Deep Learning Mass Mapping with Conformal Predictions

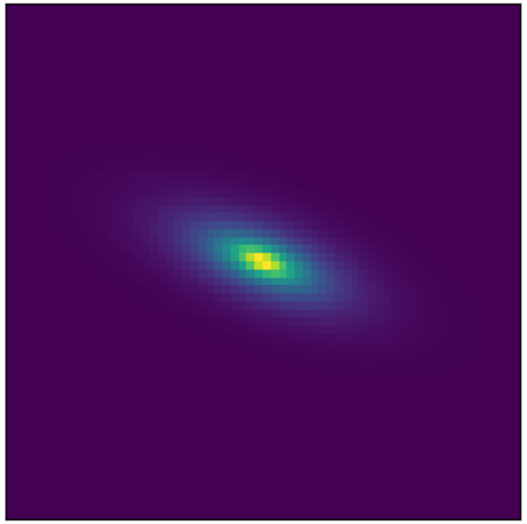
Hubert Leterme, Postdoc
Affiliated to GREYC CNRS-Ensicaen (Caen, France)
Co-supervised at CosmoStat, CEA DAp

TOSCA meeting, Nice

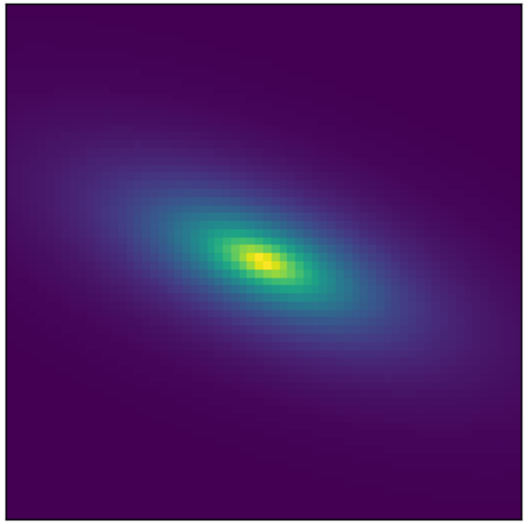
7th November 2024

Context and objectives

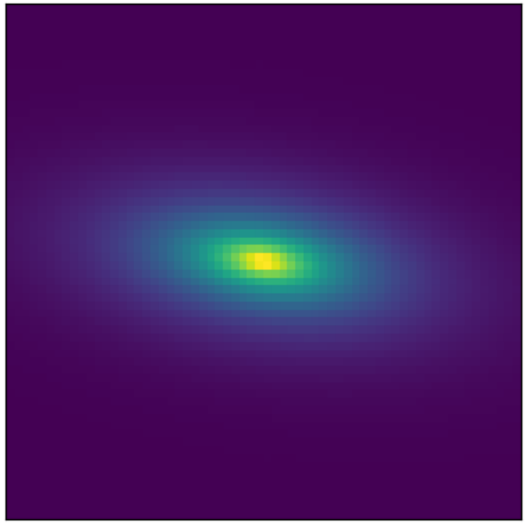
- Convergence map $\boldsymbol{\kappa} \in \mathbb{R}^K$: isotropic dilation of the galaxy image.
  - Proportional to the projected mass along the line of sight.
  - Used to constrain cosmological parameters $\Rightarrow$ **variable of interest**.
  - However, $\boldsymbol{\kappa}$ cannot be directly measured.
- Shear map $\boldsymbol{\gamma} \in \mathbb{C}^K$: anisotropic stretching of the galaxy image.
- Relationship between shear and convergence maps: $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\kappa}$, with $\mathbf{A} \in \mathbb{R}^{K \times K}$ (known).



