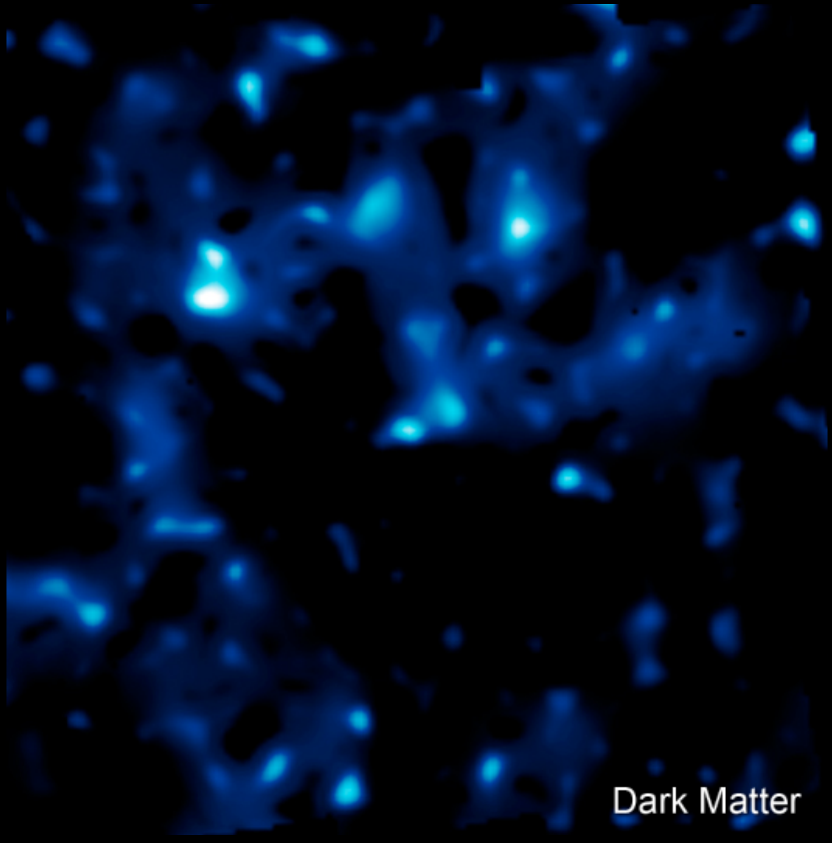
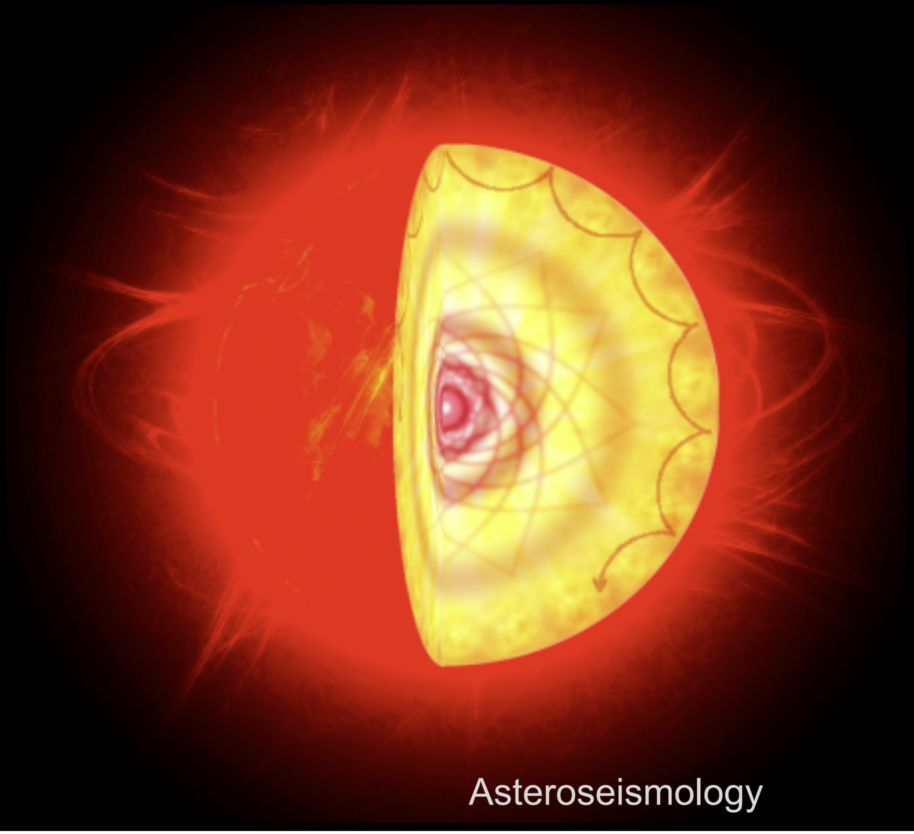
Missing Data in Astrophysics





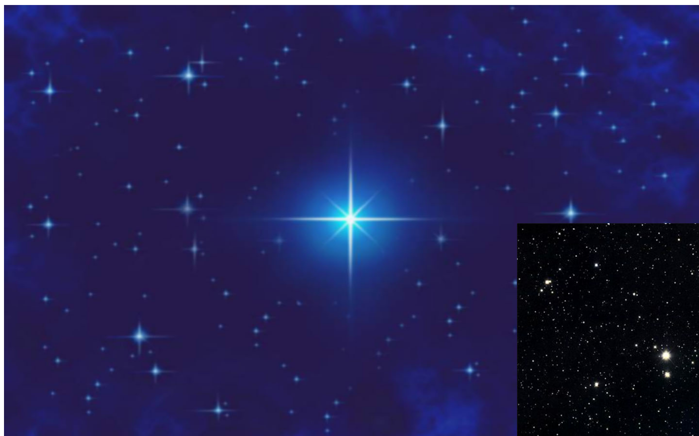
Sandrine Pires
sandrine.pires@cea.fr
Collaborators: J.L Starck, R. Garcia, S. Mathur, E. Jullo, J. Ballot, J. Fadili, R. Teyssier, A. Amara, A. Réfrégier, D. Stello

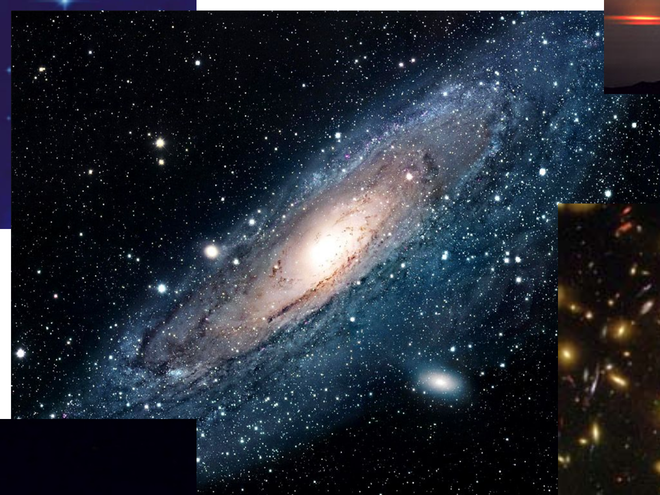
missDATA 2015, June 18-19 2015, Rennes

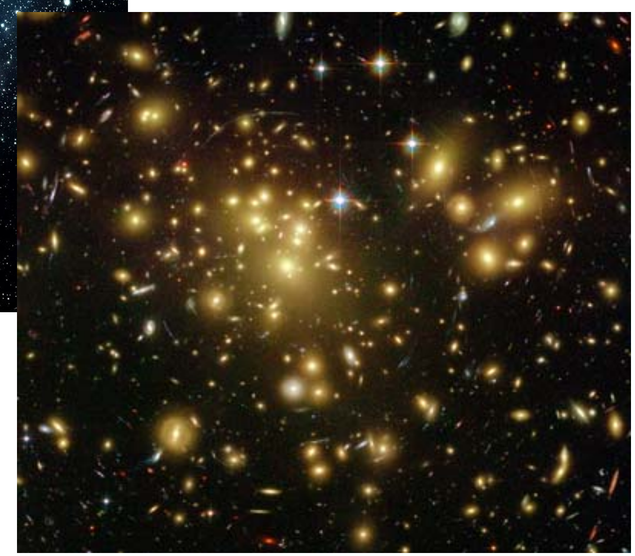
Outline

- Missing data in Astrophysics
  - Origin of missing data
  - Characteristics of missing data
- Sparse inpainting
  - Sparse and Compressible signals
  - Inpainting
- Applications
  - Application to Asteroseismology (1D signal)
  - Application to Weak Lensing (2D signal)

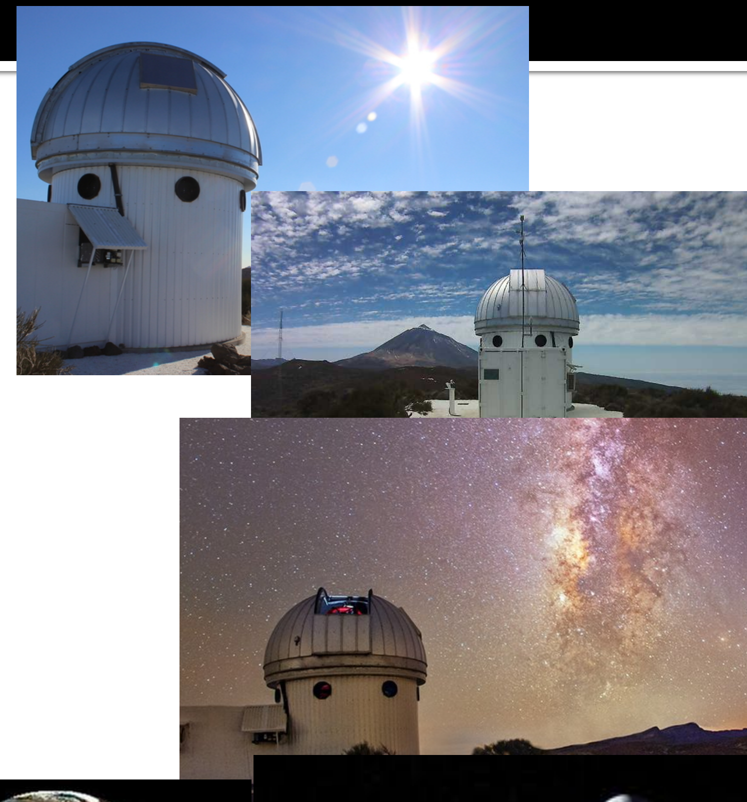
Astrophysics



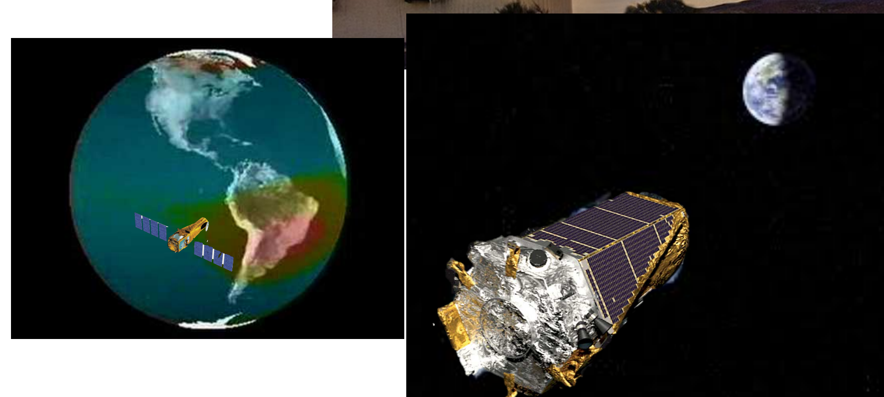




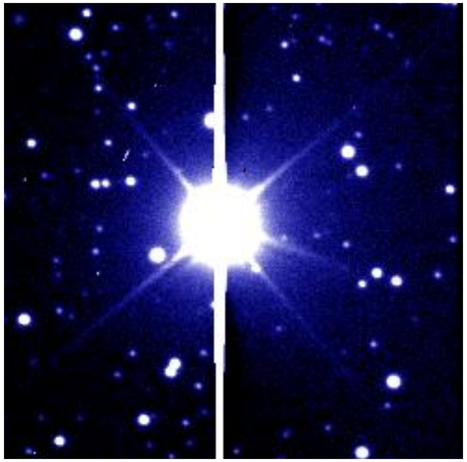
Missing Data in 1D observations



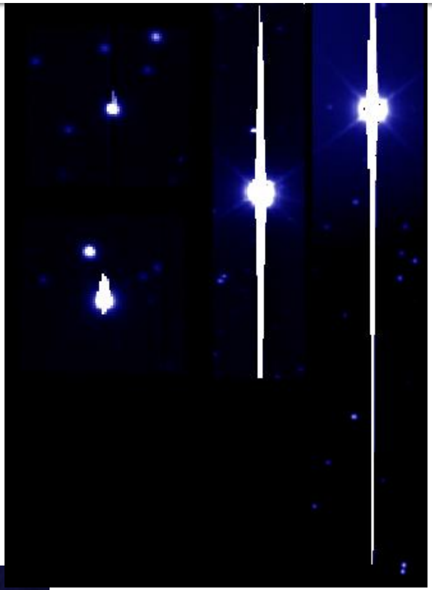
- Ground-based observations
  - Weather
  - Day/Night Cycle
  - Temporary failure of instruments
  - Other factors
- Space-based observations
  - Cosmic Rays
  - Nominal operations
  - Other factors



