IceCube Neutrino Clustering and Source Modeling

Gabriele Paganelli
University of Padova / Boston University
Master's Student in Statistics
gabrip02@bu.edu

1 Introduction

The IceCube Neutrino Observatory, situated at the South Pole, is a cubic-kilometer detector designed to register high-energy neutrinos. Identifying the origin of these particles — whether cosmic (astrophysical) or produced by cosmic-ray interactions in the Earth's atmosphere (atmospheric background) — is the primary challenge in high-energy neutrino astronomy. Astrophysical neutrinos serve as crucial messengers for extreme environments such as active galactic nuclei or supernova remnants, providing unique insights into particle acceleration mechanisms that are inaccessible through electromagnetic observations alone.

Statistical inference that can handle both energy and directional information is essential for extracting the astrophysical component from the dominant atmospheric background. Bayesian mixture models naturally accommodate such latent structures, enabling probabilistic classification of individual events.

This work implements and compares a sequence of Bayesian models on the IceCube point-source dataset. Beginning from a conceptually complete joint energy-space model, computational constraints motivate simplified approaches that retain key probabilistic features while remaining tractable for huge sample sizes.

The main goals are to (i) evaluate the models' ability to infer event-level astrophysical probabilities, (ii) assess the relative contribution of energy and spatial information, and (iii) provide a qualitative assessment of model behavior under severe class imbalance.

2 Data

The publicly available *all-sky point-source* dataset from the IceCube Neutrino Observatory (2011–2018, IC86) includes over one million track-like neutrino candidates with reconstructed direction, energy, and per-event angular uncertainty (σ_i). Track-like events are selected for their sub-degree angular precision, enabling point-source studies.

Minimal, physically motivated cuts are applied to ensure data quality. Only events from the fully deployed detector (IC86, 2011-2018) are retained, and events with poor directional reconstruction (ang_err $> 4°$) are removed. After these cuts, roughly $N \simeq 880{,}000$ events remain.

Two observables drive the inference. The reconstructed energy E (expressed as $\log_{10}(E/\mathrm{GeV})$) provides the main discrimination between astrophysical and atmospheric events. Energy distributions are modeled as Truncated Power Laws within the observed range $[E_{\min}, E_{\max}]$. Directional information is incorporated via a pre-computed spatial weight w_i for each event, quantifying proximity to $J = 110$ candidate sources from the IceCube 10-year time-integrated search.

For each event-source pair (i, j), the effective angular variance merges the measurement uncertainty with the source extent, reflecting both detector resolution and the expected angular size of the emission region:

$$\sigma^2_{\mathrm{eff},ij} = \sigma_i^2 + \sigma^2_{\mathrm{src}}, \quad \sigma^2_{\mathrm{src}} = \max\left[\sigma_{\min},\ a/(\hat{n}_s + b)^\gamma\right]^2,$$

with $a = 0.5$, $b = 1$, $\gamma = 0.5$, and $\sigma_{\min} = 0.05°$, where $\hat{n}_s$ is a frequentist estimate of source activity. This scaling keeps the Gaussian angular approximations consistent with spherical geometry and weights sources according to their inferred brightness. The unnormalized weight sums contributions from all sources:

$$\tilde{w}_i = \sum_{j=1}^{J} \exp\left[-\theta_{ij}^2/(2\sigma^2_{\mathrm{eff},ij})\right],$$

and is normalized by the dataset mean to produce $w_i = \tilde{w}_i/\bar{\tilde{w}}$, so that $\bar{w} = 1$. These two observables, energy and spatial weight, form the basis for subsequent probabilistic classification of astrophysical events.

3 Models

The physically motivated approach is a joint Bayesian model over both spatial and energetic observables, with all parameters estimated simultaneously. This Conceptual Model provides the theoretical foundation for our analysis, though computational limitations require simplified implementations. In this framework, each observed event arises from one of two latent populations: the astrophysical signal ($\mathcal{L}_{\text{Astro}}$), modeled as a mixture of Gaussian functions in spatial coordinates centered on known candidate point sources, coupled with a comparatively hard power law energy spectrum; and the atmospheric background ($\mathcal{L}_{\text{Atmo}}$), modeled as uniform over the sky, coupled with a steep power law energy spectrum.

