Correcting galaxy systematics using self-organizing maps

Gaël Alguero

ML coffee, March 29, 2024

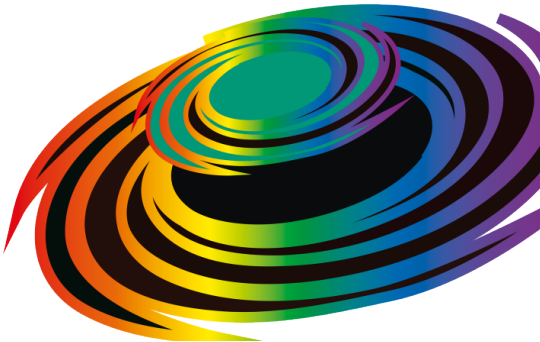
LPSC Grenoble

For the VMPZ-ID team: Jérôme Odier and Juan F. Macías-Pérez

Introduction systematic decontamination

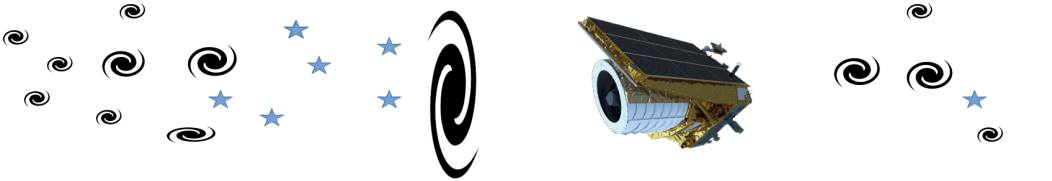
Decontaminating DES systematics

Scientific validation



Systematics on galaxy density

- Spatial variations of survey and sky properties as well as of instrumental performance will imprint on the measured galaxy density



Real galaxy distribution | Astrophysical contamination and extinction | Observational conditions and instruments performances | Distorted galaxy distribution

$\Rightarrow$ Can affect galaxy clustering and the detection of clusters of galaxies

$\Rightarrow$ Need to correct this distortion

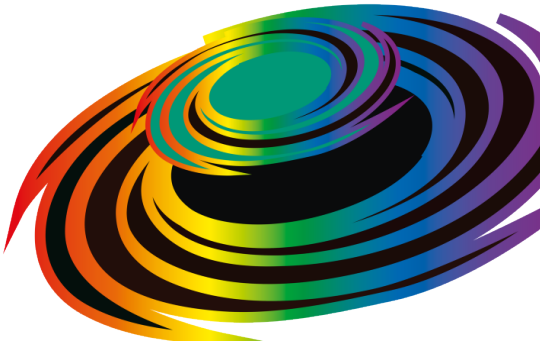
Decontamination methods

- Linear multiplicative methods in DES
  - Iterative Sytematics Decontamination (ISD): Elvin-Poole et al. [arXiv:1708.01536], Rodriguez-Monroy et al. [arXiv:2105.13540]
  - Elastic Net Regularisation (ENet): Weaverdyck, Huterer [arXiv:2007.14499]
- Linear multiplicative and additive models
  - Theoretical framework: Weaverdyck, Huterer [arXiv:2007.14499]
  - Other non-linear methods
- Non-linear method in KIDS
  - Organised randoms from Self-Organizing Maps (SOM): Johnston et al. [arXiv:2012.08467]
- Other ideas?