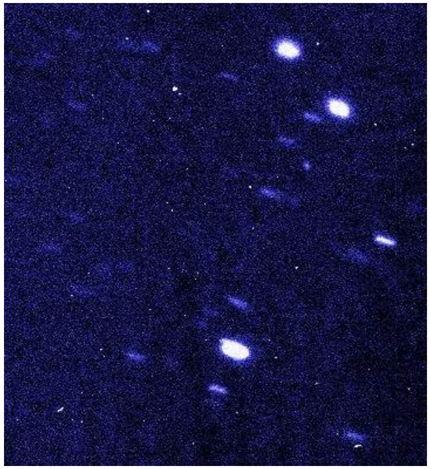
Missing Data in 2D observations



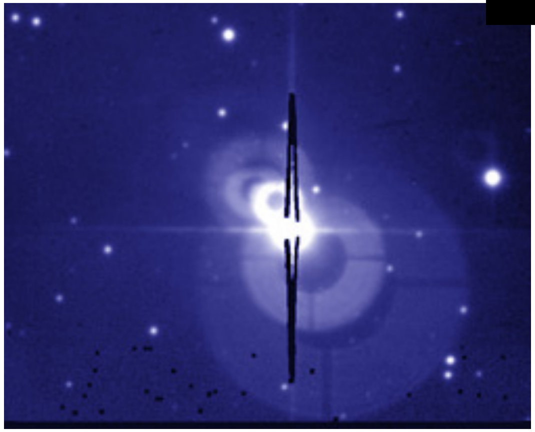
Saturation Pattern



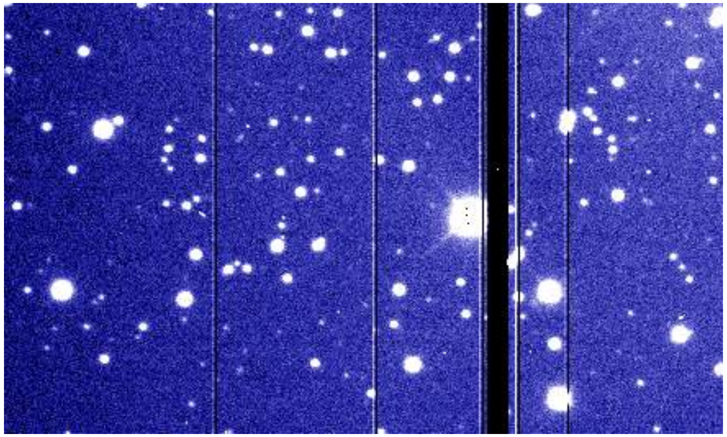
Saturation Trail



Cosmic Rays



Reflection ghosts



Bad columns

The missing data problem

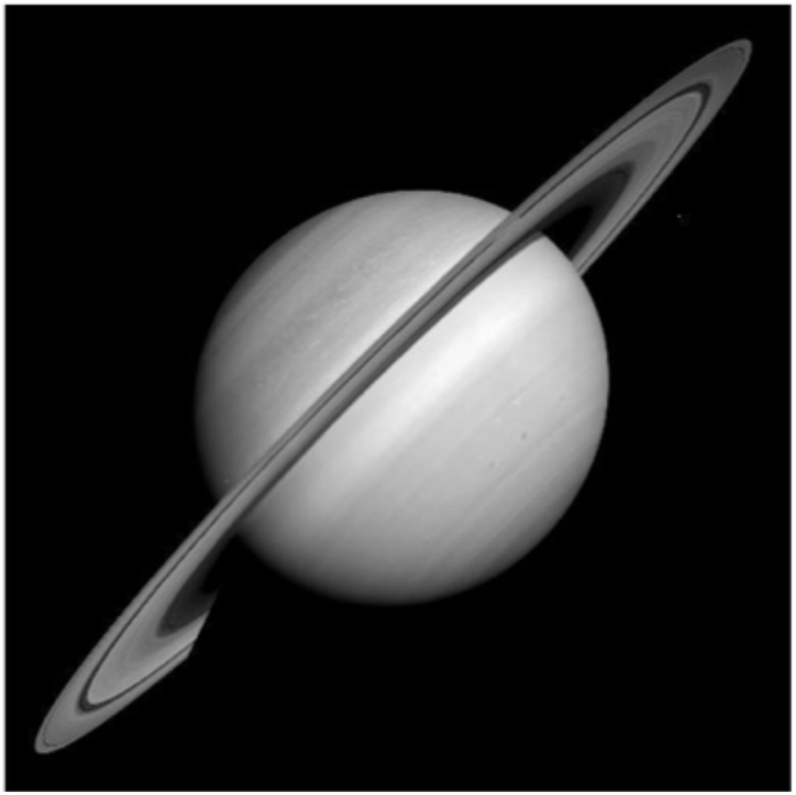
$$Y(t) = M(t)X(t)$$

- The effect of the observational window will depend on :
  - The percentage of missing data
  - The distribution of the missing data with respect to the signal
- How to correct from missing data ?
  - Correction factor
  - Mask deconvolution
  - Interpolation of the missing data

Outline

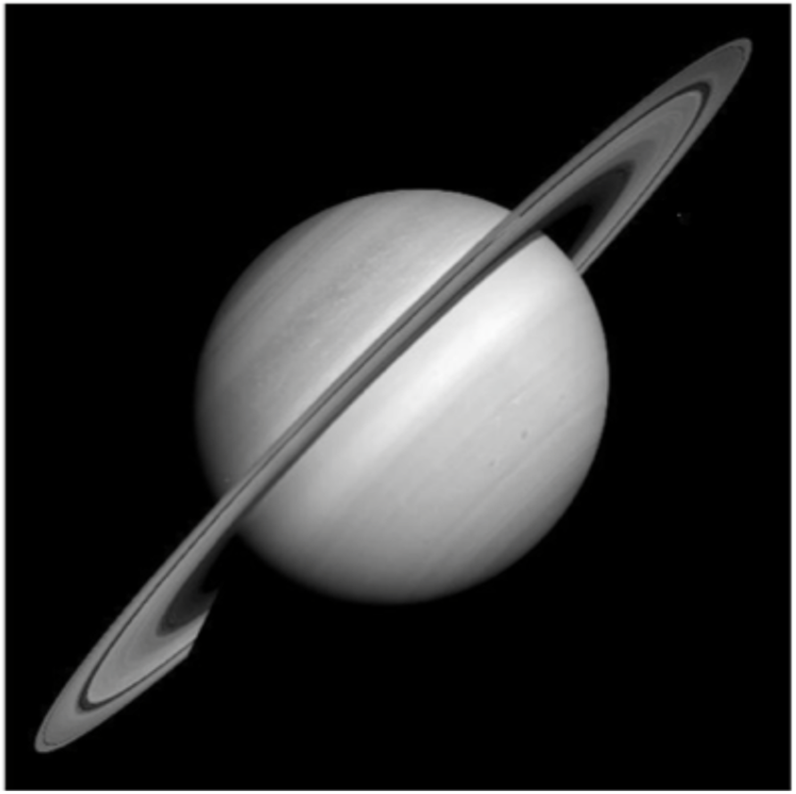
- Missing data in Astrophysics
  - Origin of missing data
  - Characteristics of missing data
- Sparse inpainting
  - Sparse and Compressible signals
  - Inpainting
- Applications
  - Application to Asteroseismology (1D signal)
  - Application to Weak Lensing (2D signal)

Compressible signals



The top 1% of the coefficients concentrate only 8.66% of the energy : not sparse in the direct space

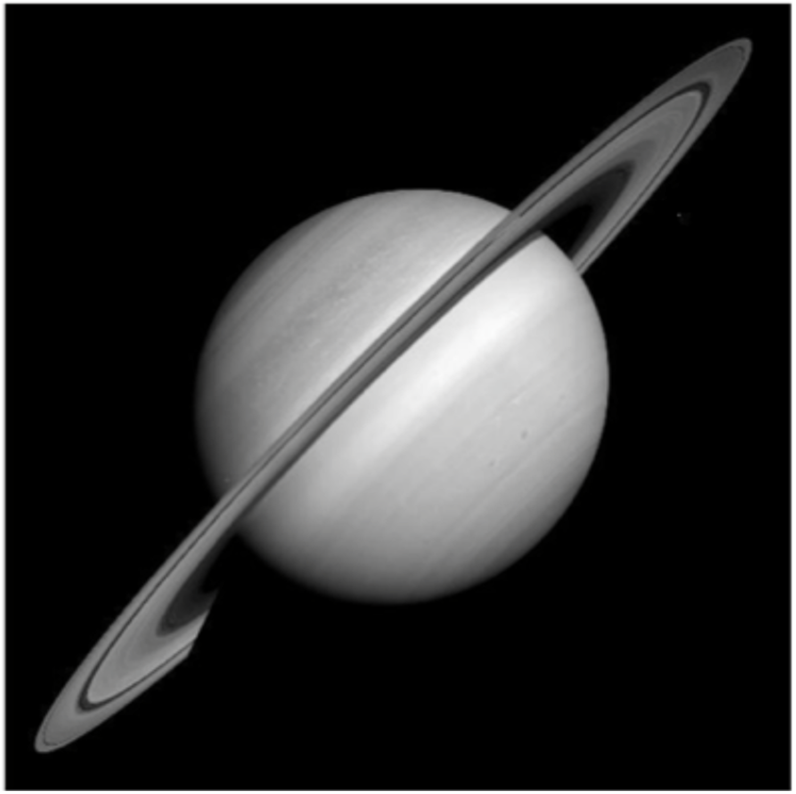
Compressible signals

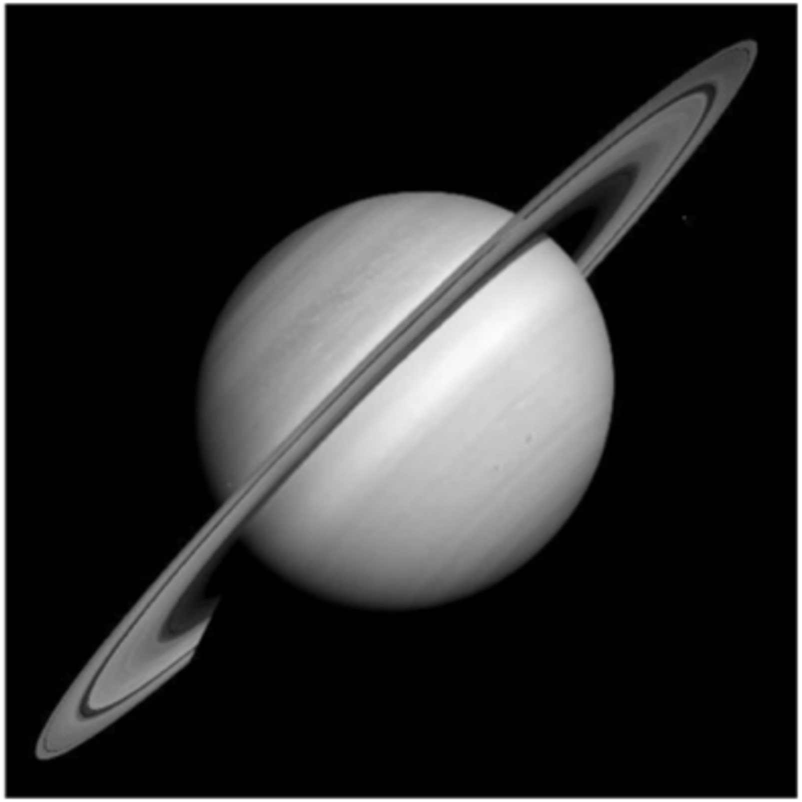


Wavelet transform

1% of the largest coefficients in the Wavelet space
(the others are set to zero)

Compressible signals

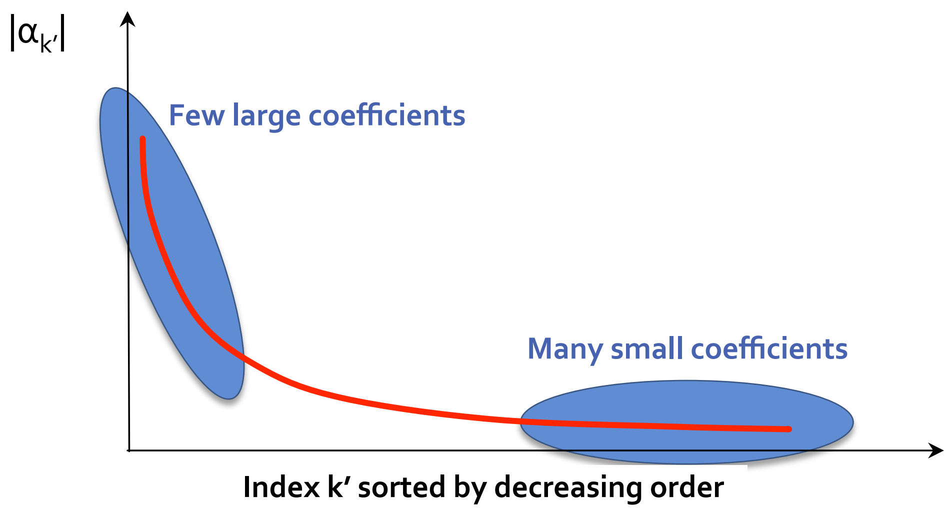




The top 1% of the coefficients concentrate 99.96% of the energy : Sparse in the wavelet space

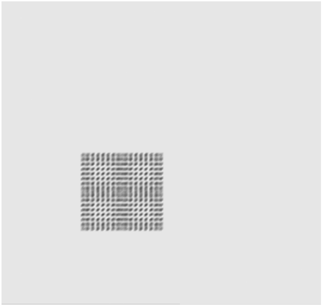
Weak sparsity prior or compressible signals

$$\overset{\alpha}{W(a,b)} = K \int_{-\infty}^{+\infty} \overset{\Phi}{\psi^*\left(\frac{x-b}{a}\right)} \overset{X}{f(x)} dx$$

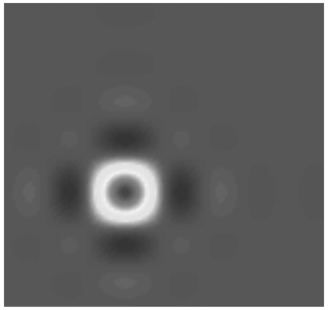


Adapted representations

- **Local DCT** ***(Ahmed, 1975)***
  - Stationary textures
  - Locally oscillatory



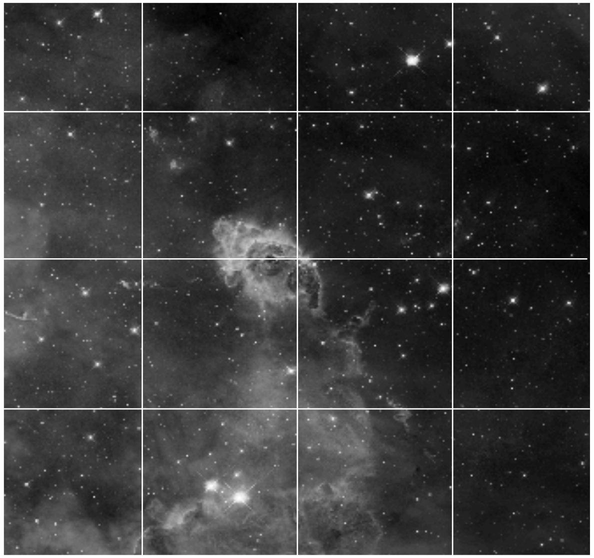
- **Wavelet transform** ***(Mallat, 1999)***
  - Piecewise smooth
  - Isotropic structures



