

Analyse statistique des distributions de sortie du Kernel-Nulling pour la détection à haut contraste d'exoplanètes dans les configurations VLTI et LIFE

Vincent Foriel, David Mary, Frantz Martinache

December 2, 2025

Abstract

L'interférométrie à annulation par noyau (Kernel-Nulling) [MI18] représente une approche prometteuse pour la détection directe d'exoplanètes. Cette technique génère des distributions d'intensité caractéristiques selon la présence ou l'absence d'un compagnon planétaire. L'analyse statistique de ces distributions est essentielle pour une détection robuste de planètes [HMM⁺11]. Nous développons et comparons plusieurs tests statistiques pour discriminer efficacement entre les hypothèses \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (système étoile-planète). Nous analysons les performances de différentes statistiques de test afin d'évaluer la pertinence de chacun d'eux en fonction de la nature des données à disposition. Nous effectuons des simulations numériques pour trois scénarios instrumentaux : les configurations VLTI au sol [C⁺23] et celle de LIFE dans l'espace ainsi qu'un scénario fictif consistant à placer le VLTI dans l'espace afin d'isoler les pertes de performance causées par l'atmosphère. Pour chaque scénario, nous générerons des jeux de données sous les deux hypothèses \mathcal{H}_0 et \mathcal{H}_1 , en tenant compte des paramètres instrumentaux spécifiques et des niveaux de bruit. Nous évaluons les performances des tests en utilisant les courbes ROC et l'analyse des valeurs P. Cette analyse statistique, actuellement appliquée à des données simulées, ouvre la voie à une détection robuste d'exoplanètes à haut contraste utilisant l'interférométrie à annulation par noyau.

0.1 Tests statistiques

En présence d'un compagnon (hypothèse \mathcal{H}_1), la distribution de sortie du Kernel-Nulling est modifiée par rapport au cas de l'étoile seule (hypothèse \mathcal{H}_0) comme le montre la figure 1.

Cette modification se manifeste principalement par un décalage de la distribution, bien que d'autres changements dans la forme de la distribution puissent également survenir. En fonction des conditions, il est par exemple possible de constater un aplatissement de la distribution. Pour détecter efficacement la présence d'un compagnon, nous avons implémenté et évalué plusieurs tests statistiques, chacun étant sensible à différents aspects des distributions.

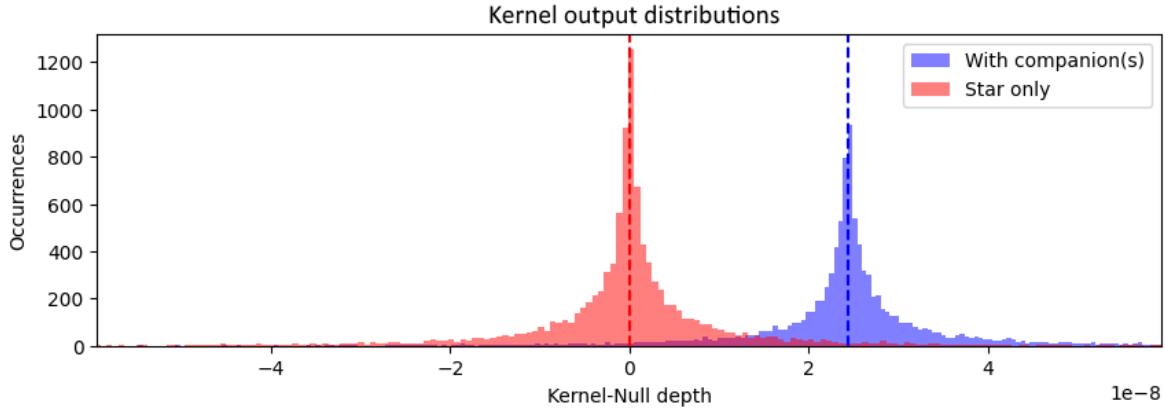


Figure 1: Example of kernel nulling depth distributions for hypotheses \mathcal{H}_0 (star alone) and \mathcal{H}_1 (with companion). This example scenario is heavily exaggerated with a companion having low contrast to induce a significant distribution shift. In practice, the two distributions are generally much closer and difficult to distinguish. Cf texte; aussi prends l'habitude d'être ultra-spécifique, c'est un papier scientifique. Là tu es bcp trop vague : "low contrast" (c'est quoi low ?) "much closer" (=...).