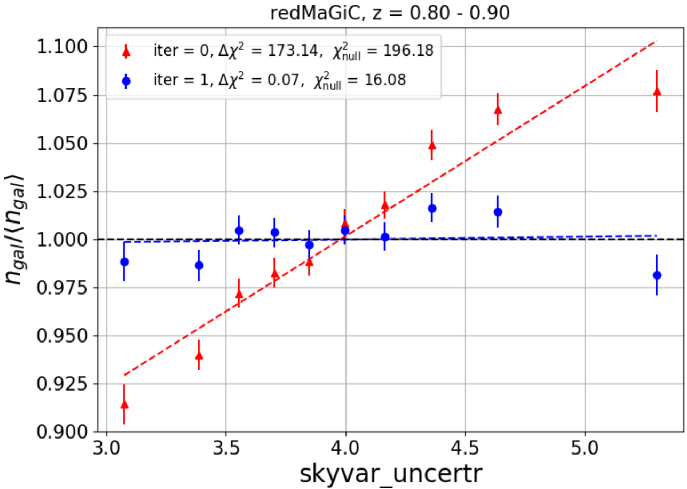
Introduction systematic decontamination

Decontaminating DES systematics

Scientific validation



Iterative systematic decontamination

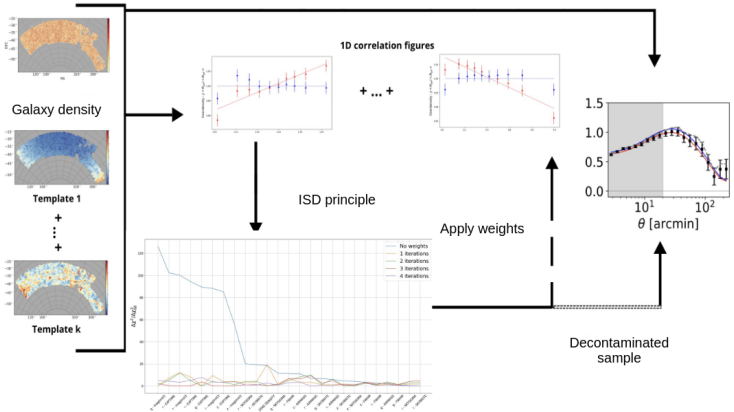


Iterative systematic decontamination

- Assuming multiplicative bias
- Compute linear weights for correction

$$n^{\mathrm{true}}(\theta,\phi) = \frac{n^{\mathrm{obs}}(\theta,\phi)}{ax(\theta,\phi)+b}$$

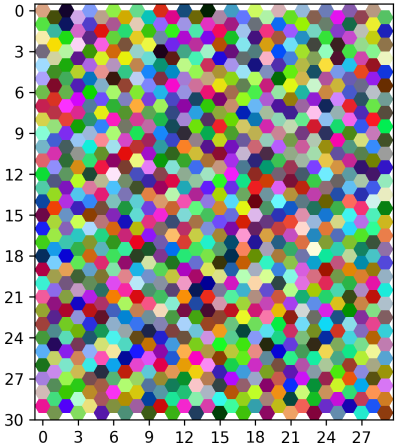
- Caveats
  - Needs fine-tuning
  - No correlations between systematics
  - Only linear effects



Example dataset

	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●
R	0	255	255	0	0	255	0	255	128	0	128	0	255	255	255	240
G	0	255	0	255	0	255	255	0	128	128	0	0	69	165	215	230
B	0	255	0	0	255	0	255	255	0	0	128	128	0	0	0	140

SOM principle



Two implementations of the SOM

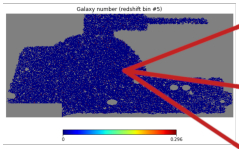
On-line SOM

- Input vectors are presented sequentially
- The SOM latent space is updated for each new vector
- Can't be parallelized

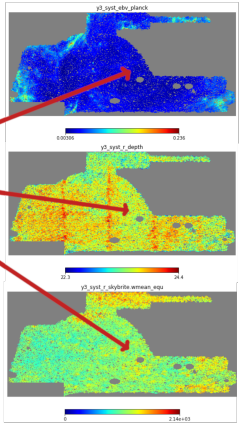
Batch SOM

- Input vectors are all presented simultaneously
- The SOM latent space is updated is updated once per epoch
- Highly parallelizable
- Ideal for a large amount of data, but do we lose information?

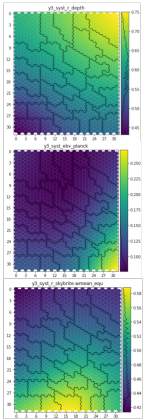
1. Training with galaxies



Galaxy catalog



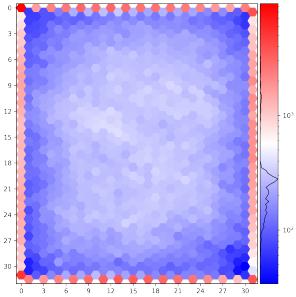
Systematic vector



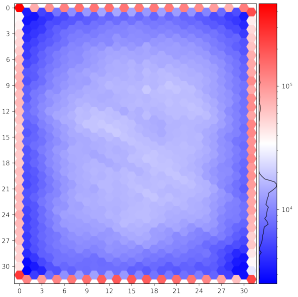
Reduced systematics

Reducing the observed sky into a small area while keeping (most of) the contamination information

2. Activation maps

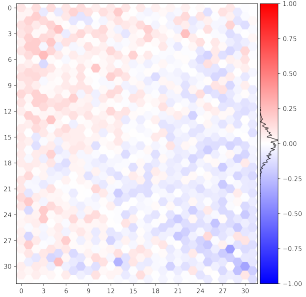


Galaxies activation map



Footprint activation map

⇒

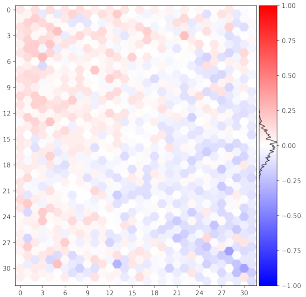


Reduced galaxy density

3. Decontaminated galaxy density

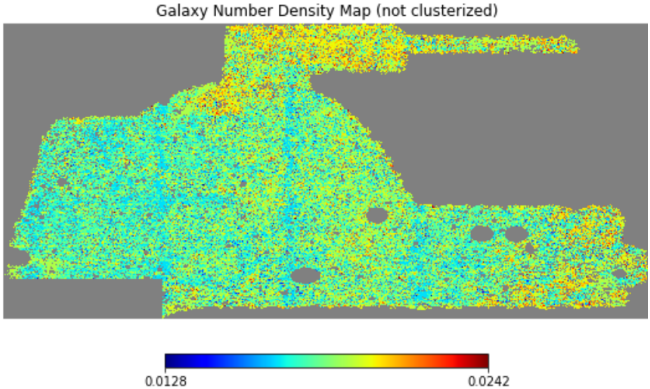
From reduced systematics

$\Rightarrow$ we reconstruct a galaxy density map containing the imprint of contamination effects



