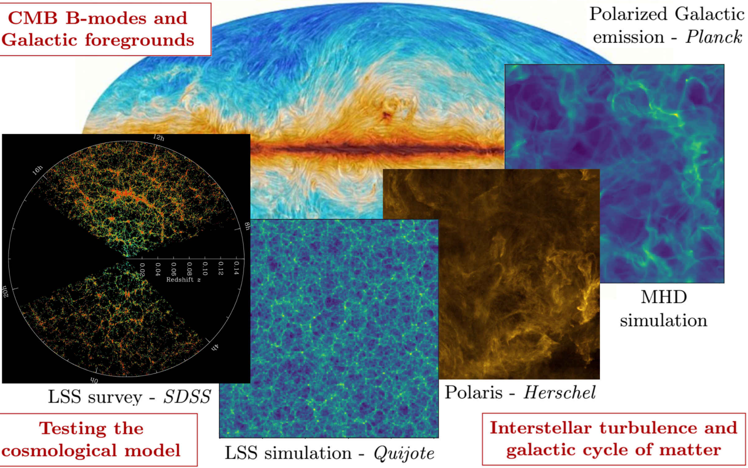
Scattering Transforms in astrophysics, application to component separation

Erwan Allys - ENS, Paris
(Physics laboratory and Center for data science)

LPNHE Seminar
Paris, May 15th 2023

Outline

CMB B-modes and Galactic foregrounds



Testing the cosmological model

Interstellar turbulence and galactic cycle of matter

Common difficulty : non-linearity $\Rightarrow$ non-Gaussian structures
$\rightarrow$ Important lever arm for a lot of astrophysical objectives

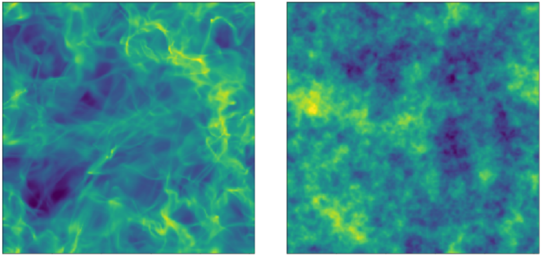
Different scientific objectives

- **Different methodological objectives**
  (with increasing subjective difficulty...)
  - Estimate physical parameters
  - Model an astrophysical process
  - Separate different components
  - Constraint a physical model

$\rightarrow$ **Which statistics for non-Gaussian astrophysics?**
$\rightarrow$ **How to use non-Gaussian information for these tasks?**

Beyond Power Spectrum statistics

- **A generic tool: the Power Spectrum**
  - Square amplitude of Fourier modes
  - Energy/Power in each Fourier mode
  - Most usual statistical tool in astrophysics



Champs de même spectre de puissance

→ Does not characterize interaction between scales
→ Need beyond Power Spectrum statistics for NL fields

Specific challenges of astrophysics

- **A limited amount of intricate observations**
  - A unique static multi-frequency sky
  - Mixture of non-stationary components
    $\rightarrow$ isolated processes are very rare

- **Almost no training ground**
  - Often no complete physical/numerical models
  - Simulations are very expensive
    $\rightarrow$ no or very limited dataset

Specific challenges of astrophysics

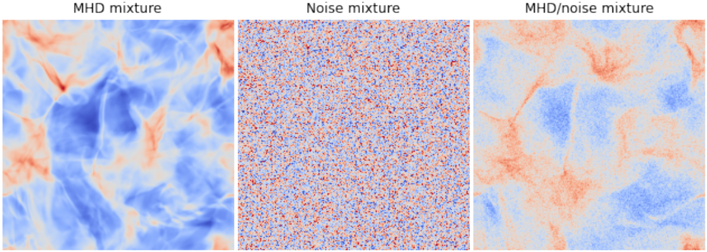
- **A limited amount of intricate observations**
  - A unique static multi-frequency sky
  - Mixture of non-stationary components
    $\rightarrow$ isolated processes are very rare

- **Almost no training ground**
  - Often no complete physical/numerical models
  - Simulations are very expensive
    $\rightarrow$ no or very limited dataset

$\rightarrow$ Work mainly from obs. data and physical knowledge
$\rightarrow$ Need to work with low-variance statistics!

