Detection of compact sources in complex microwave and submillimeter backgrounds

The Biparametric Adaptive Filter

Marcos López-Caniego
&
Patricio Vielva

Instituto de Física de Cantabria
Consejo Superior de Investigaciones Científicas – Universida de Cantabria

Statistical Issues in Searches – SLAC June 2012

Outline

Why are we looking for compact sources in radio and submillimeter

- **Interesting astrophysical objects**
- **Affect other analysis needed to do cosmology**

How are we going to detect and characterize these objects

- **Filtering and Detection techniques**
  - **Matched Filter**
  - **MHWn**
  - **Biparametric Adaptive Filter**

Conclusions

Introduction

One of the most active fields is astronomy is the study of the Cosmic Microwave Background radiation (CMB).

There are satellites, balloon-borne experiments and ground based observatories dedicated just to observe this radiation that was originated when the Universe was just 380.000 years old.

The detailed analysis of this radiation, in particular, its anisotropies, allow astronomers to learn things about the origin the Universe, it's age, matter and energy content, geometry, dynamics, etc. "condensing" all this information in less than two dozen parameters, that are known with very small uncertainties.

Most interestingly, CMB scientists are able to arrive to very similar conclusions, if not the same, that other experimental approaches in cosmology (studies of the Large Structure of the Universe, Supernovae, BAO, etc) and other areas of science (particle physics, astro-chemistry,…)

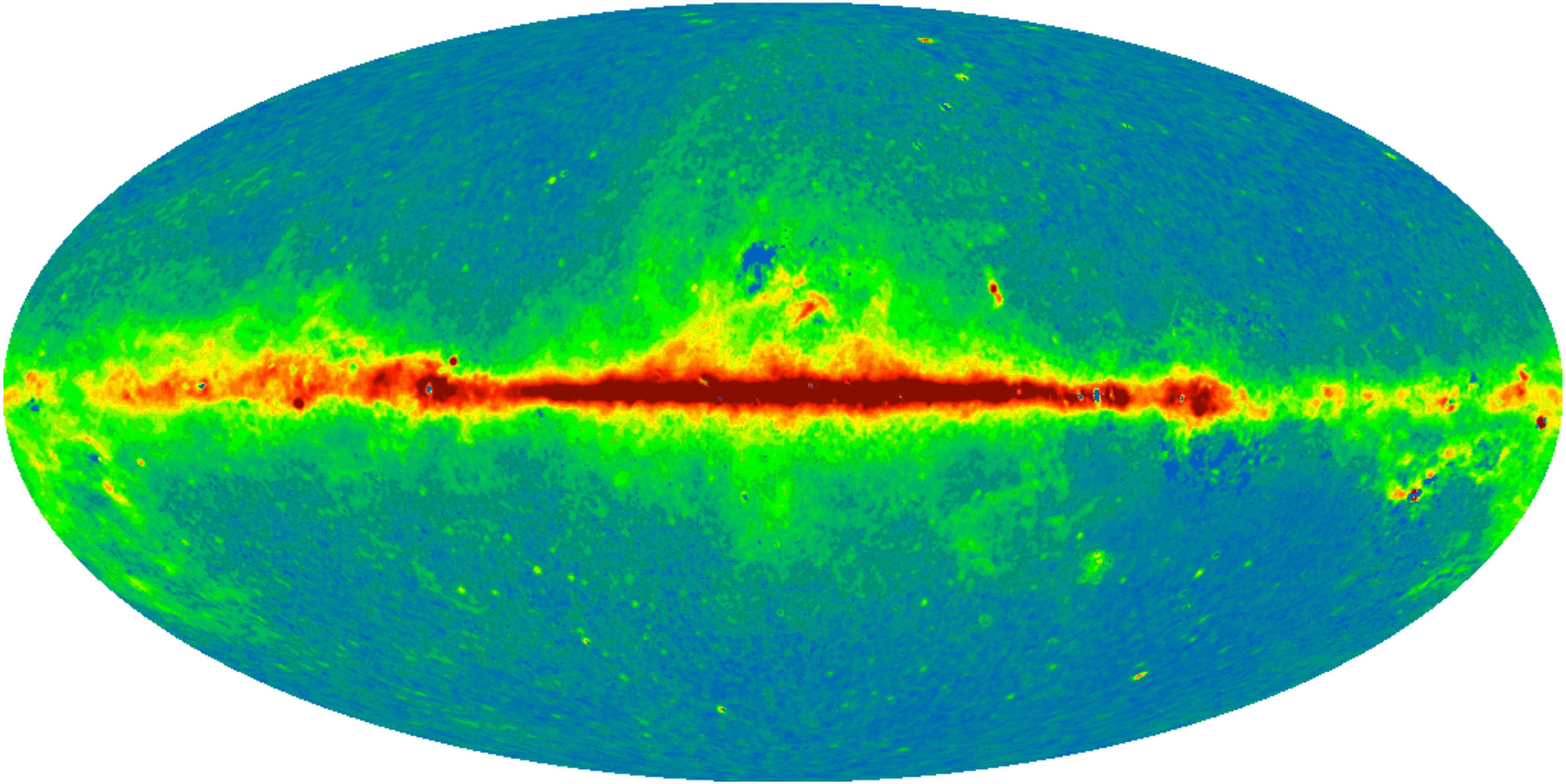
But before one can extract such a wonderful science from CMB data, scientist have to deal with very relevant problems: systematic effects and astrophysical emmisions that distort and/or contaminate the precious CMB signal.

- Systematics of the experiment: instrumental noise, knowledge of your beams (non-gaussianity and non-circularity of the detectors), scanning strategy, calibration sources, etc. The devil is in the systematics!

- From the point of view of astrophysical contributions to the CMB:
  - Galactic Compact emissions (SNr, ultra compact HII regions, etc.)
  - Galactic Diffuse emissions (synchrotron, bremstrahlung, infrared)
  - Unresolved Extragalactic emissions (distant galaxies and clusters of galaxies)
    - Radio galaxies (very bright at low freqs but invisible in the submm.)
    - Infrared galaxies (bright in the submm but invisible in the radio)
    - A mixture of them

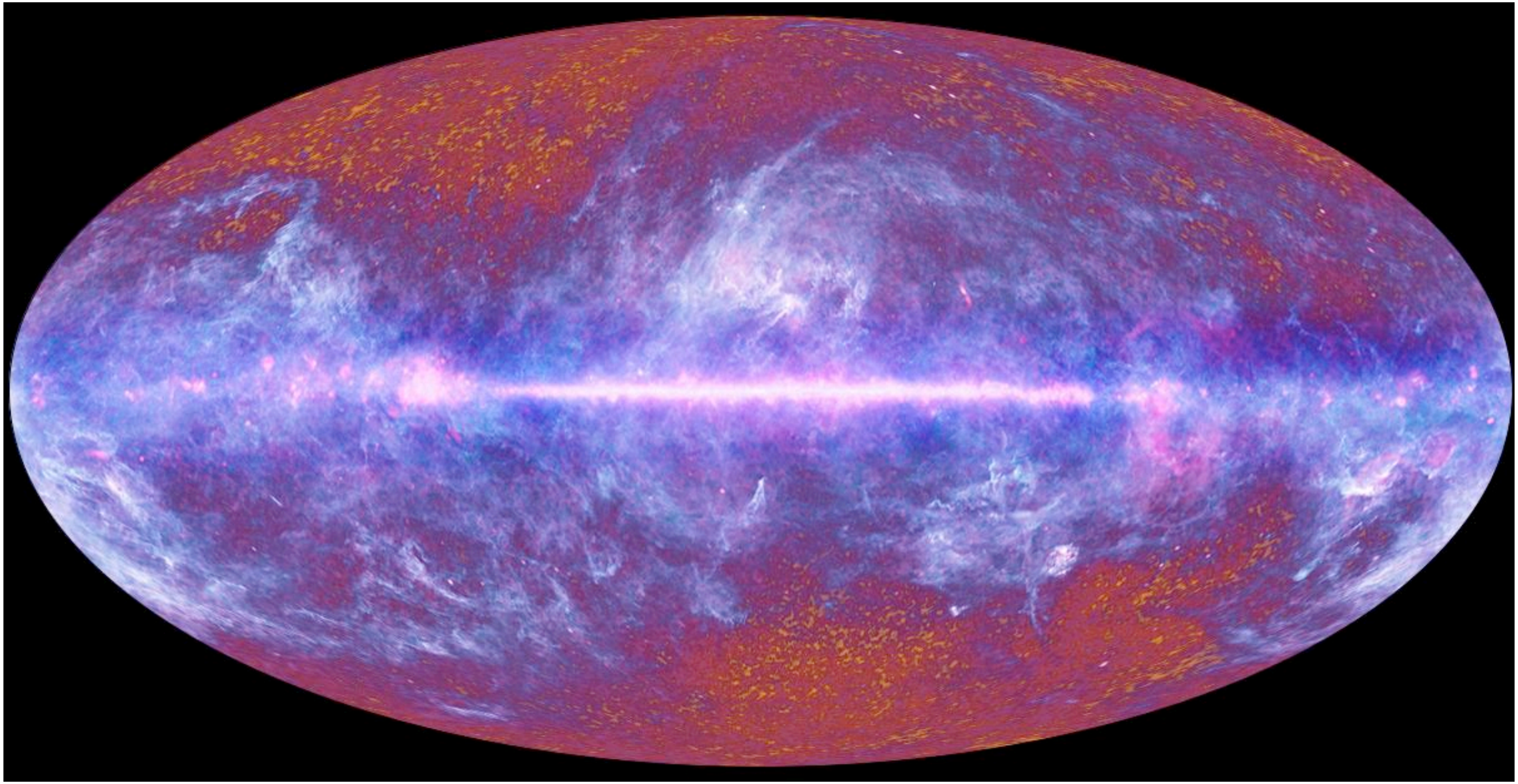
Many astronomers are not aware of the kind of backgrounds found in CMB experiments such as Planck with a frequency coverage that goes from 30 to 857 GHz: noise, cmb, bright galactic emissions, far-infrared background and unresolved radio sources and bright extragalactic sources across the whole frequency range lie on top of the CMB.