0.1.1 Likelihood Ratio (LR) et approche paramétrique

Le test du rapport de vraisemblance (LR) consiste à comparer la probabilité d'observer les données sous chaque hypothèse :

$$\Lambda(x) = \frac{p(x; \mathcal{H}_1)}{p(x; \mathcal{H}_0)} \quad (1)$$

Ce test est optimal au sens de Neyman-Pearson. Cependant, il nécessite une connaissance analytique des distributions, qui est ici inconnue. Une première approche pour approximer ce test consiste à ajuster des modèles paramétriques (Cauchy, Laplace) sur les distributions simulées, comme discuté précédemment. Cela permet de construire un test de rapport de vraisemblance basé sur ces lois ajustées.

0.1.2 Generalized Likelihood Ratio (GLR)

Une alternative puissante, qui ne nécessite pas de connaître la forme analytique exacte de la distribution des sorties du nul mais qui exploite la connaissance du modèle instrumental direct, est le rapport de vraisemblance généralisé (GLR).

Le principe est d'utiliser le modèle direct de l'instrument pour prédire le signal attendu pour une position et un contraste de planète donnés, et de comparer les résidus observés sous les deux hypothèses. Si l'on suppose que le bruit (après soustraction du modèle) suit une loi connue (par exemple gaussienne ou approximativement gaussienne en raison du théorème central limite sur un grand nombre de données, ou une loi à queues lourdes modélisée empiriquement), on peut écrire la vraisemblance.

Sous \mathcal{H}_0 (pas de planète), le signal attendu est nul (ou égal au fond résiduel). Sous \mathcal{H}_1 , le signal dépend des paramètres de la planète θ (position, contraste). Le GLR consiste à maximiser la vraisemblance sous \mathcal{H}_1 par rapport à ces paramètres inconnus :

$$\Lambda_{GLR}(x) = \frac{\sup_{\theta} p(x|\theta, \mathcal{H}_1)}{p(x|\mathcal{H}_0)} \quad (2)$$

En supposant un bruit indépendant sur chaque mesure x_i , le log-GLR devient :

$$\log \Lambda_{GLR}(x) = \sup_{\theta} \sum_i (\log p(x_i|\theta, \mathcal{H}_1) - \log p(x_i|\mathcal{H}_0)) \quad (3)$$

Cette méthode est particulièrement intéressante car elle permet non seulement la détection mais aussi l'estimation des paramètres de la planète (la valeur de θ qui maximise la vraisemblance). Elle est cependant plus coûteuse en calcul car elle nécessite une optimisation sur la grille des paramètres.

0.1.3 Moyenne

Le test le plus naturel à faire dans ces conditions consiste à comparer la moyenne à un seuil. Etant donné que la planète peut induire un décalage de la distribution dans les deux sens, on utilise la valeur absolue de la moyenne.

$$D_M = \left| \frac{1}{N} \sum_{i=1}^N x_i \right| \stackrel{H_1}{\underset{H_0}{\gtrless}} \xi$$

En pratique, on constate que la moyenne est ici une statistique de test particulièrement peu fiable (figure 2) en raison de sa sensibilité aux valeurs extrêmes des distributions à queues lourdes.

0.1.4 Médiane

La médiane, définie comme la valeur centrale d'un ensemble de données triées, est intrinsèquement plus robuste aux valeurs extrêmes.

$$D_m = \text{median}(|x|) \stackrel{H_1}{\underset{H_0}{\gtrless}} \xi$$

Surprenamment, la médiane s'avère être une statistique de test très performante dans ce contexte (figure 2), surpassant bon nombre de tests plus sophistiqués. Elle capture efficacement le décalage global de la distribution tout en ignorant les queues lourdes non-informatives ou bruitées.

0.1.5 Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov (KS) compare les fonctions de distribution cumulative (CDF) empiriques.

$$D_{KS} = \sup_x |F_{obs}(x) - F_{ref}(x)| \stackrel{H_1}{\underset{H_0}{\gtrless}} \xi$$

Le test KS est sensible à toute déformation de la distribution (décalage, aplatissement, changement de forme). Il s'avère robuste et efficace, mais nécessite un échantillon de référence fiable pour H_0 .

