Data Cleaning: Further studies on the Neck Cut.

Samuel Marchand, Neil McCauley and Richard Helmer

TRIUMF and Oxford

SNO-STR-2000-028

Abstract

In this report, the neck cut and neck events are revisited. The discussion concentrates mainly on the computation of the median time instead of the mean time of the PMTs of the lower hemisphere of the PSUP. In light of our studies on contamination and sacrifice, the conditions used to discriminate neck events are revised and new ones are suggested. The neck cut's time parameters are revised to enhance the use of a time median calculation for the PMTs of the lower hemisphere of the detector. Among the conditions studied on this cut, two new conditions stand out as minimizing the contamination without further increasing the sacrifices of neck events.

1 Introduction.

Neck events are an important background in the SNO detector. Over the past few months they have occured at a rate of about 13 events/day according to the original neck cut set in the DAMN bank. Their name stems from the fact that they are apparently generated in the neck of the AV. Therefore, the neck tubes should record a relatively high charge and should be triggered early.

The following is a brief review of the original neck cut; a full description can be found in the Neck Cut Report [1]. Neck events are eliminated in two different ways by this cut. First an event ``fails" the cut (i.e. is set to be a neck event) if it fires two neck tubes. Then, an event can be cut if a single neck tube satisfies charge and time requirements . In the latter case, the charge condition is satisfied if the charge in a neck tube is well above the single p.e. peak. If this condition is met, the algorithm then checks the time condition. The time condition is satisfied if a single neck tube is more than a set number of nanoseconds earlier than the mean time of the PMTs in the lower hemisphere of the PSUP. There are two modes for calculating the charge and time of the PMTs, either raw ADC values or ECA calibrations can be used. Since the neck tubes only have ECA calibration, when the neck cut checks the time condition it computes ECA calibration for the PMTs involved in the mean time calculation. The modes using raw ADC values are only used when the ECA calibration has failed the neck tube in that particular event.

2 The Neck Cut revisited.

This project started with the original neck cut. The results given by previous studies indicated that further studies were needed to establish the level and the nature of contamination and to find possible upgrades. The first step in this direction is to compile the program with various logical conditions, to test every change on the appropriate samples and to compare the results with previous ones.

2.1 *The data samples used for this report.*

The data samples used need to be described in order to give a meaning to the results obtained on the conditions and parameters of the neck cut. They were used to test the effects of various changes on the neck cut and give an idea of the statistical validity of any conclusions on the results. The samples are shown in table ~\ref{tab:samples}.

Sample	Total Events	Neck Events
A	958	958
B	1079	1079
C	100671	100671
D	29964	3
E	4394	~1143

Table 1 Samples used and their statistics.

Sample	Description
A	It is a standard neck data set with a few contaminating retriggers removed. It contains events activating between 10 and 5000 tubes. Additionally, this sample comes from a 54 hours run.
B	This sample is identical to A but from a different run.
C	This sample contains an important diversity of neck events that were tagged by the original neck cut. Therefore, it may hold non-neck events.
D	All events passing data cleaning for data division sets A from november 1999 to march 2000 are present in this sample. It contains missed neck events and other undesirable events among good ones.

E	This is a handscanned set assembled from good runs, and contains events with over 25 nhits. This sample is extensively referred to in the original neck cut [1]. There is also a DARN bank containing the results of the handscanning

Table 2 Description of samples and their uses.

2.2 The time median calculation.

What underlies the studies on the implementation of the time median in the neck cut algorithm are neck events missed by the original neck cut. In many cases, contamination seems to be caused by some of the tubes at the bottom of the PSUP also being hit at early times, for example by noise hits. These outlier tubes can shift the mean time calculation so that the difference between it and a neck PMT time no longer satisfies the time cut parameter. Consequently, a calculation more stable to outliers would be useful in that it would allow a more precise signature of neck events.

In this section, the arguments supporting the implementation of the time median in the neck cut are presented. There are essentially three points to cover. First, the time median calculation has a sharper spectrum than the mean time one. Second, it cuts more neck events out of various samples than the original neck cut. Finally, the median of a set of points, as a mathematical technique, is more stable to extreme points than is the mean.

