Shape analysis and combination of presupernova neutrino signal

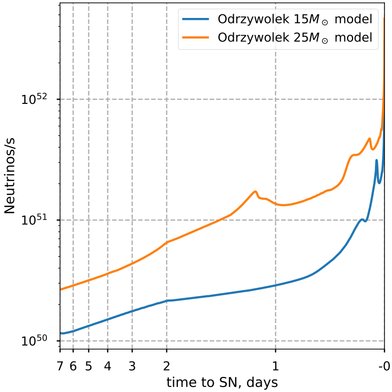
Andrey Sheshukov

Joint Institute for Nuclear Research, Dubna

SNEWS2 Meeting @ Neutrino2020

19.06.2020

Pre-supernova signal time profile



- Growing neutrino rates and energies towards the SN collapse time.
- Various model-dependent features (bumps location etc).

Typically an experiment:

- Measures $\{t_i\}$: timestamps of neutrino interaction candidates integrated within an energy range.
- Estimates B: background event rate in this energy range.

	N_{bg}	N_{sg}@$200pc$
JUNO	$\approx$128	764
KamLAND	0.142	12

Table: Expected number of interactions during 48 hours before collapse

Goal of presupernova alert for SNEWSv2 — predict supernova event, observing deviation from background.

Hypothesis test

We need to distinguish H_0: Bg only and H_1: Bg+preSN given measured data $\{t_i\}$ [1]

- Define test statistics function $X(\{t_i\})$ so that $P(X|H_0)$ and $P(X|H_1)$ are separated.
- Deviation from H_0 is measured by p-value

$$p(\{t\}) = P(x > X(\{t_i\})|H_0)$$

- Often presented as z-score, i.e. "sigmas" of standard normal distribution:

$$z(\{t_i\}) = \Phi^{-1}(p(\{t_i\}))$$

where Φ is tail probability function for Gaussian with $\mu = 0$, $\sigma = 1$.

[1] In principle, need to scan hypotheses H_1^* for various preSN models, distances, progenitors.

Log Likelihood Ratio

The best test statistics for hypotheses separation is the log likelihood ratio:

$$X(\{t_i\}) = \log \frac{P(H_1|\{t_i\})}{P(H_0|\{t_i\})} = \sum_i \log \frac{P(t_i|H_1)}{P(t_i|H_0)}$$

If hypotheses predict different rates vs. time:

$$P(t|H_0) \sim B(t)$$
$$P(t|H_1) \sim B(t) + S(t)$$

and signal rate depends on the assumed time of SN start t_{SN}:

Log likelihood ratio

$$\ell(\{t_i\}) = \sum_i \log\left(1 + \frac{S(t_i - t_{SN})}{B(t_i)}\right)$$

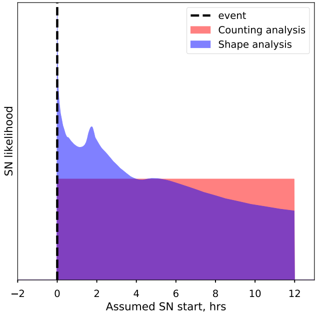


Figure: Each data event elevates the likelihood of having a supernova in the future. "Response" function depends on selected signal shape.

KamLAND counts events in 48hr window: this is also an assumption of a signal shape.

Detection example

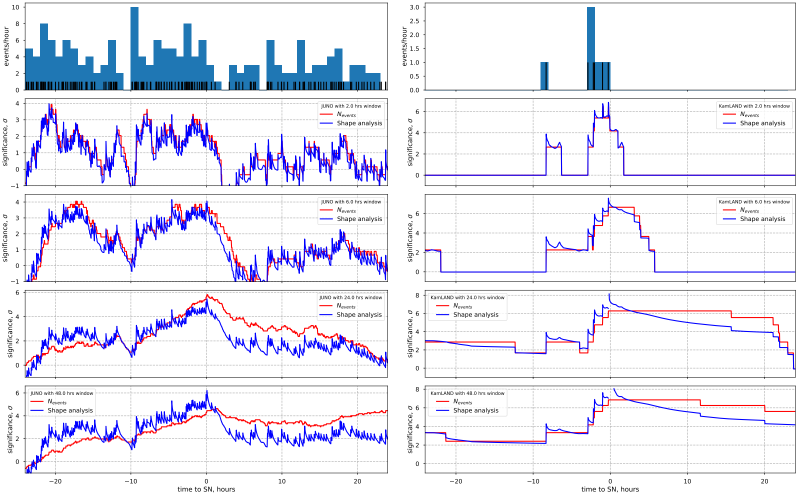
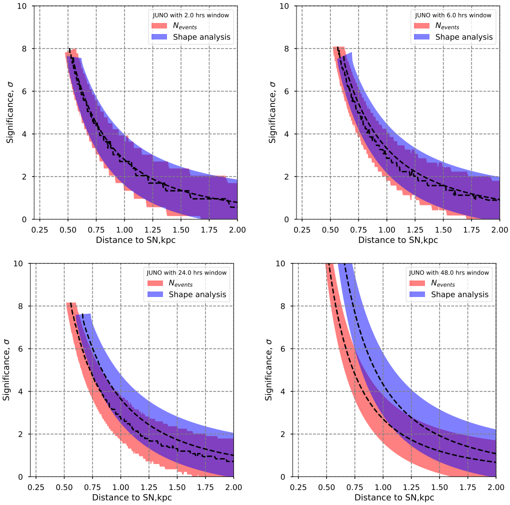


Figure: PreSN@750pc in JUNO(left) and preSN@300pc in KamLAND(right)

Significance vs. distance



Shape analysis performs at least not worse than counting analysis

No big difference for short windows (low Bg)

Counting analysis degrades, when window becomes larger (Bg grows,Sg almost the same)

Shape analysis performs better on larger windows (more information)

Summary

Shape analysis depends on:

- Expected signal shape $S(t)$
- Background rate estimation $B(t)$ and it's uncertainty.
- Signal amplitude (distance to SN) — the LLR will give the best discrimination for the given H_1 vs H_0. At other distances it will be suboptimal. However, it still works, and for lower distances the significance will be high because of statistics. So we can choose some distance where we want improvement.