Source galaxy, unlensed



Convergence only
$\kappa = 1$

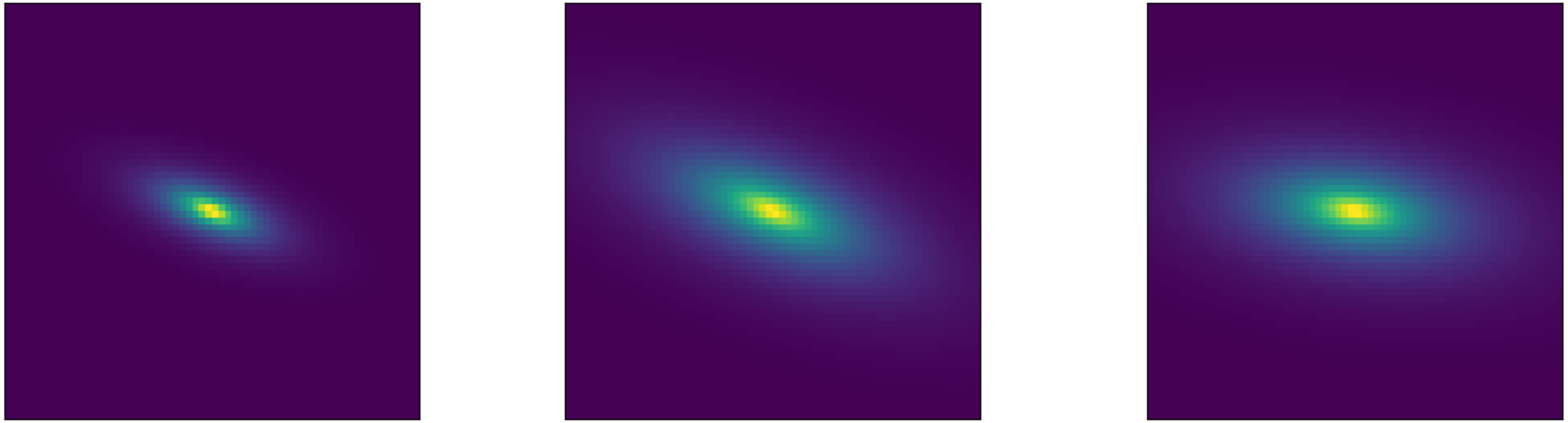


Convergence + shear
$\kappa = 1$ and $\gamma = (0.1 - 0.3\, i)$

Context and objectives

- Convergence map $\boldsymbol{\kappa} \in \mathbb{R}^K$: isotropic dilation of the galaxy image.
  - Proportional to the projected mass along the line of sight.
  - Used to constrain cosmological parameters $\Rightarrow$ **variable of interest**.
  - However, $\boldsymbol{\kappa}$ cannot be directly measured.
- Shear map $\boldsymbol{\gamma} \in \mathbb{C}^K$: anisotropic stretching of the galaxy image.
- Relationship between shear and convergence maps: $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\kappa}$, with $\mathbf{A} \in \mathbb{R}^{K \times K}$ (known).

After mean-centering (mass-sheet degeneracy)



Source galaxy, unlensed

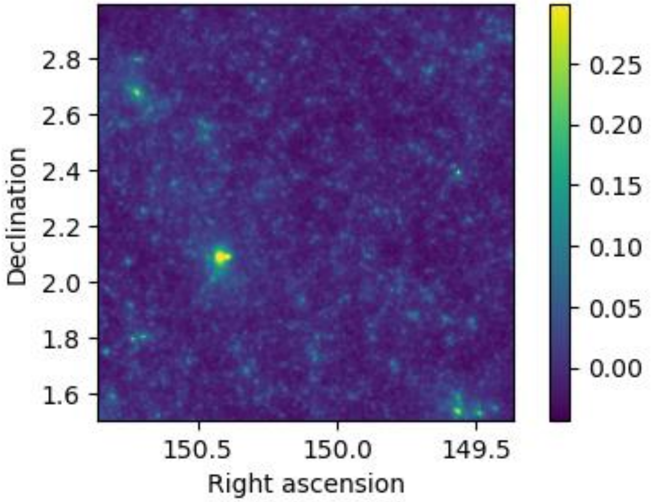
Convergence only $\kappa = 1$

Convergence + shear $\kappa = 1$ and $\gamma = (0.1 - 0.3\, i)$

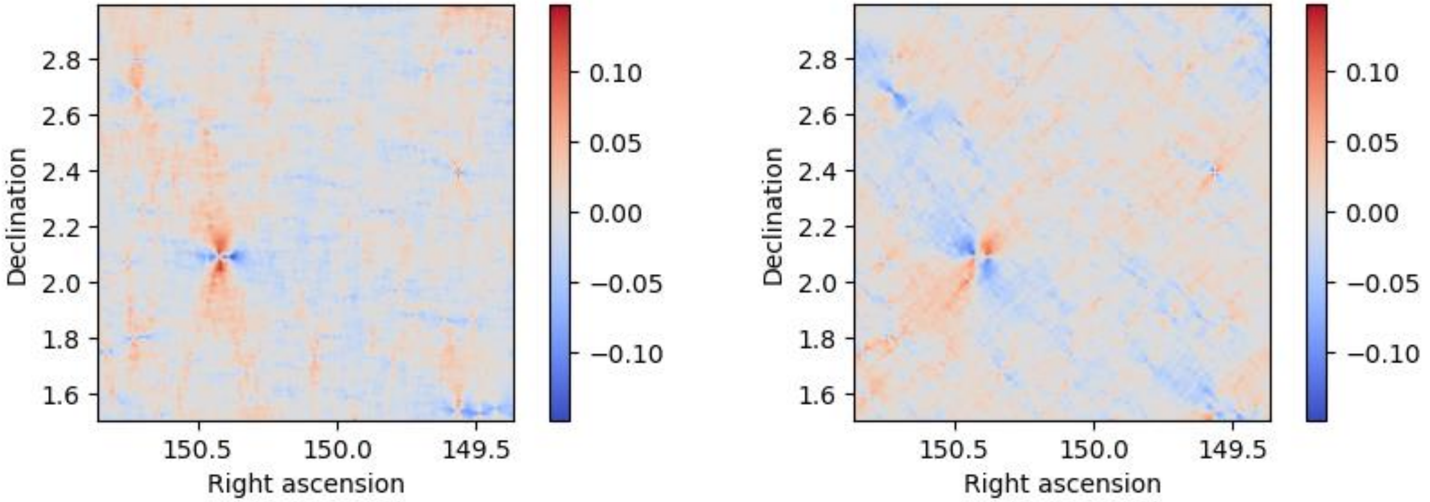
Context and objectives

Example with the κTNG simulated dataset[1]

Simulated convergence map



Corresponding shear map (real and imaginary parts)