The RMS of the time median spectrum is consistently smaller than the RMS of the mean calculation. Therefore, neck events are more uniform under this technique.This phenomenon can be seen by comparing figures 1 and 2. Other samples tested display a similar difference in favor of the median calculation. The blue lines on the figures represent the appropriate cut parameter for a given calculation. The plots shown were produced with sample C but the other samples display the same behaviour.

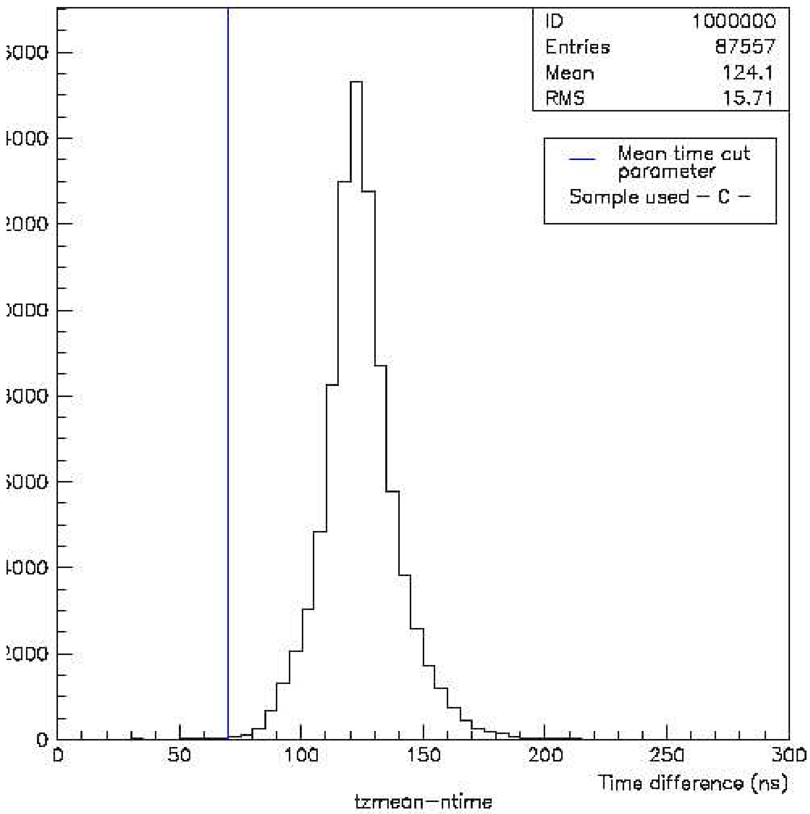

Figure 1 The difference between the mean time of the tubes of the lower hemisphere and the neck tubes earliest timing.

