

Stage/These 2025

Title Cosmological parameter inference using theoretical Wavelet statistics predictions

Laboratory: IRFU/DAP/CosmoStat, CEA Saclay

Supervisors: Jean-Luc Starck & Sandrine Codis

Contact: jstarck@cea.fr ☎ : 01 69 08 57 64 <http://jstarck.cosmostat.org>

sandrine.codis@cea.fr ☎ : 01 69 08 35 87 <https://sandrinecodis.wixsite.com/sandrinewebsite>

Keywords: weak lensing, Euclid, cosmological parameter likelihood free inference, high order statistics

Subject:

Launched in 2023, the Euclid satellite is surveying the sky in optical and infrared wavelengths to create an unprecedented map of the Universe's large-scale structure. A cornerstone of its mission is the measurement of weak gravitational lensing—subtle distortions in the shapes of distant galaxies. This phenomenon is a powerful cosmological probe, capable of tracing the evolution of dark matter and helping to distinguish between dark energy and modified gravity theories.

Traditionally, cosmologists have analyzed weak lensing data using second-order statistics (like the power spectrum) paired with a Gaussian likelihood model. This established approach, however, faces significant challenges:

- **Loss of Information:** Second-order statistics fully capture information only if the underlying matter distribution is Gaussian. In reality, the cosmic web is highly structured, with clusters, filaments, and voids, making this approach inherently lossy.
- **Complex Covariance:** The method requires estimating a covariance matrix, which is both cosmology-dependent and non-Gaussian. This necessitates running thousands of computationally intensive N-body simulations for each model, a massive and often impractical undertaking.
- **Systematic Errors:** Incorporating real-world complications—such as survey masks, intrinsic galaxy alignments, and baryonic feedback—into this framework is notoriously difficult.

In response to these limitations, a new paradigm has emerged: **likelihood-free inference via forward modelling**. This technique bypasses the need for a covariance matrix by directly comparing real data to synthetic observables generated from a forward model. Its advantages are profound: it eliminates the storage and computational burden of massive simulation sets, naturally incorporates high-order statistical information, and can seamlessly integrate systematic effects. However, this new method has its own hurdles: it demands immense GPU resources to process Euclid-sized surveys, and its conclusions are only as reliable as the simulations it uses, potentially leading to circular debates if simulations and observations disagree.

A **recent breakthrough** (Tinnaneni Sreekanth, 2024) offers a compelling path forward. This work provides the first theoretical framework to directly predict key wavelet statistics of weak lensing convergence maps—exactly the kind Euclid will produce—for any given set of cosmological parameters. It has been shown in Ajani et al (2021) that the wavelet coefficient L1-norm is extremely powerful to constraint the cosmological parameters. This innovation promises to harness the power of advanced, non-Gaussian statistics without the traditional computational overhead, potentially unlocking a new era of precision cosmology.

We have demonstrated that this theoretical prediction can be used to build a highly efficient emulator (Tinnaneri Sreekanth et al, 2025), dramatically accelerating the computation of these non-Gaussian statistics. However, it is crucial to note that this emulator, in its current stage, provides only the mean statistic and does not include cosmic variance. As such, it cannot yet be used for full statistical inference on its own.

This PhD thesis aims to revolutionize the analysis of weak lensing data by constructing a complete, end-to-end framework for likelihood-free cosmological inference. The project begins by addressing the core challenge of stochasticity: we will first calculate the theoretical covariance of wavelet statistics, providing a rigorous mathematical description of their uncertainty. This model will then be embedded into a stochastic map generator, creating realistic mock data that captures the inherent variability of the Universe.

To ensure our results are robust, we will integrate a comprehensive suite of systematic effects—such as noise, masks, intrinsic alignments, and baryonic physics—into the forward model. The complete pipeline will be integrated and validated within a simulation-based inference framework, rigorously testing its power to

recover unbiased cosmological parameters. The culmination of this work will be the application of our validated tool to the Euclid weak lensing data, where we will leverage non-Gaussian information to place competitive constraints on dark energy and modified gravity.

References

1. V. Ajani, J.-L. Starck and V. Pettorino, "[Starlet l1-norm for weak lensing cosmology](#)", **Astronomy and Astrophysics**, 645, L11, 2021.
2. V. Tinnaneri Sreekanth, S. Codis, A. Barthelemy, and J.-L. Starck, "[Theoretical wavelet l1-norm from one-point PDF prediction](#)", **Astronomy and Astrophysics**, 691, id.A80, 2024.
3. V. Tinnaneri Sreekanth, J.-L. Starck and S. Codis, "[Generative modeling of convergence maps based in LDT theoretical prediction](#)", **Astronomy and Astrophysics**, 701, id.A170, 2025.

Résumé:

Lancé en 2023, le satellite Euclid observe le ciel dans les longueurs d'onde optiques et infrarouges pour cartifier la structure à grande échelle de l'Univers avec une précision inédite. Un pilier fondamental de sa mission est la mesure du **cisaillement gravitationnel faible** — de subtiles distorsions dans la forme des galaxies lointaines. Ce phénomène constitue une sonde cosmologique puissante, capable de retracer l'évolution de la matière noire et d'aider à distinguer les théories sur l'énergie noire de celles de la gravité modifiée.