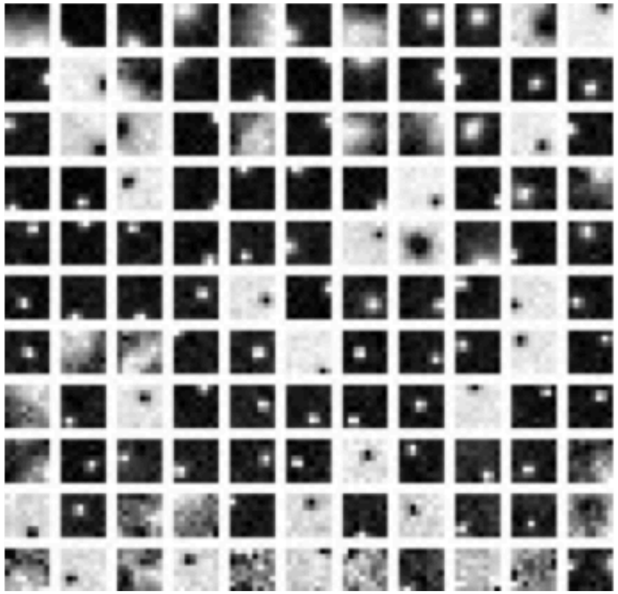
- **Curvelet transform** ***(Candès, 2006)***
  - Piecewise smooth
  - Edge structures

