

Identification of time-correlated neutrino clusters in populations of astrophysical transient sources

Mathieu Lamoureux^{a,*} and Gwenhaël de Wasseige^a

^a*Centre for Cosmology, Particle Physics and Phenomenology - CP3,
Universite Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

E-mail: mathieu.lamoureux@uclouvain.be

The detection of astrophysical neutrinos from transient sources can help to understand the origin of the neutrino diffuse flux and to constrain the underlying production mechanisms. In particular, proton-neutron collisions may produce GeV neutrinos. However, at these energies, neutrino data from large water Cherenkov telescopes, like KM3NeT and IceCube, are dominated by the well-known atmospheric neutrino flux. It is then necessary to identify a sub-dominant component due to an astrophysical emission based on time correlation across messengers. The contribution covers several methods to search for such a signal in short time windows centered on observed transient sources, including a novel approach based on the distribution of time differences. Their performance is compared in the context of subpopulations of astrophysical sources that may show prompt or delayed neutrino emissions. The outlook for the usage of such techniques in actual analyses is also presented.

The 38th International Cosmic Ray Conference (ICRC2023)
26 July – 3 August, 2023
Nagoya, Japan



*Speaker

1. Introduction

Since the discovery of high-energy astrophysical neutrinos from the IceCube Collaboration in 2013 [1], neutrino telescopes have continuously been monitoring the sky in search of the corresponding sources. Multi-messenger approaches have allowed identifying some candidates. In 2017, a high-energy neutrino from IceCube has been detected from the direction of the blazar TXS 0506+056, in time coincidence with a flare in gamma rays [2]. More recently, several tidal disruption events have also been associated with neutrinos in IceCube [3, 4].

One key feature to clearly associate neutrinos with an astrophysical source is to rely on the spatial and temporal coincidences of the signal events. For transient sources, emitting on a short time scale, such a method is very efficient at reducing the contamination from background neutrino events, such as the ones originating from atmospheric neutrinos.

In the case of the IceCube GeV sample described in [5, 6], no direction reconstruction is available at the time of writing, so it is only possible to apply time cuts. The expected background rate is at the level of 20 mHz. This sample is sensitive to all-flavor neutrinos at energies ranging from 0.5 to 100 GeV, making it a nice probe for potential GeV neutrino emission.

This work focuses on the search for a short neutrino signal from sub-populations of astrophysical transient sources. The exact delay Δt of the neutrino emission with respect to the time t_0 of the detection with another messenger is not known. Other studies often employ a conservative 1000 s time window centered on t_0 [7]. However, in the case of the IceCube GeV sample, this corresponds to ~ 20 expected background events, limiting the discovery potential for sources that may emit a flux corresponding to only a few detectable neutrinos.

As long as no clear detection is reported, the value of Δt is not known a priori. One may expect it to be of the same order within a given population and for a given emission mechanism. It is then possible to reduce the impact of background events by looking for a coherent excess of neutrinos with the same delay with respect to t_0 over a set of transients. Additionally, several timings may be relevant for a given source type, such as precursor, prompt, and delayed emissions.

The goal of this work is to identify subpopulations in a set that may contain sources with no detected signal neutrinos, sources with a signal at Δt , or sources with a signal at a different timing. The Figure 1 shows an illustration of the time distribution of detected events in toy experiments.

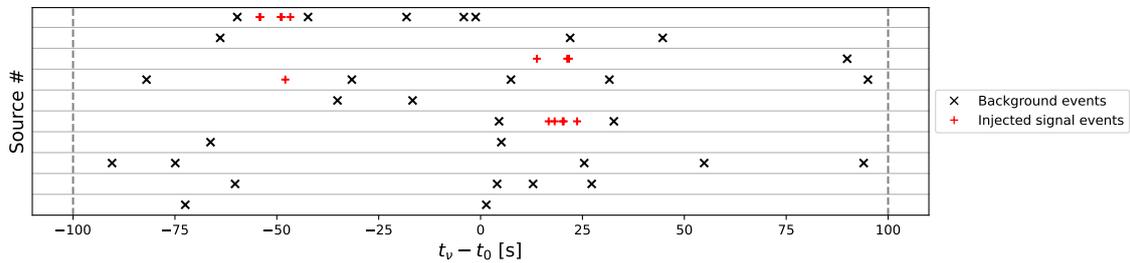


Figure 1: Illustration of the time distribution of neutrino candidate events in toy experiments within $t_0 \pm 100$ s. Each row corresponds to one realization. The black crosses indicate background events with a Poisson rate $r_{\text{bkg}} = 20$ mHz (illustrative of the IceCube GeV sample). The red pluses are injected signal events at $\Delta t = -50$ s and $\Delta t = 20$ s, with a width of 2 s.