WMAP synchrotron emmision in the Galaxy at 21 GHz



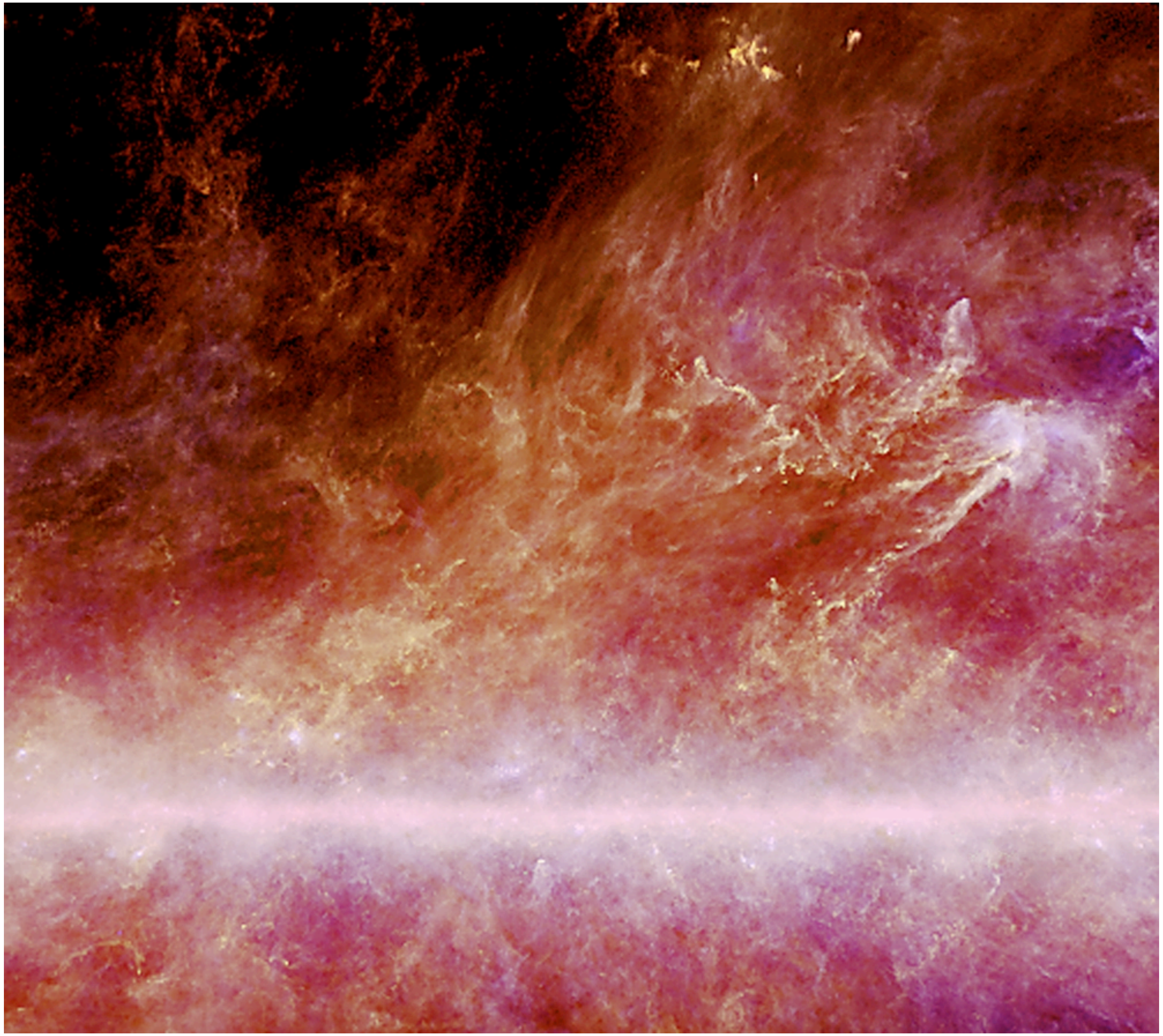
Credit: WMAP http://lambda.gsfc.nasa.gov/

Planck First all-SKy Survey Map combining different bands



Credit: ESA-Planck http://sci.esa.int/planck

Galactic dust in the vicinity of the Galactic Center as observed by Planck during the first survey



Credit: ESA-Planck http://sci.esa.int/planck

Compact source detection

One has to distinguish between the **detection** and the **characterization** of a compact sources (shape, position angle, flux density estimation,etc.)

Some methods can be optimized to detect sources and others to characterize them.

We try to develop techniques that can do both: improve the detection process in order to produce more complete and reliable catalogs (very important for number counts, follow up's, etc.) and simultaneously compute unbiased flux densities.

Our Goal: detect, characterize and separate the compact sources and diffuse astrophysical components from the CMB before doing any science with it.

In particular we focus on the separation of compact sources:

- Interesting *per se* (simultaneous observation across frequency bands)
- Largely affect the separation of the diffuse emission and other important analysis of the CMB (angular power spectrum estimation,, lensing and detectability of gravitational waves, non-Gaussianity studies, etc.), and we need to subtract or mask.

In practice… how do we do build these catalogs of compact sources?

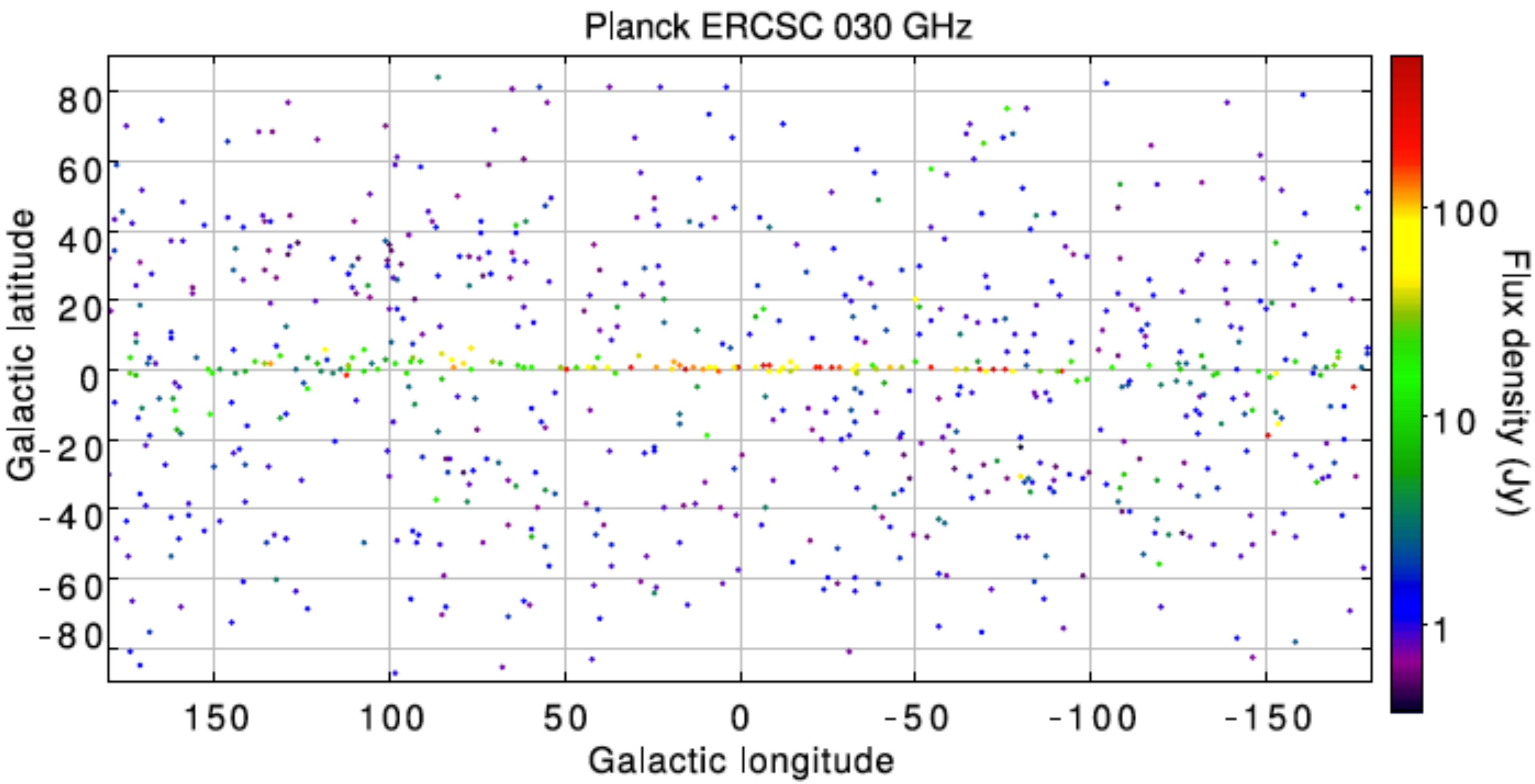
Not long ago the detection of compact sources in a WMAP map would take days to do:
produce flat patches from the all-sky maps, apply a filtering algorithm (MHW, BAF, etc.), apply the detection algorithm (threshold, Bayesian detector, etc.), obtain a subcatalog for every patch, merge together all these subcatalogs and produce a final list of detections above some threshold.

