REPORT ON JINST_014P_0220

DATE: FEBRUARY 21, 2020

AUTHOR(S): ICECUBE COLLABORATION

TITLE: In-situ calibration of the single-photoelectron charge response of the Ic

RECEIVED: 2020-02-03 18:24:40.0

Report of referee 1

This contribution describes an improved experimental approach to identify the single photoelecton charge distribution of the IceCube PMTs. Several examples are given of how results of this technique can be used to evaluate the performance of different technological choices as well as to describe better the detector in data analysis. However, there are several points needed to be addressed further. This contribution should be published after the authors have made the revisions mentioned in the following.

Section 1.2
A reference for the "WaveDeform" (waveform unfolding process)is needed

Page 7, first paragraph: The weight factor has been set to the specific value of 28.4 in order to compensate for the different sizes of the MinBias and BeaconLaunch datasets ? Please clarify.

Section 2.1
Page 7, first paragraph: What are the "late pulses from the trigger" ?

Page 8, last paragraph: How this double occupancy (6.5%) has been estimated? (by fitting the charge distributions, as in Fig. 5?). Please explain.

Section 2.2

Page 9, Fig. 3: The term "cumulative distribution" has a specific statistical meaning. Please rephrase.
The MinBias and BeaconLaunch histograms, shown in this Fig., are normalized to the same livetimes (i.e. using the weight factor mentioned in Section 1.2)?
Why the lowering of the WaveDeform (modified WaveDeform) threshold changes the BeaconLaunch distribution at large charge values?
How the normalization factor M is evaluated ?

Page 9, last paragraph: The statement that "The modified WaveDeform datasets show a minimal increase in the contribution of noise to the low-charge region" is not supported by Fig 3.

Section 2.3

Page 10, 4th paragraph: The assumption that "This analysis assumes the same shape of the steeply falling exponential component (Exp1) for all DOMs in the detector to avoid large fluctuations in the DOM-to-DOM efficiencies." needs confirmation (e.g. by fitting several average charge distributions, corresponding to samples of randomly selected DOMs, estimate the parameters of exp1 and then examine their spread).
How is possible to avoid large DOM-to-DOM efficiency fluctuations by the above data-analysis assumptions?

Page 10, 4th paragraph: "we background-subtract the BeaconLaunch distribution from the MinBias data". Obviously the BeaconLaunch distribution is dominated by noise. Nevertheless, to be consistent with the previous Section, the noise distribution evaluated in Section 2.2 (i.e. blue histo in Fig 3) should be used for background subtraction. Anyway you should comment on your choice.

Page 10, 5th paragraph: In principle, a first, biased fit will influence the second fit as well (by attributing more weight to the "wrong" data points). Further explanation is needed to support that the applied fit-technique produces unbiased estimations.

Page 10, 6th paragraph: Define better the term "average residual". Do you mean a correction function as the one shown in Fig.4?
The residual correction cannot be used as a scaling factor. Do you mean

an additive correction (which is proportional to the SPE template, with the proportionality factor given by the function of Fig 4) ?

Section 2.4

Page 11, 1st paragraph: The estimated values of P_{e1} and w_1 are in principle correlated. Please give the correlation factor.
Page 11, 2nd paragraph and last paragraph: Please avoid the use of the term "cumulative".

Page 11, Table 1: When fitting the SPE charge distribution of an individual DOM, the estimated parameter values are mutually correlated. A comment on the size of the correlations is needed.
In principle, because of these correlations, it is not correct to treat the estimated parameters ($Pe_{1,2}$ w_2, μ and σ) independently, i.e. to form independently their averages, as in this Table. Please comment.

Page 12, Fig. 5: The blue curve, representing the SPE template, differs from the red curve (the result of the fit) because it does not include the contribution of multiple pes. However, the Gaussian component of the SPE template is systematically lower than the data even at charges where the 2PE contribution is negligible. An explanation is needed.

Section 3.1

Page 12, 1st paragraph: The numbers given in this paragraph to quantify the differences between technologies, are not supported by the numbers given in Table 1. Please clarify.

Page 13, 1st Paragraph: The numbers given in this paragraph are not supported by the numbers given in Table 1.

Page 14, Fig 8: Please explain how the confidence intervals have been evaluated.

