

## Euclid

---

**Title** Optimal Transport and Deep Learning to model the Euclid Point Spread Function

**Laboratory:** IRFU/DAP/CosmoStat, CEA Saclay

**Supervisors:** Jean-Luc Starck and Martin Kilbinger

**Contact:** [jstarck@cea.fr](mailto:jstarck@cea.fr) ☎ : 01 69 08 57 64 <http://jstarck.cosmostat.org>

**Keywords:** Optimal transport, deep learning, graph, Euclid, Cosmology, PSF, cosmology, weak lensing.

### Subject:

The Euclid satellite, to be launched in 2022, will observe the sky in the optical and infrared, and will be able to map large scale structures and weak lensing distortions out to high redshifts. Weak gravitational lensing is thought to be one of the most promising tools of cosmology to constrain models. Weak lensing probes the evolution of dark-matter structures and can help distinguish between dark energy and models of modified gravity. Thanks to the shear measurements, we will be able to reconstruct a dark matter mass map of 15000 square degrees. These shear measurements are derived from the galaxy shapes, which are blurred by the PSF (point-spread function) of the optical imaging system. One of the main problems to achieve the scientific goals is therefore the need to model the point spread function (PSF) of the instrument with a very high accuracy. The PSF field can be estimated from the stars contained in the acquired images. It has to take into account the spatial and spectral variation of the PSF. An additional problem to take care of is the subsampling of the images. Once the PSF is correctly modelled, we need to derive the shear from galaxy shapes.

In a recent paper (Schmitz et al 2018) we shown that optimal transport (OT) techniques allow to extremely well represent the evolution of the PSF with the wavelength and on-going work (Morgan et al, 2018) consists in building a 3D Euclid PSF modelling, which takes into account both the spatial variation of the PSF and the PSF wavelength dependency. However even if OT produces beautiful results, its use is extremely limited in practice due to a prohibitive computational cost, and we cannot consider to use our OT PSF modeling for the huge Euclid set.

The goal of the PhD consists first in finding an efficient way to build such a 3D PSF model. A solution could be to use the Deep Wasserstein Embedding technique (Courty, Flamary and Ducoffe, 2017) to get an approximation mechanism that allows to break the complexity. The second step will be to interpolate, from the reconstructed 3D PSFs at stars position, the PSF at any position in the field. This will done by extending to the third dimension the 2D interpolation on a Graph Laplacian we proposed in (Schmitz, Starck and Ngole, 2018), which allows us to interpolate the PSF on the adequate manifold. The final step will be to quantify the modelling errors by studying using simulations the propagation of the reconstructed PSFs errors to cosmological parameters.

### Sujet:

Le projet spatial Euclid, dont le lancement est prévu en 2020, observera le ciel en optique et en infrarouge et permettra de construire des cartes de très grandes échelles afin mesurer les distorsions gravitationnelles jusqu'à des redshifts très élevés. Grâce à ces mesures de cisaillement gravitationnel, nous pourrons reconstruire des cartes de matières noires de 15000 degrés carrés, soit presque la moitié du ciel. Ces mesures de shear sont dérivées de l'analyse des formes de galaxies, qui sont floutées dues l'optique du télescope. L'un des problèmes majeurs pour atteindre les objectifs scientifiques est donc la nécessité de modéliser la fonction d'étalement du point (Point Spread Function (PSF) en anglais) du satellite, et de mesurer la forme des galaxies avec une très grande précision et corrigée de la PSF. Le champ de PSF peut être calculé à partir des étoiles contenues dans les images observées. Il doit prendre en compte la variation spatiale et spectrale de la réponse de l'instrument. Une difficulté supplémentaire vient du problème de sous échantillonnage des images.

Nous avons récemment montré (Schmitz et al 2018) que les techniques de transport optimal (OT) nous permettent de bien modéliser la variation de la PSF avec la longueur d'onde et des travaux en cours (Schmitz et al 2018) consistent à construire une modélisation tridimensionnelle de la PSF, qui tient en compte à la fois de la variation spatiale de la PSF et de sa dépendance à la longueur d'onde. Si le transport optimal permet d'obtenir de très beaux résultats, son utilité est limitée en pratique, pour des raisons de temps de calcul trop importants dans le cas de grands volumes de données comme ceux liés au projet Euclid.

Le but de cette thèse est de trouver une solution efficace pour construire un tel modèle tridimensionnelle de PSF. Une solution pourrait être d'utiliser la technique *Deep Wasserstein Embedding* (Courty, Flamary and Ducoffe, 2017) afin d'obtenir une approximation de la solution, mais avec une complexité bien meilleure. La seconde étape sera d'interpoler, à partir des PSFs 3D reconstruites aux positions des étoiles dans le champ, la PSF à n'importe quelle position spatiale. Ceci se fera en étendant à la troisième dimension l'interpolation 2D basée sur *Graph Laplacian* (Schmitz, Starck and Ngole, 2018), qui permet d'effectuer l'interpolation sur la variété adéquate. La dernière étape sera de quantifier l'erreur de modélisation en étudiant à partir de simulation la propagation de ces erreurs de reconstruction sur l'estimation des paramètres cosmologiques.

### **References:**

1. M. A. Schmitz et al, "[Wasserstein Dictionary Learning: Optimal Transport-based unsupervised representation learning](#)", *SIAM Journal on Imaging Sciences*, 11(1), 643–678, 2018.
2. M.A. Schmitz et al, "3D Euclid PSF Modeling using Optimal Transport", in preparation.
3. N. Courty, R. Flamary, M. Ducoffe, "Learning Wasserstein Embeddings", International Conference on Learning Representations (ICLR), 2018.
4. M.A. Schmitz, J.-L. Starck and F.M. Ngolè, "Euclid Point Spread Function field recovery through interpolation on a Graph Laplacian", submitted, 2018.