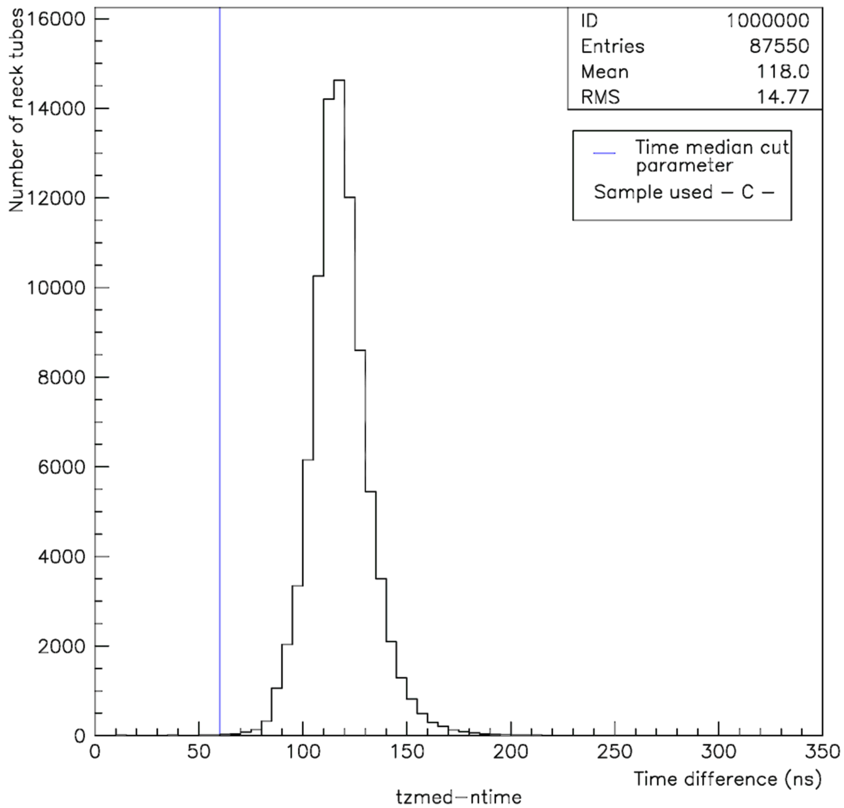

Figure 2 The difference between the median time of the tubes of the lower hemisphere and the earliest neck tubes timing.

Note that in figure 2 the threshold has been lowered. This change is needed because the mean value of the median time distribution is slightly smaller (from 5 to 10 ns depending on the sample tested) than the mean value of the mean time distribution. Therefore, the median time is closer to the neck PMTs timing by the same amount of time. The difference between the two calculation is displayed on figure 3. This effect can be easily countered by lowering the time difference parameter by a certain value which is displayed on figure 2.

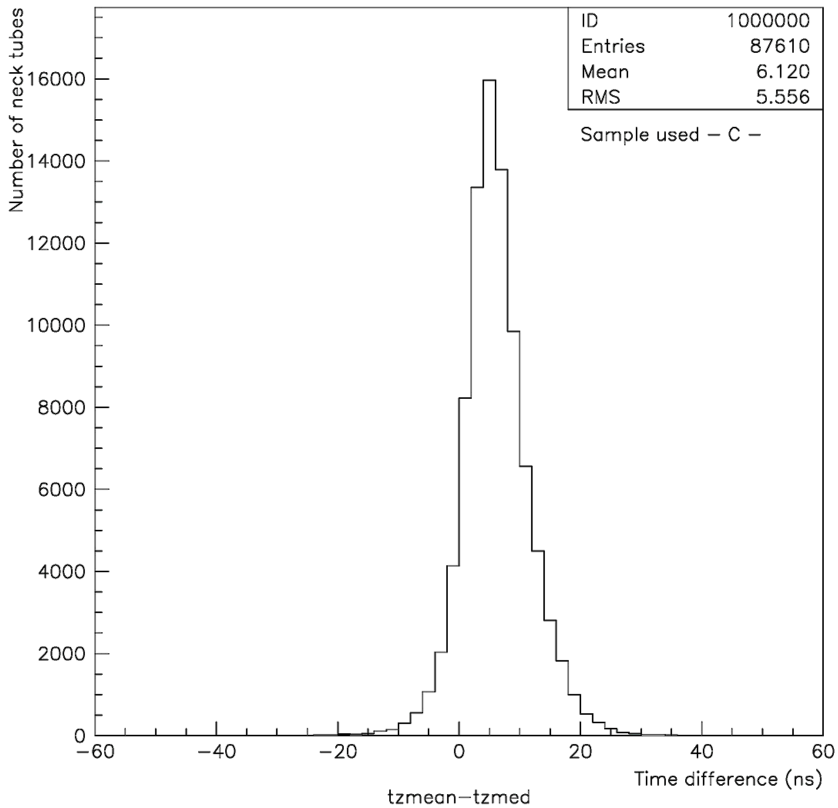

Figure 3 The difference between the mean timing and the median timing of the tubes of the lower hemisphere.