Section 3.2

Page 14, 1st Paragraph: Rephrase the sentence "The largest deviation ...

in Ref. [9]."

Section 3.3: The whole Section 3.3 (including 3.3.1) should be rewritten.

Page 15: Define better the physical meaning of the function f(q) (is the probability density function of the SPE charge?).
What is meant by the sentence "Generally, f(q) and ... for a dim source"?
What is q_{pk}? (the mean value of the Gaussian peak of the SPE charge distribution?)
What is meant by "multiply observed charges by q_{pk}"?

Page 16, 1st paragraph: It is not clear if the data defined as "Semi-bright source measurements" are the MinBias events, especially selected for calibration purposes, or normal data collected just with the trigger condition described in Section 1. In the latter case the trigger pulse (which must be above threshold, i.e. > 0.23PE) is included in the collected data. Why then the lower integration limit in eq. 3.4 is not $0.23q_{pk}$?
Even if the applied threshold is $0.13q_{pk}$, why the low integral limit is $0.10q_{pk}$?
The software termination condition is mentioned in Section 2.2 (and not 2.1). The statment "For light levels that are large, the trigger is satisfied regardless of the response to individual photons..." is not supported by the trigger requirement described in Section 1.2, please explain.

Section 3.3.1

Page 16 : The meaning of the sentences "A natural question to ask ... on the shape of f(q) ...in Table 2" is not clear. What is the purpose of this Section. It is obvious, from the definitions given in Section 3.3, that the ratios tabulated in Table 2 are independent of η_0 and they only depend on the shape of the SPE charge distribution. However, the importance of these ratios is not obvious. Please explain.
Finally, it is very confusing that the entries in Table 2 are called "measurements".

Section 3.4
Page 17: The significance of this Section to this work is not apparent. Please explain.

Section 3.5

Page 17, 2nd paragraph: The sentence "Two simulation sets consisting of the same events were processed through the IceCube Monte Carlo simulation chain to the final analysis level of an update to the IC86.2011 sterile neutrino analysis [24]" is very confusing. Do you mean that two different simulation sets of events are employed, which differ only on the SPE templates used to describe the detector?

Page 17, 3rd paragraph: Please explain further what are the data points (trigger condition etc.).

Conclusion
Page 18, 1st paragraph: The statement that "The result of this measurement was shown to be useful for improving the overall data/MC agreement" applies to a particular analysis only. Please rephrase.

Report of referee 2

Re: Preprint: JINST_014P_0220
Authors: IceCube Collaboration
Title: In-situ calibration of the single-photoelectron charge response of the IceCube photomultiplier tubes

Dear Editor,

We reviewed this interesting paper that describes an improved in-situ calibration of the single-photoelectron charge distributions for each of the in-ice photomultiplier tubes (PMTs) of the IceCube Neutrino Observatory. The characterization of the individual charge distributions is important for PMT calibration and for the understanding of the detector performance. In the paper, the single photoelectron identification procedure and the method to extract the single-photoelectron charge distribution are discussed in detail. The method to extract the single-photoelectron charge distribution uses a deconvolution of the multiple-photoelectron distributions.

The paper is of interest and contains sound information and valid scientific methods, thus proving the validity of the results obtained. Therefore, we support the publication of the scientific content. However, we find that

the discussion can be substantially improved by complementing the text with clarifications and a systematic organization of the presentation and discussion of the content. A major revision of the text and discussion of the results is required. A few suggestions are listed below.

with my best regards

A Reviewer

>>>

a) Introduction:

- This section contains detailed information and becomes very technical. It contains the technical details of the detector structure and components that are needed later. It would be better to start with an overview of the main goals discussed in this paper, i.e. further expanding on the PMT calibration and why this is relevant to the experiment. It would certainly benefit if a sentence explains or justifies what are the resolution values that are needed to reach a specific measurement. Later, and in a different section called "Experimental setup" or "Detector", the detailed description of the detector can be summarized.

- A sketch of the detector array where the names of the different parts are presented would be a useful and helpful addition for the reader.

- "There are two versions of AC coupling". If AC coupling is explicitly mentioned, it would be useful to explain what/why they are and/or how those affect the detector response.

- please explain what is "...bifilar wound 1:1 ...". It is very technical as it is.