2. Method

The main ingredient of the search described here is the relative time $\{t_i\}$ of all detected neutrino candidates with respect to the considered transient time t_0 . For convenience, the analysis is restricted to a time window $[T_{\min}, T_{\max}]$ large enough to contain any kind of signal. To be consistent with other searches, $T_{\min}, T_{\max} = -500, 500$ s will be employed.

2.1 MLE approach

A usual technique to identify a flaring excess is to define the following likelihood:

$$\mathcal{L}(N, \{t_i\}; n_s, \Delta t, \sigma_t) = \prod_{i=1}^N \left[\frac{n_s}{N} \mathcal{S}(t_i, \Delta t, \sigma_t) + \left(1 - \frac{n_s}{N}\right) \mathcal{B}(t_i) \right], \quad (1)$$

where N is the total number of events, $\{t_i\}$ are their detection times, n_s is the expected number of signal events, $\Delta t, \sigma_t$ is the signal central time and its width, and $\mathcal{S}(\mathcal{B})$ is the signal (background) pdf. In the following, \mathcal{S} is a Gaussian distribution with $\mu = \Delta t, \sigma = \sigma_t$, and $\mathcal{B}(t) = 1/(T_{\max} - T_{\min})$. It is then possible to write the Maximum Likelihood Estimator

$$\text{MLE} = 2 \log \left(\frac{\mathcal{L}(N, \{t_i\}; \hat{n}_s, \hat{\Delta}t, \hat{\sigma}_t)}{\mathcal{L}(N, \{t_i\}; n_s = 0)} \right), \quad (2)$$

where the hatted terms are the values maximizing the likelihood, while the denominator is the background-only likelihood. The maximization is done with the `iminuit` package [8].

The background distribution of MLE $P_B(\text{MLE})$ is obtained using background-only experiments, as illustrated in Figure 2. The significance of a measurement with $\text{MLE} = \text{MLE}_{\text{observed}}$ is then characterized by the p-value $p = \int_{\text{MLE}_{\text{observed}}}^{\infty} P_B(\text{MLE}) d\text{MLE}$.

For each transient, one may then compute p and consider an observation to be significant if $p < 3 \times 10^{-3}$, this would correspond to a $\sim 3\sigma$ deviation from the background expectation. The values of $\hat{n}_s, \hat{\Delta}t, \hat{\sigma}_t$ may also be recovered to characterize the potential signal. The selection efficiency of the approach can be evaluated using simulations as shown in Figure 2.

Although such a method has proven its liability for such flare searches, it does not take into account the correlations of the signal times across several time windows but only performs independent probes for individual transient.

2.2 PCA approach

As the objective is to separate the full dataset into different subpopulations based on the presence or the absence of a neutrino signal, clustering techniques are appropriate for the task. Principal Components Analysis (PCA) is well suited to reduce the dimensionality of the problem before performing clustering, as initially presented in [9].

The data is summarised by the set $\{t_i\}$ of observed relative times. To make it suitable for PCA, they are binned in K slices such that $n_j = \sum_{i=1}^N H(t_i - \tau_j)H(\tau_{j+1} - t_i)$ for $j = \{0, \dots, K-1\}$, where H is the Heavyside step function, and $\tau_j = T_{\min} + j \frac{T_{\max} - T_{\min}}{K}$. The overall dataset has then a dimension $M \times K$, where M is the number of analyzed time windows.