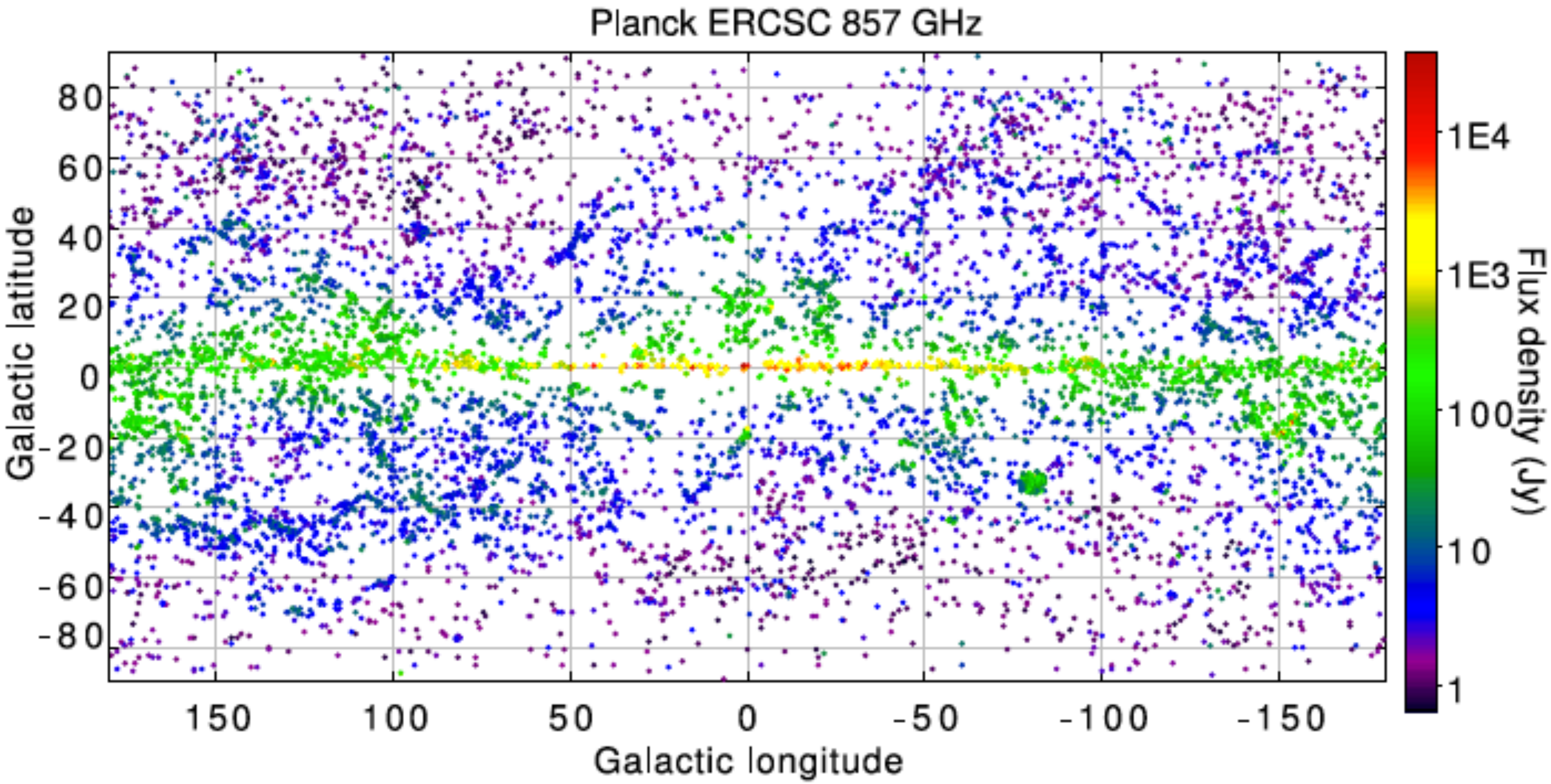
One could also try to work directly on the sphere, but detection is not always optimal.

Nowadays, the tools that we use have been optimized and this analysis can take just few minutes for WMAP and up to an hour for Planck higher resolution LFI and HFI channels.

This not only allows one to do the analysis real quick, but most important, allows to do Monte Carlo simulations to asses the errors in the photometry, in the positions, as well as predict what will be the compleness and reliability of the catalog.

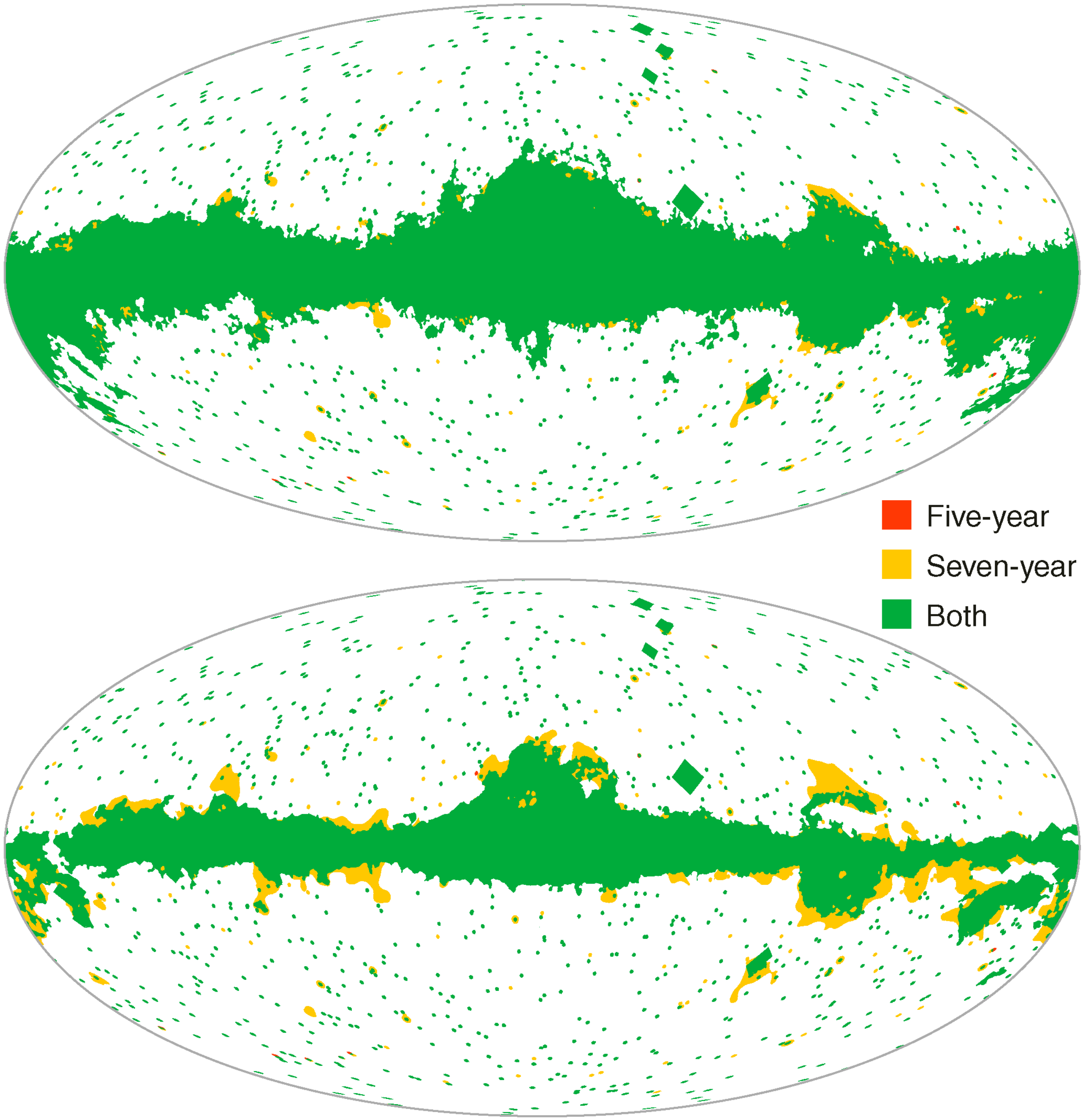
Of course, after this one still has to do an extensive validation work of the catalog, for example looking for associations in other catalogs, which can take weeks or even months if follow up observations are needed.


Planck ERCSC 030 GHz
Galactic latitude
80
60
40
20
0
-20
-40
-60
-80
150
100
50
0
-50
-100
-150
Galactic longitude
Flux density (Jy)
100
10
1



M. López-Caniego SIS 2012

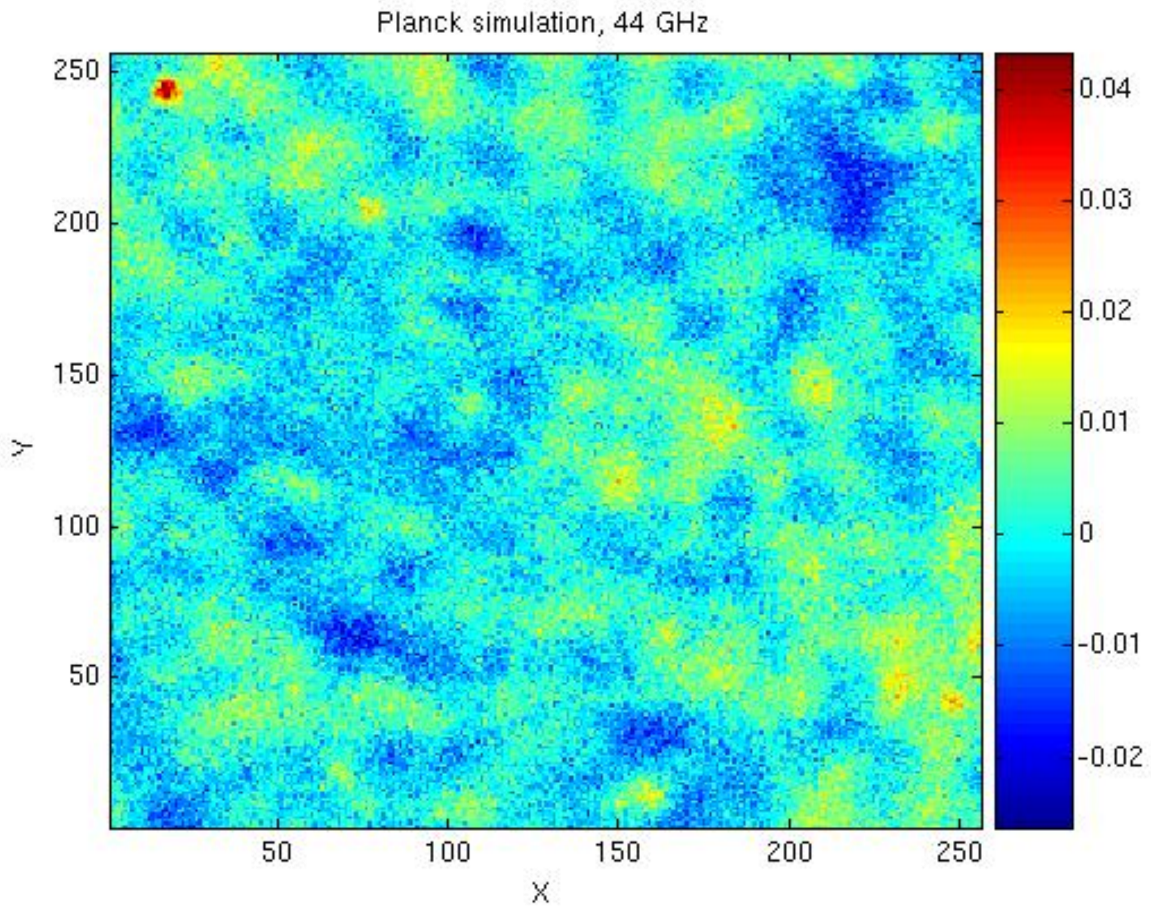
WMAP point source and galactic diffuse emmssion masks used to do science