Counting analysis is not "model independent" - it is approximating the shape with flat region. Any simple approximation can do better.

Better time localization of the SN likelihood (sharper peaks) - using the fact that the largest flux of neutrinos if very near to the SN time.

Further plans

- Check how the prediction time changes with shape analysis
- Use correct expected signals (this example was just neutrino flux, normalized to needed number of interactions)
- Check various models

Methods for combining measurements

Problem

We have N detectors with signal and background rates $S_n(t)$, $B_n(t)$, each measured data $\{t_i\}_n$. SNEWS network should calculate the joint significance for this data:

$$Z_{joint} = f(\{t_i\}_1, ...\{t_i\}_N)$$

providing the good separation between H_0 and H_1

Case 1: Combining significances

Detectors calculate significance on their sides, and send us z_n. We should calculate $Z(z_1, ..., z_N)$.

- Tippet's combination: $X = n$-th maximal of $\{z_n\}$. This is case of SNEWSv1 (n=2).
- Stouffer's combination: $Z = \sum_n w_n z_n$
- Fisher's combination: $X = -2\sum_n \ln p_n$

Methods for combining measurements

Problem

We have N detectors with signal and background rates $S_n(t)$, $B_n(t)$, each measured data $\{t_i\}_n$. SNEWS network should calculate the joint significance for this data:

$$Z_{joint} = f(\{t_i\}_1, ...\{t_i\}_N)$$

providing the good separation between H_0 and H_1

Case 2: Combining data

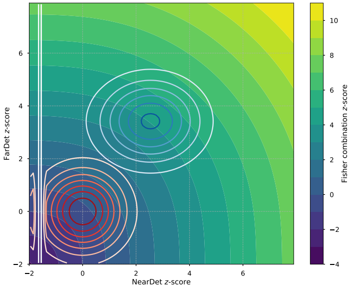
If possible perform a combined analysis on the data, using the joint log likelihood ratio:

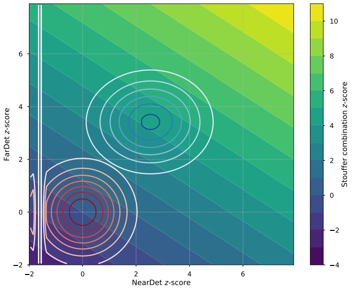
$$\ell(\{t_i\}_1, ...\{t_i\}_N) = \log \frac{P(\{t_1\}, ...\{t_N\}|H_1)}{P(\{t_1\}, ...\{t_N\}|H_0)} = \sum_{n=1}^{N} \ell_n(\{t_i\}_n)$$

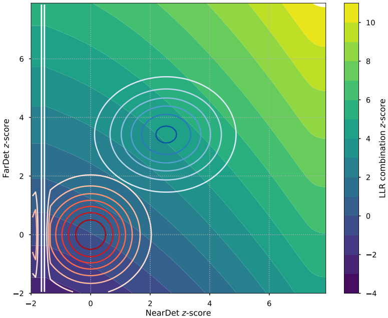
So the LLR test statistics is additive for various experiments, *if they use the same H_0 and H_1, and their measurements are independent*.

Getting the significance requires calculation of the joint distribution $P(\ell|H_0)$.
This is the ideal case: if the detectors are within one project, or different detection channels in one detector, or different energy bins.

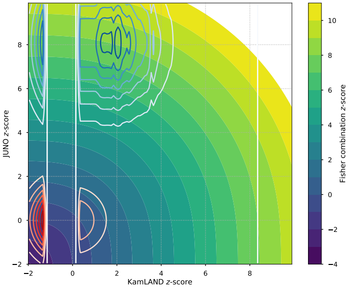
Combination: high background

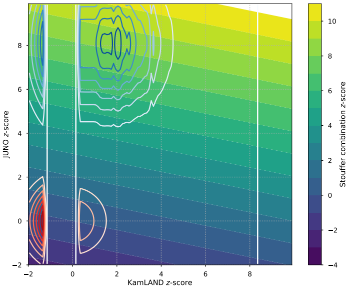


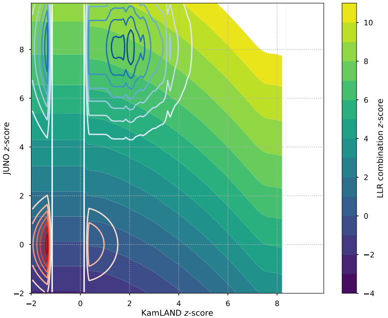




Combination: low background







Summary

I propose shape analysis for calculation of the presupernova significance.

- Log likelihood ratio is additive for both data measurements and detectors combination.
- In ideal case we can combine the data and all the information on the detectors.
- In other cases we can combine significance, without intrinsic detector knowledge.
- Additional studies need to determine the best way to combine.

For the practical use in SNEWSv2:

- Collect data (event timestamps) from the clients and calculate significance.
- Or provide the application which would calculate significance on the client side.
- We can use the same pipeline for preSN and ccSN.