Traditionnellement, les cosmologues analysent les données de cisaillement faible à l'aide de statistiques du second ordre (comme le spectre de puissance) couplées à un modèle de vraisemblance gaussien. Cette approche établie rencontre cependant des défis significatifs :

1. **Perte d'information** : Les statistiques du second ordre ne capturent toute l'information disponible que si la distribution de matière sous-jacente est gaussienne. En réalité, la toile cosmique est une structure complexe, composée d'amas, de filaments et de vides, ce qui rend cette approche intrinsèquement incomplète.
2. **Covariance complexe** : La méthode nécessite l'estimation d'une matrice de covariance, qui est à la fois dépendante de la cosmologie et non-gaussienne. Ceci exige de réaliser des milliers de simulations numériques de type N-corps, extrêmement coûteuses en calcul, pour chaque modèle cosmologique, un effort souvent prohibitif.
3. **Effets systématiques** : L'intégration des complications observationnelles — telles que les masques de survey, l'alignement intrinsèque des galaxies, et les effets de rétroaction baryonique — dans ce cadre théorique est notoirement difficile.

Face à ces limitations, un nouveau paradigme a émergé : **l'inférence sans vraisemblance par modélisation directe (forward modelling)**. Cette technique contourne le besoin d'une matrice de covariance en comparant directement les données observées à des observables synthétiques générés par un **modèle direct** (forward model). Ses avantages sont profonds : elle élimine le fardeau de stockage et de calcul lié aux vastes ensembles de simulations, intègre naturellement l'information statistique d'ordre supérieur, et permet d'inclure de manière transparente les effets systématiques.

Cependant, cette nouvelle méthode présente ses propres obstacles : elle demande des ressources de calcul (GPU) immenses pour traiter des surveys de l'envergure d'Euclide, et ses conclusions ne sont aussi fiables que les simulations sur lesquelles elle s'appuie, ce qui peut mener à des débats circulaires si les simulations et les observations divergent.

Une percée récente (Tinnaneni Sreekanth, 2024) ouvre une voie prometteuse. Ces travaux fournissent le premier cadre théorique permettant de prédire directement les principales statistiques en ondelettes des cartes de convergence — exactement le type de cartes qu'Euclide produira — pour un jeu de paramètres cosmologiques donné. Il a été démontré dans Ajani et al. (2021) que la norme L1 des coefficients en ondelettes est extrêmement puissante pour contraindre les paramètres cosmologiques. Cette innovation promet d'exploiter la puissance des statistiques non-gaussiennes avancées sans le surcoût computationnel traditionnel, ouvrant potentiellement la voie à une nouvelle ère de cosmologie de précision.

Nous avons démontré que cette prédiction théorique peut être utilisée pour construire un émulateur hautement efficace (Tinnaneni Sreekanth et al., 2025), accélérant considérablement le calcul de ces statistiques non-gaussiennes. Il est crucial de noter qu'à son stade actuel, cet émulateur ne fournit que la statistique moyenne et n'inclut pas la variance cosmique. En l'état, il ne peut donc pas encore être utilisé seul pour une inférence statistique complète.

Objectif de cette thèse de doctorat

Cette thèse de doctorat vise à révolutionner l'analyse des données de cisaillement faible en construisant un cadre complet et intégré pour l'inférence cosmologique sans vraisemblance. Le projet commence par adresser le défi fondamental de la stochasticité : nous calculerons d'abord la **covariance théorique des statistiques en ondelettes**, fournissant une description mathématique rigoureuse de leur incertitude. Ce modèle sera ensuite intégré dans un **générateur de cartes stochastiques**, créant ainsi des données synthétiques réalistes qui capturent la variabilité intrinsèque de l'Univers.

Pour garantir la robustesse de nos résultats, nous intégrerons une suite complète **d'effets systématiques** — tels que le bruit, les masques observationnels, les alignements intrinsèques et la physique baryonique — dans le modèle direct. Le pipeline complet sera intégré et **validé au sein d'un cadre d'inférence basée sur les simulations**, en testant rigoureusement sa capacité à retrouver des paramètres cosmologiques non biaisés. L'aboutissement de ce travail sera l'application de notre outil validé aux **données de cisaillement faible d'Euclide**, où nous exploiterons l'information non-gaussienne pour poser des contraintes compétitives sur l'énergie noire et la gravité modifiée.

References

4. V. Ajani, J.-L. Starck and V. Pettorino, "[Starlet l1-norm for weak lensing cosmology](#)", **Astronomy and Astrophysics**, 645, L11, 2021.
5. V. Tinnaneni Sreekanth, S. Codis, A. Barthelemy, and J.-L. Starck, "[Theoretical wavelet l1-norm from one-point PDF prediction](#)", **Astronomy and Astrophysics**, 691, id.A80, 2024.
6. V. Tinnaneni Sreekanth, J.-L. Starck and S. Codis, "[Generative modeling of convergence maps based in LDT theoretical prediction](#)", **Astronomy and Astrophysics**, 701, id.A170, 2025.