Credit: WMAP http://lambda.gsfc.nasa.gov/

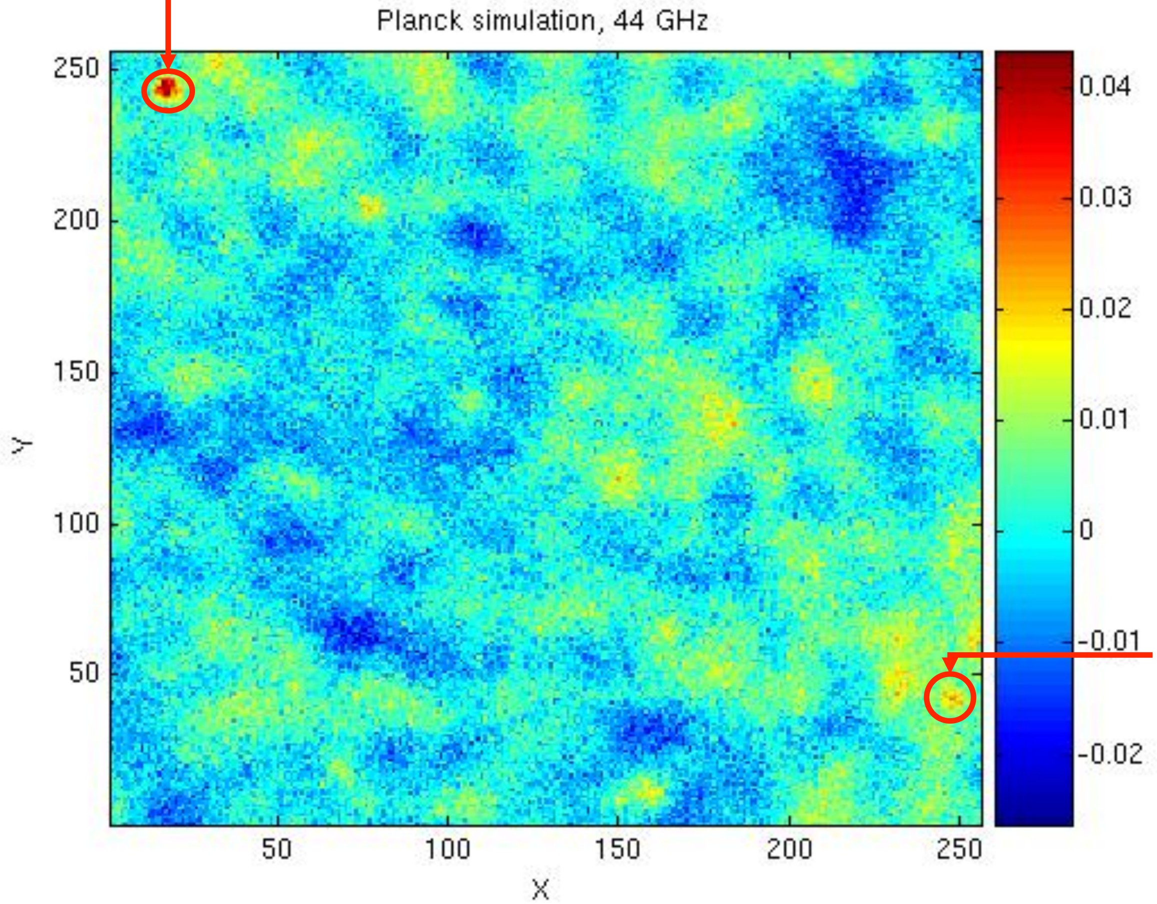
M. López-Caniego SIS 2012

Detection Techniques



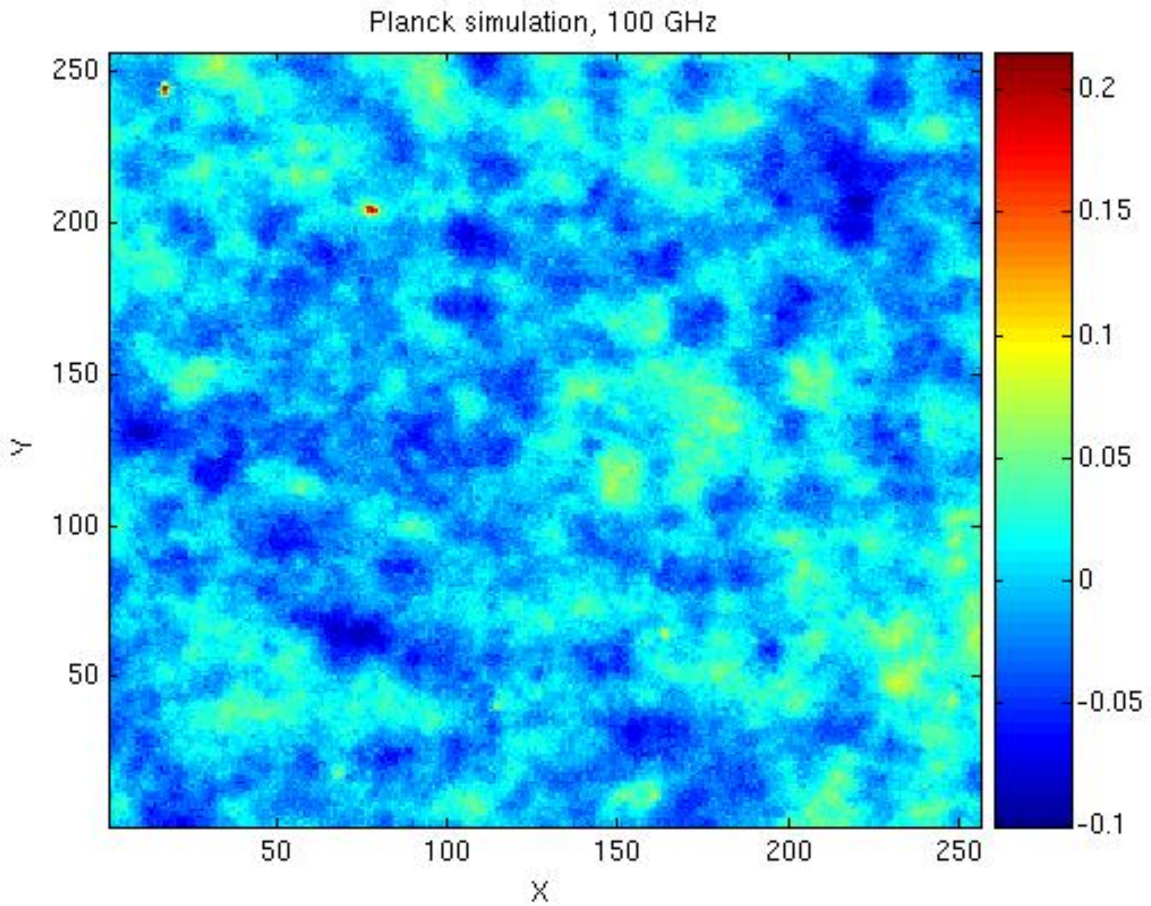
Detection Techniques

An obvious candidate



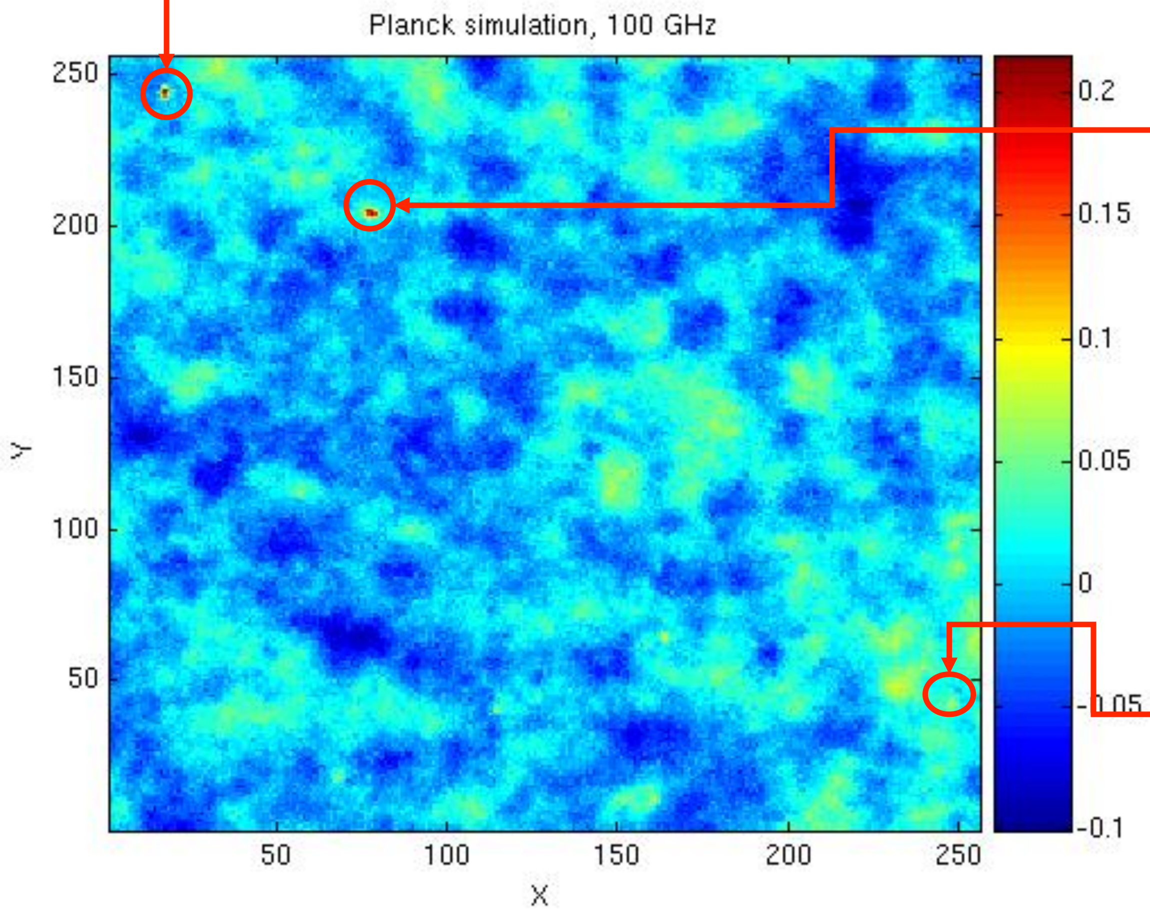
Is it a source, a peak of the CMB or a combination of CMB+Galactic emmission?