Let the class-conditional energy likelihood be defined as a truncated power law,

$$p(E \mid z) = \frac{E^{-\alpha_z}}{\int_{E_{\min}}^{E_{\max}} E^{-\alpha_z}\, dE}, \qquad z \in \{\text{astro}, \text{atmo}\},$$

and all subsequent models make use of this expression.

The Conceptual Model likelihood for a single event is then written as

$$\mathcal{L}_{\text{Conceptual}}(E_i, \mathbf{x}_i \mid \Theta) = \pi_{\text{astro}} \cdot \left[\mathcal{L}_{\text{Astro, Space}}(\mathbf{x}_i) \cdot \mathcal{L}_{\text{Astro, Energy}}(E_i) \right] \\ + (1 - \pi_{\text{astro}}) \cdot \left[\mathcal{L}_{\text{Atmo, Space}}(\mathbf{x}_i) \cdot \mathcal{L}_{\text{Atmo, Energy}}(E_i) \right]$$

where π_{astro} is the global astrophysical mixing fraction and Θ denotes the full parameter set. The astrophysical spatial likelihood is modeled as a mixture of Gaussian kernels centered on the J known candidate sources,

$$\mathcal{L}_{\text{Astro, Space}}(\mathbf{x}_i) \propto \sum_{j=1}^{J} \mathcal{N}\left(\mathbf{x}_i \mid \mu_j, \sigma_{\text{eff},ij}^2 \right),$$

where $\sigma_{\text{eff},ij}^2$ combines detector angular resolution and intrinsic source extension. The atmospheric spatial likelihood is assumed to be uniform over the sky, $\mathcal{L}_{\text{Atmo, Space}}(\mathbf{x}_i) \propto 1$. The Gaussian approximation for the spatial signal component is justified under the assumption of point-like sources and small angular uncertainty. A more physically accurate treatment would employ a von Mises-Fisher distribution on the sphere.

In practice, implementing this full unbinned spatial-energetic likelihood on a dataset of $N \simeq 8.8 \times 10^5$ events proved computationally infeasible with the available resources. This limitation necessitated a sequence of progressively simplified models. An initial approximation (Model 1) replaces the explicit spatial likelihood with a pre-computed spatial compatibility weight w_i, which is incorporated directly into the likelihood. A further simplification (Models 2 and 3) reduces the Bayesian inference to the energy domain alone, introducing spatial information only *a posteriori* through odds reweighting to compute the final classification score γ_i^*.

Model 1 constitutes the closest practical approximation to the fully generative conceptual framework that could be implemented within the available computational resources. In this model, both reconstructed energy and directional information enter the likelihood function, with the spatial component incorporated through pre-computed spatial weights. Under the assumption of conditional independence between energy and direction within each class, the class-conditional likelihood is approximated as

$$p(E_i, \mathbf{x}_i \mid z_i) \;\propto\; p(E_i \mid z_i)\, w_i^{\mathbb{I}(z_i = \text{astro})},$$

where $w_i > 0$ is the normalized spatial compatibility weight of event i with the catalog of candidate astrophysical sources. The weights are rescaled to have unit mean over the dataset, so that $w_i > 1$ indicates above-average spatial compatibility with the source population, while $w_i < 1$ indicates below-average compatibility. Atmospheric events are assumed to be isotropically distributed, and therefore no spatial modulation is applied to the atmospheric component. The resulting event-level mixture likelihood is therefore

$$p(E_i, \mathbf{x}_i) = \pi_{\text{astro}}\, w_i\, p(E_i \mid \text{astro}) + (1 - \pi_{\text{astro}})\, p(E_i \mid \text{atmo}).$$

Priors for all three models are summarized in Table 1.

Models 2 and 3 discard directional information entirely and retain only the reconstructed energy as a discriminating variable. The likelihood for each event reduces to a standard two-component mixture model,

$$p(E_i) = \pi_{\text{astro}}\, p(E_i \mid \text{astro}) + (1 - \pi_{\text{astro}})\, p(E_i \mid \text{atmo}).$$

	α_{astro}	α_{atmo}	π_{astro}
Model 1	$\mathcal{N}(2.2, 0.5^2)$	$\mathcal{N}(3.7, 0.3^2)$	$\text{Beta}(1, 99)$
Model 2	$\mathcal{N}(2.2, 0.5^2)$	$\mathcal{N}(3.7, 0.3^2)$	$\text{Beta}(2, 35)$
Model 3	$\text{Uniform}(1, 4)$	$\text{Uniform}(1, 5)$	$\text{Beta}(2, 3)$

Table 1: Prior distributions for the three simplified models.