Dictionary Learning

Beckouche et al., A&A, 2013



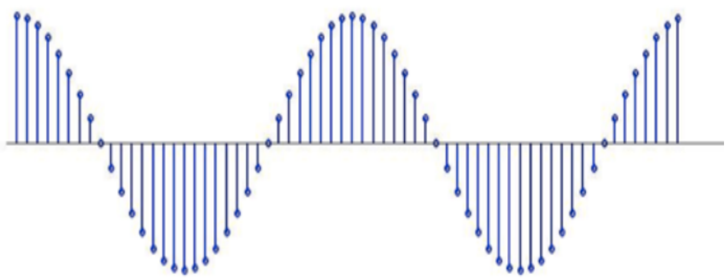
Data



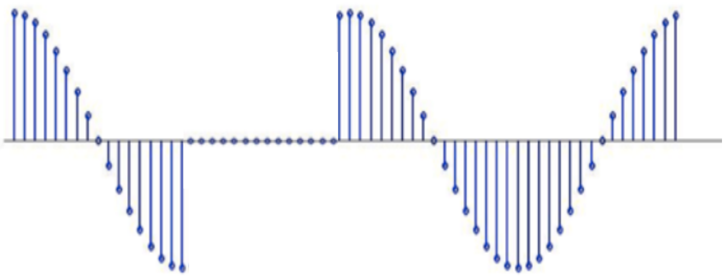
Learned dictioanary

Sparse Inpainting

Elad et al., JACHA, 2005

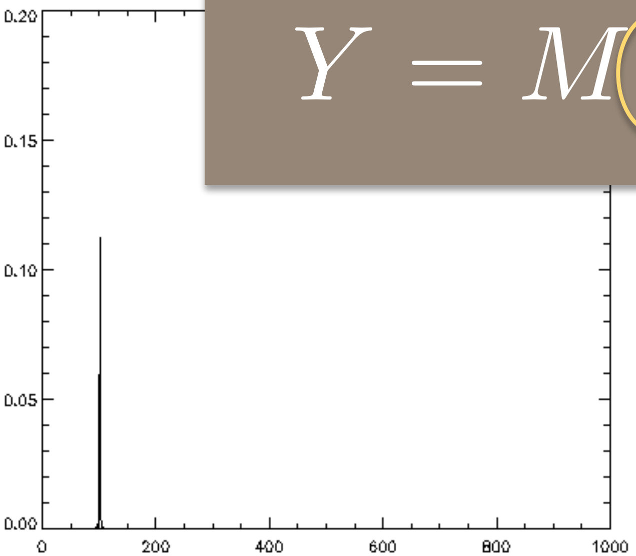


Complete Sine

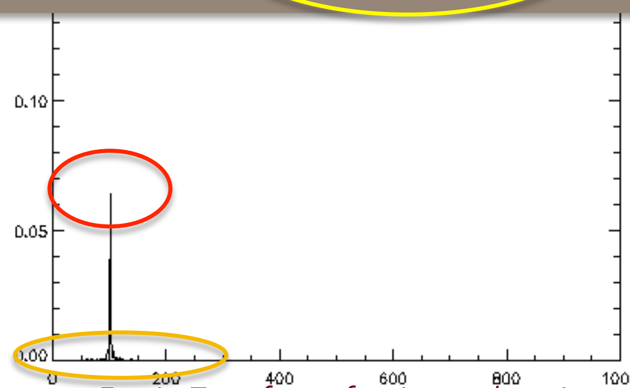


Incomplete Sine

$$Y = M\Phi\alpha \text{ with } \min_{\alpha} ||\alpha||_1$$

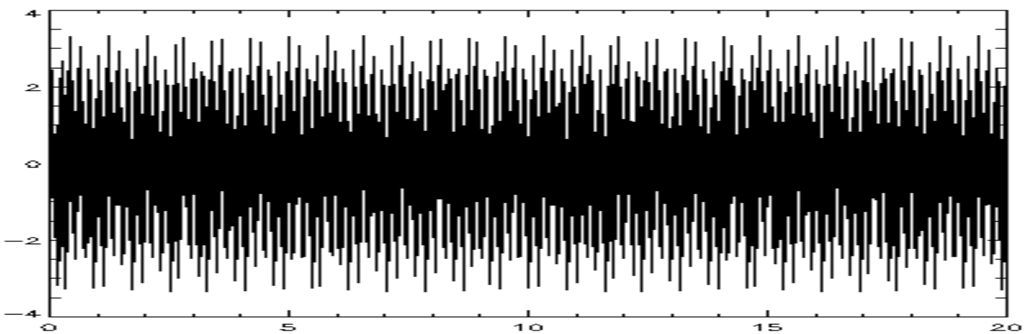


Fourier Transform of a complete sine

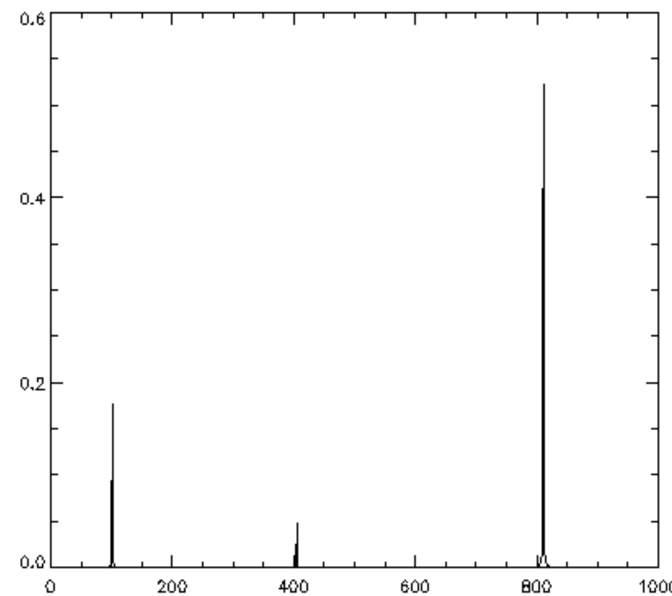


Fourier Transform of an incomplete sine

Sparse Inpainting Algorithm

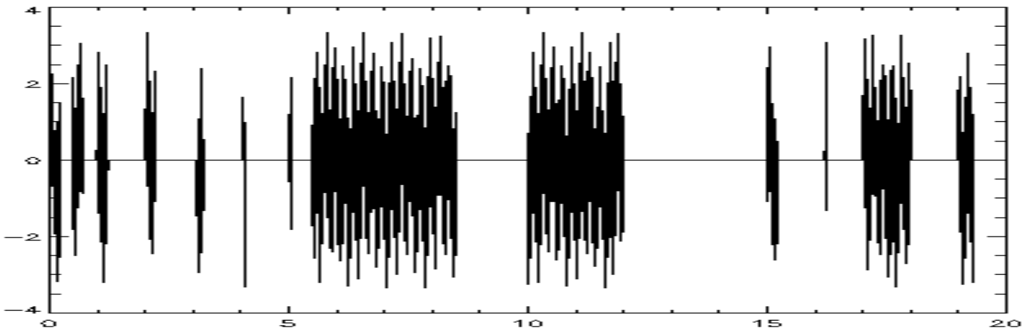


Complete Signal

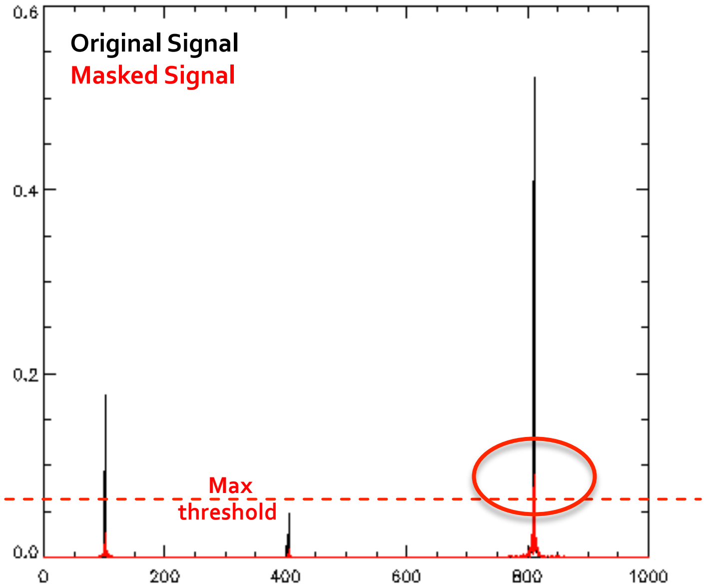


Fourier Transform

Sparse Inpainting Algorithm

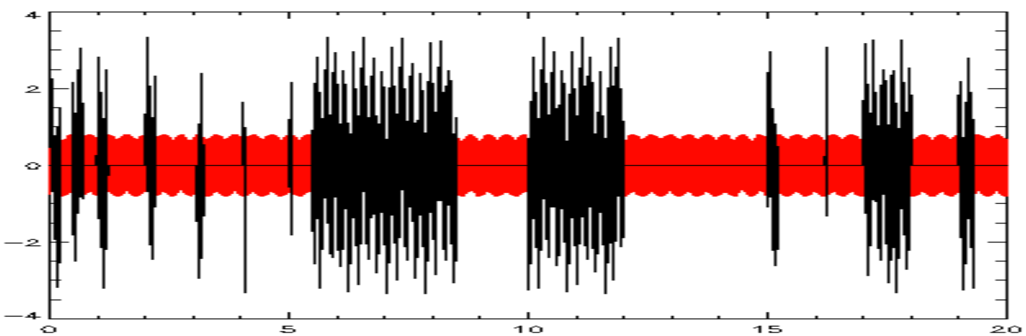


Incomplete Signal

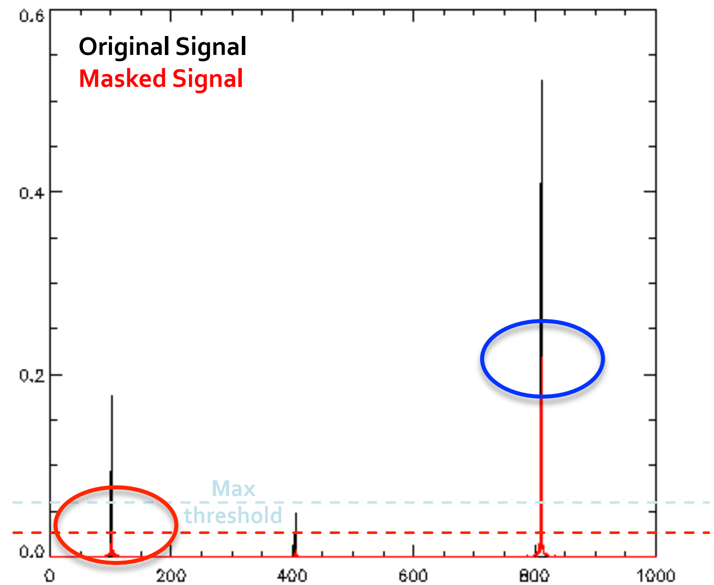


Fourier Transform

Sparse Inpainting Algorithm

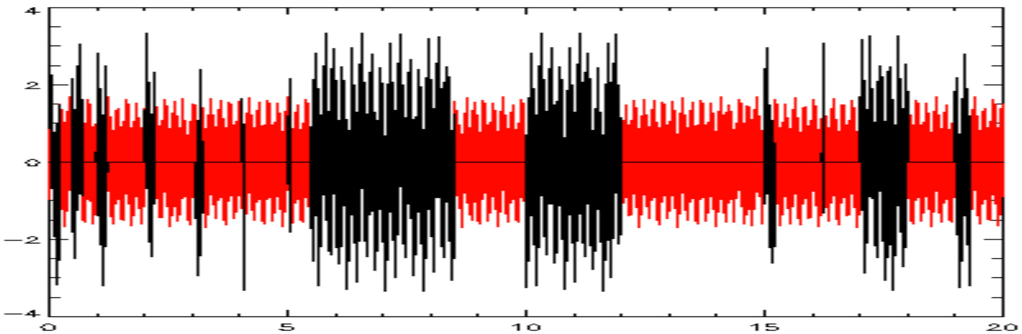


Reconstructed signa at iteration 1

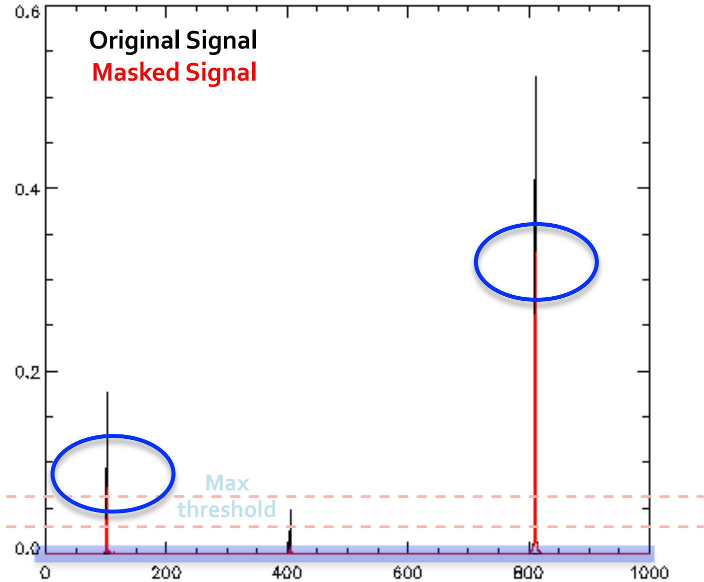


Fourier Transform

Sparse Inpainting Algorithm

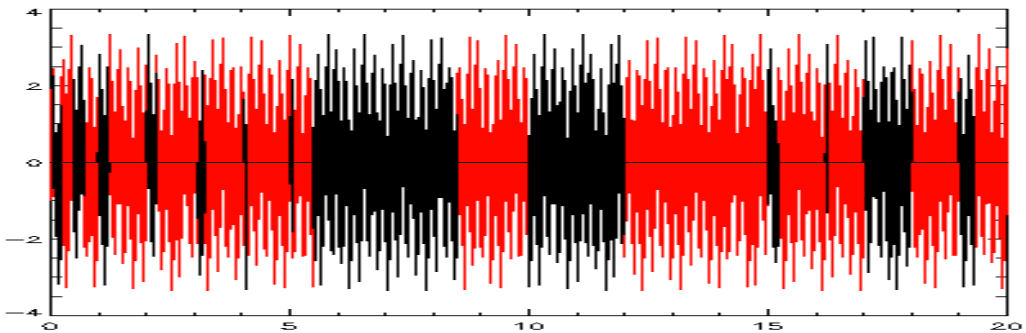


Reconstructed signal at iteration 2

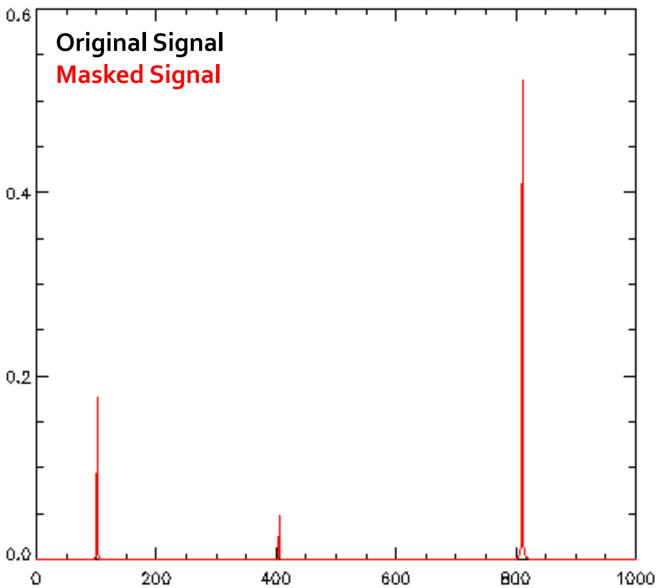


Fourier Transform

Sparse Inpainting Algorithm



Reconstructed signal at iteration N

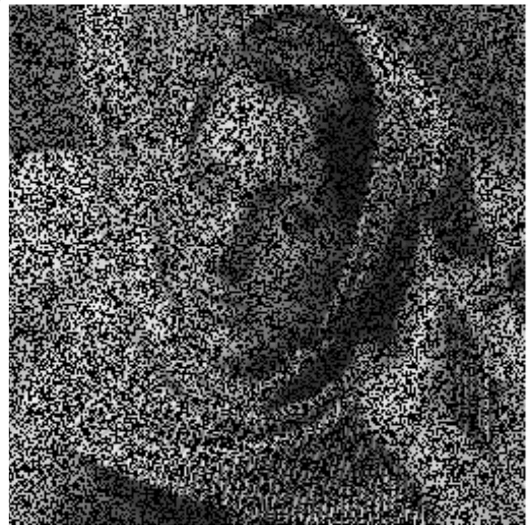


Fourier Transform

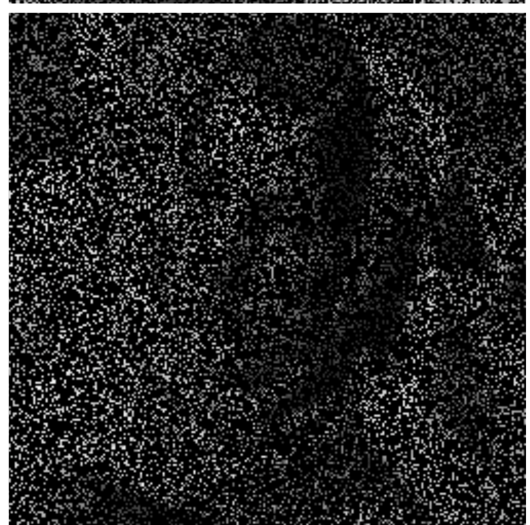
Sparse Inpainting

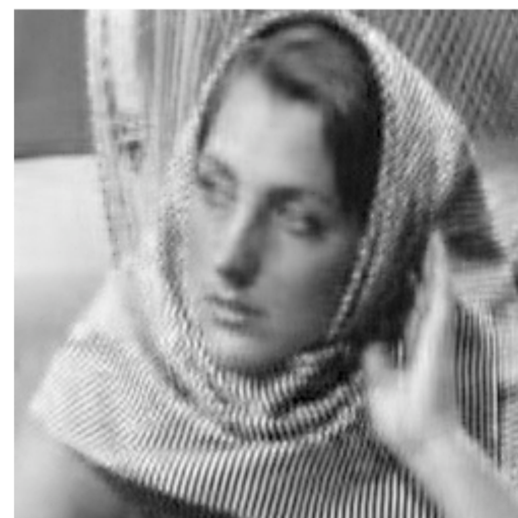
Elad et al., JACHA, 2005

50%



80%

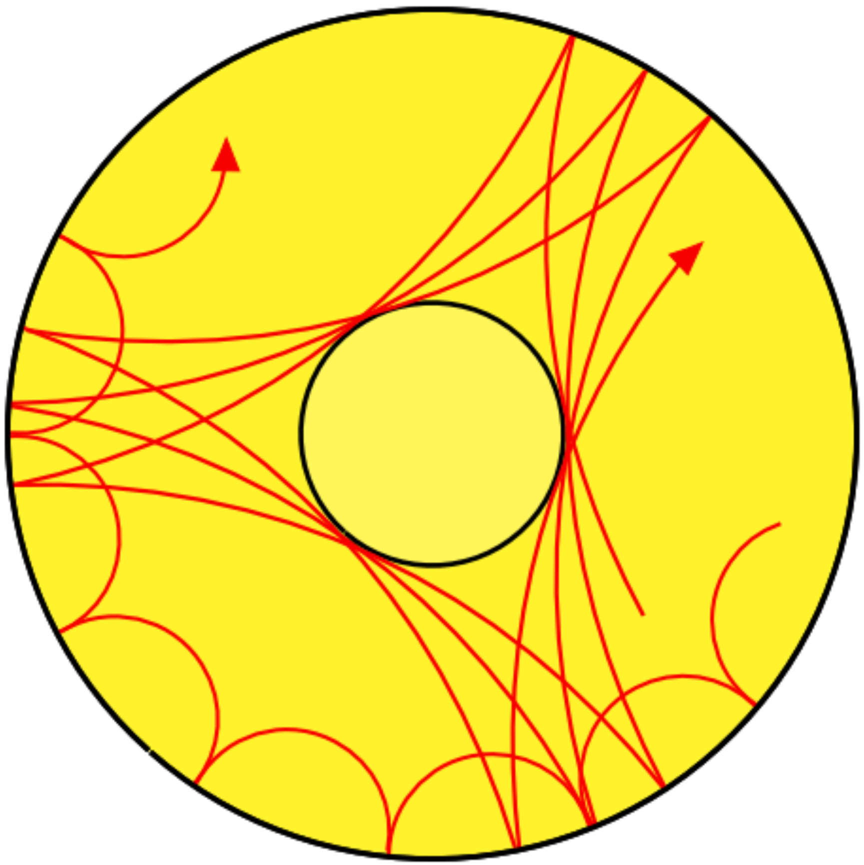




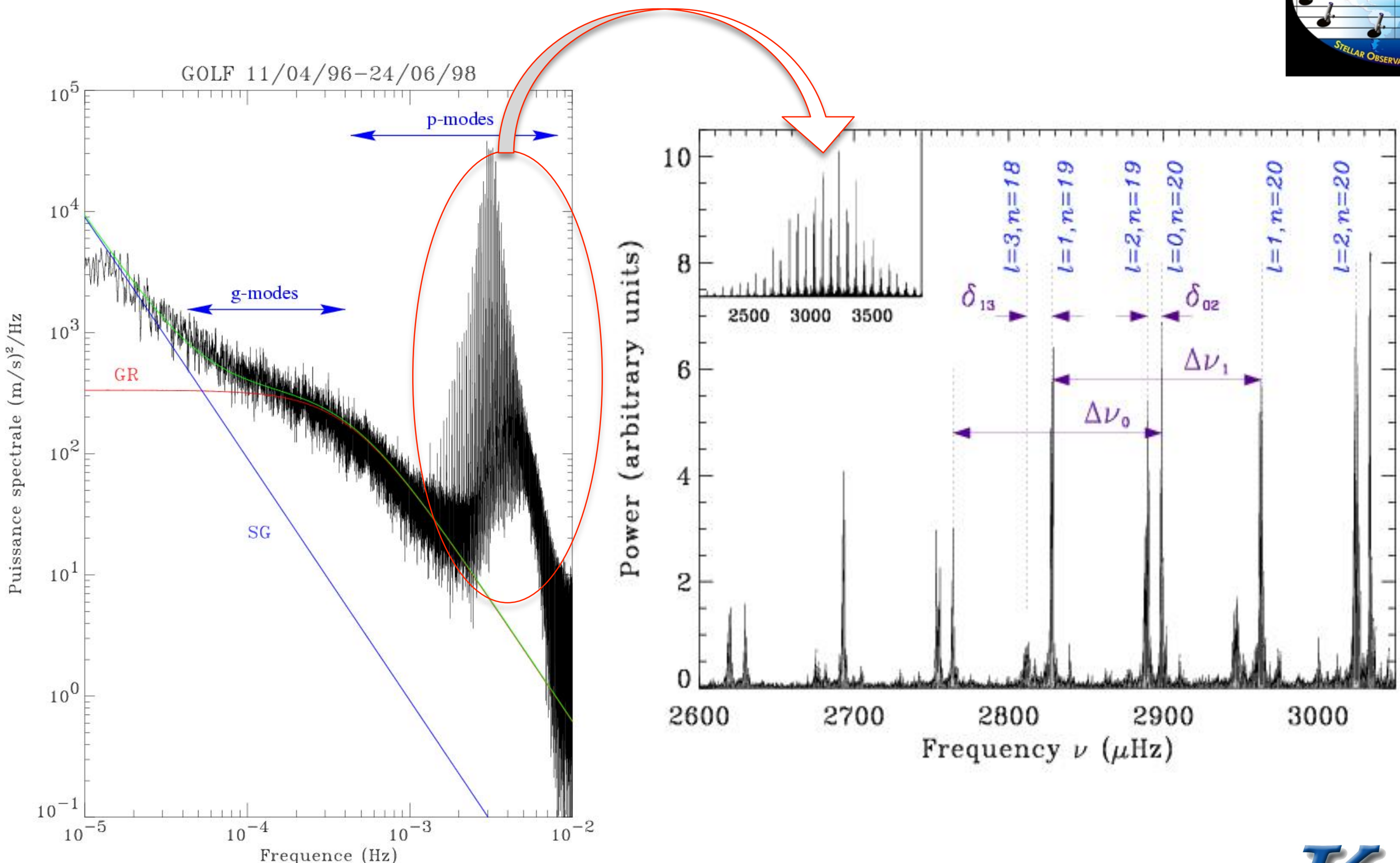
Outline

- Missing data in Astrophysics
  - Origin of missing data
  - Characteristics of missing data
- Sparse inpainting
  - Sparse and Compressible signals
  - Inpainting
- Applications
  - Application to Asteroseismology (1D signal)
  - Application to Weak Lensing (2D signal)

Asteroseismology

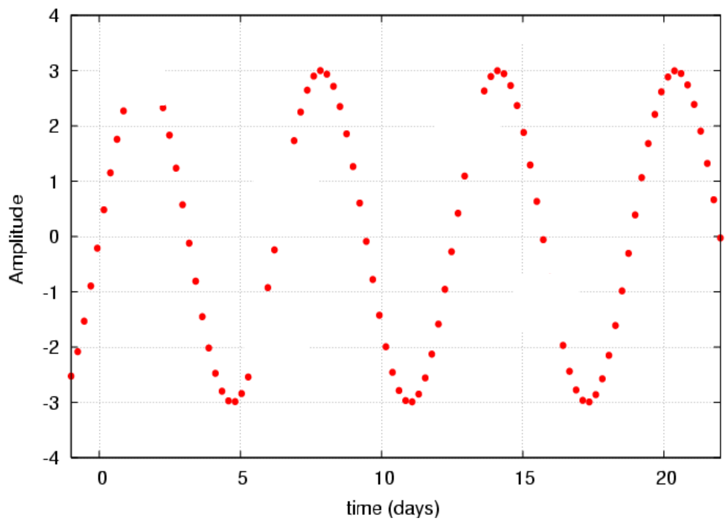


Asteroseismology



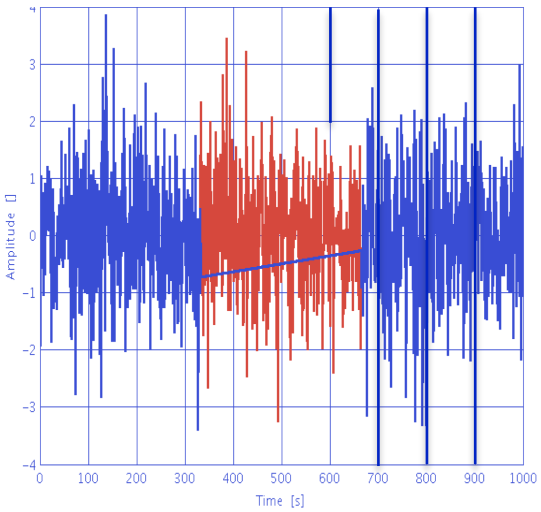
Standard Methods

- Direct Power spectrum estimation
  - Sine Wave Fitting $y = a\cos wt + b\sin wt$
  - CLEAN