Detection Techniques



Detection Techniques

At 100 GHz the source is still there, but the flux density has decreased significantly.



New sources start to appear as we increase the freqeuncy

What happened with that blob? Was it a source with steep spectrum that is not visible anymore or was it a peak of the background

Detection Techniques

The detector can be refered to as the criteria used to decide whether or not a signal belongs to a source or it is just a maximum of the background.

The most common detector used in astronomy is the "sigma threshold" (e.g., how many objects have a signal-to-noise ratio greater than something, typically 5).

There are other detectors: one can define a hypothesis test to try to find the threshold that, for example, maximizes the number of detections for a given percentage of false detections.

In order to do this one has to assume some properties of the background (e.g., Rice's 1954 expected number of maxima in a Gaussian random field in the presence and in the absence of real sources with a known profile). Then one can introduce things as the curvature to "add" information to the usal threshold.

There are methods that use priors about the distribution of sources, templates of diffuse emmission or even maps of the CMB, to improve their decision rule. These methods can be more sensitive to assumed properties of the background than other methods.

There are also multifrequency detection/filtering techniques that work really well.

But beware: for historical reasons it is common to hear people speaking of of Nσ detections, when the background is non-Gaussian!!!

Filtering Techniques

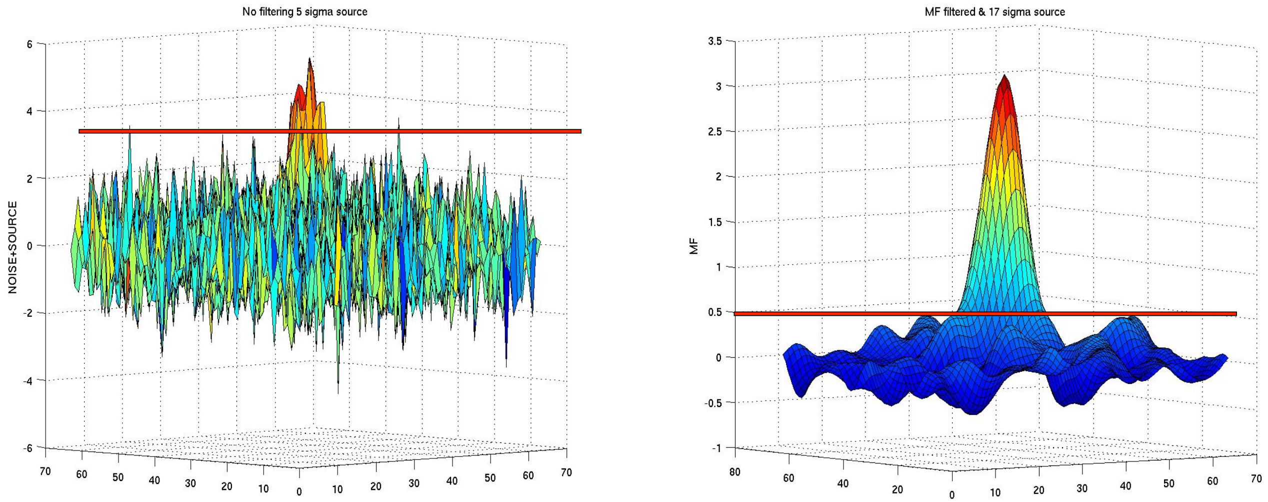
Why would one want to modify the properties of your data in such a dramatic way?

To Filter or not Filter

Sometimes it is not necessary to filter the maps, however

3σ before filtering

17σ **after filtering**



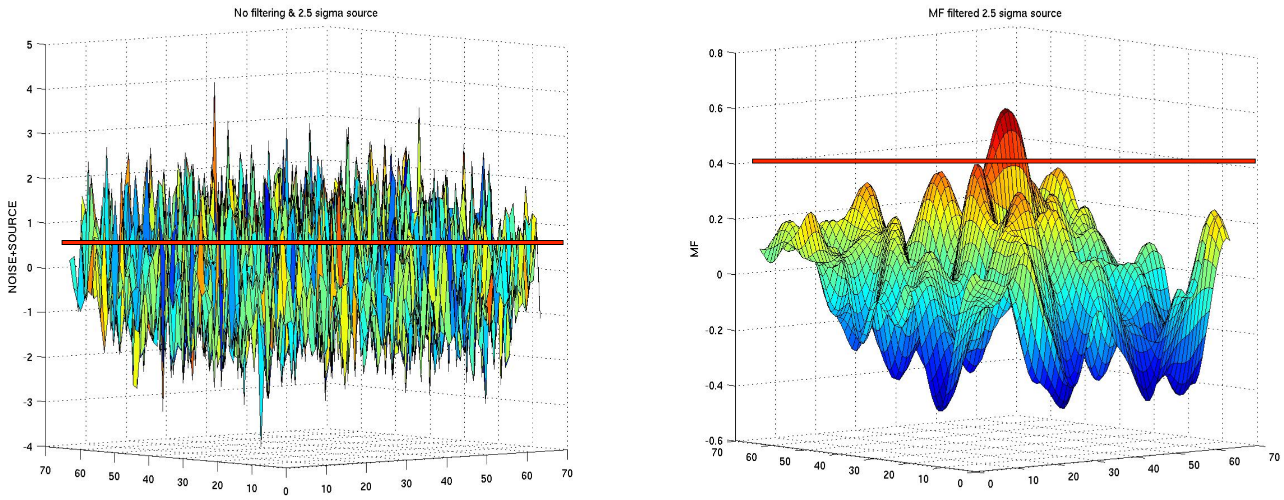
What if the sources are very weak?

$\sigma_{sf} = 1$

0.5σ before filtering

$\sigma_f = 0.17$

3σ after filtering



Filtering Techniques

Data Model $$y(\vec{x}) = s(x) + n(\vec{x})$$

The observations:

Image (on the sphere or in flat patch) obtained for a given frequency

The signal (the source):

$s(x)=A\tau(x)$

$\tau(x)$:instrument response to a point source (PSF)

A is the amplitude or intensity

The noise:

Everything else!

It is common to assume that the noise is homogeneous and isotropic and characterized by a Power Spectrum

$P(q) = \langle n(q)n^*(q)\rangle$

$\Psi(x)$ is a linear filter such that when convolved with the observations (***y***), the filtered field at the position ***b*** is given by:

$$\boxed{w(\vec{b}) = \int d\vec{x}\, y(\vec{x})\Psi(\vec{x}-\vec{b}) = \int d\vec{q}\, y(\vec{q})\psi(q)e^{-i\vec{q}\vec{b}}}$$

Filtering Techniques

There are many filtering techniques used to detect compact sources:

One can try to clasify them in terms of flat/sphere

- Most of them work with flat patches of the sky
  - MF’s, MHW’s (IFCAMEX), BAF, IFCAPOL, MMF, PowellSnakes, Sextractor, etc.
- Some can operate directly on the sphere
  - SMHW, MF, possibly others

One can also try to to clasify them in terms of single/multi frequency.

- Most of them do single frequency detections:
  - MF’s, MHW’s (IFCAMEX), BAF, IFCAPOL, PowellSnakes, Sextractor, etc.
- Some do multifrequency
  - Matrix Filters, MMF (both SZ and sources), Powellsnakes, MADX,etc.

The Matched Filter

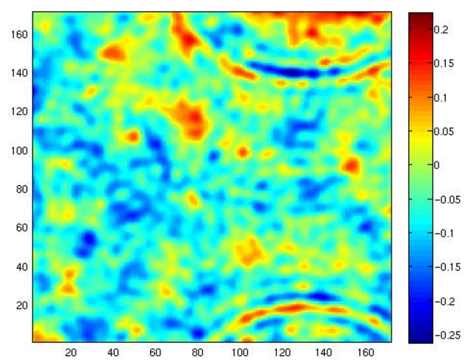
It is obtained imposing the following conditions:

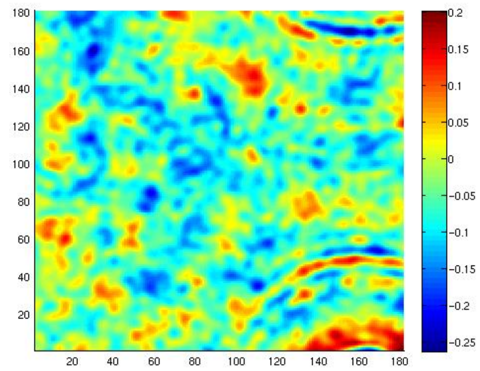
i) The filtered field, at the position of the source, is an unbiased measure of its amplitude
ii) The variance of the filtered field is minimum (maximum efficiency)

$$\psi_{MF}(q) = \frac{1}{a}\frac{\tau(q)}{P(q)} \qquad a = 2\pi\int_0^\infty dq\,q\,\frac{\tau^2(q)}{P(q)}$$

Requirements: the profile of the source is known and the noise is known or can be estimated from the data.

Known problems: when the noise is neither homogeneous nor isotropic. Is very sensitive to the determination of the power spectrum (e.g., in the presence of bright galactic structures or crowded fields things complicate).





However, some of these problems can be reduced with a careful implementation of the algorithm.