PCA is applied to reduce the number of features to one or two using `scikit-learn` existing methods [10]. This (these) feature(s) may then be used to easily cut between background events

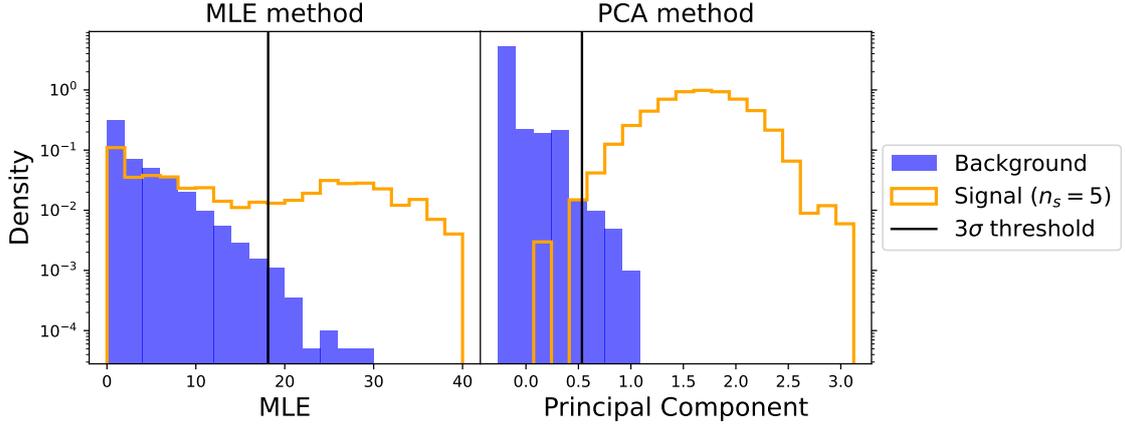


Figure 2: Illustration of the MLE (left) and PCA (right) methods. The blue histogram corresponds to the expected distribution for background events with constant rate $r_{\text{bkg}} = 20$ mHz. The orange histogram is the distribution for an injected signal in the search window, with the number of signal events following a Poisson distribution with $\lambda = 5$ and a time distribution following a Gaussian distribution with $\sigma = 2$ s. The black line indicates the 3σ threshold derived from the background distribution. In this example, there are only two sub-populations, those with only background and those with an injected signal, so the PCA threshold is defined by looking at the first component and using the same strategy as for MLE.

and potentially interesting time windows. To evaluate the performance of the search, it is necessary to make use of a partially-labeled dataset. The following procedure has been applied:

1. A dataset of pure background events \mathcal{D}_{bkg} is formed. In the following example, this will consist in pseudo-experiments where no signal is injected but only the expected background. For actual measurements, one may consider time windows in real data where no transients were detected.
2. \mathcal{D}_{bkg} is appended to the dataset of time windows of interest $\mathcal{D}_{\text{search}}$ (corresponding to actual data or to simulations with some injected signal): $\mathcal{D} = \mathcal{D}_{\text{bkg}} \cup \mathcal{D}_{\text{search}}$.
3. PCA reduction is performed on the full dataset: $\mathcal{D} \rightarrow \mathcal{D}^* = f(\mathcal{D})$.
4. The distribution of the remaining features is plotted only for the samples originating from \mathcal{D}_{bkg} . It is then possible to define a threshold such that any observation exceeding this threshold would correspond to a 3σ excess.
5. The observed $f(\mathcal{D}_{\text{search}})$ are compared with this threshold to search for any significant event.

A simple example is shown in Figure 2. The performance will be compared with other methods in Section 3. As compared to the likelihood approach, the PCA approach properly covers the needs for the search of subpopulations with similar timings with respect to the detected transient.

2.3 PeANuTS method

Going back to the basics, the current search consists in searching for consistent deviations from the Poisson expectation (that correctly described the background). The available database

consists of a succession of neutrino candidate events characterized by their relative detection times with respect to the considered transient source detection time, across time windows covering a population of such sources.

When considering the event times as our only input, a divergence from the Poisson expectation may be identified by plotting the inter-arrival time $\delta t_i = t_i - t_{i-1}$ distribution and looking for deviations from an exponential distribution $p(x) = 1/r \exp(-x/r)$, where r is the expected rate. This also applies if one computes the delay between a random time and the next (or previous) event.