For both Models 2 and 3, the combined posterior is obtained by a compact Bayesian odds update:

$$\mathcal{O}_{E,i} = \frac{\gamma_{E,i}}{1 - \gamma_{E,i}}, \qquad \mathcal{O}_i^* = \mathcal{O}_{E,i}\, w_i, \qquad \gamma_i^* = \frac{\mathcal{O}_i^*}{1 + \mathcal{O}_i^*}.$$

This provides a concise expression of the energy-to-odds transformation, multiplicative spatial Bayes factor, and back-transformation to posterior probability.

The models were estimated using Stan via the RStan interface. We ran 4 Markov chains for 4,500 iterations each, discarding the first 1,500 iterations as warm-up, yielding 12,000 post-warm-up samples for inference.

4 Results

Across all three models, Markov Chain Monte Carlo diagnostics indicate excellent convergence and sampling efficiency. Autocorrelation functions drop to approximately zero within the first few lags, the four independent chains display mutually consistent traceplots and marginal posterior densities, effective sample sizes are satisfactory, and the Gelman-Rubin statistic remains essentially equal to one. These results indicate that posterior summaries are not limited by poor mixing or lack of convergence. Model 3 diagnostics are shown in Figure 1; notably, model 3 exhibited the worst MCMC performance among the three.

A striking empirical feature shared by all models is the near-identical posterior concentration of the spectral indices. In Model 1 the astrophysical and atmospheric indices concentrate at low values (approximately $\alpha_{\text{astro}} \approx 1.22$ and $\alpha_{\text{atmo}} \approx 1.12$), whereas in Models 2 and 3 both indices collapse to an almost identical value around $\alpha \approx 1.12$. This practical degeneracy implies that, within the explored parameter region, the energy distributions of the two components become empirically indistinguishable. As a consequence, the energy likelihood alone provides no discriminative power between astrophysical and atmospheric origins.

The inferred astrophysical fraction π_{astro} exhibits a strong dependence on prior specification. This prior sensitivity arises directly from the spectral degeneracy: when $\alpha_{\text{astro}} \approx \alpha_{\text{atmo}}$, the energy likelihoods $p(E \mid \text{astro})$ and $p(E \mid \text{atmo})$ become nearly indistinguishable, rendering the data uninformative about the true mixture proportion. In the absence of discriminative information from the energy distribution, π_{astro} remains effectively non-identifiable, and the posterior is dominated by the prior. In Model 1, with a strongly informative prior Beta(1,99), the posterior concentrates at 0.00247. Model 2, under moderately informative priors, yields $\pi_{\text{astro}} \approx 0.0417$, while Model 3, with a weak Beta(2,3) prior, produces a substantially larger estimate of 0.3732 — close to the prior mean of 0.4. Notably, Model 1 exhibits considerably reduced posterior variance relative to its prior, whereas in Models 2 and 3 the posterior distributions of π_{astro} remain nearly identical to their respective priors. This contrast reflects the fact that Model 1 incorporates spatial information directly in the likelihood, allowing the data to constrain π_{astro} despite the energetic degeneracy, while Models 2 and 3 retain spatial weights only for post-hoc reweighting and thus extract no information about the global mixture fraction from the observed data. Despite these large differences in the inferred global mixture fraction, the event-level posterior class probabilities remain tightly constrained below 0.4 in all models.