Wavelets (used as filtering kernels)

The wavelet transform is giving us information about the scale and location of interesting structures present in the maps (e.g., point sources)

$$w(R,\vec{b}) = \int d\vec{x}\, y(\vec{x})\Psi(R;\vec{b},\vec{x})$$

$\int d\vec{x}\psi = 0,$ Mean=0

$\int d\vec{x}\psi^2 = 1,$ Square Norm = 1

$C_\psi = (2\pi)^2 \int dq q^{-1} \hat{\psi}^2(q) < \infty,$ Square integrable

$$\Psi(R;\vec{b}) = \frac{1}{R^2}\psi\left(\frac{|\vec{x}-\vec{b}|}{R}\right)$$

traslation

scale

Wavelet coefficient at the position b and with scale R

"mother wavelet"

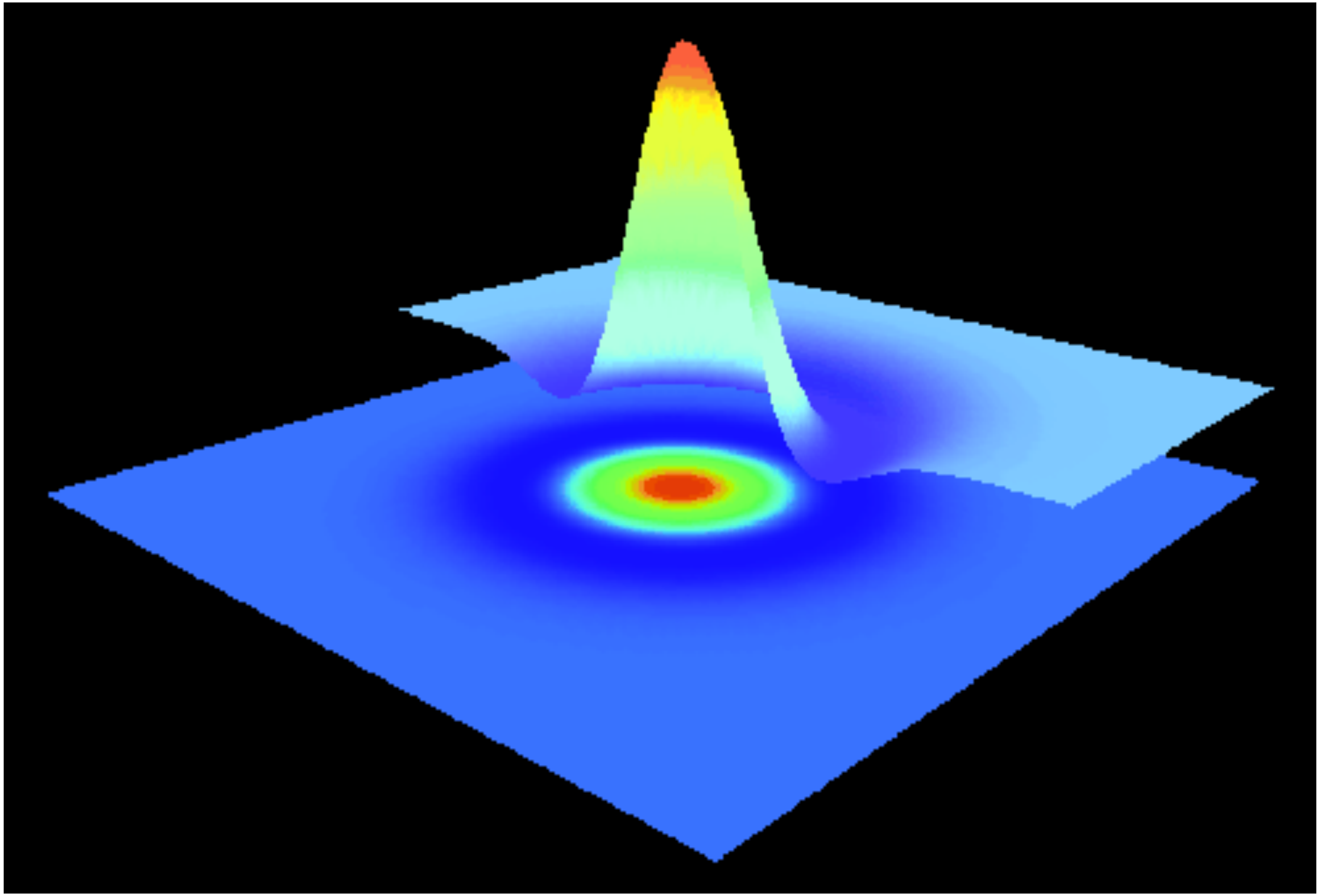
The Mexican Hat Wavelet Family: set of isotropic filters that can be easily obtained iteratively applying the Laplacian operator over the 2D Gaussian function.

$$\psi_n(x) = \Delta^n g(x), \quad n = 1,2,\ldots$$

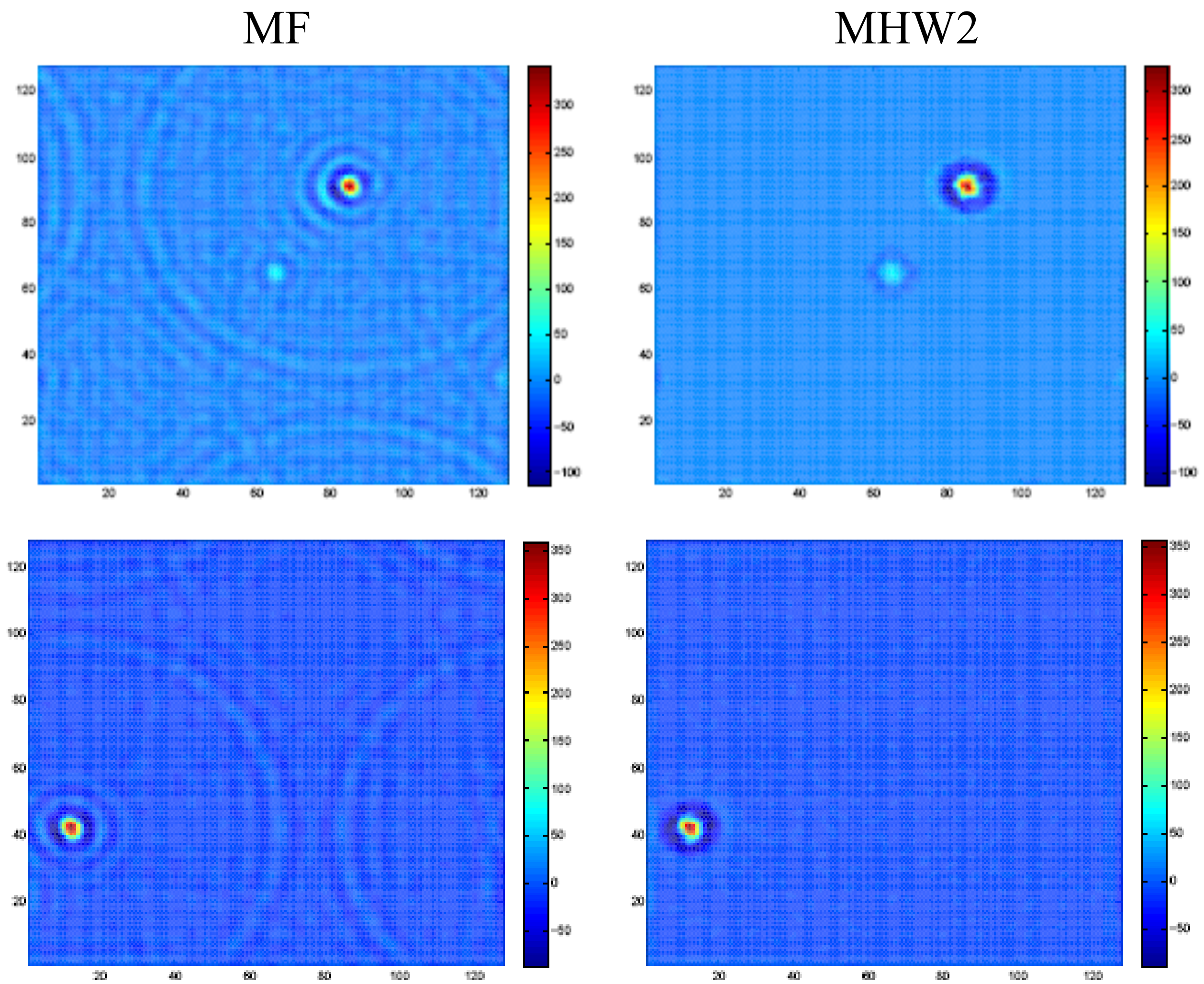
$$g(x) = \frac{1}{2\pi}\exp\left(-x^2/2\right)$$

$$\boxed{\psi_n(qR) = \frac{(qR)^{2n}}{2^n n!} e^{-\frac{(qR)^2}{2}}}$$

- Adapts very well to gaussian-like PSFs
- Fixing n fixes the shape of the filter
- The scale R can be easily optimized to maximize the amplitude of the sources
- Quick and easy implementations
- Has been used with WMAP, Herschel and Planck data (González-Nuevo et al. 2006, LC et al. 2006, Massardi et al. 2009, etc..)



Matched Filter vs MHW2 in the presence of a very bright source



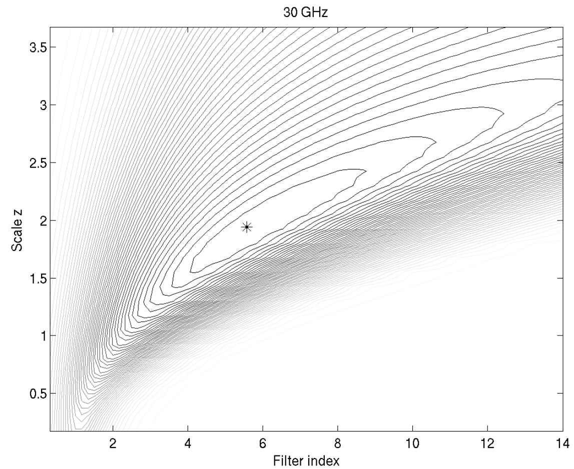
Some of the problems found with the MF can be solved doing a proper estimation of the Power spectrum of the image (e.g., masking or substracting the bright source, apodization, etc.)

Biparametric Adaptive Filter (LC & Vielva, 2012, MNRAS, 421,2139)

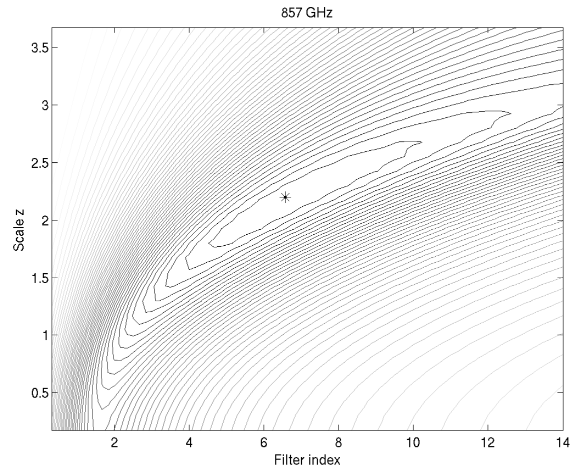
It is a generalization of the idea behind the MHW family, but:

- Optimizing the scale of the filter
- Optimizing the index g (related to power spectrum of the local background)
- Does not impose a profile for the source allowing to introduce non-gaussian PSF's.

Amplification



Amplification



The optimal values of R and g fix the shape.

By construction, the Amplification of the BAF is >= MHW, MHW2, etc.