- As for the convergence map $\boldsymbol{\kappa}$, the true shear map $\boldsymbol{\gamma}$ cannot be directly measured.
- Unbiased estimator of $\boldsymbol{\gamma}$, obtained by measuring galaxy ellipticities: $\boldsymbol{\gamma} \longleftarrow \boldsymbol{\epsilon} - \langle \boldsymbol{\epsilon} \rangle$
- Relation between $\boldsymbol{\gamma}$ (observable) and $\boldsymbol{\kappa}$ (quantity of interest):

$$\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\kappa} + \boldsymbol{n},$$

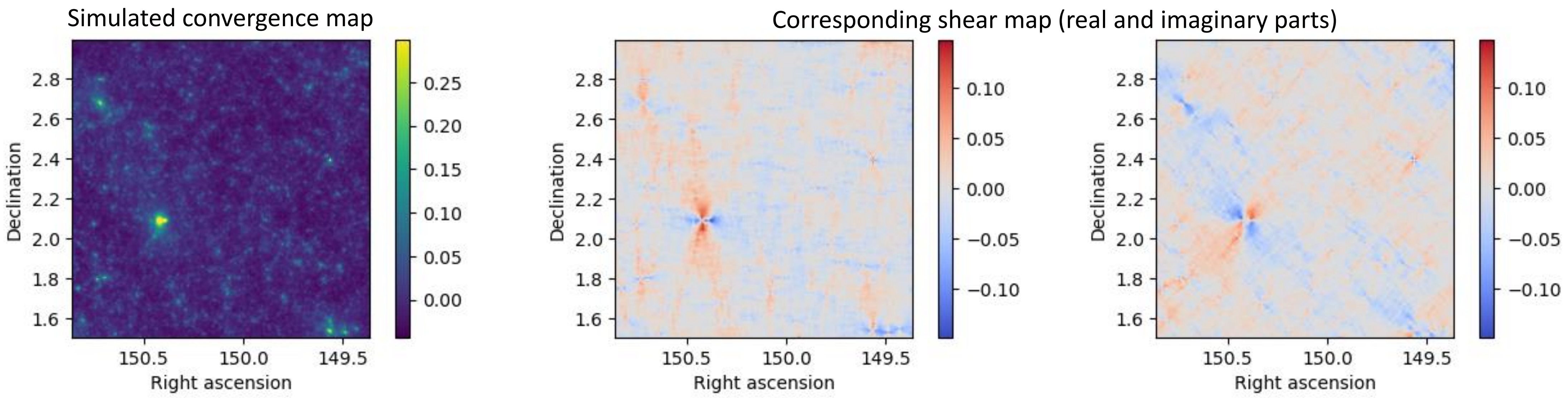
with noise $\boldsymbol{n}$ assumed Gaussian, zero-centered and with diagonal covariance matrix $\boldsymbol{\Sigma}$.

- Noise level (standard deviation per pixel): $\boldsymbol{\Sigma}[k,k] = \sigma / N_k$.

[1] K. Osato, J. Liu, and Z. Haiman, “κTNG: effect of baryonic processes on weak lensing with IllustrisTNG simulations,” Monthly Notices of the Royal Astronomical Society, vol. 502, no. 4, pp. 5593–5602, Apr. 2021

Context and objectives

Example with the κTNG simulated dataset[1]



- As for the convergence map $\boldsymbol{\kappa}$, the true shear map $\boldsymbol{\gamma}$ cannot be directly measured.
- Unbiased estimator of $\boldsymbol{\gamma}$, obtained by measuring galaxy ellipticities: $\boldsymbol{\gamma} \longleftarrow \boldsymbol{\epsilon} - \langle \boldsymbol{\epsilon} \rangle$
- Relation between $\boldsymbol{\gamma}$ (observable) and $\boldsymbol{\kappa}$ (quantity of interest):

$$\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\kappa} + \boldsymbol{n},$$

with noise $\boldsymbol{n}$ assumed Gaussian, zero-centered and with diagonal covariance matrix $\boldsymbol{\Sigma}$.

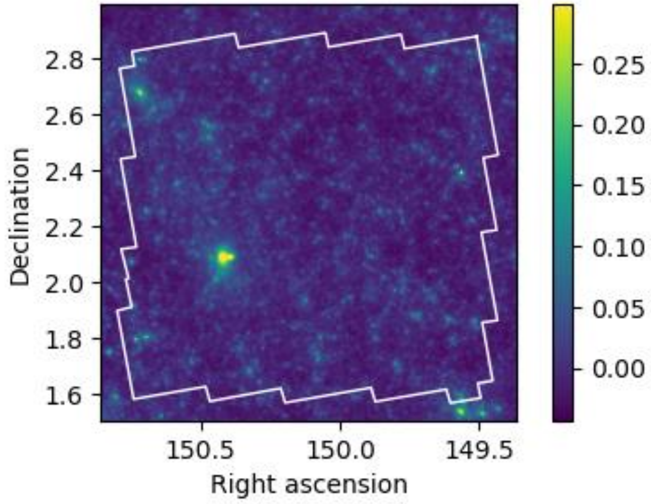
- Noise level (standard deviation per pixel): $\boldsymbol{\Sigma}[k,k] = \sigma / N_k$. ← Nb measured galaxies; Intrinsic ellipticity (std)