0.1.6 Flattening

Nous avons défini une statistique "flattening" mesurant la dispersion absolue autour de la médiane :

$$D_f = \frac{1}{N} \sum_{i=1}^N |x_i - \tilde{x}| \stackrel{H_1}{\underset{H_0}{\gtrless}} \xi$$

Ce test vise à détecter l'aplatissement de la distribution induit par le compagnon. Bien que moins performant seul, il confirme que la modification de la forme de la distribution est une information exploitable.

0.1.7 Median of Absolute Values (MAV)

Cette statistique combine la robustesse de la médiane et la sensibilité au décalage et à l'étalement :

$$D_{MAV} = \text{median}(|x_i|) \stackrel{H_1}{\underset{H_0}{\gtrless}} \xi$$

C'est l'une des statistiques les plus performantes sur nos simulations, bénéficiant à la fois du décalage de la moyenne (via les valeurs absolues) et de la robustesse de la médiane.

0.1.8 Autres tests considérés

Nous avons également exploré d'autres tests (Cramer-von Mises, Wilcoxon-Mann-Whitney, Anderson-Darling, Brunner-Munzel) qui n'ont pas montré de gain significatif par rapport aux méthodes plus simples présentées ci-dessus.

1 Généralisation

Jusqu'à présent, nous avons considéré l'analyse d'une seule sortie de Kernel-Nulling (un seul noyau) pour une seule configuration instrumentale. Cependant, les instruments réels comme le VLTI ou LIFE produisent plusieurs sorties Kernel simultanées (K1, K2, K3, etc.) [LCM20] et l'observation se fait souvent sur plusieurs poses (rotation de la ligne de base) pour couvrir le plan (u,v).

La généralisation des tests statistiques à ces cas multidimensionnels est cruciale.

1.1 Combinaison de plusieurs Kernels

Les différentes sorties Kernel sont généralement corrélées car elles proviennent des mêmes entrées photométriques perturbées. Cependant, dans une première approximation ou si l'on utilise des décorrélateurs (comme la matrice de covariance du bruit), on peut combiner les statistiques. Pour le GLR, la généralisation est naturelle : la vraisemblance totale est la somme des log-vraisemblances de chaque sortie (en tenant compte de la matrice de covariance Σ si nécessaire) :

$$\log \Lambda_{tot} = \sum_k \log \Lambda_k(x^{(k)}) \quad (4)$$

Pour des statistiques comme la médiane ou la MAV, une approche consiste à calculer la statistique sur chaque Kernel individuellement, puis à combiner ces valeurs (par exemple, somme des carrés ou somme pondérée) pour former une statistique globale.

1.2 Combinaison de plusieurs Poses

Les observations à différents angles de rotation sont temporellement décorrélées (si le temps entre poses est supérieur au temps de cohérence atmosphérique). L'information planétaire varie d'une pose à l'autre (modulation du signal). Le GLR exploite parfaitement cette modulation : le modèle direct $p(x|\theta, \mathcal{H}_1)$ prédit exactement cette variation en fonction de la position θ testée. C'est ici que le GLR prend tout son sens par rapport aux tests scalaires simples (comme la médiane globale) qui pourraient "moyennner" le signal planétaire variable.

2 Résultats

2.1 Courbes ROC

Les courbes ROC (Receiver Operating Characteristic) permettent de comparer l'efficacité de différentes statistiques de test.

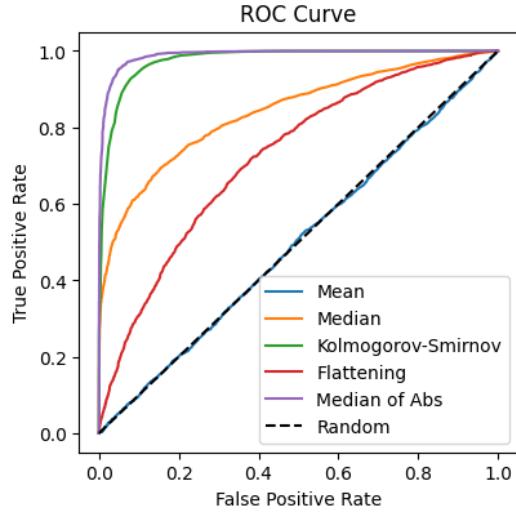


Figure 2: Courbes ROC pour différentes statistiques de test sur les distribution simulées dans le contexte du VLTI avec un compagnon de contraste 10^{-2} .