Non-Gaussian foregrounds models

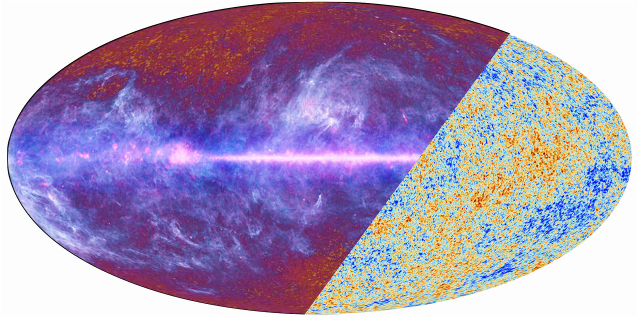
- **Non-Gaussianity is not our enemy!**



  - Important lever arm for components separation
  - Even from a small amount of data

→ Challenge of using non-Gaussian information
→ Should be possible to work with (very) small dataset

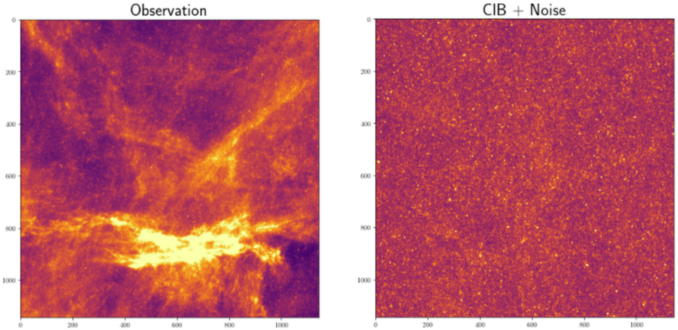
Example I: CMB/Galactic foregrounds



- **Cosmic Microwave Background (CMB) polarized B-modes**
  - Gaussian process, know spatial/spectral distribution
  - Signature of primordial Universe (inflation epoch)
  - Beyond much brighter ($\simeq 10^{2-3}$) Galactic foregrounds, no model

$\rightarrow$ **Detection and measurement of CMB B-modes?**

Example II: CIB/Galactic dust emission



- **Galactic dust emission and Cosmic Infrared Background (CIB)**
  - Thermal dust emission in the interstellar medium
  - Same emission from Milky Way and other galaxies
  - Cosmic background dominates a smaller scales

→ **Characterization of Galactic dust on those scales?**

Outline

Which statistics for non-Gaussian fields?

- **High-order correlation functions**
  - e.g., third order correlation function and bispectrum

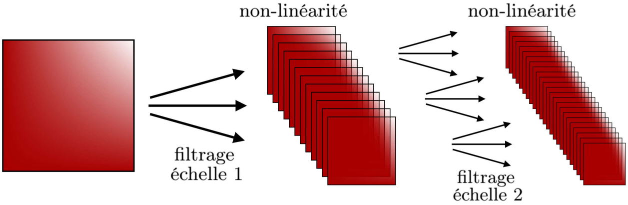
$$\langle s(\vec{x})s(\vec{x}+\vec{\tau}_1)s(\vec{x}+\vec{\tau}_2)\rangle_{\vec{x}} \longrightarrow B(\vec{k}_1,\vec{k}_2,\vec{k}_3)$$

  - Link with dynamics in perturbative regime
  - Theoretically contains all information...
    $\rightarrow$ but numerous terms with high variance

$\rightarrow$ Explicit mathematical form and interpretability
$\rightarrow$ Not suitable for highly NG fields with limited data...

Which statistics for non-Gaussian fields?