- page 3, line 3: what is a "scaled single PE pulse"? is it a "superimposition of single PE pulses"? Please clarify.

- page 3, 1st paragraph: 1.2mV is equivalent to ~0.23PE. How is the equivalence set? please explain. Also, later in the text, a different equivalence of 2V and 560PEs is mentioned (beginning of page 6). It would be useful to clarify the source of the apparent inconsistency.

- p.3, 2nd paragraph: high-gain and low-gain. It is not clear how many and where these different stages are.

- are the three channels separately amplified with different gains? and why? perhaps the sentence should read "are each subject to different levels of separate amplification"?

- three channels are mentioned but then there is a reference to only one high-gain channel. Please clarify and structure the sentence such that a complete description is given.

It is incomplete and confusing as it is. Also, it does not serve any purpose unless there is a clear indication of the reason why this detailed information is relevant. Rather, it would be useful to clarify the reason why the signal is split into four channels.

- p.3, 3rd paragraph: It may be intuitive, but please define "SPE", as it is the first time it appears.

Section 1.1:
- line 2: it would be better to replace "corresponds" with " is calibrated to correspond"

- line 3: "that create a spurious structure"

- of the items listed there are a few that are intrinsic properties/imperfections of the PMTs and can therefore be testes separately in the lab. Have those effects been tested before? are similar results obtained to those presented here?

- p.5, 5th paragraph: the PMT used for the standard calibration has a SQE. Why not making it clear explicitly here? it would help the reader, instead of having to look for it a few pages back. Also, are the other PMTs responding in the same way?

- "-32C": what is the temperature at which the in-ice PMTs operate? It could be clarified in the text. What is the temperature dependence response of the PMTs?

- page 5, "electronic noise": ATWD is not defined, except that probably is an acronym for some unspecified ADC. As the acronym is used several times later, please explain/clarify what that is.

- p.5, 6th paragraph: Please add the Ref to the 2013 laboratory measurements. Please spell out "laboratory".

- A second exponential component has been added with respect to the standard laboratory calibration. What is the reason of adding the 2nd exponential? What is the "low-charge contribution" due to?

- Eq.1.1: It would help the clarity of the paper to clearly define the different parts of Eq.1.1, provide a justification for the choice, and state how each component is estimated For example, do exp1 and exp2 provide a description of the background? if so, why not stating this clearly? Please clarify.

- Eq.1.1: Please provide the explicit definition of "Erfc" used. This can be done in the text (preferably) or in a footnote.

- Sec.1.2, 3rd paragraph: The sentence including "droop and undeshoot" and "expected reaction" is technical/jargon and it should be clarified. What is the "expected reaction"?

- Sec.1.2, "MinBias dataset" subsection: it is not clear what "1:1000 events" is. Does it refer to a dataset that collects 1:1000 of the randomly-selected events? Please clarify.
- "16 triggered DOMs": are there 16 DOMs that trigger the events? are they adjacent? please clarify.

- Sec.1.2: This is an example that occurs elsewhere and in similar acronyms in the text. Usage of acronyms like "WaveDeform" or "BeaconLaunch" that are used several times throughout the text could be avoided. Those acronyms are jargon, they are not intuitive, and do not correspond to the names given in parentheses, thus making hard to connect to the actual meanings. It is certainly a straightforward correspondence for those in the IceCube collaboration, but it is hard for the average reader that has to constantly refer to the prior and non-intuitive explanation. Please consider using alternative names, or provide correct acronyms.

- Fig.1, caption. "OTS": same as the comment above.

- Sec. 2.1, 3rd paragraph: "ATWD time window"? does it correspond to the "time integration window"?

- Sec.2.1: 3rd paragraph: "light grey region". It is hard to identify the different areas. Sometimes it is referred to the color, other times to the time span. Letters for each area would simplify the identification, remove ambiguities, and make the text uniform.

- Sec. 2.2, 1st paragraph: "threshold at ... 0.13PE" is hard to see/justify in Fig.2 (right) is referred as due to a termination, and it is referred to in the Fig.2 caption as "fall off in charge ... 0.13PE". Please give consistent explanations.

- Sec.2.3: 2nd paragraph: "...contamination from two-photon events...". Why is this more important than other effects? is this due to what? why is this effect highlighted or important? Please clarify in the text.
The next paragraph mentions "three photon contribution". Same as above.