[1] K. Osato, J. Liu, and Z. Haiman, "κTNG: effect of baryonic processes on weak lensing with IllustrisTNG simulations," Monthly Notices of the Royal Astronomical Society, vol. 502, no. 4, pp. 5593–5602, Apr. 2021

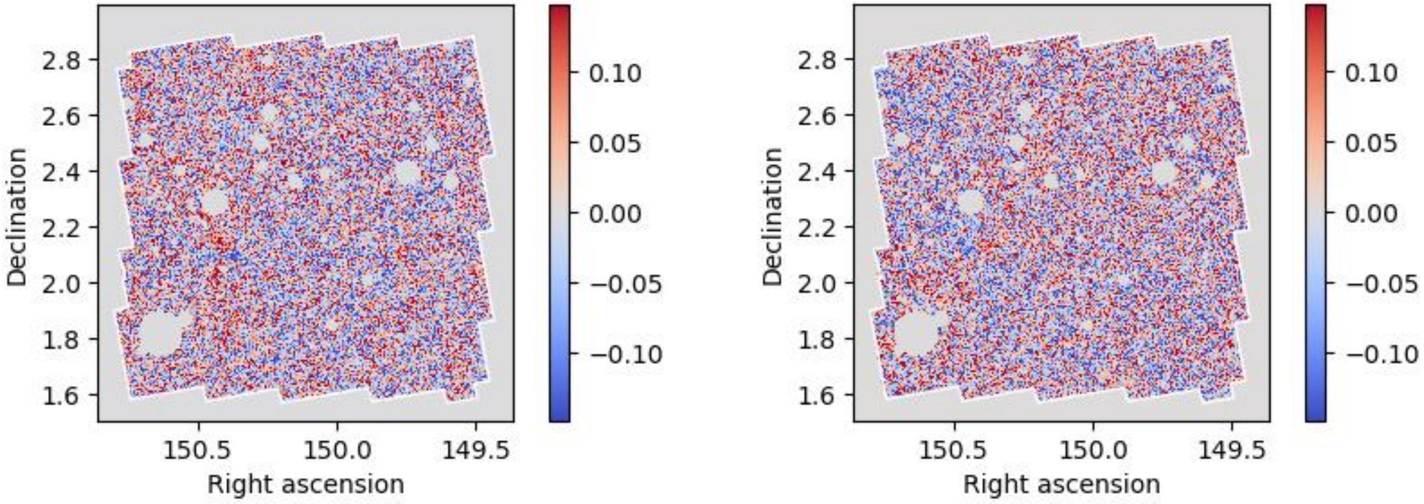
Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])

Simulated convergence map



Corresponding shear map with noise (real and imaginary parts)



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^{-}$ and $\widehat{\boldsymbol{\kappa}}^{+}$ such that

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^{-}, \widehat{\boldsymbol{\kappa}}^{+})] \leq \alpha.$$

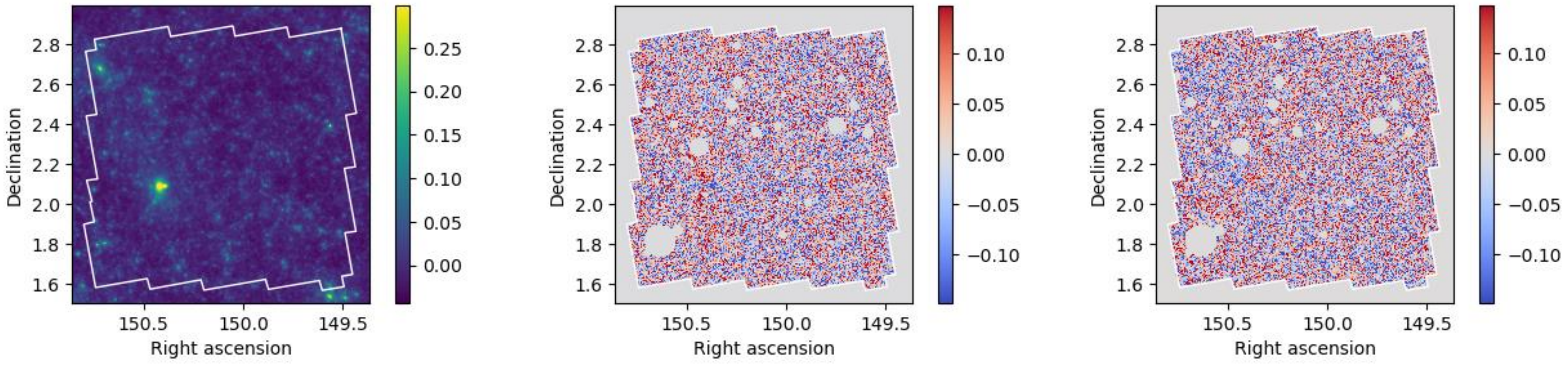
- Over which uncertainties the expected value is calculated?

[1] https://astro.uni-bonn.de/en/m/schrabba/research

Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])

Simulated convergence map   Corresponding shear map with noise (real and imaginary parts)



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^-$ and $\widehat{\boldsymbol{\kappa}}^+$ such that

Expected miscoverage rate
(% of pixels outside the bounds)

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^-, \widehat{\boldsymbol{\kappa}}^+)] \leq \alpha.$$

- Over which uncertainties the expected value is calculated?