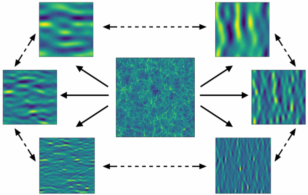
- **Machine learning and neural network**
  - e.g., convolutional network structure with learned weights
  - Extremely efficient to deal with complex images...
    → but need specific tasks and good quality data



→ Need of data, uncertain transfer, weak interpretability...
→ Can we take an intermediate path?

Scattering transform (ST) statistics

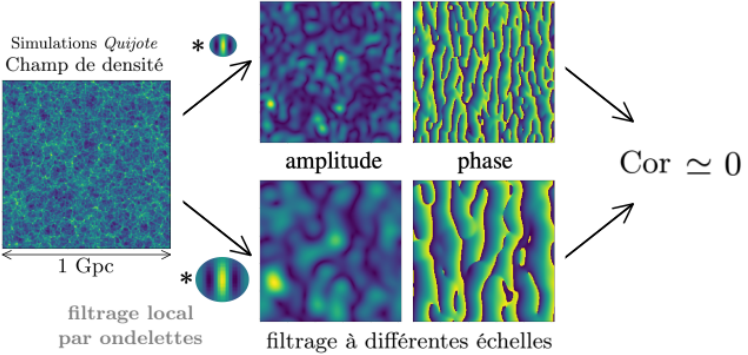
- **Scattering transform statistics** (Mallat+, 2010+)
  - Initially developed in data science
  - Inspired from neural networks
    $\rightarrow$ efficient characterization and reduced variance
  - Do not need any training stage
    $\rightarrow$ explicit mathematical form and interpretability



$\rightarrow$ Wavelet filters separating the different scales
$\rightarrow$ Coupling between scales with non-linearities

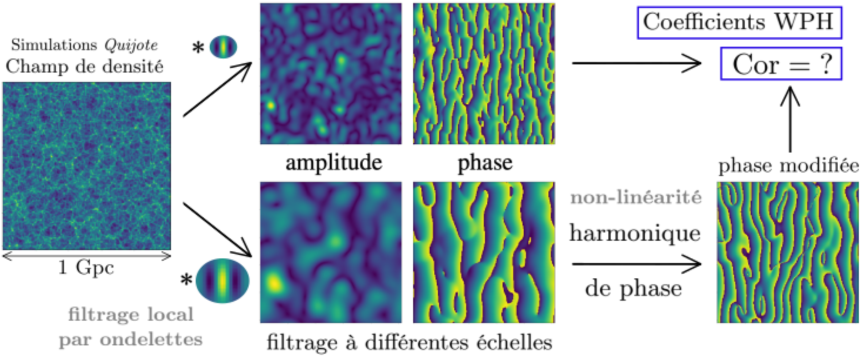
Scattering Transform (ST) statistics

- **Wavelet Phase Harmonics and phase alignment** *(EA+, 20)*



Scattering Transform (ST) statistics

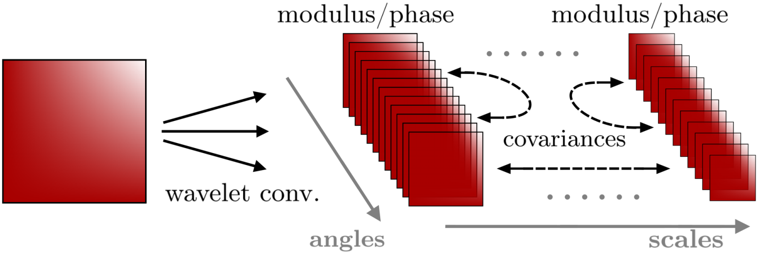
- **Wavelet Phase Harmonics and phase alignment** *(EA+, 20)*



→ **1 coeff / pair of scales / type of interaction**
→ **Can be extended to cross-statistics between maps**

Scattering Transform (ST) statistics

- **Final network structure for scattering statistics**



→ 2 initial convolutions + 1 translation ⇒ triplets of scales
→ 1 coefficient per covariance and per type of coupling