Although Model 1 directly incorporates spatial weights into the likelihood and Models 2 and 3 apply spatial information only *a posteriori*, the resulting event-level classification scores are nearly perfectly correlated across all approaches. An equally strong correlation is observed between these posterior probabilities and the corresponding spatial weights. The inferred spectral indices collapse to nearly identical values, effectively removing the discriminating power of the energy likelihood. As a result, in Models 2 and 3 the energy-based posterior probability $\gamma_{E,i}$ is nearly uniform across events, and the only mechanism capable of introducing event-level variability in the final classification score γ_i^* is the spatial weight w_i. The inclusion of spatial information mainly induces a smooth rescaling of otherwise nearly uniform classification scores. Spatial information therefore acts more as a mild reshuffling mechanism than as a strong classifier, and the overall ability to separate astrophysical candidates from atmospheric background remains limited.
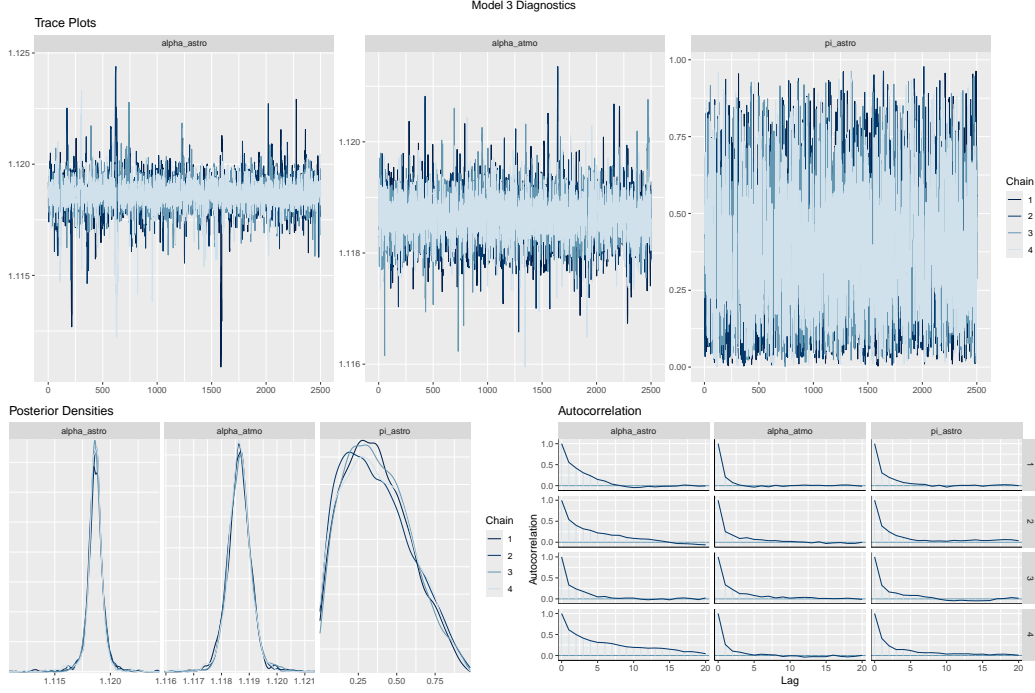
Figure 1: MCMC diagnostics for Model 3. Trace plots for the 3 parameters across four independent chains. Posterior densities, showing stable estimates for $\alpha_{astro}, \alpha_{atmo}$ and broad variability for π_{astro}. Autocorrelation functions: rapid decay confirms effective chain mixing. The variability in π_{astro} reflects the limited ability of energy alone to constrain the astrophysical fraction.

Figure 2a displays the celestial distribution of the top 5% events ranked by astrophysical posterior probability ($\gamma > 0.013$ in Model 1), overlaid with the 110 candidate sources. A clear spatial association emerges: the most probable astrophysical candidates are systematically those in close angular proximity to known source positions. This validates the spatial weighting scheme and confirms that directional information, despite its limited overall discriminative power, does capture the expected geometric correlation between neutrino arrival directions and source locations. The clustering pattern provides the primary evidence for source-associated signal in the dataset.

The relationship between reconstructed energy and posterior astrophysical probability is weak and slightly negative (correlation of order -0.1). Rather than enhancing the astrophysical classification at high energies, increasing energy appears marginally associated with a lower posterior signal probability. This behavior aligns with the collapse of the two spectral indices and reflects the effective loss of energetic discriminative power within the fitted mixture models. Once the component spectra become nearly indistinguishable, variations in reconstructed energy no longer promote a systematic separation between the two physical classes.

Figure 2b visually confirms the spectral degeneracy: with $\alpha_{\text{astro}} \approx \alpha_{\text{atmo}} \approx 1.12$, the model reduces to a single-component mixture, producing unimodal predictive distributions despite the bimodal structure in observed data. This reinforces that energy provides minimal discriminative power, and spatial information drives the event-level classifications.

5 Conclusions

We developed and evaluated a sequence of Bayesian mixture models to classify high-energy neutrino events as astrophysical or atmospheric, spanning a conceptual joint spatial-energetic framework, a practical implementation incorporating spatial weights (Model 1), and energy-only models (Models 2–3) with *a posteriori* spatial reweighting.