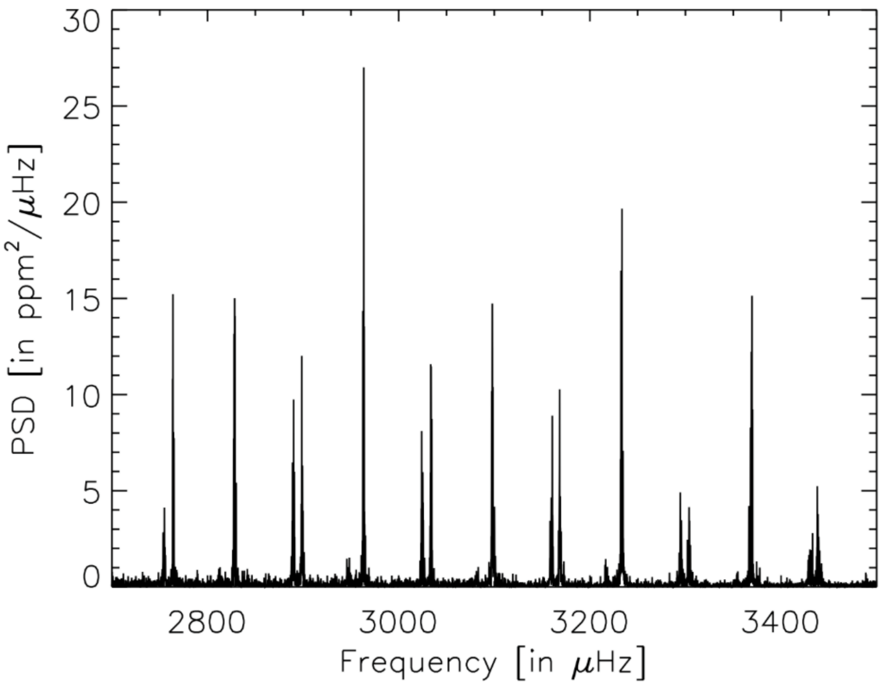


Standard Methods

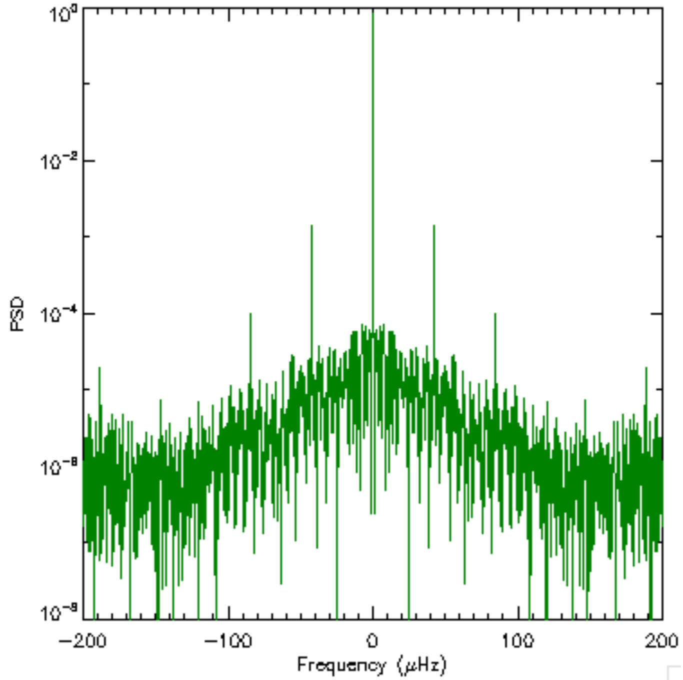
- Direct Power spectrum estimation
  - Sine Wave Fitting
  - CLEAN
- Power spectrum estimation based on gap filling
  - Linear interpolation
  - ARMA
  - Sparse Inpainting



Simulated data



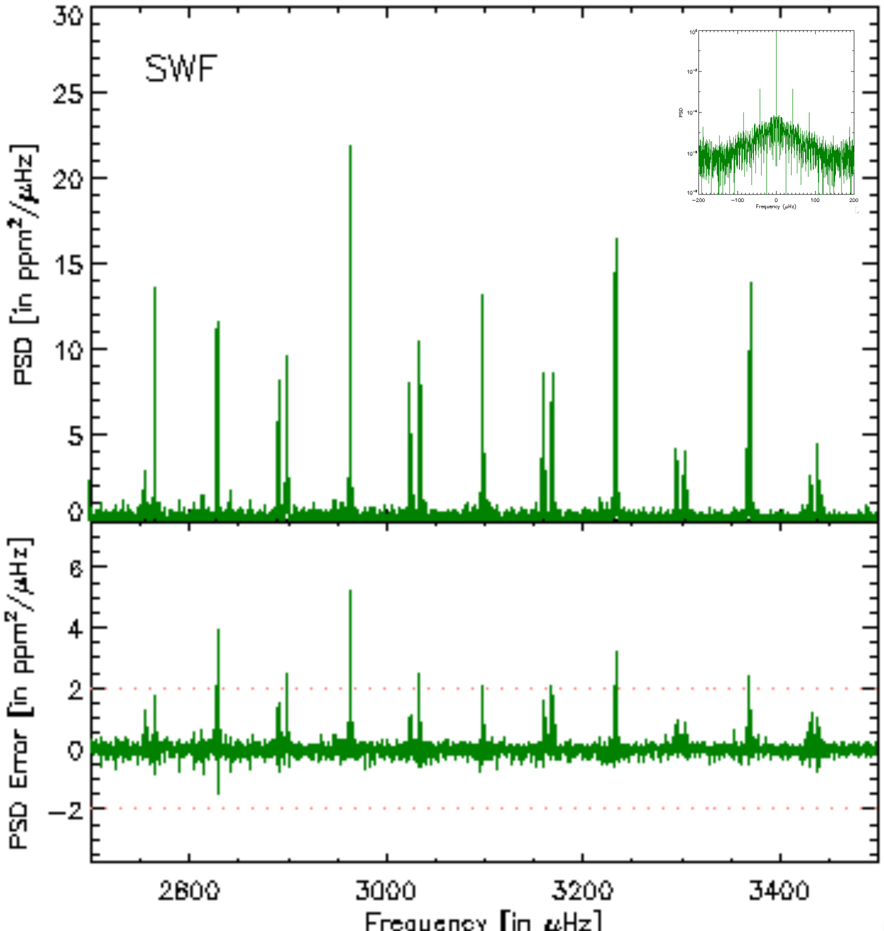
Power Spectrum of the complete time series

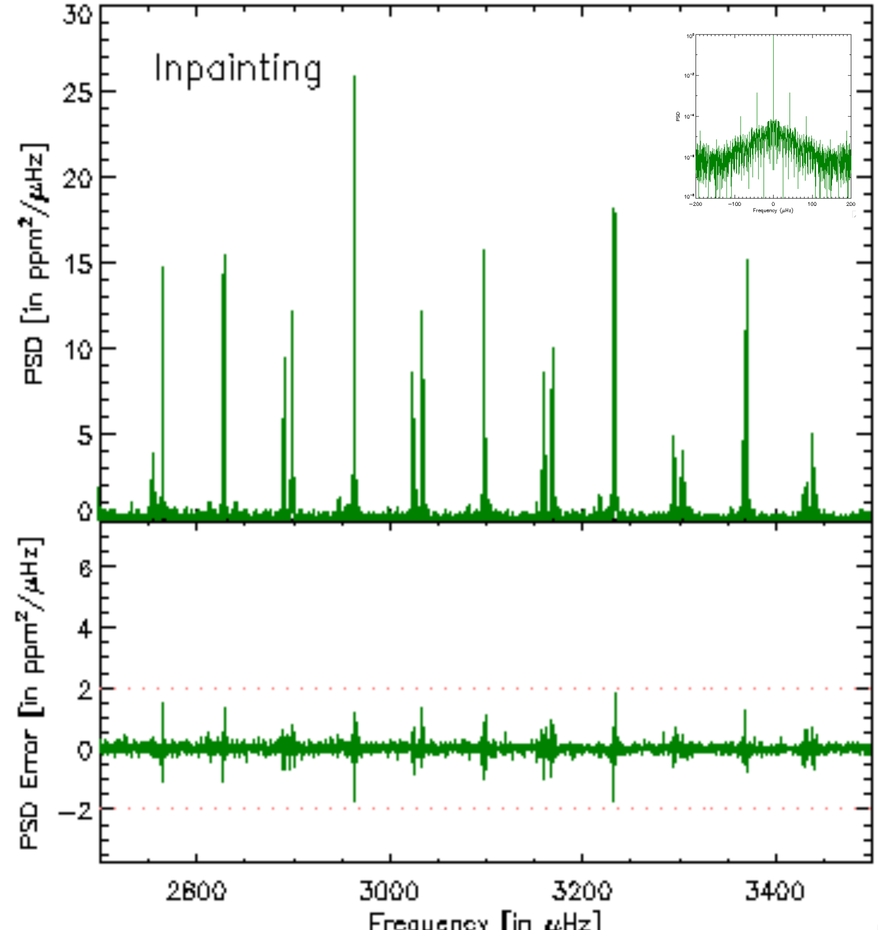


Quasi-regular mask
15% of missing data

Space-based like data : SWF vs Inpainting

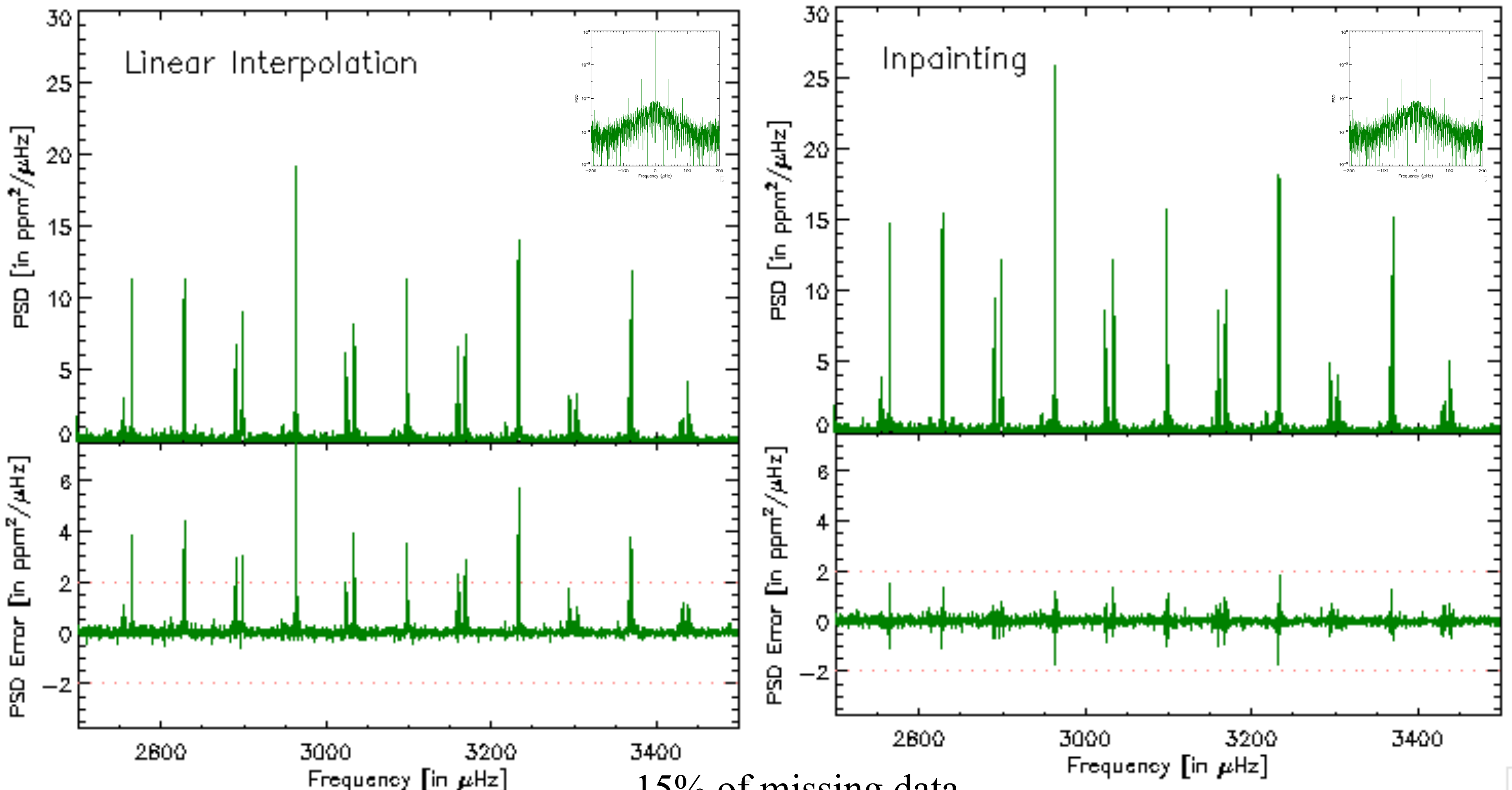
Pires et al., A&A, 2015





15% of missing data

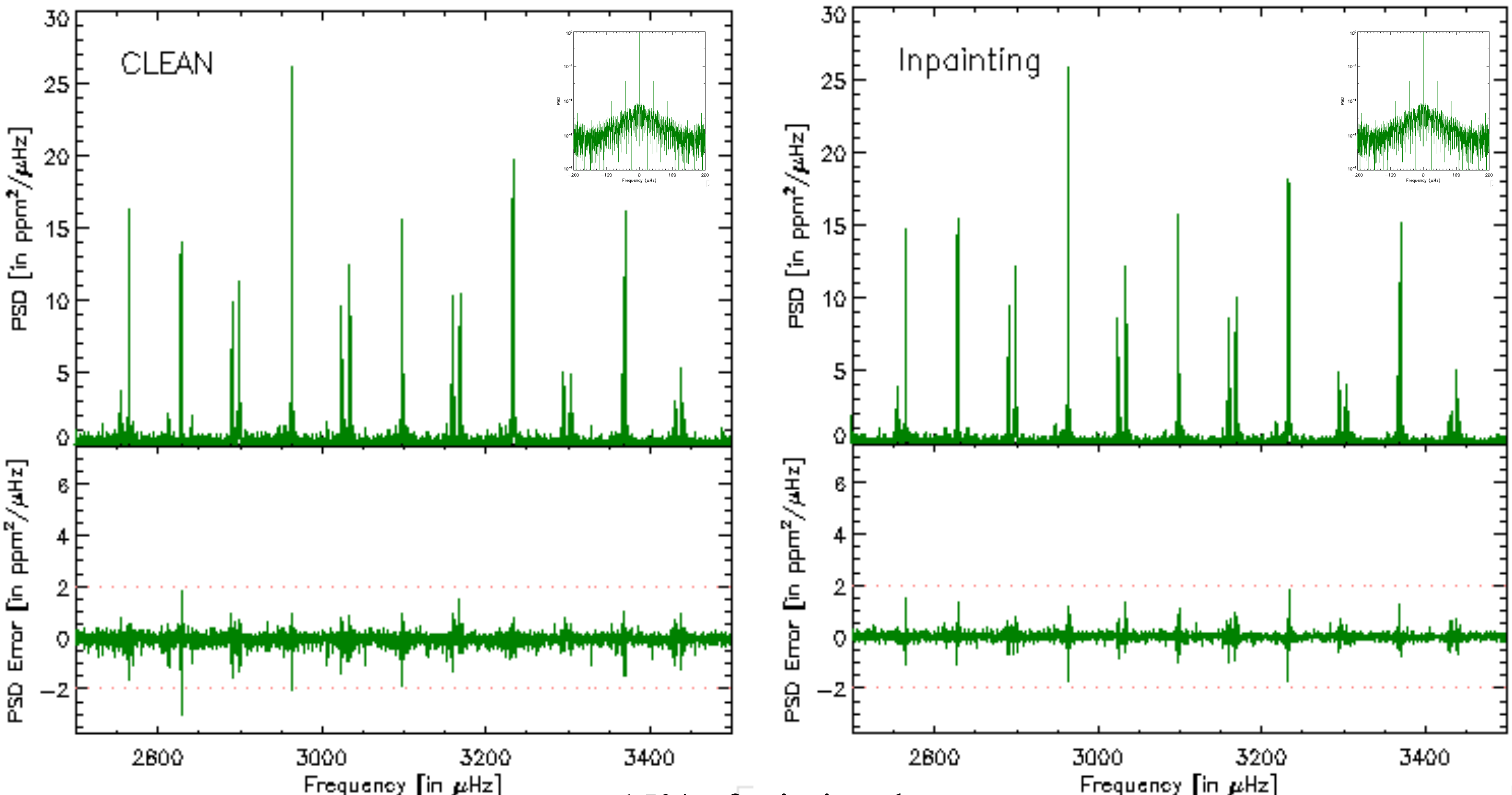
Space-based like data : LI vs Inpainting