Scattering Transform (ST) statistics

- **A family of statistics**
  - Different generations of statistics
    → Wavelet Scattering Transforms (WST) ***(EA+, 19)***
    → Wavelet Phase Harmonics (WPH) ***(EA+, 20)***
    → Scattering covariances/spectra ***(Cheng+, in prep.)***
  - All share the same framework

Scattering Transform (ST) statistics

- **A family of statistics**
  - Different generations of statistics
    $\rightarrow$ Wavelet Scattering Transforms (WST) ***(EA+, 19)***
    $\rightarrow$ Wavelet Phase Harmonics (WPH) ***(EA+, 20)***
    $\rightarrow$ Scattering covariances/spectra ***(Cheng+, in prep.)***
  - All share the same framework

- **Characterization and parameter inference**
  - Interstellar medium ***(EA+ 19, Regaldo+20, Saydjari+, 20, Lei+, 22)***
  - Weak lensing ***(Cheng+, 20, 21)***
  - Large scale structures ***(EA+, 20, Eickenberg+, 22, Valogiannis+, 22a, 22b)***
  - 21cm epoch of reionization ***(Greig+, 22)***
  - ...

$\rightarrow$ Very informative (sometimes on par with CNN!)
$\rightarrow$ Wide range of applicability (generic, training-less)

Generative models from Scattering transforms

- **Generative model from ST statistics** *(Bruna, Mallat, 19)*
  - Generative model from the ST statistics $\Phi(s)$ of a map s
  - Maximum entropy microcanonical model
  - Generate new maps $\tilde{s}$ with same ST statistics
  - Non-gaussian properties quantitatively reproduced

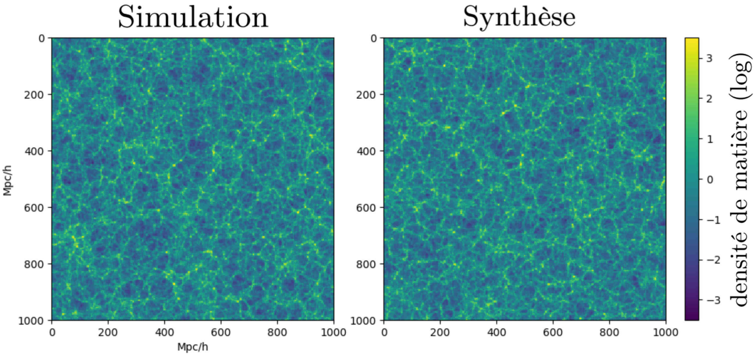
Generative models from Scattering transforms

- **Generative model from ST statistics** *(Bruna, Mallat, 19)*
  - Generative model from the ST statistics $\Phi(s)$ of a map s
  - Maximum entropy microcanonical model
  - Generate new maps $\tilde{s}$ with same ST statistics
  - Non-gaussian properties quantitatively reproduced

- **Practical implementation**
  - Constraints $\Phi(s)$ from a (set of) data s
  - Sampled with a gradient-descent algorithm
    $\rightarrow$ from a whrite noise realization
    $\rightarrow$ optimizing $\tilde{s}$ such that $\Phi(\tilde{s}) \simeq \Phi(s)$

Generative models from Scattering transforms

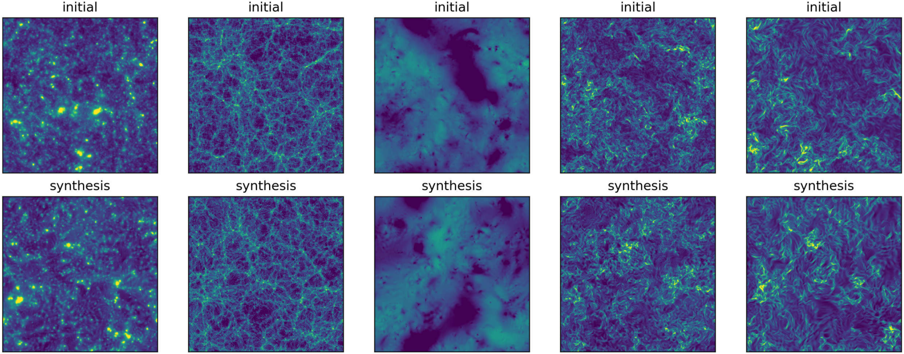
- **Quantitative validation of syntheses** *(EA+, 20)*
  - Large scale structures density field, Wavelets Phase Harmonics