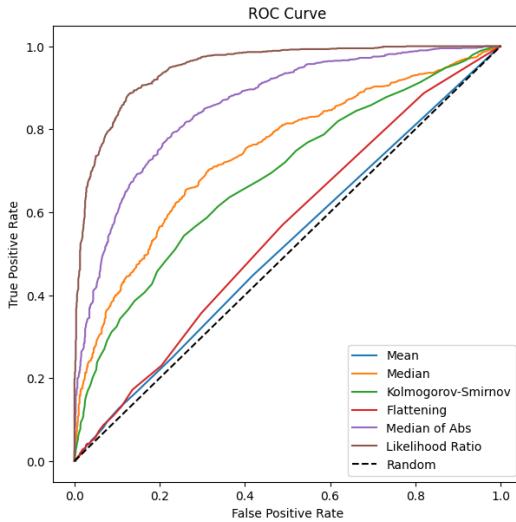


Figure 3: Courbes ROC pour différentes statistiques de test comparées au test optimal de Neyman-Pearson, dans le contexte du VLTI avec un compagnon de contraste 10^{-3} .

2.2 Analyse des valeurs P

Les valeurs P fournissent une mesure de la probabilité d'obtenir une statistique de test au moins aussi extrême que celle observée, sous l'hypothèse nulle.