The primary finding is a severe spectral degeneracy: inferred power-law indices collapse to nearly identical values ($\alpha_{\text{astro}} \approx \alpha_{\text{atmo}} \approx 1.12$), rendering energy information non-discriminative. Event-level classification is driven almost entirely by spatial proximity to candidate sources rather than spectral characteristics. This degeneracy manifests across all model formulations and is robust to prior specification, though the global astrophysical fraction π_{astro} remains prior-dependent when energy provides no constraint. Posterior predictive checks confirm that the fitted model reduces to an effective single-component mixture, unable to capture the
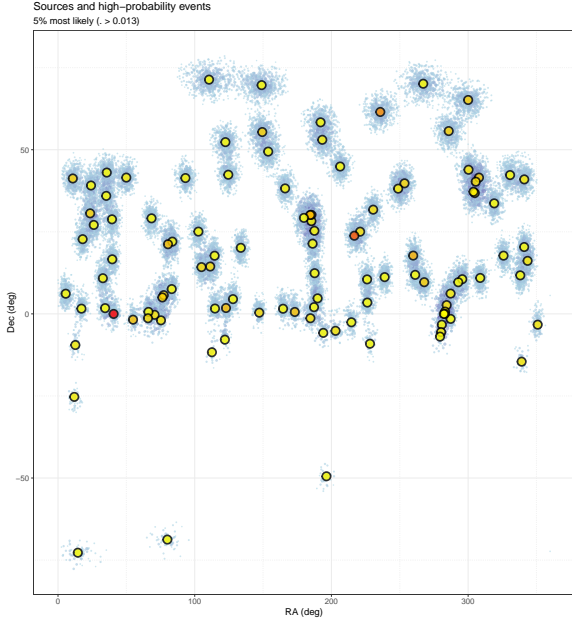
(a) Spatial distribution: top 5% events by γ
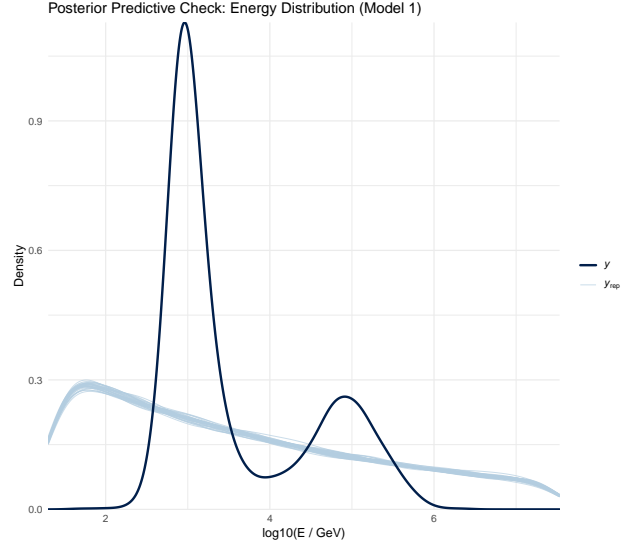
(b) Posterior predictive check

Figure 2: Spatial and energetic diagnostics for Model 1. (a) Events with highest astrophysical posterior probability ($\gamma > 0.013$, top 5%) overlaid with $J = 110$ candidate sources. Colors indicate γ (neutrinos) and $\hat{n}_s$ (sources). Spatial clustering confirms directional information drives classification. (b) Posterior predictive draws (blue) fail to reproduce observed bimodal energy structure (black), confirming spectral degeneracy ($\alpha_{\text{astro}} \approx \alpha_{\text{atmo}} \approx 1.12$).

bimodal structure observed in the empirical energy distribution.

Despite limited discriminative power, the analysis demonstrates a coherent Bayesian workflow for large-scale neutrino event classification, yielding well-calibrated posterior probabilities and stable MCMC inference. Future work should address the spectral degeneracy through more flexible parameterizations (broken power laws, non-parametric spectral models), incorporate additional observables beyond energy and direction, and pursue hierarchical source-level modeling to capture population heterogeneity. Advanced computational approaches such as variational inference may enable the full joint spatial-energetic model at scale, recovering the complete generative framework that current constraints preclude.