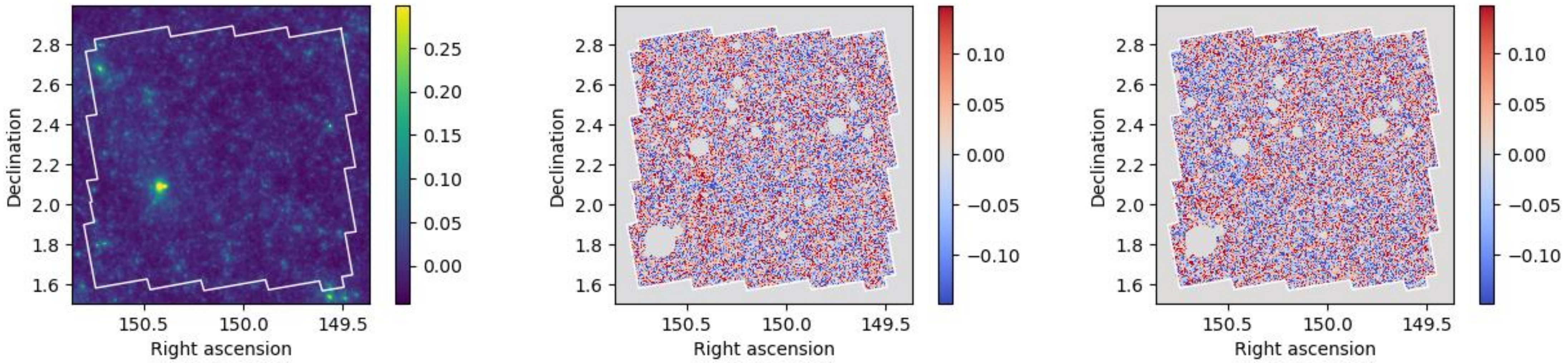
[1] https://astro.uni-bonn.de/en/m/schrabba/research

Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])

Simulated convergence map

Corresponding shear map with noise (real and imaginary parts)



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^{-}$ and $\widehat{\boldsymbol{\kappa}}^{+}$ such that

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^{-}, \widehat{\boldsymbol{\kappa}}^{+})] \leq \alpha.$$

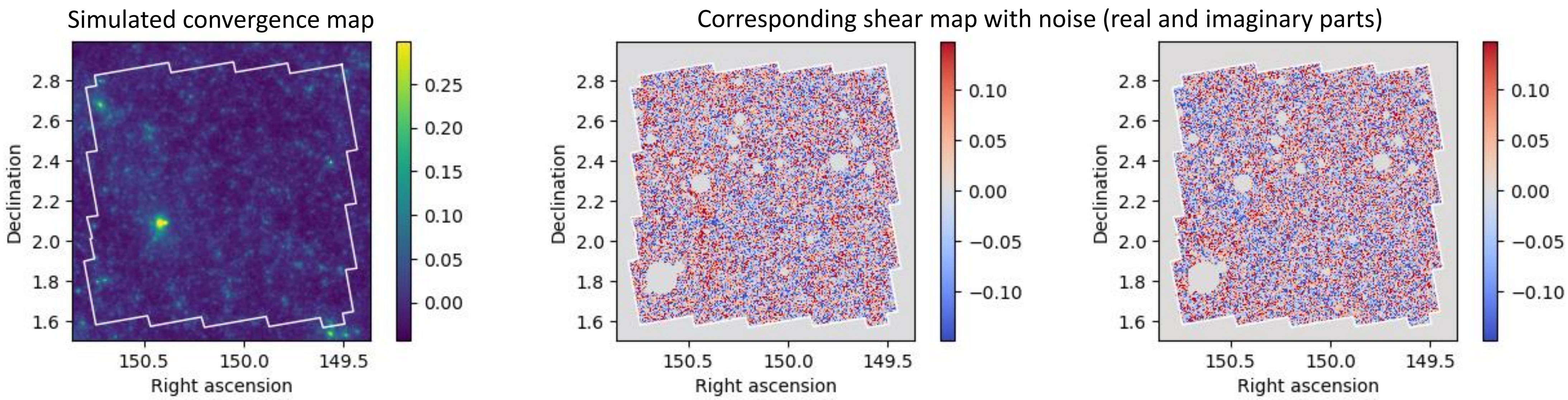
Confidence level $\in]0, 1[$

- Over which uncertainties the expected value is calculated?

[1] https://astro.uni-bonn.de/en/m/schrabba/research

Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^-$ and $\widehat{\boldsymbol{\kappa}}^+$ such that

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^-, \widehat{\boldsymbol{\kappa}}^+)] \leq \alpha.$$