→ **Usual (NG) statistics very well reproduced (up to 1-10 %)**

Generative models from Scattering transforms

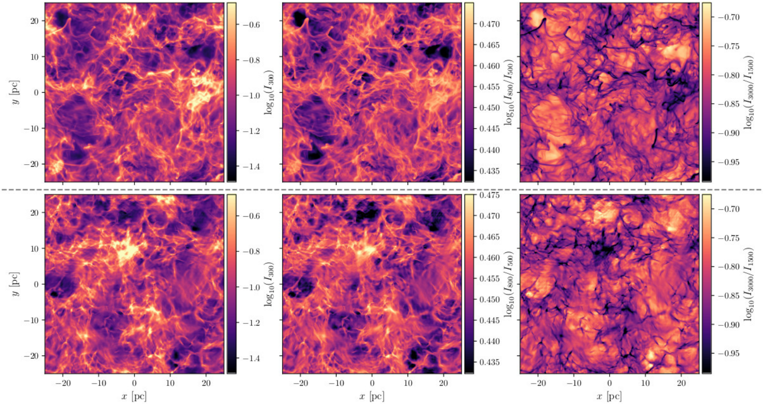
- **Syntheses from a single image** *(Cheng+, in prep.)*
  - Scattering spectra + physical dimensionality reduction



→ **Realistic NG models from a few hundreds coefficients!**

Generative models from Scattering transforms

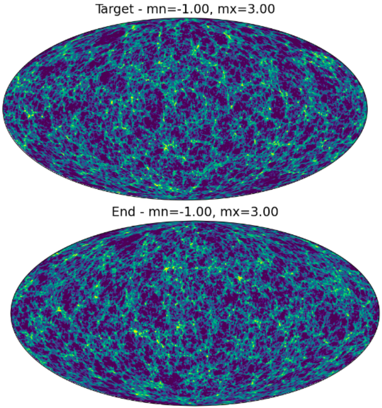
- **Multi-frequency dust syntheses** *(Regaldo+, 22)*
  - Cross-WPH, simulated dust intensity, 300/500/800/1500/3000GHz



→ **Dispersion of MBB parameters reproduced!**

Generative models from Scattering transforms

- **Scattering transform on the sphere** *(Delouis, Mousset+, in prep.)*



→ **Public tools should be available soon**

Outline

Modelisation and separation of components

- **Framework of the problem**
  - We observe a mixture $d = s + c$
    $\rightarrow$ d data, s signal of interest, c contamination
  - Use prior knowledge to recover properties of s
  - Typical for astrophysical observations

Modelisation and separation of components

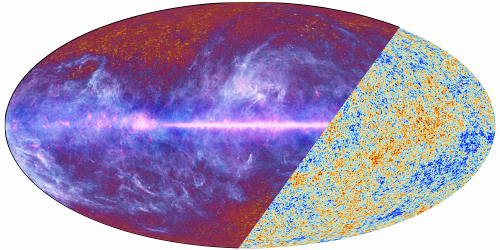
- **Framework of the problem**
  - We observe a mixture $d = s + c$
    $\rightarrow$ d data, s signal of interest, c contamination
  - Use prior knowledge to recover properties of s
  - Typical for astrophysical observations

- **If we can model, can we separate?**
  - Assume we have a model for s and c
    $\rightarrow$ we can generate a data set of $d_i = s_i + c_i$
    $\rightarrow$ we can train a neural network to recover s from d

$\rightarrow$ Does it work from a single image?