15% of missing data

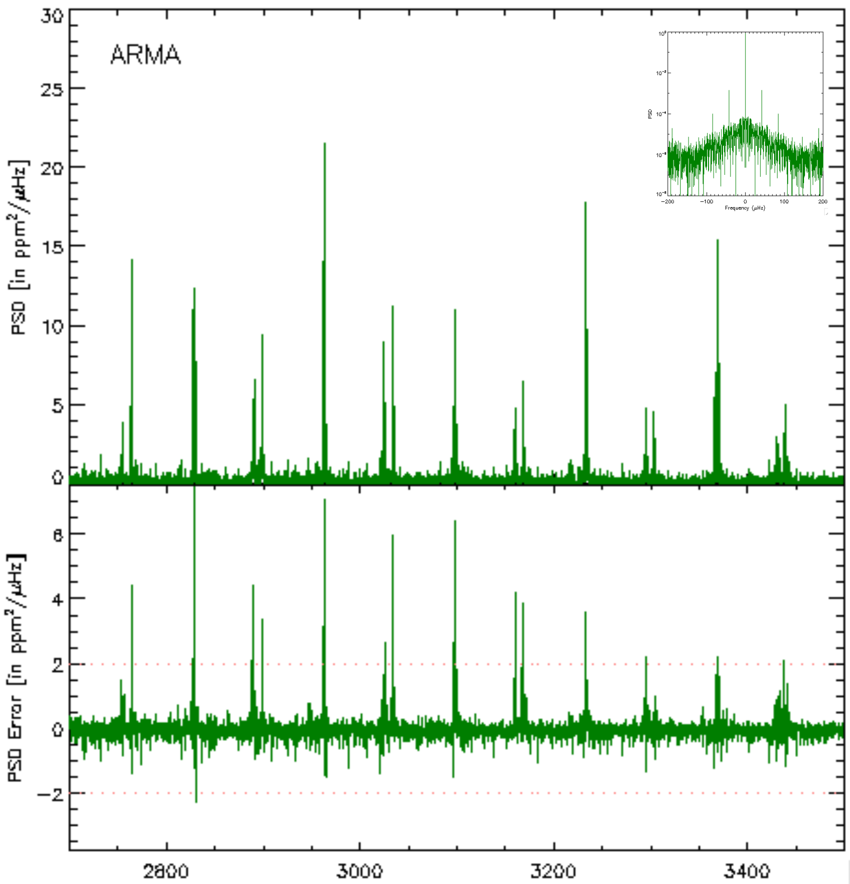
Space-based like data : CLEAN vs Inpainting

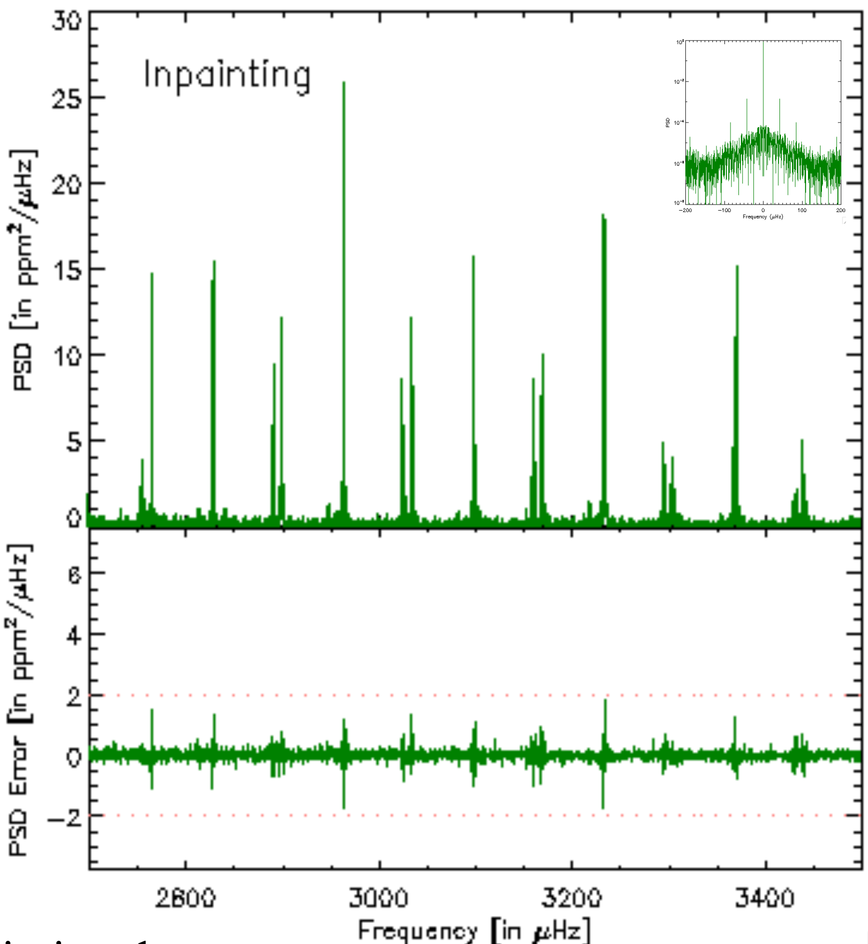
Pires et al., A&A, 2015



15% of missing data

Space-based like data : ARMA vs Inpainting





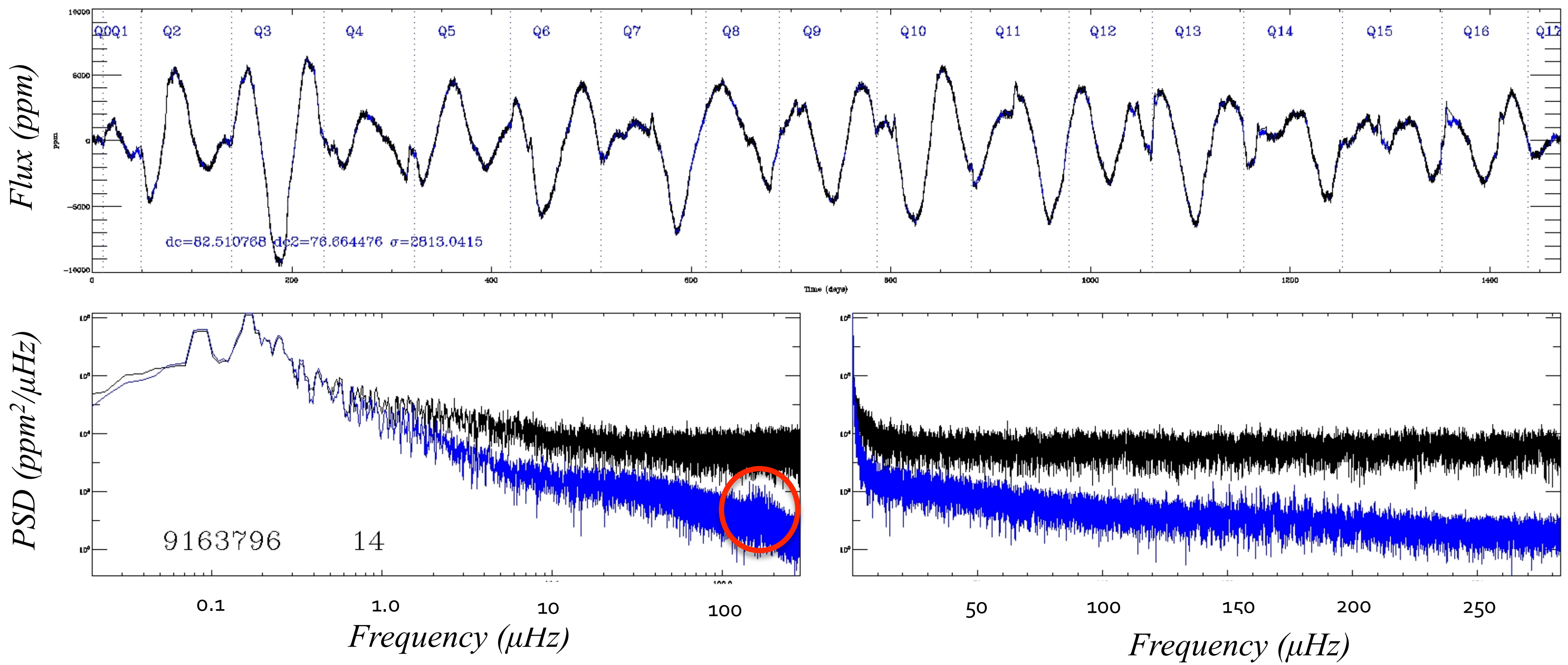
15% of missing data

Processing time

For a time series of 50 days with a sampling of 32s (duty cycle = 50 %):

- SWF
  - 4 hours to compute one tenth of the full power spectrum
- CLEAN
  - about 4 days to compute one tenth of the full power spectrum
- Linear Interpolation
  - Few seconds to compute the full power spectrum
- Sparse Inpainting
  - 4 min to compute the full power spectrum
- MIARMA
  - Few hours to compute the full power spectrum

Application on real data :
Unveiling modes



K-Inpainting software

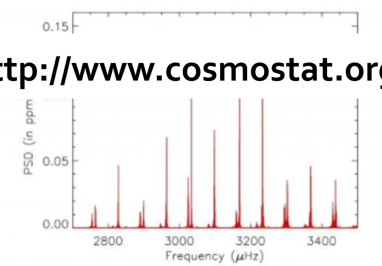
CosmoStat
An AIM laboratory @CEA Saclay

Home News Research People Projects Publications Software Products Events Education Jobs

K-Inpainting

Welcome to the K-Inpainting web page.
This page introduces the K-Inpainting software (Version 1.0)

In asteroseismology, the observed time series often suffers from incomplete time coverage due to repeated gaps. The presence of periodic gaps may generate spurious peaks in the power spectrum that limit the analysis of the data. Various methods have been developed to deal with gaps in time series data. We propose a new approach to handle the problem, the so-called inpainting method. This technique, based on a sparsity prior, enables to judiciously fill-in the gaps in the data, preserving the asteroseismic signal, as far as possible. This method can be applied both on ground and space-based data. It appears that the inpainting technique improves the oscillation modes detection and estimation. The impact of the observational window function is reduced and the interpretation of the power spectrum is simplified. Additionally, it can be used to study very long time series of many stars because its computation is very fast.



http://www.cosmostat.org/software/k-inpainting/

Power Density Spectrum for a duty cycle of 83% computed using an FFT on the inpainted time series.

Description

K-Inpainting (Inpainting for Kepler) is a software written in C++ with an IDL interface.