Reduced galaxy density

$\Rightarrow$

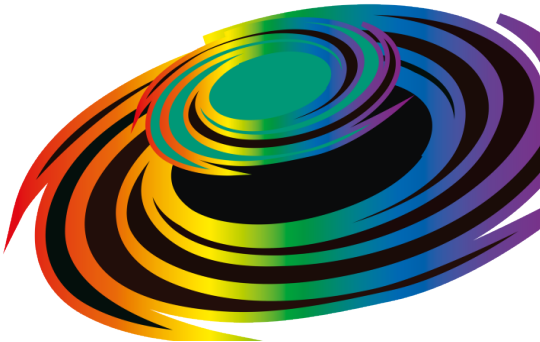


Projected back into physical space

Introduction systematic decontamination

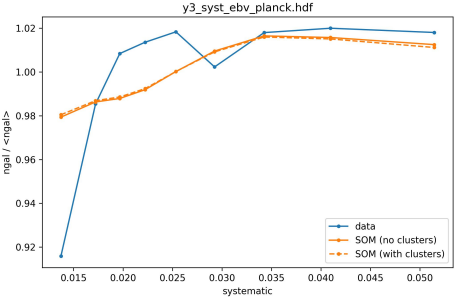
Decontaminating DES systematics

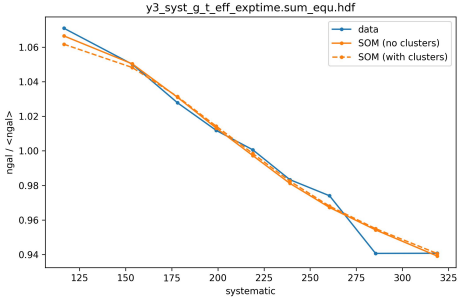
Scientific validation



Data-to-systematic correlations

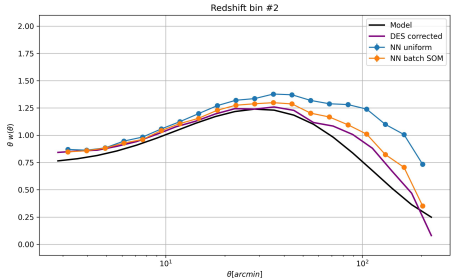
The algorithm is able to fit linear and non-linear effects

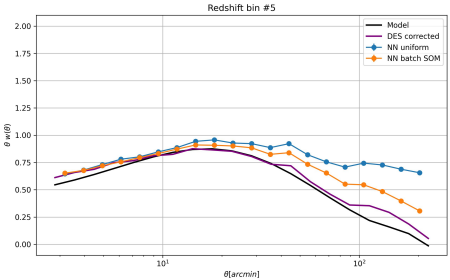




Two-points angular correlation functions

- Performed SOM decontamination with 35 DES systematics (4 photometric bands)
  - Decontamination is not perfect: DES selection cuts ?





Thanks Nicolas Tessore for the help with GLASS* and TreeCorr†!

*GLASS: Generator for Large Scale Structure, Tessore N. et al. [arXiv:2007.14499]

†TreeCorr, Jarvis M., Bernstein G. and Jain B . [ascl]

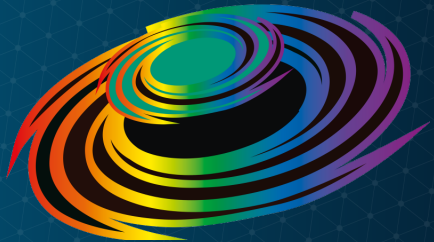
Conclusions

- Self-organizing maps as a promising alternative for systematic decontamination
  - On-line and batch implementations
  - The real challenges :
    - A huge amount of data, about 15 000 deg^2, billions of galaxies
    - Handling missing data
- Scientific validation ongoing on DES data
  - Two-points angular correlations
    - Need to tune the SOM parameters?
    - Apply the DES selections?
    - Project systematics into a principal component basis?
  - Other algorithms to implement
    - Iterative systematic decontamination
    - ElasticNet regularization (ongoing)
    - Principal component analysis (ongoing)
- First tests on Euclid PV phase data

Our decontamination tool is publicly available at lpsc-euclid.github.io/decontamination

Dedicated Slack channel: #sgs-ou-le3-id-ymnz

Thanks for your attention!



SOM equations

On-line:

$$\vec{w}_k(t+1) = \vec{w}_k(t) + \alpha(e) h_{ck}(e) \left[\vec{x}(t) - \vec{w}_k(t)\right]$$

Batch:

$$\vec{w}_k(t+n) = \frac{\sum_{t'=t}^{t+n} h_{ck}(t') \vec{x}(t')}{\sum_{t'=t}^{t+n} h_{ck}(t')}$$

Two-points angular correlation functions