The goal is to adapt the filter to the background in the vicinity of the source

$$\boxed{\hat{\psi}_g\left(qR\right) = \frac{1}{\pi}\frac{1}{\Gamma\left(\frac{2+g}{2}\right)}\left(qR\right)^g\,\tau(qR)}$$

Given a profile (e.g., 2D Gaussian)

$$\hat{\psi}_g\left(qR\right) = \frac{1}{\pi}\frac{1}{\Gamma\left(\frac{2+g}{2}\right)}\left(qR\right)^g\,e^{-\frac{1}{2}(qR)^2}$$

$$\omega_g\left(R\right) = \frac{I_0 2^{\frac{g+2}{2}} z^g}{\left(1+z^2\right)^{\frac{g+2}{2}}},$$

$$\sigma_\omega^2\left(R,g\right) \equiv 2\pi\int_0^\infty dq\;qP\left(q\right)\hat{\psi}_g^2(Rq)$$

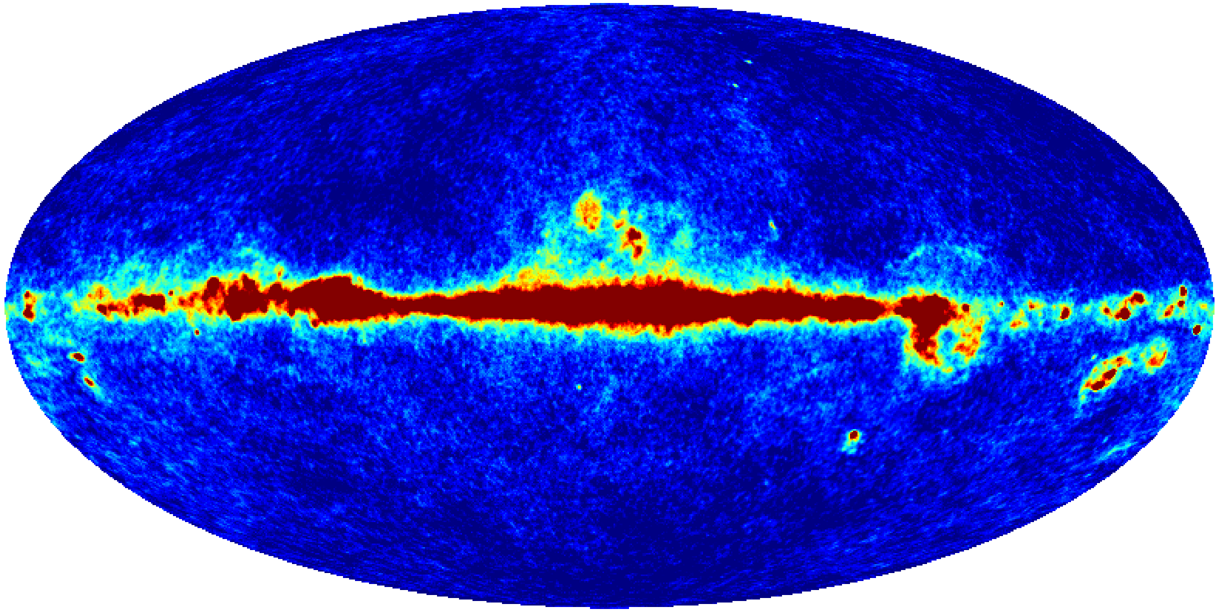
The index g can take any value >=0

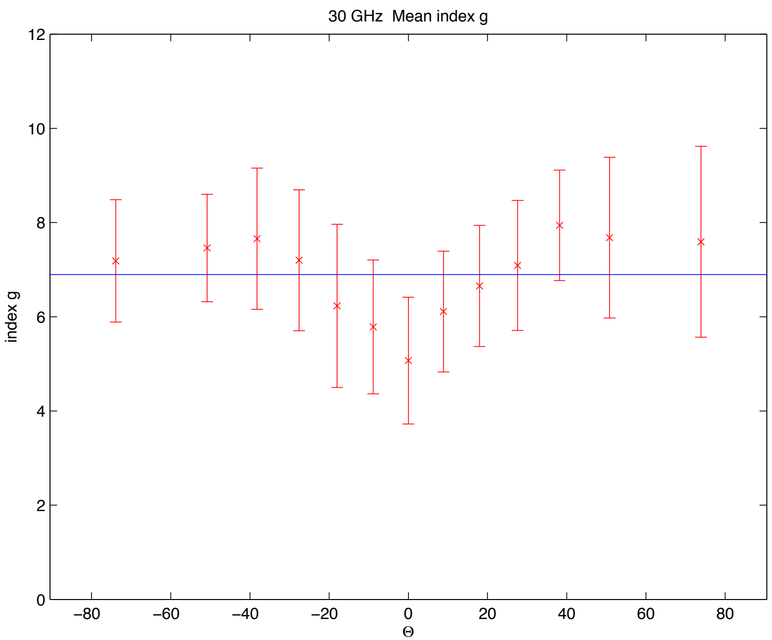
This filter defaults to:

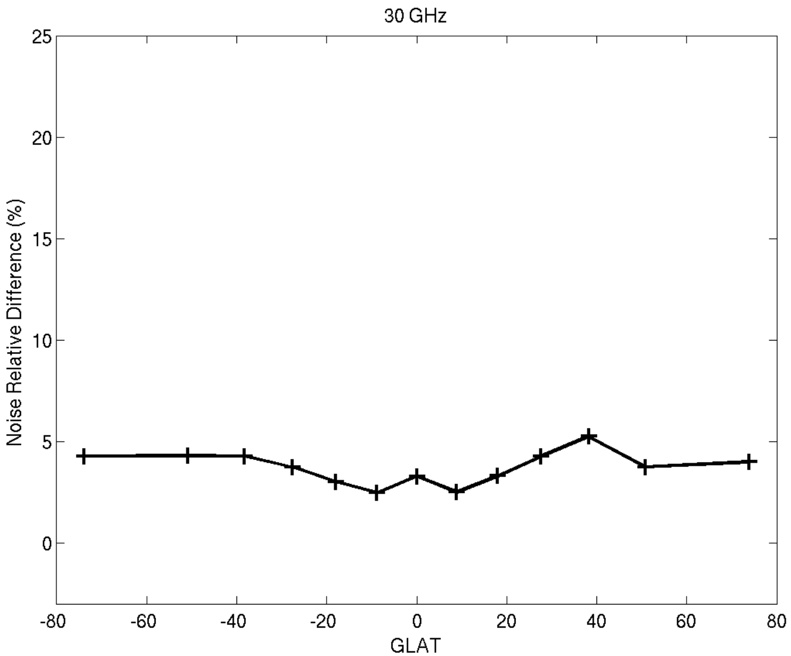
Gaussian	g=0
MHW	g=2
MHW2	g=4

Planck Simulation of CMB, noise, compact sources and diffuse Galactic emission

30 GHz







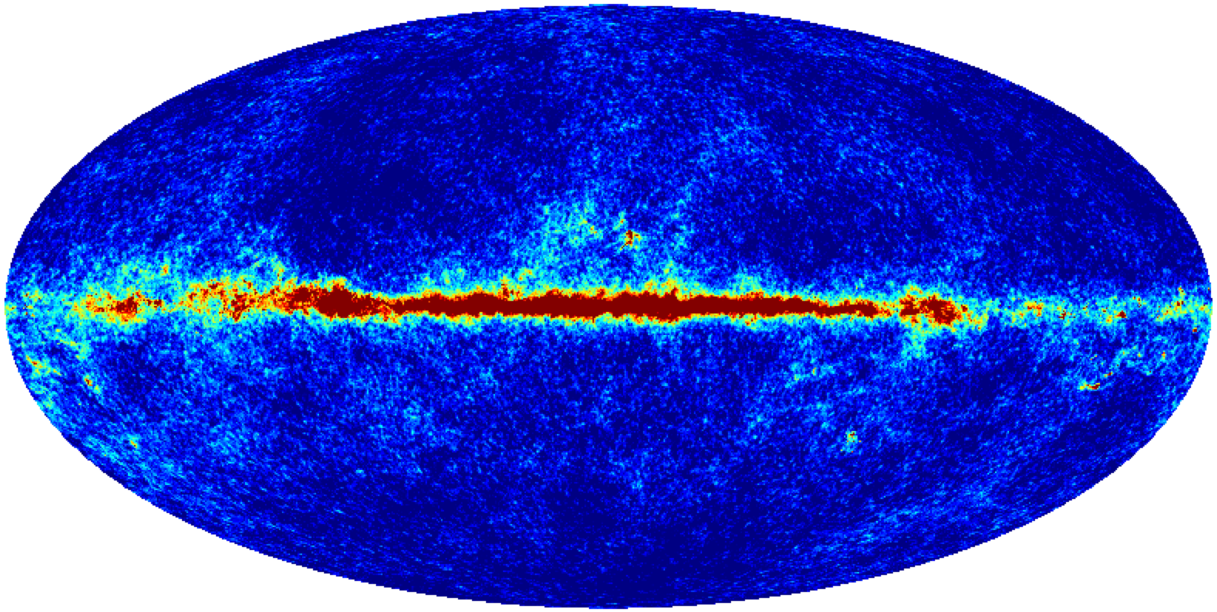
(LC & Vielva, 2012, MNRAS, 421,2139)

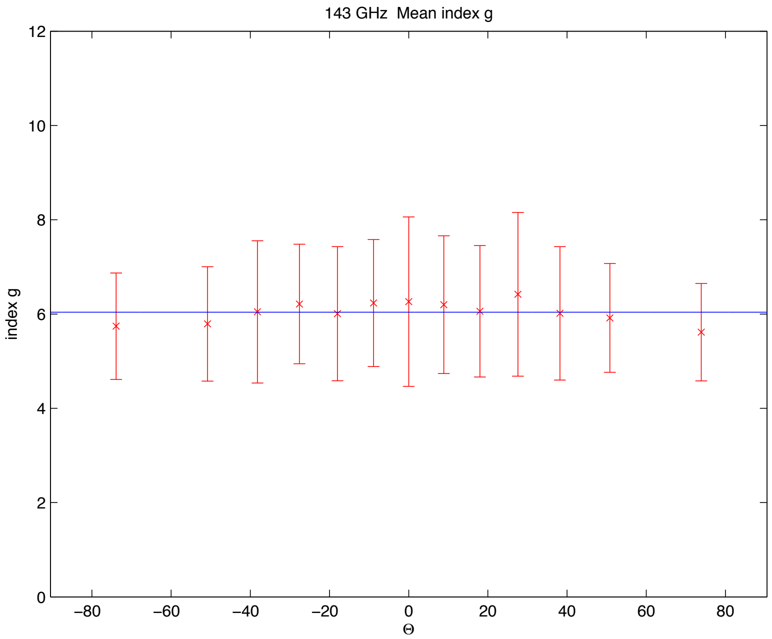
R and g have been optimized for each of the 1500 flat patches