In the following, M time windows of interest (for instance ± 500 s around each selected transient source) are considered and the time of each event is noted $t_{m,i}$, where m is the label of the time window and i is indexing the event in the time window. For each event (m, i) , one may compute

$$\Delta t_{m,i}^{n,-} = \begin{cases} t_{m,i} - t_{m,i-1}, & \text{for } n = m, \\ \min [t_{m,i} - t_{n,j}]_{t_{n,j} < t_{m,i}}, & \text{for } n \neq m, \end{cases}$$

$$\text{and } \Delta t_{m,i}^{n,+} = \begin{cases} t_{m,i+1} - t_{m,i}, & \text{for } n = m, \\ \min [t_{n,j} - t_{m,i}]_{t_{n,j} > t_{m,i}}, & \text{for } n \neq m, \end{cases}$$

This set is noted $\mathbf{D}_{m,i} = \{\Delta t_{m,i}^{n,-}, \Delta t_{m,i}^{n,+} \forall m \in \{1, \dots, M\}\}$ and contains $2M$ different values, as illustrated in Figure 3. For ‘‘border’’ events (i.e., first and last events of any time window), the events right before/after the search time window are used as fill values (or the related value may be set to ∞ , which would have no negative impact on the final sensitivity).

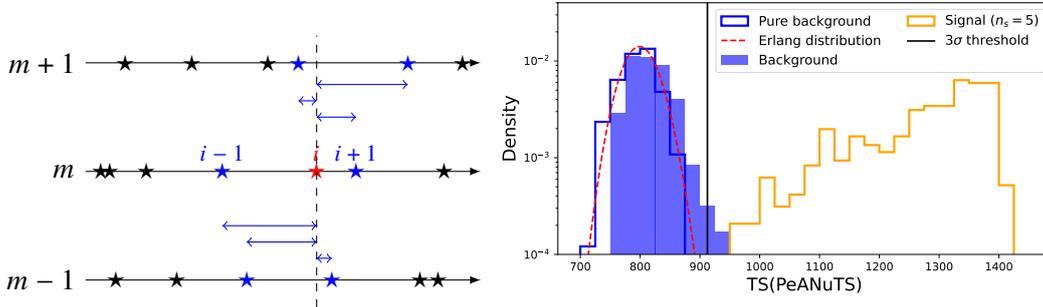


Figure 3: Left: Illustration of the definition of the $\Delta t_{m,i}^{n,-}$ and $\Delta t_{m,i}^{n,+}$ values. Right: The blue step histogram corresponds to pure background, well fitted by an Erlang distribution shown in dashed red. The blue-filled and orange step histograms are the background and signal distributions with the same inputs as in Figure 2.

From the discussion above, if data is well described by a Poisson background with rate r_{bkg} , it is expected that $\mathbf{D}_{m,i}$ values follow an exponential distribution with $\lambda = 1/r_{\text{bkg}}$. This check may be formalized by computing, for each event (m, i) ,

$$\text{TS}_{m,i} = - \sum_{\Delta t \in \mathbf{D}_{m,i}} \log [1 - \exp(-r_{\text{bkg}} \Delta t)]. \quad (3)$$

The test statistic TS follows an Erlang distribution (Erlang($x; k$) = $x^{k-1} e^{-x} / (k-1)!$) with $k = 2M$. A right tail might appear for background events that would accidentally fall close to where signal clusters are located, as illustrated in the right panel of Figure 3. This method is called PeANuTS in the following, for Poisson-expectation Anomaly for Neutrino Transient Source Search.

As opposed to the MLE and PCA approaches, one value per event is computed instead of one per time window. It is then necessary not only to define a threshold TS_{thr} to select interesting events, but also to choose the minimum number of events N_{min} passing this threshold in a given time window to characterize an interesting excess. The 3σ threshold TS_{thr} can be found using the Erlang inverse survival function. The cut N_{min} is then useful to reduce the impact of the right tail mentioned in the above paragraph; $N_{\text{min}} = 2$ is sufficient to remove all such background events.