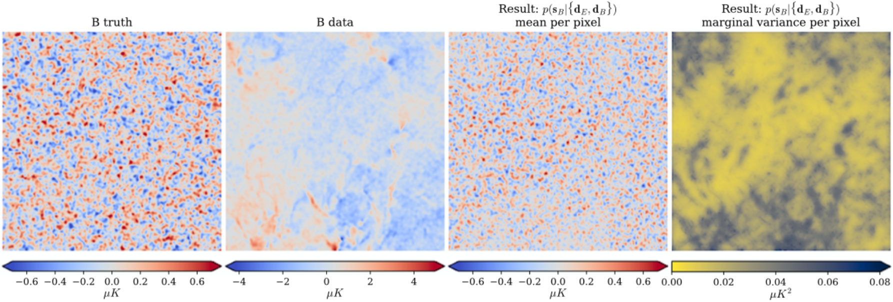
Modelisation and separation of components

- **CMB B-mode/dust foregrounds at 143GHz** *(Jeffrey+, 22)*
  - Application with simulated foregrounds
    → Assume we have a single foreground map
    → Construct a model from this image
  - Gaussian model with prior distribution for CMB
  - Train a neural network to perform foreground removal
    → Moment network for Bayesian framework



Modelisation and separation of components

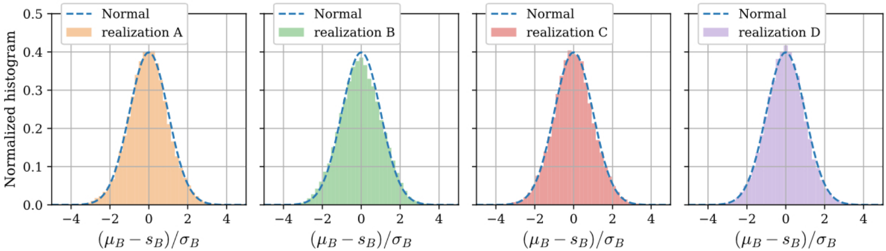
- **Validation on independent simulated foregrounds** *(Jeffrey+, 22)*



→ Marginal posterior distributions per pixel well recovered

Modelisation and separation of components

- **Validation on independent simulated foregrounds** *(Jeffrey+, 22)*



→ Successful separation in a Bayesian framework
→ Validate the foregrounds model learned from one image

Modelisation and separation of components

- **But this problem was easy...**
  - We rarely have a given "clean" realization
  - Have to deal with unknown components
    $\rightarrow$ we have at best a model for contamination

Modelisation and separation of components

- **But this problem was easy...**
  - We rarely have a given "clean" realization
  - Have to deal with unknown components
    $\rightarrow$ we have at best a model for contamination

- **Scientific objective of the components separations?**
  - Recovering s: minimizing $\mathrm{MSE}(s, \tilde{s})$
    $\rightarrow$ filtering effect at low SNR
    $\rightarrow$ statistics of $\tilde{s}$ not well constrained
  - Recovering s: minimizing $\mathrm{MSE}(\phi(s), \phi(\tilde{s}))$
    $\rightarrow$ model and statistics of s are recovered
    $\rightarrow$ most of the time scientific target

$\rightarrow$ ***Statistical* components separation**
$\rightarrow$ **Can we extend the generative framework?**

Statistical components separation

- **Maximum entropy model from available sample**
  - Estimate $\phi(\bar{s})$ from sample $\bar{s}$
  - Generate a map such that

$$\Phi(\tilde{s}) \simeq \Phi(s)$$

  - Sampled with gradient descent from white noise

Statistical components separation

- **Maximum entropy model from available sample**
  - Estimate $\phi(\bar{s})$ from sample $\bar{s}$
  - Generate a map such that

$$\Phi(\tilde{s}) \simeq \Phi(s)$$

  - Sampled with gradient descent from white noise

- **Indirect observation with know contamination**
  - $d = s_0 + c_0$, assume we have $\{c_i\}_i$
  - Generate a map such that