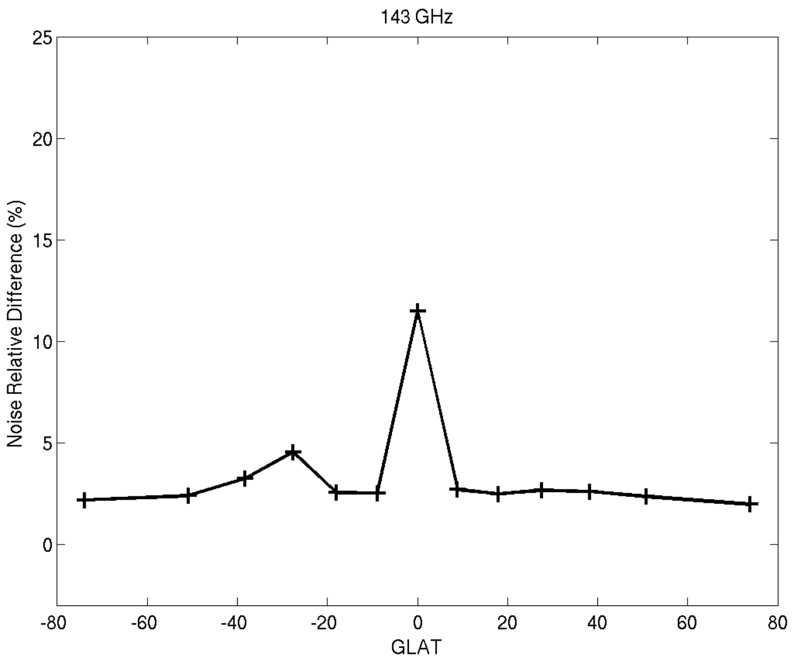
M. López-Caniego SIS 2012

Planck Simulation of CMB, noise, compact sources and diffuse Galactic emission

143 GHz







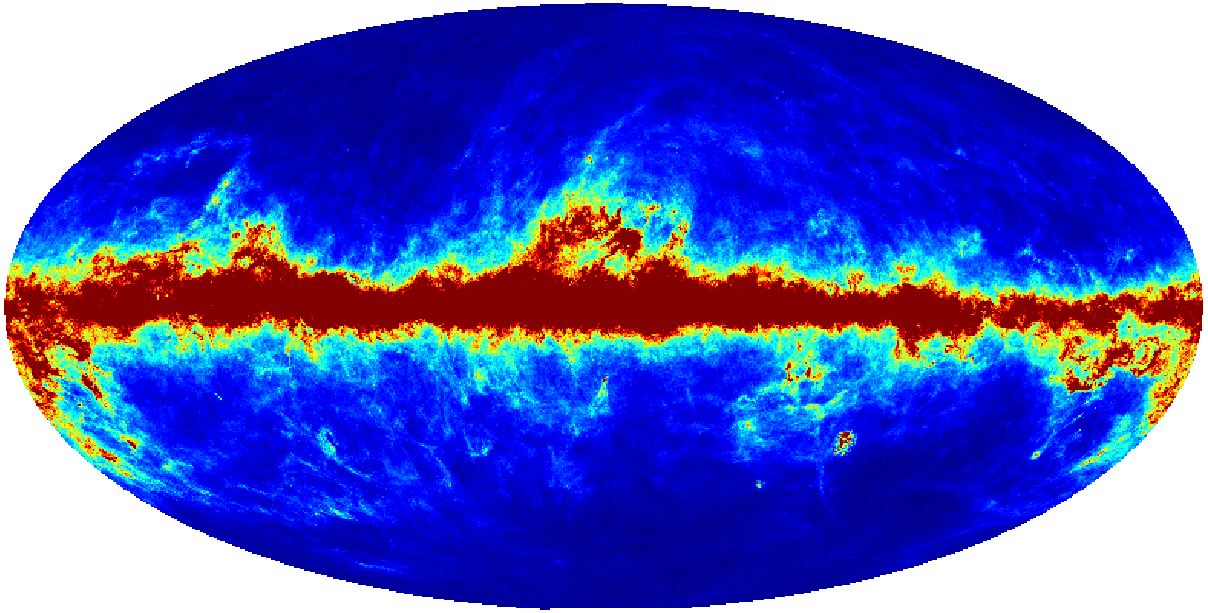
(LC & Vielva, 2012, MNRAS, 421,2139)

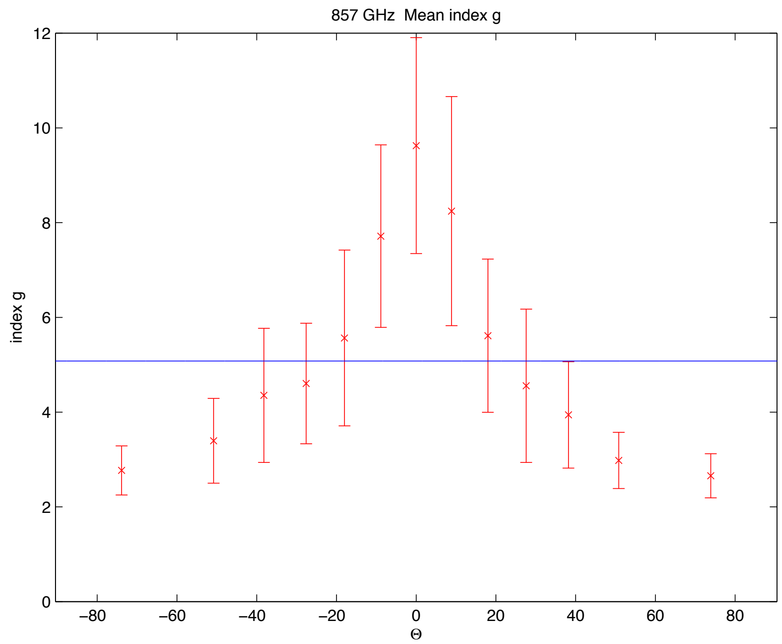
R and g have been optimized for each of the 1500 flat patches

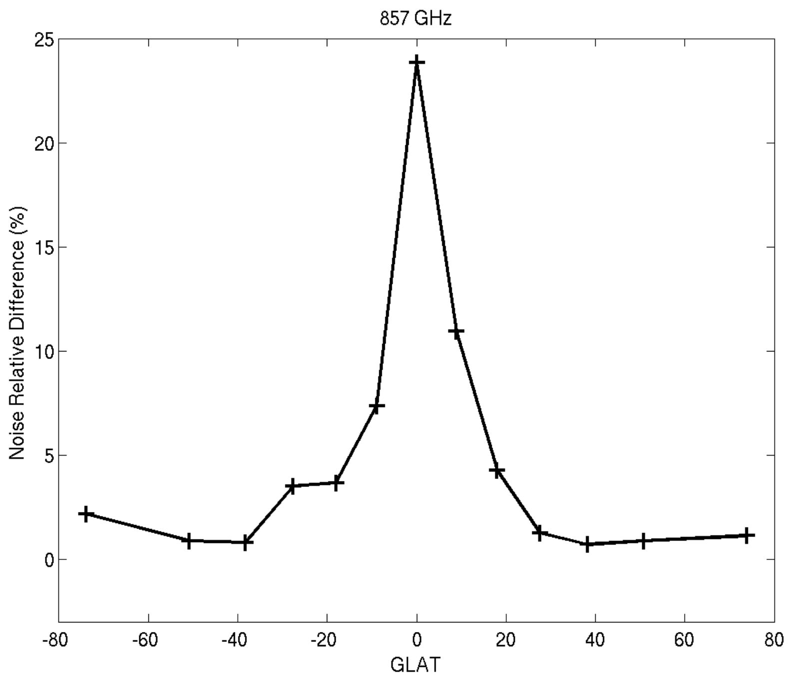
M. López-Caniego SIS 2012

Planck Simulation of CMB, noise, compact sources and diffuse Galactic emission

857 GHz







(LC & Vielva, 2012, MNRAS, 421,2139)

R and g have been optimized for each of the 1500 flat patches

M. López-Caniego SIS 2012

Conclusions

- The detection of compact sources in CMB experiments requires special tools that can deal with complex backgrounds that combine instrumental noise, CMB, compact and diffuse emission from the Galaxy and/or radio and far-infrared backgrounds.

- Several techniques have been proposed (MF, MHWn, PowellSnakes, IFCAPOL,Sextractor, BAF).

- The BAF tries to incorporate the best of these worlds in a filter that adapts to the local properties of the background in the vicinity of a source, improving the detection process.

- In particular, and by construction, this filter will be always equal or better than the MHWn, which is already a very robust tool used for WMAP and Planck analysis.

- Photometry, both tools work equally well, specially in the Galactic plane where they provide unbiased estimation of the flux density of compact sources with less than a few percent errors.

- The MHW2 is a good compromise for the different types of backgrounds that one can find in an experiment like Planck (30 to 857 GHz).

- The BAF is able to detect more real sources than the MHW2 in cleaner regions of the sky and, at the same time, is more reliable in complex regions, in particular at low Galactic latitudes where many other methods tend to find many spurious deections.

The standard cosmological model also predicts that the CMB radiation is linearly polarized. The polarization signal and its cross-correlation with the temperature anisotropies constitute an important consistency check and help in breaking the degeneracies among some cosmological parameters.

A net value of the Stokes parameters Q and U is expected from Thomson scattering during decoupling of photons and baryons. Since Q and U are not invariant quantities on the sphere it is convenient to transform them in a gradient field called "E-mode" and a rotational field called "B-mode".

The most important property of the E and B-mode decomposition is that from their measurement we can distinguish between primordial scalar perturbations (density) and primordial tensor perturbations (gravitational waves).

More specifically, both types of perturbations can generate E-mode polarization, however only gravitational waves can produce B-mode polarization. It is this property that makes polarization a key tool for the detection of the primordial gravitational wave background (GWB) which is expected to be generated during the inflationary period of the universe.

Moreover, a detection of the B-mode would directly provide the energy scale of inflation (as measured by the ratio *r* of tensor to scalar perturbations, see Fig. 0).

Impacto de las fuentes en el modo-B