May be random

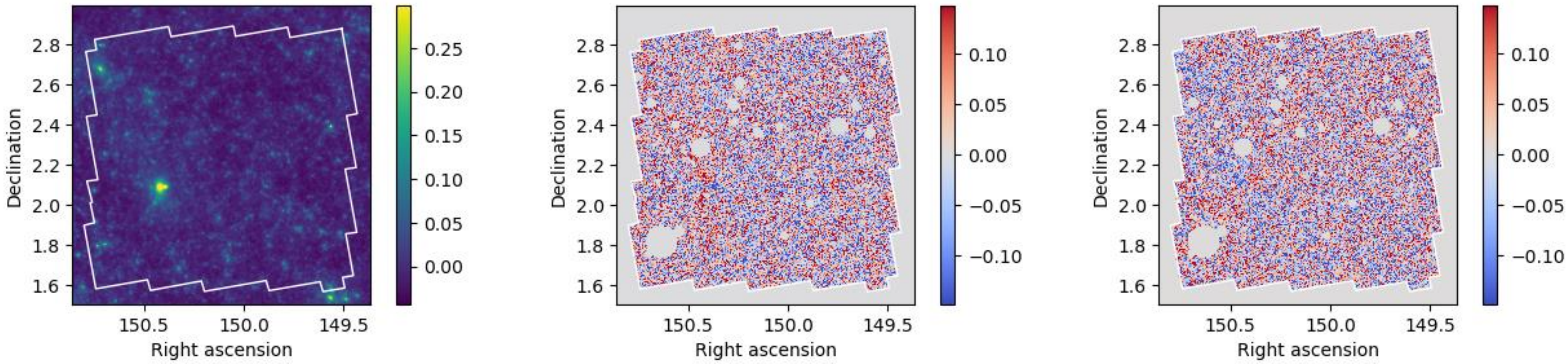
- Over which uncertainties the expected value is calculated?

[1] https://astro.uni-bonn.de/en/m/schrabba/research

Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])

Simulated convergence map — Corresponding shear map with noise (real and imaginary parts)



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^-$ and $\widehat{\boldsymbol{\kappa}}^+$ such that

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^-, \widehat{\boldsymbol{\kappa}}^+)] \leq \alpha.$$

Depends on $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\kappa} + \boldsymbol{n}$

- Over which uncertainties the expected value is calculated?

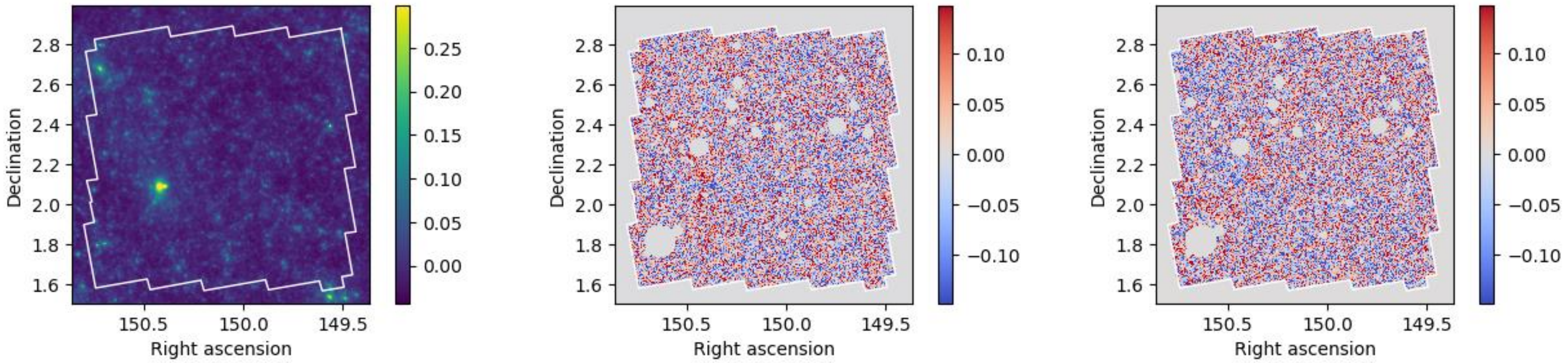
[1] https://astro.uni-bonn.de/en/m/schrabba/research

Context and objectives

Noisy shear maps (noise variance taken from the COSMOS shape catalog[1])

Simulated convergence map

Corresponding shear map with noise (real and imaginary parts)



Objective: given $\boldsymbol{\gamma}$, estimate $\widehat{\boldsymbol{\kappa}}^-$ and $\widehat{\boldsymbol{\kappa}}^+$ such that

Two sources of randomness

$$\mathbb{E}[L(\boldsymbol{\kappa}, \widehat{\boldsymbol{\kappa}}^-, \widehat{\boldsymbol{\kappa}}^+)] \leq \alpha.$$

Depends on $\boldsymbol{\gamma} = \boldsymbol{A\kappa} + \boldsymbol{n}$

- Over which uncertainties the expected value is calculated?

[1] https://astro.uni-bonn.de/en/m/schrabba/research

Proposed approach

1. Compute a point estimate $\hat{\boldsymbol{\kappa}}$ and a residual $\hat{\boldsymbol{r}}$ using two families of mass mapping methods:
   a. **Model-driven** methods: Kaiser-Squires inversion,[1] proximal Wiener filtering,[2] MCALens;[3]
   b. **Data-driven** (deep-learning-based) methods: DeepMass,[4] DLPosterior,[5] other method?
2. Set initial bounds:

$$\hat{\boldsymbol{\kappa}}^- := \hat{\boldsymbol{\kappa}} - \hat{\boldsymbol{r}} \quad \text{and} \quad \hat{\boldsymbol{\kappa}}^+ := \hat{\boldsymbol{\kappa}} + \hat{\boldsymbol{r}}$$