Download

[Download the K-Inpainting software]

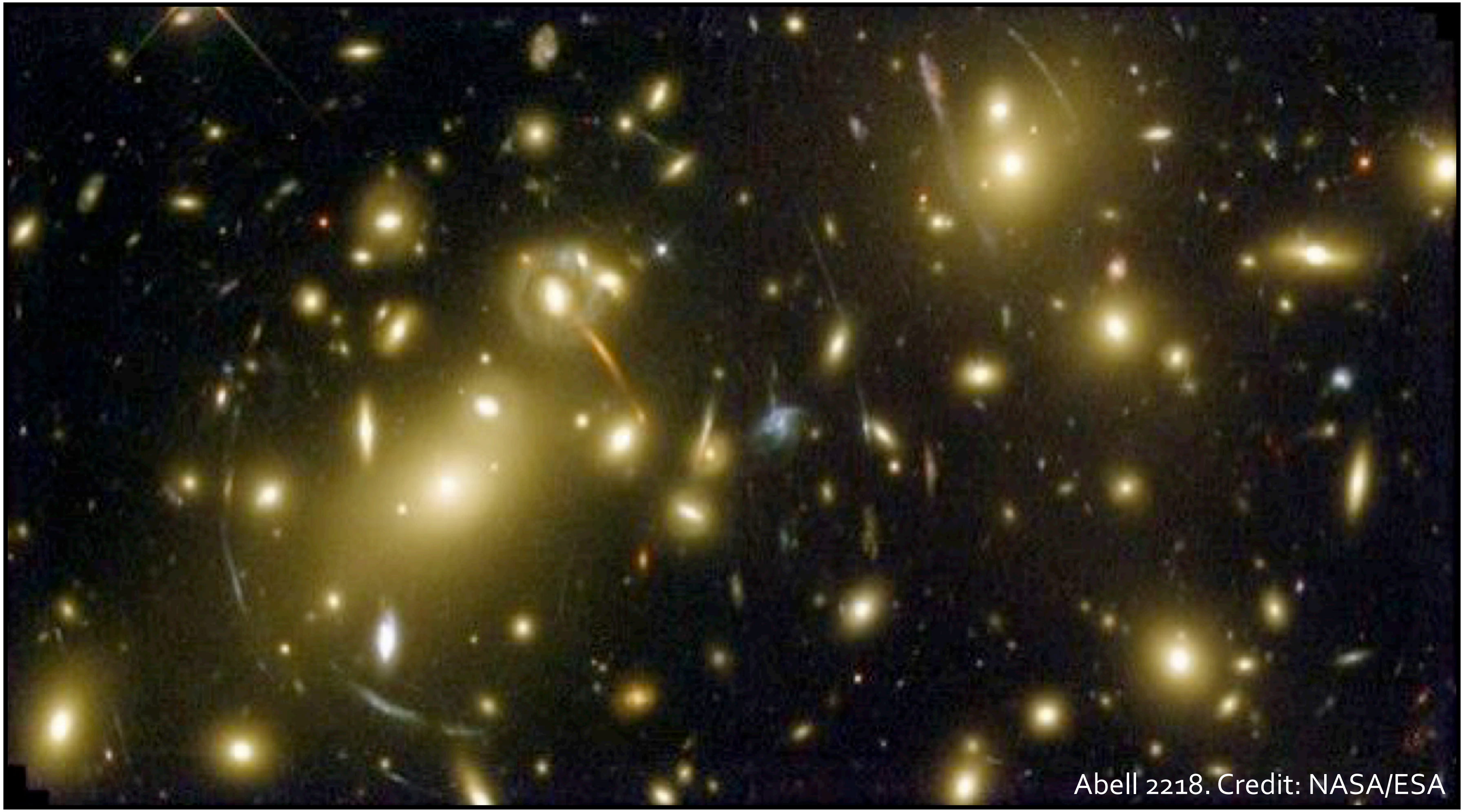
Application on real data

- The **Kepler Asteroseismic pipeline** developed by the Kepler AsteroScientic Consortium (KASC) uses the software K-Inpainting to correct the light curves from missing data.
- The **official CoRoT pipeline** will use K-Inpainting software to correct the missing data in both asteroseismic and exoplanet channel.
  - Deadline production code delivery: end of May 2015

Outline

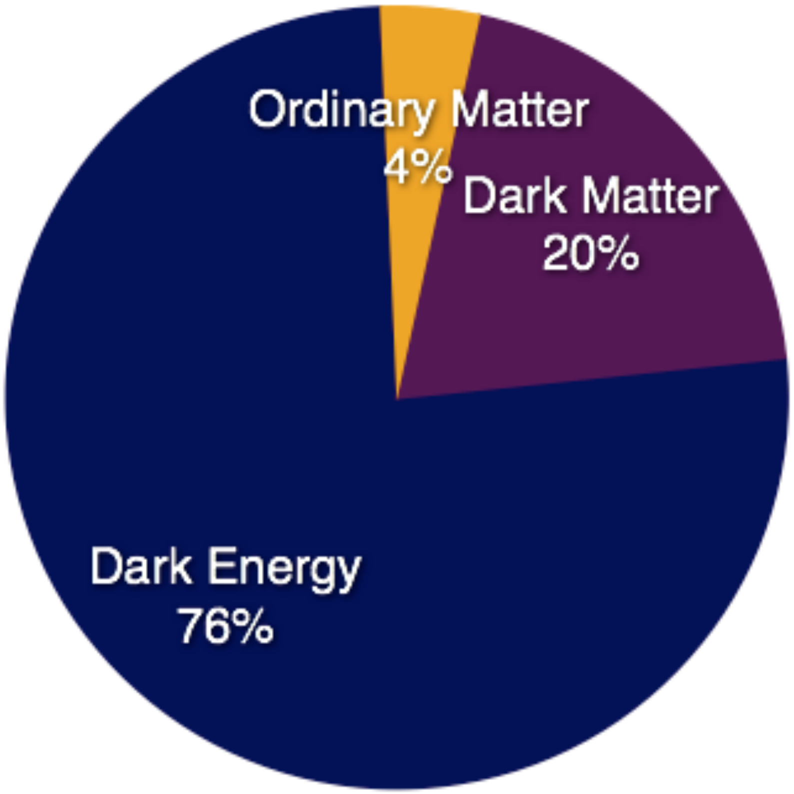
- Missing data in Astrophysics
  - Origin of missing data
  - Characteristics of missing data
- Sparse inpainting
  - Sparse and Compressible signals
  - Inpainting
- Applications
  - Application to Asteroseismology (1D signal)
  - Application to Weak Lensing (2D signal)

Gravitational Lensing effect

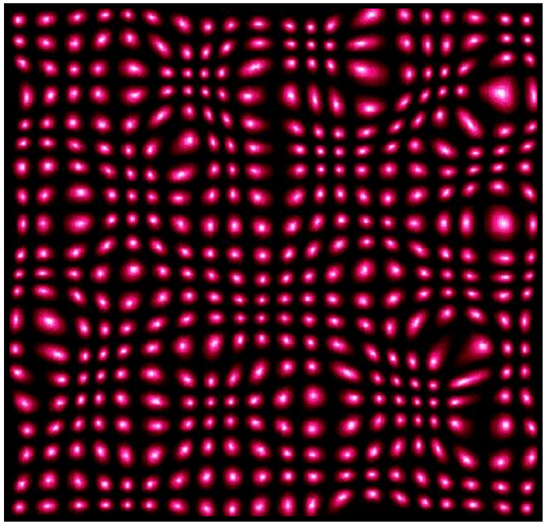


Abell 2218. Credit: NASA/ESA

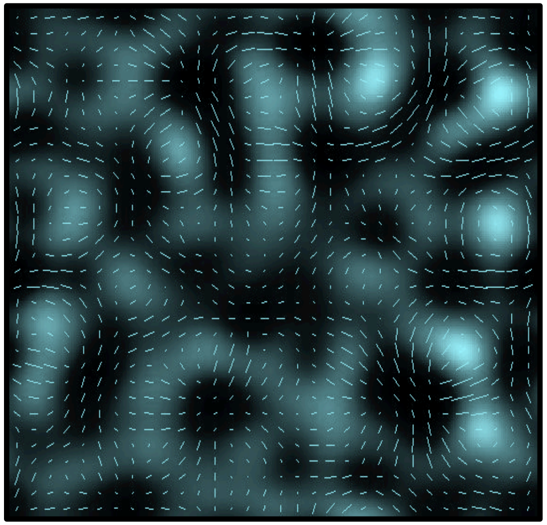
Cosmological model



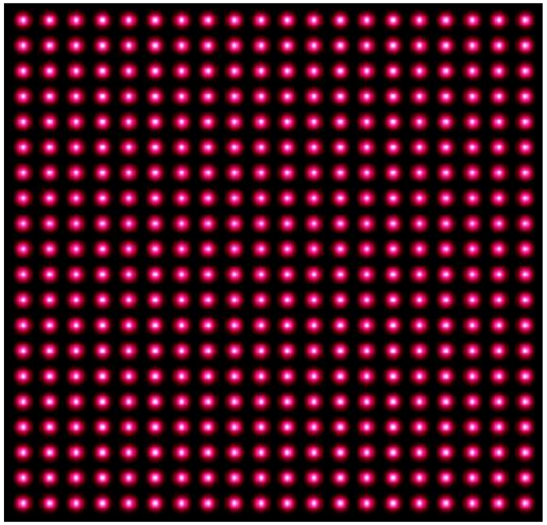
Weak Gravitational Lensing



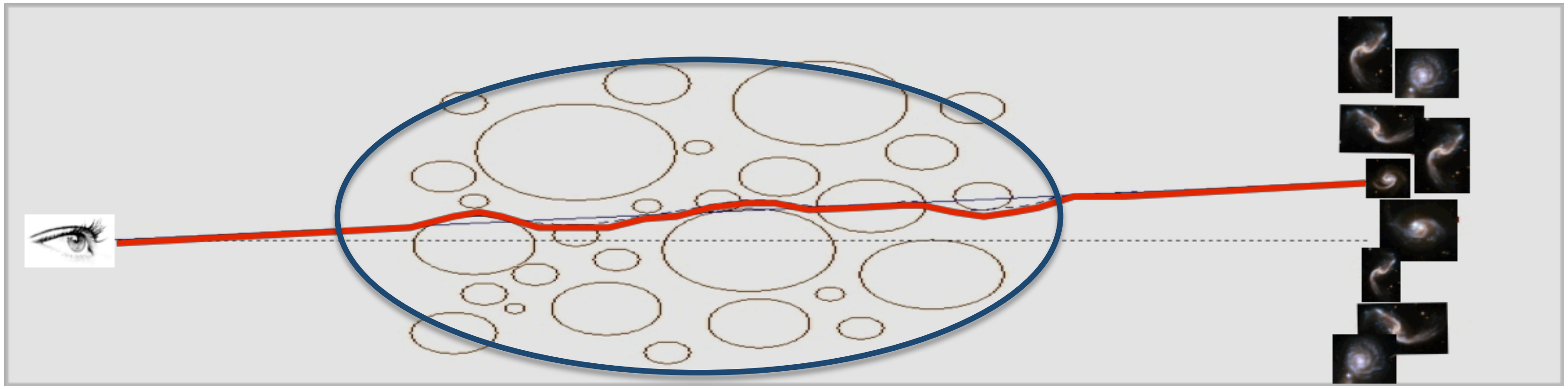
Observer



Gravitational lens

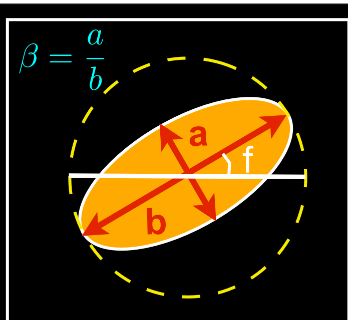


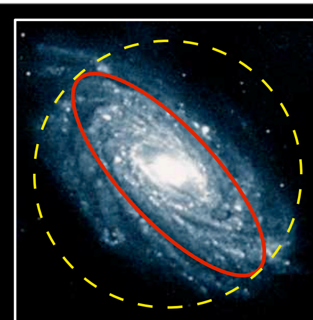
Background galaxies

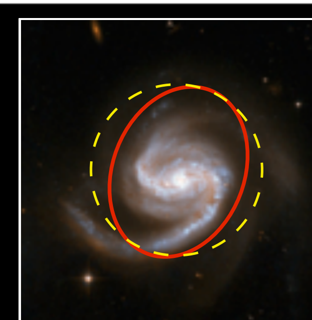


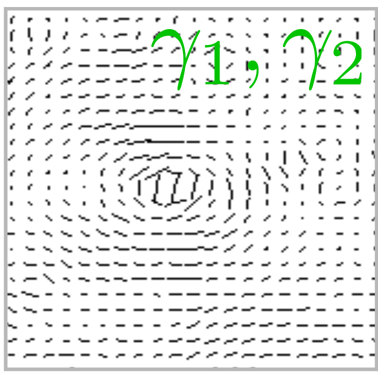
Gravitational lens

Inversion Equations





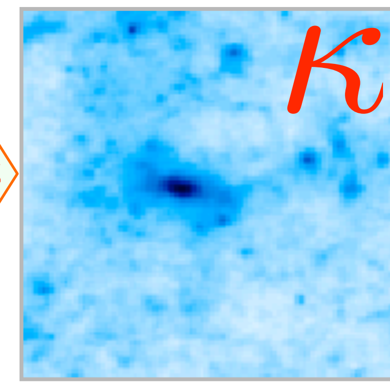




$$\gamma_1 = \frac{1}{2}\left(\partial_1^2 - \partial_2^2\right)\psi$$
$$\gamma_2 = \partial_1 \partial_2 \psi$$

$$\psi$$

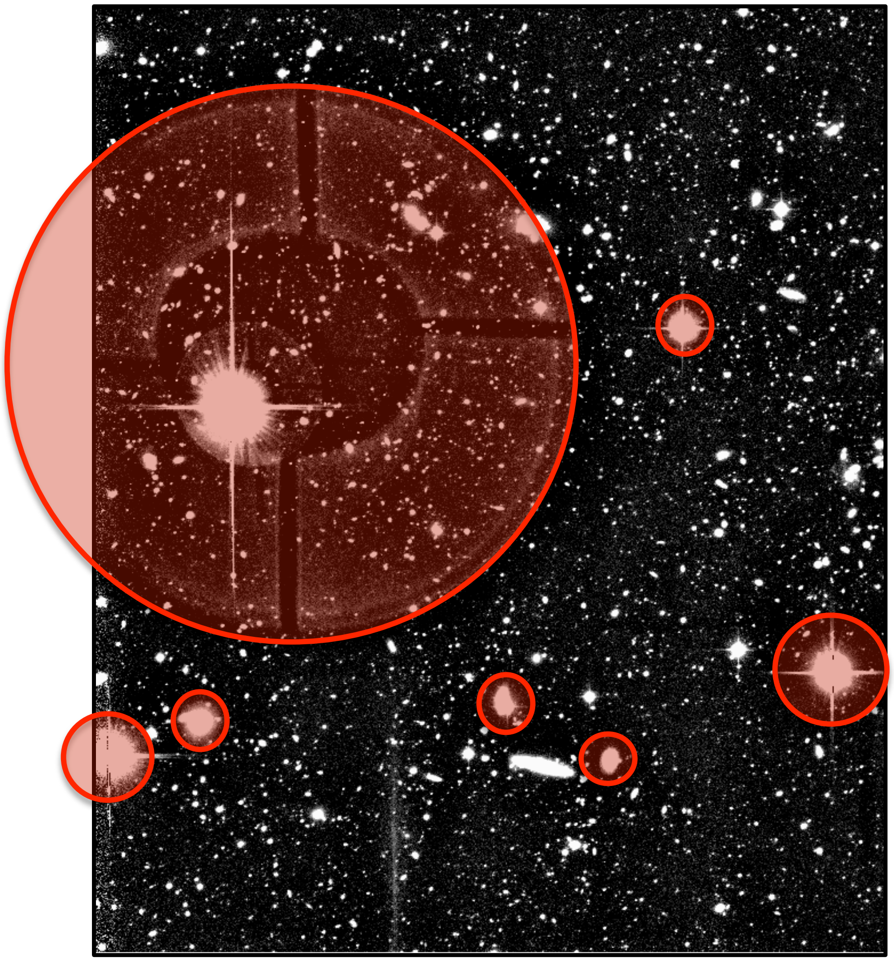
$$\frac{1}{2}\left(\partial_1^2 + \partial_2^2\right)\psi = \kappa$$



$$\gamma_i = \hat{P}_i \kappa$$
$$\kappa = \hat{P}_1 \gamma_1 + \hat{P}_2 \gamma_2$$

$$\hat{P}_1(k) = \frac{k_1^2 - k_2^2}{k^2}$$
$$\hat{P}_2(k) = \frac{2 k_1 k_2}{k^2}$$