3. Results

To compare the three methods presented in [Section 2](#), a toy scenario is considered :

- Background rate is fixed to 20 mHz, representative of the expected background in IceCube GeV neutrino sample [5] taken as example.
- The injected signal in a given time window is characterized by a strength n_S that is the Poisson mean expected value. A Gaussian temporal profile, with mean T_0 and sigma σ_T , is considered. The typical values used for T_0 and σ_T are 0 s and 0.5/2/5 s, respectively.
- A total of $M = 200$ time windows is considered, and a signal is injected in 20% of them.

The comparison is then performed by simulating 10000 times such a scenario, referred to as pseudo-experiments. The efficiency of a given method is then evaluated as the fraction of time windows over all pseudo-experiments that are above the 3σ threshold as defined in [Section 2](#) for each method. For the new TS approach, the value of m_{min} is set to 2. The left panel of [Figure 4](#) shows the results as a function of the injected signal strength, for different signal widths.

As expected, the MLE method is relatively robust with respect to the value of σ_T , as there is a corresponding fit parameter. However, as this approach only considers time windows independently, it fails to achieve high efficiencies, especially for a low number of events, this is also shown by the presence of a long tail on the signal distribution in [Figure 2](#). The PCA method is performing very well for short signals, where it is easier to identify a principal component within the dataset \mathcal{D} . For $\sigma_T < 1$ s, the latter may simply be the number of events in the bin corresponding to the signal window. For longer signals, the performance worsens significantly. Then, the PeANuTS approach proposed here generally outperforms both MLE and PCA. Though it also has the same preference towards small signals, it is still efficient for larger σ_T . Compared to the two other methods, the signal strength threshold (corresponding to $> 50\%$ selection efficiency) is lowered by factor 1.2 to 3 which, in turn, would correspond to a similar improvement in the flux constraints.

Two additional scenarios are then considered to cover cases where several timings are involved:

- (a) 20% of the time windows have $T_0 = 0$ s, and another 20% have $T_0 = 100$ s.
- (b) 20% of the time windows have $T_0 = 0$ s, 20% have $T_0 = 100$ s, and 20% have $T_0 = 300$ s.

In both cases, the rest of the parameters do not change. For the PCA approach, as multiple clusters are expected to form for these new scenarios, the two first components are used (instead of only the first one) and the threshold is based on the bi-dimensional z-score of a point with respect to the background expected distribution (obtained from \mathcal{D}_{bkg}). The related efficiencies are shown in the right panel of [Figure 4](#). The MLE results are relatively insensitive to these different scenarios

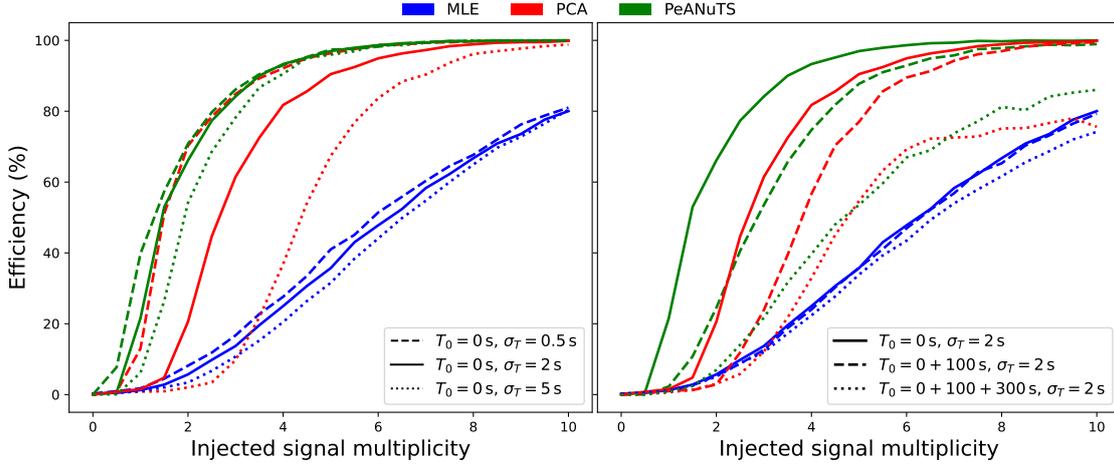


Figure 4: Comparison of the signal selection efficiencies for the three methods (MLE in blue, PCA in red, PeANuTS in green) varying n_S from 0 to 10, for the toy scenario with $T_0 = 0$ s (left) and for the two additional scenarios involving multiple signals (right). In the left panel, the line styles correspond to different values for σ_T . In the right panel, they are used to distinguish the different sets of T_0 values.

as T_0 corresponds to a free parameter in the fit. On the contrary, the performance of the two other methods degrades significantly, as the signal is less precisely defined and therefore more difficult to identify. Nevertheless, the PeANuTS method is still favored with respect to MLE and PCA.