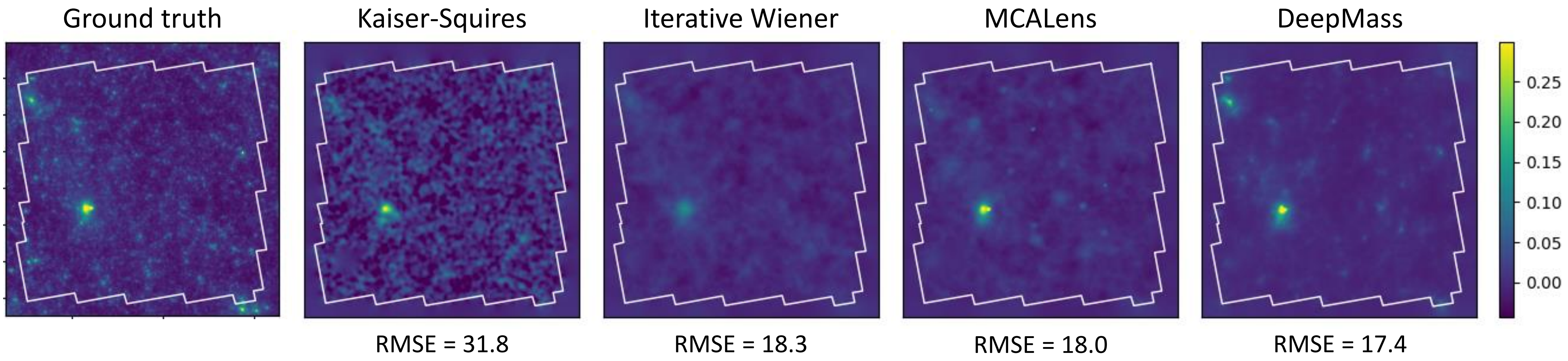
3. Post-processing: adjust residual $\hat{\boldsymbol{r}}$ using a **calibration set**.

→ Distribution-free UQ, does not assume any prior distribution on $\boldsymbol{\kappa}$.

→ Works for any blackbox prediction method, including deep learning.

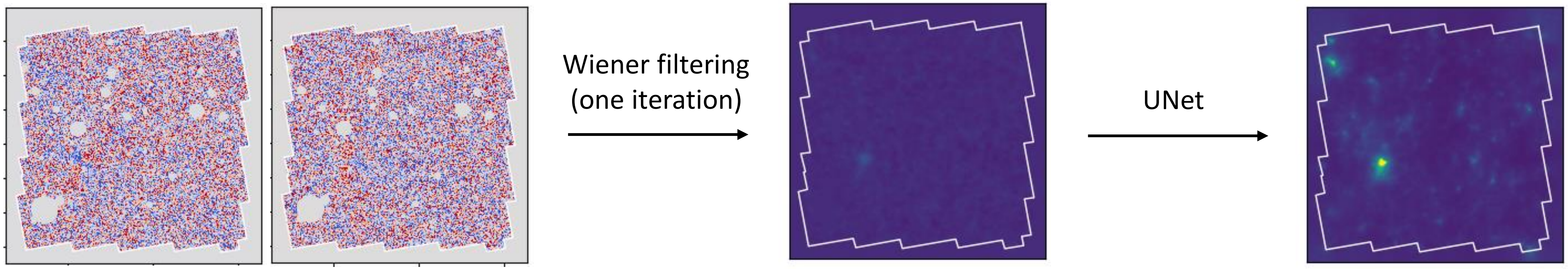
[1] Kaiser, N. & Squires, G. Astrophysical Journal 404, 441–450 (1993)
[2] Bobin, J., Starck, J.-L., Sureau, F. & Fadili, J. Advances in Astronomy 2012, e703217 (2012)
[3] Starck, J.-L., Themelis, K. E., Jeffrey, N., Peel, A. & Lanusse, F. A&A 649, A99 (2021)
[4] Jeffrey, N., Lanusse, F., Lahav, O. & Starck, J.-L. MNRAS 492, 5023–5029 (2020)
[5] Remy, B. et al. A&A 672, A51 (2023)

Reconstruction accuracy



Deep-learning-based methods

DeepMass



- Minimizing the MSE $||F_{\mathbf{\Theta}}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}||_2^2$ evaluated on the training set → DeepMass approximates the **posterior mean**:

$$F_{\widehat{\mathbf{\Theta}}}(\boldsymbol{\gamma}) \approx \widehat{\boldsymbol{\kappa}} := \iint \boldsymbol{\kappa}' \, p(\boldsymbol{\kappa}' \,|\, \boldsymbol{\gamma}) \, d\boldsymbol{\kappa}'.$$

- Remark about DLPosterior: MCMC sampling, prior learned from data → $\widehat{\boldsymbol{\kappa}}$ can be approximated by **averaging over samples**.

Deep-learning-based methods

Strengths and weaknesses

	Fast rec. + UQ	Trained once	Acc.	Comments
DeepMass	✓	✗	✓	Point estimate + UQ
DLPosterior	✗	✓	✓	Posterior sampling

- **Objective:** implement a DL mass mapping method, satisfying:
  - Fast inference → we need a point estimate instead of sampling from the full posterior.
  - Does not need re-training for each new noise covariance matrix or mask.
- **Proposed solution:** iterative algorithm with plug-and-play (PnP).