Furthermore, the median neck cut, as we will now refer to it, cuts more neck events than the original one which uses a mean calculation. This increased aggressiveness can be seen on a touchstone sample where the median neck cut found two neck events that had been missed by the original version of the cut. No effects were gathered from sample A or B since neither contamination nor sacrifice were changed. However, it can be seen that the median neck cut is able to cut more neck events than the original neck cut.

This is independant of the samples on which the median neck cut is tested. The reader should note that those results are only valid with the appropriate change of the time parameters.

This comparative study can be summarized in by the two following tables.

Sample	Events cut	Missed neck events
A	943	15
B	1072	7
C	100616	55
D	2	1

Table 3 Results on the neck cut using a time median calculation.

Sample	Events cut	Missed neck events
A	943	15
B	1072	7
C	100616	55
D	0	3

Table 4 Results on the neck cut using a mean time calculation.

Finally, to calculate a median, its necessary to class the elements in order. Therefore, a sorting algorithm was added to the neck cut which sorts the tubes time in increasing order in a card sorting fashion. As mentioned before, for those changes to take effect, the values of the time parameters need to be revised to take into account the characteristics of the median calculation. The optimal values that were found in this study are shown in table 5.

Time parameter	**Value**
Old (run # < 10220)	60 ns
New (run # > 10220)	10 ns

Table 5 Suggested values for the neck cut’s time parameter.

They take into account the smaller time difference between the neck tubes timing and the time median of the lower hemisphere tubes.

2.3 Revision of the neck cut conditions.

In an attempt to reduce the number of missed neck events, an effort was made to design new conditions using different characteristics of neck events. Conditions have been designed to use pipebox tubes, OWLs, and the timing of both neck tubes.

Most importantly, Neck events often fire a pipebox tube in addition to the neck tubes. That fact can be used effectively to cut neck events. A condition cutting events if a pipebox tube records a time early relative to the PMTs in the lower hemisphere cuts one more neck event out of sample B with either calculation. In addition, it finds 29 neck events out of sample E with the mean time calculation but 51 with the median time calculation. Once again, the median time computation shows a positive feature by cutting 22 additionnal neck events. However, 2 non-neck events were selected as neck events from a touchstone sample. These sacrifices appear only when this conditions is enabled. The results from this condition are resumed in tables 6 and 7.

Sample	Neck events	Non-neck events
A	0	0
B	1	0
E	29	N/A
Touchstone	0	0

Table 6 Effect of the condition using the presence of pipebox tubes and a time requirement (mean time calculation).

Sample	Neck events	Non-neck events
A	0	0
B	1	0
E	51	N/A
Touchstone	0	2

Table 7 Effect of the condition using the presence of pipebox tubes and a time requirement (median time calculation).

A condition which cuts an event if a neck and a pipebox tube are both present, but without a time requirement, gave identical results. It cut the same events out of the same samples but could increase sacrifice since there is no time requirement. Therefore, random noise in the neck accompanied by a good event could be failed by this cut. It is less likely that this will happen if there is a time requirement added to the presence of a pipebox tube. Although studies of this phenomenon are needed, a brief discussion on the topic has been made previously [1]. In addition, a bottom tube can send light through the detector and be detected in the neck. For these reasons, this condition and the alternative condition considering a single pipebox tube will not be considered further.

A last condition has been tested concerning the presence of both pipebox tubes. However no effects from this more stringent condition have been seen on any of the samples used, which at least informs us that missed neck events don't possess this feature.

The OWL tubes are ofen triggered in neck events, usually by light passing through the neck into the outer regions of the cavity. According to this they would usually trigger late in an event. However, observations confirms the opposite for neck events. This could mean that the light triggering the neck tubes also hits the OWLs. Consequently the OWL tubes are often early in neck events. That fact was used to try to cut more neck events. The corresponding condition asked a OWL tube to also satisfy the time cut to cut an event. However, this attempt failed to limit contamination in any way on sample A,B and C. The same results were obtained when considering the presence of a single OWL tube, ignoring the neck tubes. Again missed neck events are probably not distinguisable by the presence of owl tubes. They are usually single neck tube events with very low charge and late trigger time. At last, a condition looking only for OWL tubes would compromise the neck cut because many events won't satisfy the first neck cut conditions but will contain OWL tubes.