- Sec.2.3, 4th paragraph: "...assumes the same shapes ... for all DOMs...". The assumption is based on what? This assumes that the dependence is purely due to the PMT intrinsic characteristics and not to local in-situ differences? Also, the PMTs are of different nature/characteristics. Does the assumption remain valid? Please clarify in the text.

- Sec.2.3, last paragraph: "predetermined values" from what/where? what does it refer to? Is the value fixed within certain predetermined boundaries? or does it refer to values determined earlier from tests in the laboratory? Please clarify.

- Fig.3: It would help the reader to extend the range of the horizontal axis to 4. As it is, it is not clear what is the current boundary. Please adjust.

- page 11, 3rd paragraph: This paragraph is confusing. A few points: "failed fits". Are they due inefficiency or to what? what is the "number of valid pulses"? what is a "valid pulse"? "goodness" by large chi-square? Please clarify.

- page 3, 3rd paragraph: "between 107 and 111 DOMs": In caption of Fig.1, it says that the DOMs shown in white have been removed from service. However, here it says that DOMs have been removed depending on the season. The two statements seem to be contradicting. Please clarify.

- Table 1:

- why only a few and not all parameters of the fit are indicated in the Table?

- What is NQE?

- "width": does it refer to "sigma" of the distribution? Please clarify.

- column 2: does it refer to Exp1 or Exp2?

- caption: what is the "standard error"? please clarify.

- Fig.5:
- what is "validity", "PVR", etc? please clarify in the caption.
- legend: add something that indicates the data points.
- The fit results correspond to "convolution fit" in the legend? Please clarify in the caption.
- caption: "black histogram" is rather "black points"
- caption: "Pe1 and w1 are fixed as discussed at the end of Sec.2.2." Where exactly is this discussed? where are these values indicated? the discussion in Sec.2.2 does not indicate the values. If so, please clearly discuss and rephrase the Section accordingly.

- Sec.3.1, 1st paragraph: "it is evident that... depends on the DOM hardware". Results in Table 1 are all within one sigma from each other. They are all consistent with each other. Please clarify.

- Sec.3.1, 1st paragraph: how is the "peak-to-valley ratio" defined? Please clarify in the text.

- page 13, 2nd paragraph: " a ... change of the fit parameters... according to their PMT serial number...". Given the difference between PMTs, is it reasonable to use the assumption in Sec.2.3 ("...this analysis uses the same shape...")? Please clarify in the text.

- Sec.3.2, 1st paragraph: "excluding Exp1". Why is Exp1 excluded? Please explain.

- Sec.3.2, 1st paragraph: "randomly": Seasons are not random- but time/temperature-dependent. Please justify the choice.

- Sec.3.3, 4th paragraph: "Q_0.25". what does it mean? Is 0.25 the threshold level? Please clarify.

- Sec. 3.5, 1st paragraph: "we must implement". Does it mean that it has been implemented already? if so, please correct. Also, how is this implemented in simulation?

- Sec.3.5: perhaps, you may want to replace: "between describing" ->"with".

- Eq.3.4: is it 0.10 or 0.13? What is "Q_0.10"? please clarify.

- Sec.3.3.1, 1st paragraph: What is a "bright" or a "dim" event? please clarify.

- Sec.3.3.1, 1st paragraph: what does the "bright-to-dim ratio" represent? please clarify.

- Table 2: the errors seem to be extremely small. How are they calculated? do they include systematical as well as statistical uncertainties?
For example, by looking at Fig.8, it seems the uncertainties are rather large and not as tiny as those indicated in Table 2. If those are the errors on the mean, it is not

- Fig.10: The original "TA0003 model" differs from the "SPE template"

- caption: is it simulation or from the SPE template fit?

- Fig.10, caption: a few suggestion to clarify/correct the text:
- ->"...to the 2012 IceCube season (data points)".
- ->"...data are shown in black."
- ->"...number of DOM channels (NChan) ...

- page 18, 2nd paragraph: why HQE DOMs have a smaller peak-to-valley ratio with respect to standard QE DOMs? From where is this information taken? Is it clear why? Naively, one would expect the opposite. Please clarify.

- a few lines below: "...corresponding in an average increased peak-to-valley ratio...". What does it mean? Please clarify.

>>>