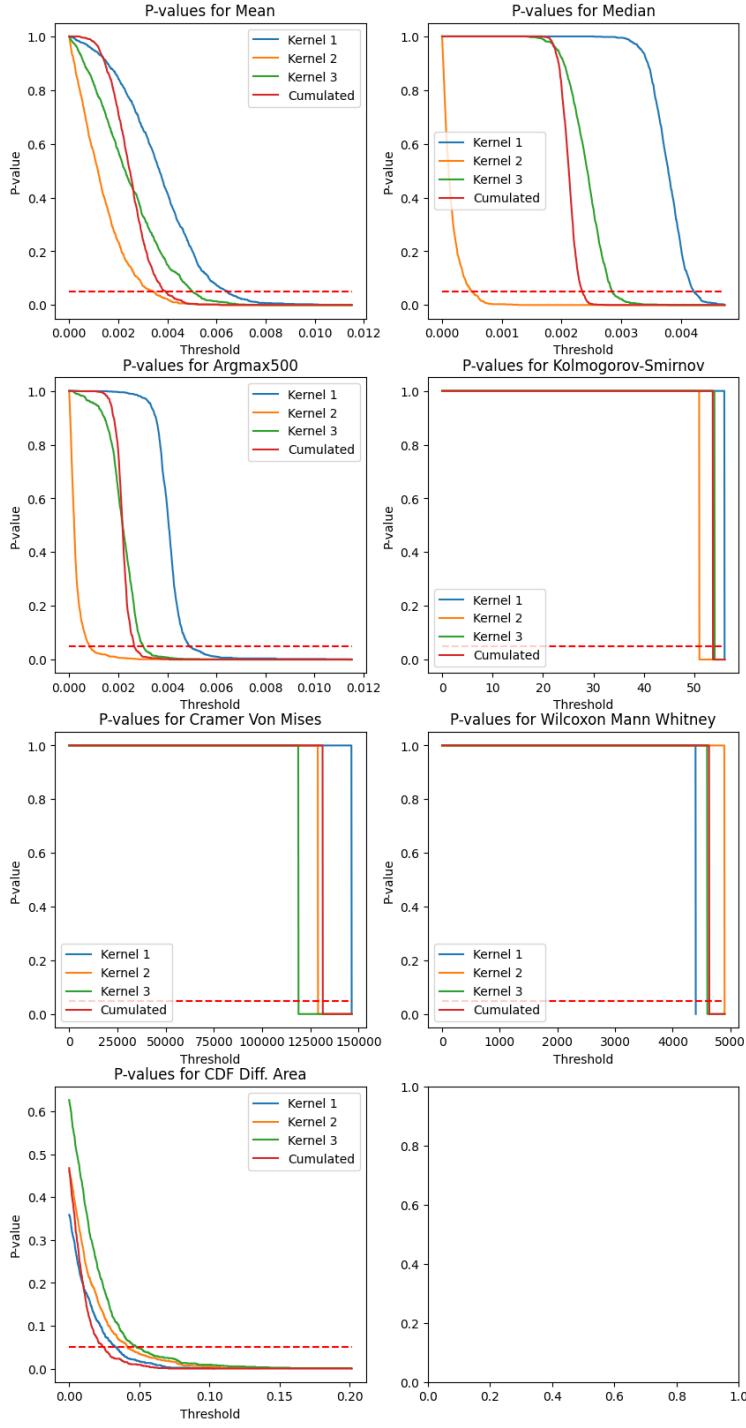


Figure 4: Évolution des valeurs P en fonction du seuil pour différentes statistiques de test.

3 Discussion

3.1 Performance comparative des tests

Nos résultats montrent une hiérarchie claire dans les performances des tests :

1. **GLR et MAV** : Offrent les meilleures performances. Le GLR est optimal car il utilise toute l'information du modèle, tandis que la MAV est une approximation robuste très efficace.

2. **Médiane et KS** : Très performants et robustes, mais nécessitent plus de données pour converger.
3. **Moyenne** : Peu performante en raison de la sensibilité aux queues de distribution.

3.2 Sensibilité au bruit

La robustesse des tests basés sur la médiane (Médiane, MAV) face aux valeurs aberrantes (queues lourdes) est un atout majeur dans le régime de fort bruit atmosphérique (VLTI). Pour LIFE (bruit instrumental dominé, plus stable), les écarts entre moyenne et médiane tendent à se réduire, mais la médiane reste une valeur sûre.

4 Conclusion

Dans cette étude, nous avons analysé les propriétés statistiques des sorties de Kernel-Nulling et évalué plusieurs méthodes de détection.

Bilan des méthodes :

- **Moyenne** :
 - + Simple à calculer.
 - Très sensible aux valeurs aberrantes, peu fiable pour les distributions à queues lourdes.
- **Médiane / MAV** :
 - + Très robuste, excellentes performances empiriques.
 - Nécessite un tri des données (coût $O(N \log N)$), demande un nombre d'échantillons suffisant.
- **Kolmogorov-Smirnov** :
 - + Non-paramétrique, capture toutes les déformations.
 - Moins puissant que les tests ciblés (GLR) pour un signal spécifique.
- **GLR (Generalized Likelihood Ratio)** :
 - + Approche théoriquement optimale, permet l'estimation des paramètres, gère naturellement la fusion de données (multi-kernel, multi-pose).
 - Coût de calcul élevé (optimisation), nécessite un modèle direct précis.

En conclusion, pour une détection rapide et robuste, la **Médiane des Valeurs Absolues (MAV)** est recommandée comme "first look". Pour une analyse approfondie et l'estimation des paramètres de l'exoplanète, le **GLR** constitue la méthode de référence, permettant de combiner rigoureusement toutes les informations disponibles.

References

- [C⁺23] P. M. Chingaip et al. High-contrast detection of exoplanets with a kernel-nuller at the vlti. *Astronomy & Astrophysics*, 676:A43, 2023.
- [HMM⁺11] C. Hanot, B. Mennesson, S. Martin, K. Liewer, F. Loya, D. Mawet, P. Riaud, O. Absil, and E. Serabyn. Improving interferometric null depth measurements using statistical distributions: Theory and first results with the palomar fiber nuller. *The Astrophysical Journal*, 729(2):110, 2011.
- [LCM20] Romain Laugier, Nick Cvetojevic, and Frantz Martinache. Kernel nullers for an arbitrary number of apertures. *Astronomy & Astrophysics*, 642:A202, October 2020.
- [MI18] Frantz Martinache and Michael J. Ireland. Kernel-nulling for a robust direct interferometric detection of extrasolar planets. *Astronomy & Astrophysics*, 619:A87, 2018.