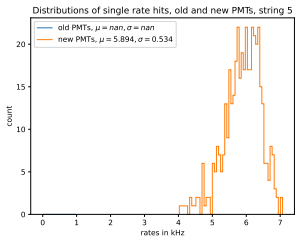


(d) String 21

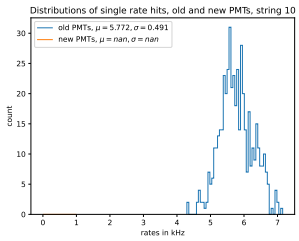


(e) String 25

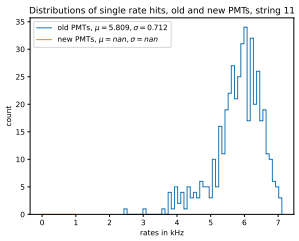
Figure 6.9: *Distribution of the rates. All zero-values have been excluded as to not skew the data.*



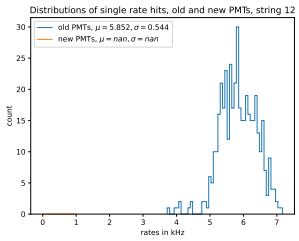
(a) *String 9*



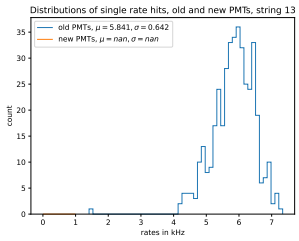
(b) *String 10*



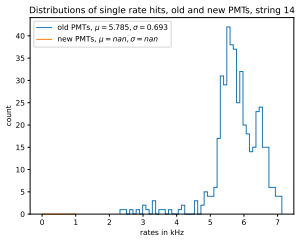
(c) *String 11*



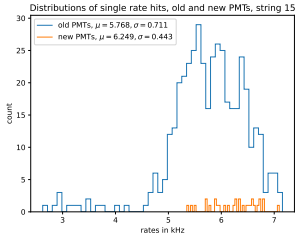
(d) *String 12*



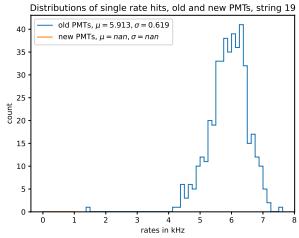
(e) *String 13*



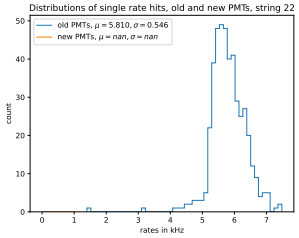
(f) *String 14*



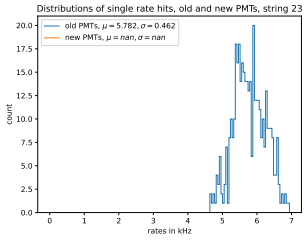
(g) *String 15*



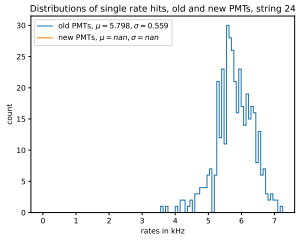
(h) *String 19*



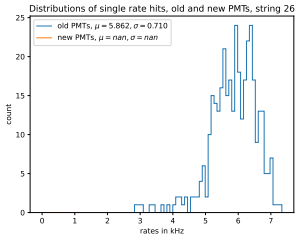
(i) *String 22*



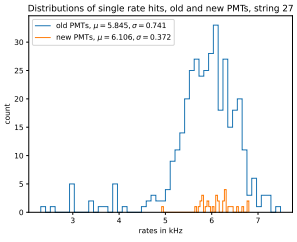
(j) *String 23*



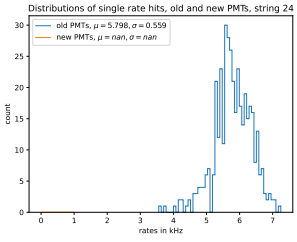
(k) *String 24*



(l) *String 26*



(m) *String 27*



(n) *String 28*

Figure 6.10: *Distribution of the rates. All zero-values have been excluded as to not skew the data.*