Zones de neutrinos : empilement de signatures multi-longueurs d'onde

Recherche non supervisée à l'échelle d'une population d'alertes IceCube

Projet de recherche M1 (10 semaines ; données publiques uniquement) · Yvonne Becherini — Université Paris Cité / APC

Question scientifique et objectifs

Les régions de localisation des alertes IceCube présentent-elles *collectivement* des propriétés gamma et X différentes de celles de régions aléatoires du ciel ? L'*empilement* peut révéler une faible signature commune, invisible à l'échelle d'une alerte, sans présélectionner de population astrophysique.

Trois objectifs : (1) Découvrir une éventuelle signature multi-longueurs d'onde collective ; **(2) interpréter**, en cas de signal, les observables qui portent la différence et proposer des populations candidates à confirmer indépendamment ; **(3) mesurer la sensibilité**, y compris en cas de résultat nul, et vérifier si l'apprentissage non supervisé apporte un gain par rapport à l'empilement classique.

Semaine 1 : lecture critique encadrée (avant toute analyse)

Lire et discuter : **(i)** IceCat-1 (alertes, cartes de vraisemblance, incertitudes) ; **(ii)** IceCube, corrélations avec les catalogues et Giommi et al. (résultats antérieurs et essais multiples) ; **(iii)** 4FGL-DR4 et Swift-BAT, 157 mois (observables et biais des relevés) ; **(iv)** Gretton et al. (2012) (*Maximum Mean Discrepancy*, MMD). **Livrable :** synthèse critique annotée de 1–2 pages (résultats, limites, hypothèse nulle), puis discussion avec l'encadrante.

Données publiques

Utiliser une version identifiée d'IceCat-1 (alertes de traces Gold/Bronze et cartes de *vraisemblance*), Fermi 4FGL-DR4 et Swift-BAT 157 mois ; 2RXS (X mous) reste facultatif. Inclure les sources associées et non associées. Documenter les masques et biais des relevés ; valider la conversion éventuelle des cartes en poids positionnels.

Méthode : empilement et apprentissage non supervisé

1. **Décrire les régions.** Extraire les sources et leurs propriétés (détection gamma/X, flux, spectres, dureté, variabilité) ainsi que leur compatibilité positionnelle avec chaque alerte. Gérer les données manquantes ; définir d'abord des variables simples et interprétables.
2. **Construire le fond.** Produire des ensembles complets d'alertes brouillées en ascension droite (déclinaisons et formes conservées), en gardant fixes catalogues et masques ; contrôler les effets de couverture du ciel. Chaque pseudo-expérience contient le même nombre d'alertes que l'échantillon réel.
3. **Apprendre sans étiquettes.** Ajuster sur les seules régions brouillées une représentation multivariée : ACP d'abord, petit autoencodeur si utile. Garder en référence les caractéristiques physiques sans ML ; figer le modèle avant l'analyse des positions réelles.
4. **Empiler, tester et interpréter.** Comparer la distribution des représentations de toutes les régions IceCube à celle des témoins brouillés par MMD. Calibrer *une valeur p globale* sur des pseudo-expériences entières et indépendantes. Si une différence apparaît, identifier les variables qui y contribuent ; ces pistes restent exploratoires avant confirmation indépendante.
5. **Évaluer la puissance.** Injecter des associations astrophysiquement plausibles (sources gamma seules, X seules, gamma+X) ; mesurer la probabilité de détection selon la fraction d'alertes associées. Comparer comptages empilés, MMD sur variables physiques et MMD sur représentation apprise ; préenregistrer la procédure avant levée de l'aveugle.

Calendrier (10 semaines) et résultats attendus

S1 : bibliographie et discussion. **S2–3 :** catalogues, localisations et variables régionales. **S4 :** brouillage et empilement classique. **S5–6 :** MMD, ML non supervisé et injections. **S7 :** calibration, puissance, gel et préenregistrement. **S8–9 :** levée de l'aveugle, analyse de la signature éventuelle et robustesse. **S10 :** rapport, figures et code reproductible.

Résultat scientifique : rechercher une signature collective gamma/X et, si elle est significative, identifier les propriétés susceptibles de faire émerger des populations candidates. En l'absence de signal, quantifier les scénarios détectables par la méthode. Dans tous les cas : test calibré et comparaison de puissance avec l'empilement classique. Une corrélation ne prouve pas, à elle seule, l'origine des neutrinos.