4. Summary and outlooks

Two common techniques using a Maximum Likelihood Estimator or Principal Component Analysis are compared with a new approach called PeANuTS in the context of a search for neutrino emission from subpopulations of transient sources. The latter achieves higher efficiencies as it strongly benefits from the potential correlations of signal timings in several time windows.

The new method may be used for searches in conservative time windows meant to catch any precursor, prompt, or slightly delayed signal. Notably, it may strongly enhance the discovery potential for neutrino samples for which the background rate is not low enough to allow claiming a discovery as soon as one or very few events are observed in this window. For instance, it is relevant for GeV neutrinos at large Cherenkov telescopes, as such samples are dominated by optical noises and by irreducible atmospheric neutrinos in this energy range. Moreover, the low number of Cherenkov photons emitted by GeV neutrinos does not allow for any precise direction reconstruction with current methods, limiting the possibility to benefit from using neutrino-source spatial correlations. This is then suited for the IceCube GeV neutrino sample [5] or for the under-development similar sample in KM3NeT [11].

This method may be applied for binary mergers as detected with gravitational waves or for gamma-ray bursts seen throughout the electromagnetic spectrum. For both these categories, large catalogs (e.g., [12, 13], respectively) are already available and may therefore be employed. For these objects, the existence and timing of the neutrino emission are currently unknown as it has never been observed, making it an interesting exploratory field.

As the PeANuTS test statistic is quite general and only relies on arrival times, it could also be applied well beyond the scope presented here. Furthermore, adjustments may be possible to cover for potential non-Poissonian effects in the data or to include additional inputs in the analysis.

Acknowledgments

M.L. is a Postdoctoral Researcher of the Fonds de la Recherche Scientifique - FNRS.

References

- [1] ICECUBE collaboration, *Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector*, *Science* **342** (2013) 1242856 [1311.5238].
- [2] ICECUBE collaboration, *Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert*, *Science* **361** (2018) 147 [1807.08794].
- [3] R. Stein et al., *A tidal disruption event coincident with a high-energy neutrino*, *Nature Astron.* **5** (2021) 510 [2005.05340].
- [4] S. Reusch et al., *Candidate Tidal Disruption Event AT2019fdr Coincident with a High-Energy Neutrino*, *Phys. Rev. Lett.* **128** (2022) 221101 [2111.09390].
- [5] ICECUBE collaboration, *Search for GeV neutrino emission during intense gamma-ray solar flares with the IceCube Neutrino Observatory*, *Phys. Rev. D* **103** (2021) 102001 [2101.00610].
- [6] ICECUBE collaboration, *Probing neutrino emission at GeV energies from compact binary mergers with the IceCube Neutrino Observatory*, 2105.13160.
- [7] B. Baret et al., *Bounding the Time Delay between High-energy Neutrinos and Gravitational-wave Transients from Gamma-ray Bursts*, *Astropart. Phys.* **35** (2011) 1 [1101.4669].
- [8] H. Dembinski et al., *scikit-hep/iminuit*, Dec, 2020. 10.5281/zenodo.3949207.
- [9] G. de Wasseige, *Multi-detector approach to enhance the sensitivity of neutrino telescopes to low-energy astrophysical sources*, *JINST* **16** (2021) C12012 [2108.07067].
- [10] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825 [1201.0490].
- [11] KM3NET collaboration, *Improving the sensitivity of KM3NeT to MeV-GeV neutrinos from solar flares*, *PoS ICRC2023* (these proceedings) 1294.
- [12] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*, 2111.03606.
- [13] “GRBweb.” https://user-web.icecube.wisc.edu/~grbweb_public/.