$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$

  - Gradient descent from d

Statistical components separation

- **Maximum entropy model from available sample**
  - Estimate $\phi(\bar{s})$ from sample $\bar{s}$
  - Generate a map such that

$$\Phi(\tilde{s}) \simeq \Phi(s)$$

  - Sampled with gradient descent from white noise

- **Indirect observation with know contamination**
  - $d = s_0 + c_0$, assume we have $\{c_i\}_i$
  - Generate a map such that

$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$

  - Gradient descent from d

→ New framework for components separation
→ Can include various statistical constraints

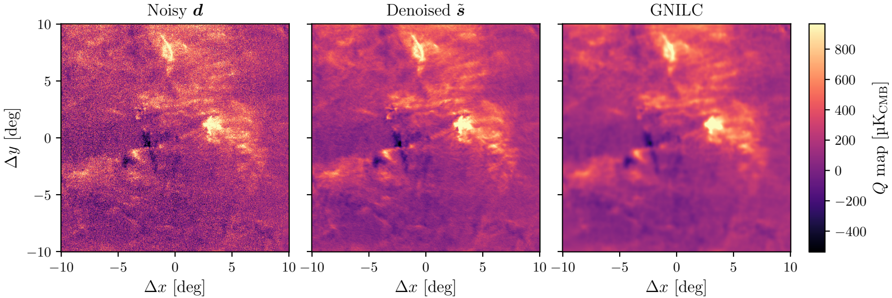
Statistical components separation

- **Application to dust polarized emission and noise (Regaldo+, 21)**
  - $d = s + c$ *Planck* polarization data at 353GHz
  - s polarized dust emission, c inhomogeneous noise
  - 300 noise realizations c_i from Planck team
  - Optimization done from d to keep largest scales

$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$

Statistical components separation

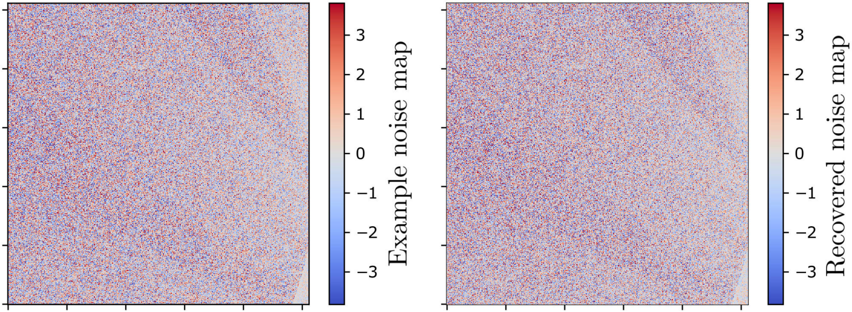
- **Application to Chameleon-Musca region** (Régaldo+ 21)



→ **Transition btw. deterministic and statistical**
→ **Conceptual validation of the method**

Statistical components separation

- **Recovered contamination** (Régaldo+ 21)

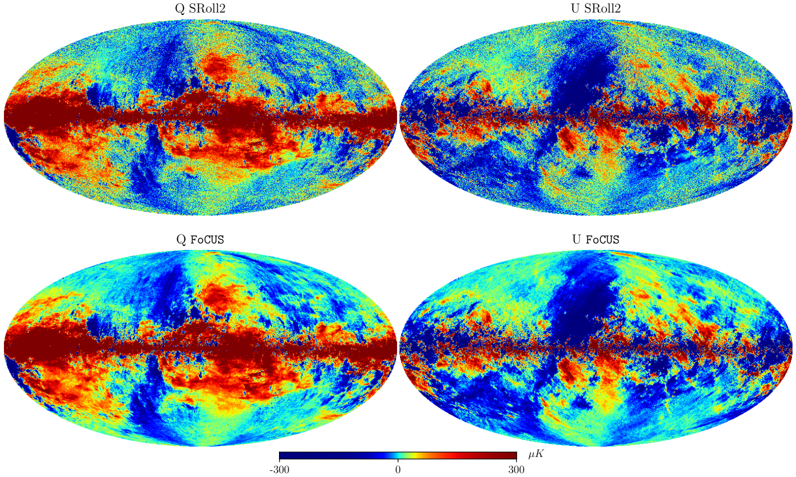


→ Statistical separation of components
→ Residual structures could also be constrained