We also investigated the effects of imposing the time condition from the case in which only one neck tube fired to the case in which both neck tubes fired. The results are shown in the next two tables. It can be seen that this additional condition caused 3 more neck events to be missed in sample A, 15 more in sample B, and 55 more in sample C. The reason is that usually the neck tubes fire non-simultaneously, often with a significant time difference. Hence one of the tubes is typically late and so the event pass the filter. Therefore the idea of introducing this extra condition was discarded, and we reverted to marking as a neck event any event in which both neck tubes fire. As said in the previous report on neck events, other types of events will unlikely fire two neck tubes.

Sample	Neck events	Missed neck events
A	943	15
B	1071	8
C	100671	0

Table 8 Results obtained with a condition considering both tubes only.

Sample	Neck events	Missed neck events
A	940	18
B	1056	23
C	100616	55

Table 9 Results obtained with a condition considering boht neck tubes and their respective timing difference with the bottom tubes.

Finally, an attempt was made to use the distribution of times of the hits at the bottom of the PSUP to optimise further the filter. The attempt consisted of placing cuts on the vertical axis of the PSUP to probe this distribution. With this condition, the behaviour of the filter using the median calculation was compared with the behaviour of the present filter. It can be seen that both filter versions have a similar behaviour with respect to this cut. This is partly the result of the bottom tubes all having similar trigger times, as determined from the mean of those times staying practically the same as the cut is varied along the vertical axis. The results are presented in tables 10 and 11. In addition, the number of neck events is found to be almost independent of the vertical cut. In fact, any value between 0 and 800 cm has no effect on the number of neck events and their characteristics.

Value of the vertical parameter (cm)	Mean (ns)	RMS (ns)
0.0	120.9	17.85
-50.0	120.9	17.87
-100.0	120.8	17.83
-200.0	120.7	17.79
-400.0	120.8	17.77
-600.0	121.0	17.59
-800.0	120.3	17.03

Table 10 The effect of the variation of the vertical parameter on the mean of the mean timing and the RMS of the bottom tubes.

Value of the vertical parameter (cm)	Mean (ns)	RMS (ns)
0.0	115.9	18.31
-50.0	115.9	18.29
-100.0	115.8	18.24
-200.0	115.7	18.28
-400.0	115.8	18.23
-600.0	116.0	17.89
-800.0	115.2	17.66

Table 11 The effect of the variation of the vertical parameter on the mean of the median timing and the RMS of the bottom tubes.

The data do show the widths of the distributions are reduced as the cut on the vertical axis is tightened. However, the effect is small and cannot be used to optimise the code. Furthermore, since the radius of the detector is 8.5 m, the effect has reached its limit.

Conclusions.

To insure that the median filter previously described works optimally, it is necessary to slightly modify the neck cut's time parameters. In fact, the median time calculation is closer to the neck tubes timing than the mean time calculation is. The new suggested parameters appears in the following table. The charge, mode and pc number parameters well described in the first neck cut report [1] are unmodified.

Parameter	**Value**
New time difference (run # > 10220)	10
Old time difference (run # < 10220)	60

Table 12 Suggested time parameter's values.

Certain parameters were experimented but don't need modifications. Its the case with the cut on the vertical axis of the geodesic sphere.

In summary, the revision of logical conditions maps an important part of the condition space. More than 8 configurations were experimented with and 3 of them were found to improve in a sensible way the neck cut. In addition, the two parameters of the time condition were tested and optimised to fit with the change of the time calculation. At last, some inconclusive results were obtained and documented.

References

[1] *Data cleaning* : *The Neck Cut* N. McCauley, Oxford University

Acknowledgements

One of us (S.M.) would like to thank the many members of the SNO collaboration who helped me understand the detector and the data cleaning cuts in general, and SNOMAN. In particular, I would like to thank Rick Van Berg, Richard Van de Water, Jacques Farine, Thomas Kutter, Christian Nally and my coauthors who helped me understand neck events and the ways in which they might be cut.