PnP forward-backward algorithm

- Objective: find the MAP estimate $\widehat{\boldsymbol{\kappa}}$ satisfying:

$$\hat{\boldsymbol{\kappa}} \in \arg\min_{\boldsymbol{\kappa}'} \left\{ \frac{1}{2} \| \boldsymbol{\gamma} - \mathbf{A}\boldsymbol{\kappa}' \|^2_{\boldsymbol{\Sigma}_n^{-1}} - \log p(\boldsymbol{\kappa}') \right\}$$

$$\hat{\boldsymbol{\kappa}} \in \arg\min_{\boldsymbol{\kappa}'} \left\{ f_1(\boldsymbol{\kappa}') + f_2(\boldsymbol{\kappa}') \right\}$$

- Iterative forward-backward algorithm:

$$\boldsymbol{\kappa}_{k+1} = \mathrm{prox}_{\mu f_2}\Big(\boldsymbol{\kappa}_k - \nabla f_1(\boldsymbol{\kappa}_k) \Big)$$

- PnP: replace the proximal operator by a deep denoiser trained on a dataset of simulated convergence maps, corrupted by a white noise of variance μ.

PnP forward-backward algorithm

- Objective: find the MAP estimate $\hat{\boldsymbol{\kappa}}$ satisfying:

$$\hat{\boldsymbol{\kappa}} \in \arg\min_{\boldsymbol{\kappa}'} \left\{ \frac{1}{2} \|\boldsymbol{\gamma} - \mathbf{A}\boldsymbol{\kappa}'\|^2_{\boldsymbol{\Sigma}_n^{-1}} - \log p(\boldsymbol{\kappa}') \right\}$$

$$\hat{\boldsymbol{\kappa}} \in \arg\min_{\boldsymbol{\kappa}'} \left\{ f_1(\boldsymbol{\kappa}') + f_2(\boldsymbol{\kappa}') \right\}$$

- Iterative forward-backward algorithm:

$$\boldsymbol{\kappa}_{k+1} = \operatorname{prox}_{\mu f_2}\left(\boldsymbol{\kappa}_k - \nabla f_1(\boldsymbol{\kappa}_k) \right)$$

Acts like a denoiser for images corrupted by a white noise of variance μ

- PnP: replace the proximal operator by a deep denoiser trained on a dataset of simulated convergence maps, corrupted by a white noise of variance μ.

Uncertainty estimation before calibration

- How to get a first estimation of the residual $\hat{\boldsymbol{r}}$?
- **Model-driven methods**: propagate noise realizations through the pipeline.
- **DLPosterior:** uncertainty embedded in posterior sampling.
- **DeepMass:** possible to use moment networks.[1] Idea: minimizing the MSE $\left\| G_{\boldsymbol{\Omega}}(\boldsymbol{\gamma}) - \left(\boldsymbol{\kappa} - F_{\widehat{\boldsymbol{\Theta}}}(\boldsymbol{\gamma})\right)^2 \right\|_2^2$ evaluated on the training set.
- **New method:** use a similar PnP FB algorithm with adapted denoiser?

[1] Jeffrey, N. & Wandelt, B. D. Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020)

Uncertainty estimation before calibration

- How to get a first estimation of the residual $\hat{\boldsymbol{r}}$?
- **Model-driven methods**: propagate noise realizations through the pipeline.
- **DLPosterior:** uncertainty embedded in posterior sampling.
- **DeepMass:** possible to use moment networks.[1] Idea: minimizing the MSE $\left\| G_{\Omega}(\boldsymbol{\gamma}) - \left(\boldsymbol{\kappa} - F_{\widehat{\boldsymbol{\Theta}}}(\boldsymbol{\gamma})\right)^2 \right\|_2^2$ evaluated on the training set.
- **New method:** use a similar PnP FB algorithm with adapted denoiser?

UNet to be trained

[1] Jeffrey, N. & Wandelt, B. D. Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020)

Uncertainty estimation before calibration

- How to get a first estimation of the residual $\hat{\boldsymbol{r}}$?
- **Model-driven methods**: propagate noise realizations through the pipeline.
- **DLPosterior:** uncertainty embedded in posterior sampling.
- **DeepMass:** possible to use moment networks.[1] Idea: minimizing the MSE $\left\| G_{\Omega}(\boldsymbol{\gamma}) - \left(\boldsymbol{\kappa} - F_{\widehat{\Theta}}(\boldsymbol{\gamma})\right)^2 \right\|_2^2$ evaluated on the training set.
- **New method:** use a similar PnP FB algorithm with adapted denoiser?

Already trained UNet
(point estimate)

[1] Jeffrey, N. & Wandelt, B. D. Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020)

Conclusion

- New deep-learning-based mass mapping method, fast at inference and generalizable to any noise covariance matrix / any mask.
- Includes initial uncertainty estimation.
- To be implemented; new results coming soon.
- Distribution-free UQ for mass mapping: provides coverage guarantees with a limited number of calibration examples.
- Works for any mass mapping method, including deep learning-based approaches.
- Next steps:
  - train on several cosmologies → CosmoSLICS;
  - extend results to the sphere;
  - UQ: focus on high-density regions.