Statistical components separation

- **Refinements of this work on the whole sky** **(Delouis+, 22)**
  - Introduce additional constraints
    $\rightarrow$ 3 constraints including cross-statistics
  - Educated normalization of each constraint
    $\rightarrow$ constraints normalized by variance over $\{c_i\}$
  - Introduce local constraints for non-stationary
    $\rightarrow$ 4 selected regions for Galactic heterogeneity

- **Full sky results at 353GHz** (Delouis+, 22)

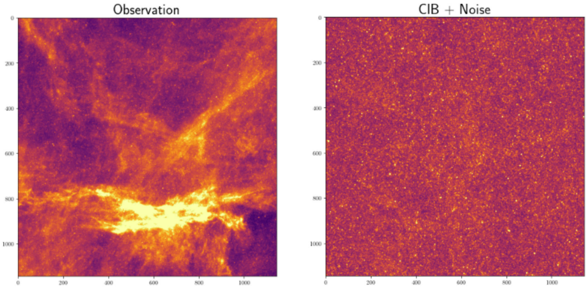


→ Deterministic up to SNR≃0.1, statistical up to SNR≃0.01
→ Efficient and versatile framework for statistical comp. separation

Separation solely from observational data?

- **But this problem was easy...**
  - Astrophysical components more difficult to model
  - Instrumental noise not always modeled
  - Can we characterize an unknown component
    $\rightarrow$ without prior model for contamination?
    $\rightarrow$ relying only on observational data?

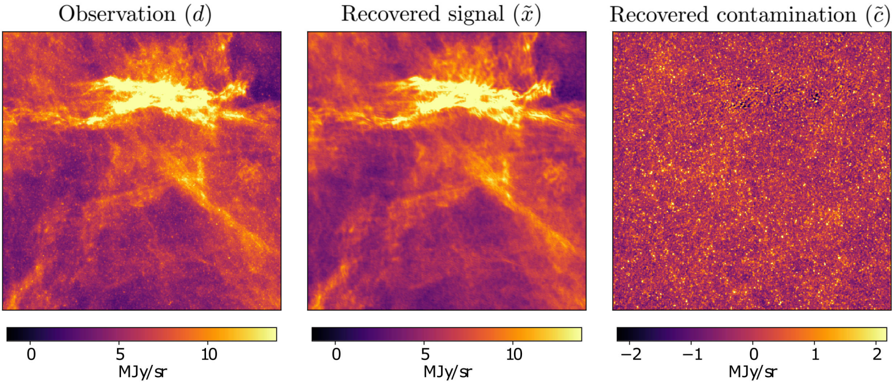
Separation solely from observational data?



- **Dust emission/Cosmic Infrared Background (Auclair+, sub.)**
  - $d = s + c_1 + c_2$, s thermal dust emission, c_1 CIB, c_2 noise
  - CIB model from separate observation (cosmological $\Rightarrow$ homogeneous)
  - Noise estimated from two observations of the same region
  - Two constraints, with $\{c_i^{\text{tot}}\}_i$ from models above

$$\left\langle \Phi(\tilde{s} + c_i^{\text{tot}}) \right\rangle_i \simeq \Phi(d), \qquad \Phi(\tilde{c}^{\text{tot}}) = \Phi(c^{\text{tot}})$$

- **Recovered components** (Auclair+, sub.)



→ **Statistical components separation solely from obs. data**
→ **Thermal dust is recovered at an unprecedented resolution**

Conclusion

- **Scattering Transforms**
  - → Efficient non-Gaussian statistics inspired from neural network
  - → Characterize interaction between scales in non-linear processes
- **New tools for (astro-)physics**
  - → Generative models and component separations
  - → Ability to work with a very limited amount of data
- **Applications to come are very exciting! :-)**

Thanks for your attention!