Missing data in Weak Lensing



Inpainting of Weak Lensing data

$$\gamma_i^{obs} \rightarrow \boxed{\min_{\kappa} ||\Phi^T \kappa||_1 \quad \text{subject to} \quad \sum_i ||\gamma_i^{obs} - M(P_i * \kappa)||_2^2 \leq \epsilon} \rightarrow \kappa$$

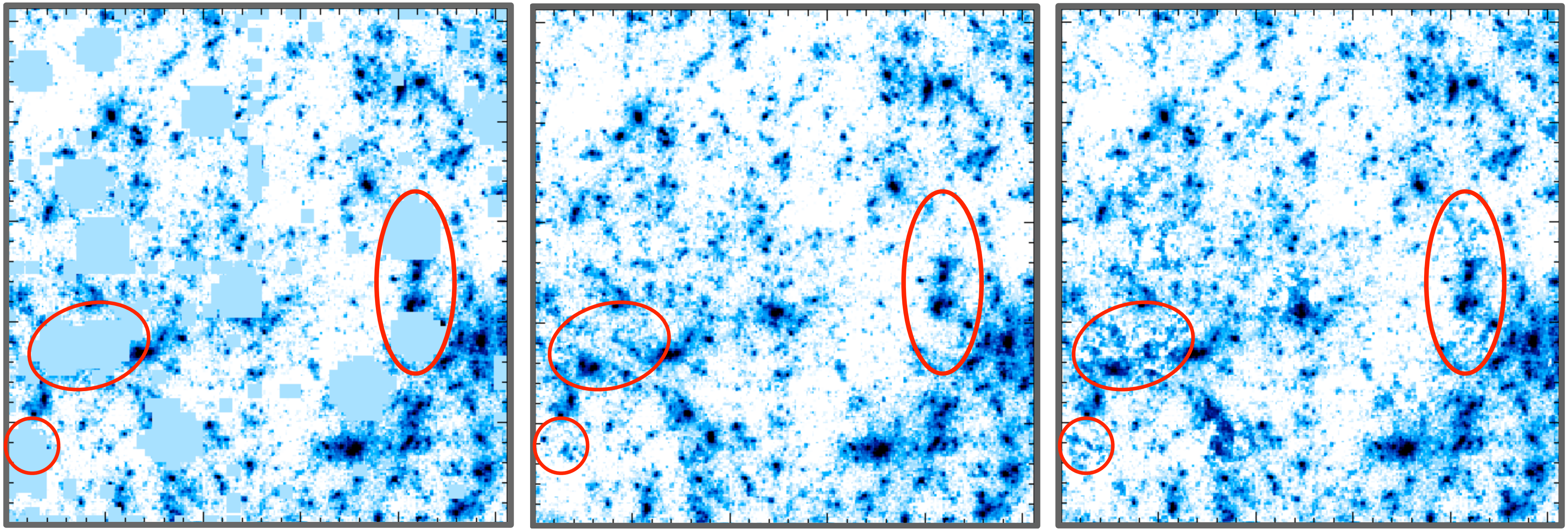
Physical priors

$$\gamma_i^{obs} = M.\gamma_i$$

$$\kappa = P_1 * \gamma_1 + P_2 * \gamma_2$$

Φ^t is the DCT

Missing Data in Weak Lensing

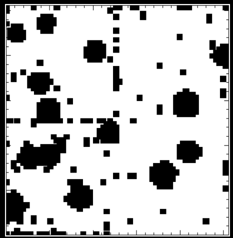


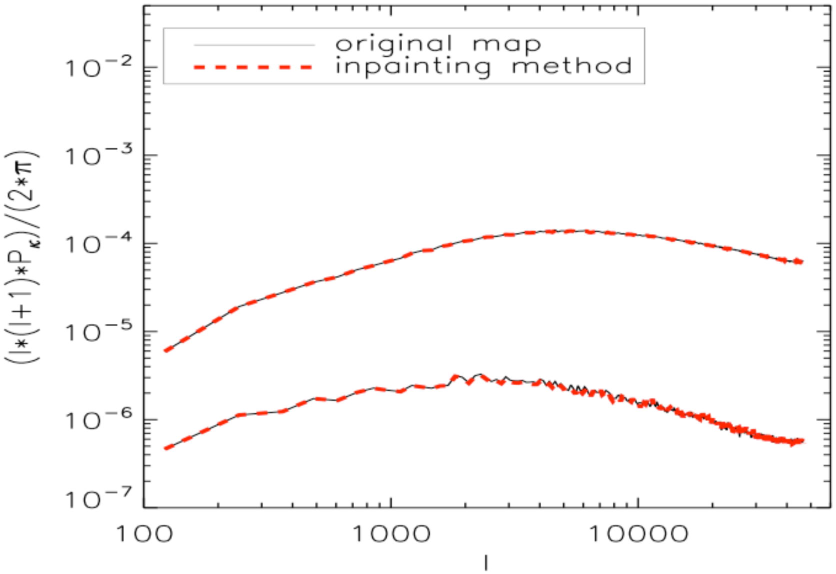
Mask pattern of the CFHTLS survey of 1° x 1° field with 20% of missing data
Courtesy to J. Berge

Which Image is the Original One ?

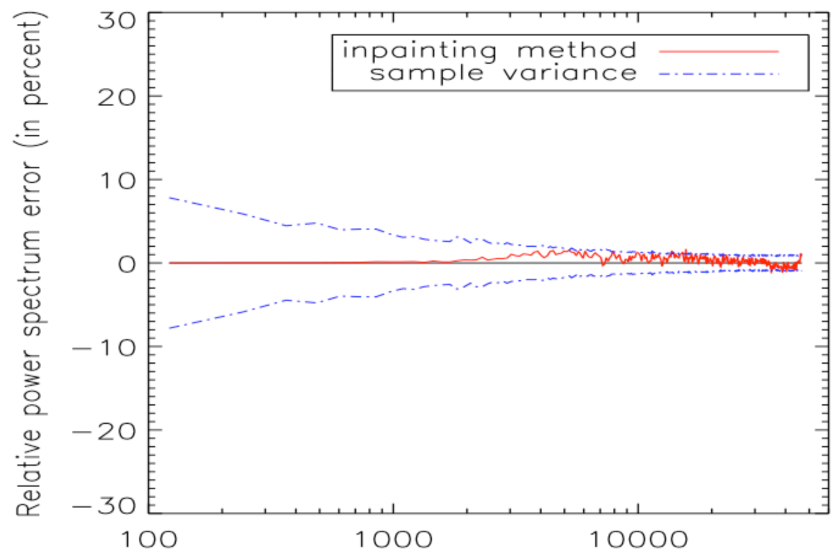
Power spectrum estimation

Pires et al., MNRAS, 2009



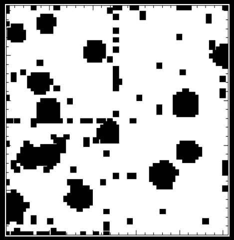


Mean Power spectrum estimated from complete shear maps (black) and inpainted shear maps (red)

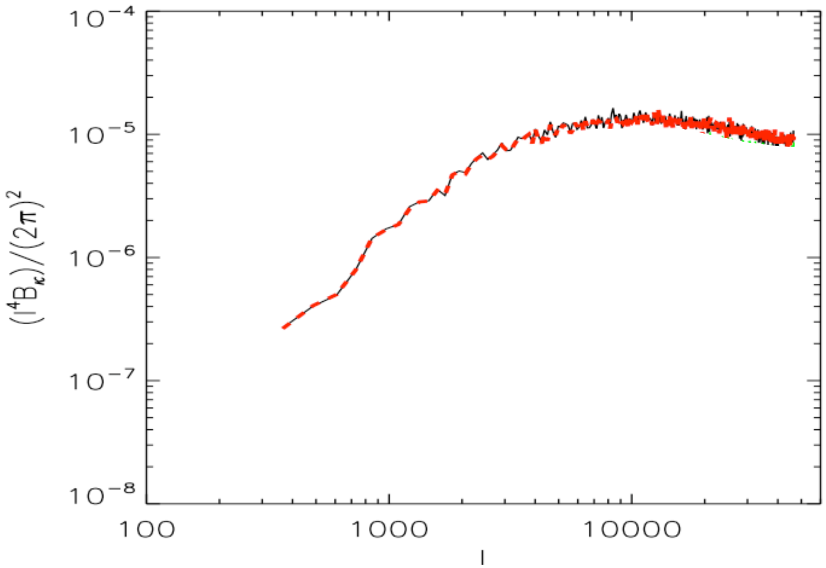


Relative Power spectrum error i.e. the normalized difference between the two upper curves of the left panel

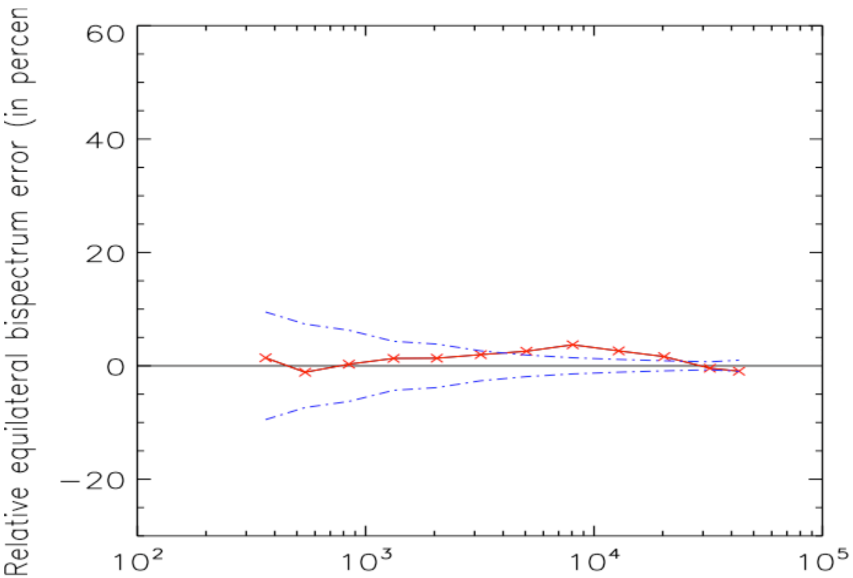
Equilateral bispectrum estimation



Pires et al., MNRAS, 2009



Mean bispectrum estimated from complete shear maps (black), incomplete shear maps (green) and inpainted shear maps (red)



Relative bispectrum error

FASTLens Code

http://www-irfu.cea.fr/Ast/fastlens_software.php

irfu cea saclay

Service d'Astrophysique
Laboratoire AIM

Grand public | Scientifiques | Étudiants | Multimédia | Contact | À propos

FASTLENS (FAST STATISTICS FOR WEAK LENSING) : FAST METHOD FOR WEAK LENSING STATISTICS AND MAP MAKING

S. Pires, J.L. Starck, A. Amara, A. Réfrégier and J. Fadili

The analysis of weak lensing data requires to account for missing data such as masking out of bright stars. To date, the majority of lensing analyses uses the two point-statistics of the cosmic shear field. These can either be studied directly using the two-point correlation function, or in Fourier space, using the power spectrum. The two-point correlation function is unbiased by missing data but its direct calculation will soon become a burden with the exponential growth of astronomical data sets. The power spectrum is fast to estimate but a mask correction should be estimated. Others statistics can be used but these are strongly sensitive to missing data.

The solution that is proposed by FASTLens is to properly fill-in the gaps with only NlogN operations, leading to a complete weak lensing mass map from which we can compute straight forwardly and with a very good accuracy any kind of statistics like power spectrum or bispectrum. We propose also a new method to compute fastly and accurately the power spectrum and the bispectrum with a polar FFT algorithm.

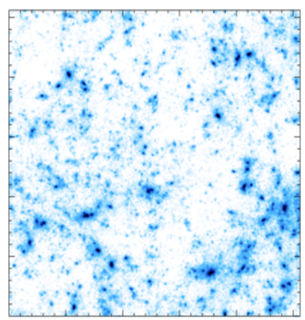
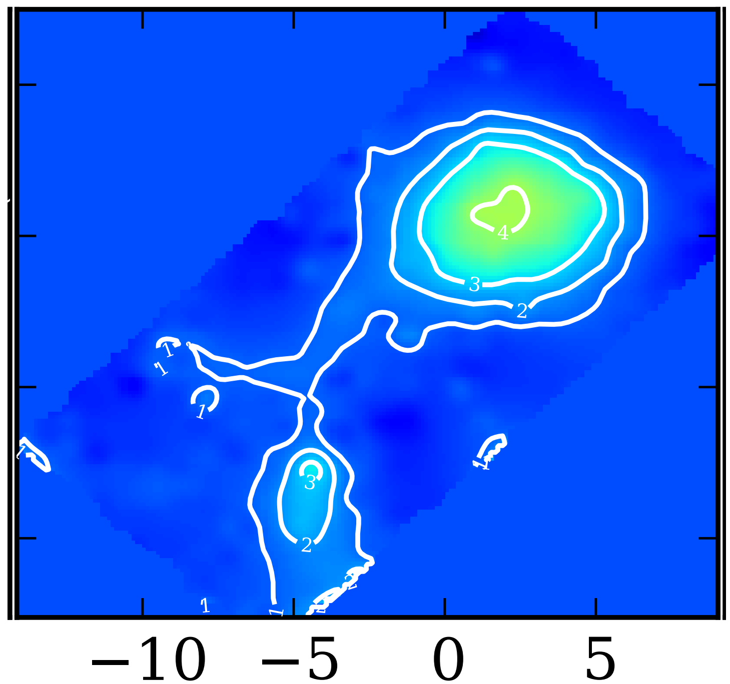


Fig. 1: Simulated weak lensing mass map for a ΛCDM cosmological model with σ8= 0.9 and Ωm = 0.3. The region shown is 1° x 1°.

Application on real data

Jullo, Pires et al., MNRAS, 2014



Reconstruction of the MACS J0717+3745 galaxy cluster field using Flens code
The filamentary structure (discovered in Jauzac et al.) is revealed

END

http://www.cosmostat.org/people/sandrine-pires/
sandrine.pires@cea.fr