Compact source extraction

R. Belén Barreiro (IFCA, Santander)

Outline of the talk

- Motivation
- Techniques for point source extraction
- Extraction of the thermal SZ effect
- Extraction of the kinematic SZ effect
- Extraction of statistical information from undetected sources
- Final remarks

The microwave sky

CMB + contaminants + noise

Contaminants

- Galactic emission (synchrotron, free-free, dust)
- Extragalactic point sources
- Thermal and kinematic Sunyaev-Zeldovich effects

Component separation techniques

- Global separation (MEM, WF, ICA, MDMC-SMICA)
  ⇨ see talk by J.-F. Cardoso
- One component extraction techniques

Extragalactic point sources

Two main populations

Radio sources (below ~ 200 GHz)

Toffolatti et al. 1998, Tucci et al. 2004

Infrared sources (above ~ 200 GHz)

Guiderdoni et al. 1998, Granato et al. 2001

- Point-like objects ⇨ beam profile
- Different frequency dependence for each source ⇨ not suited for global separation techniques
- Great uncertainty at microwave frequencies

Sunyaev-Zeldovich effect

Thermal effect

- Inverse Compton scattering of CMB photons by hot electrons in the intra-cluster medium
- Distinct spectral signature

Kinematic effect

- Due to the radial peculiar velocity of the cluster
- ~1 order of magnitude lower than the tSZ
- Same spectral dependence as the CMB

Imprint on the CMB at scales of ~ a few arcmin ⇨ beam + cluster profile

Cosmological probe ⇨ H_0, Ω_m, Ω_Λ, σ_8

Extraction of point sources

Linear filtering techniques

A filter amplifies the signal with respect to the background

The Matched Filter

- It gives maximum amplification of the source

$$\psi(q) \propto \frac{\tau(q)}{P(q)}$$

$\tau(q)$: source profile
$P(q)$: power spectrum of the background

- Estimation of p.s. catalogue for Planck from ~650 (@ 70 GHz) to ~38000 (@ 857 GHz) in ~2/3 of the sky (Tegmark & de Oliveira-Costa 1998)

The Scale-adaptive filter (Sanz et al. 2001)

- It introduces an additional constraint: a maximum at the scale of the source

$$\psi(q) \propto \frac{\tau(q)}{P(q)}\left[b + c - (a+b)\frac{d\ln\tau}{d\ln q}\right]$$

where a,b,c = f(P,τ)

- It has been applied to Planck simulated TOD (Herranz et al. 2002a)

The Adaptive Top Hat filter (Chiang et al. 2002)

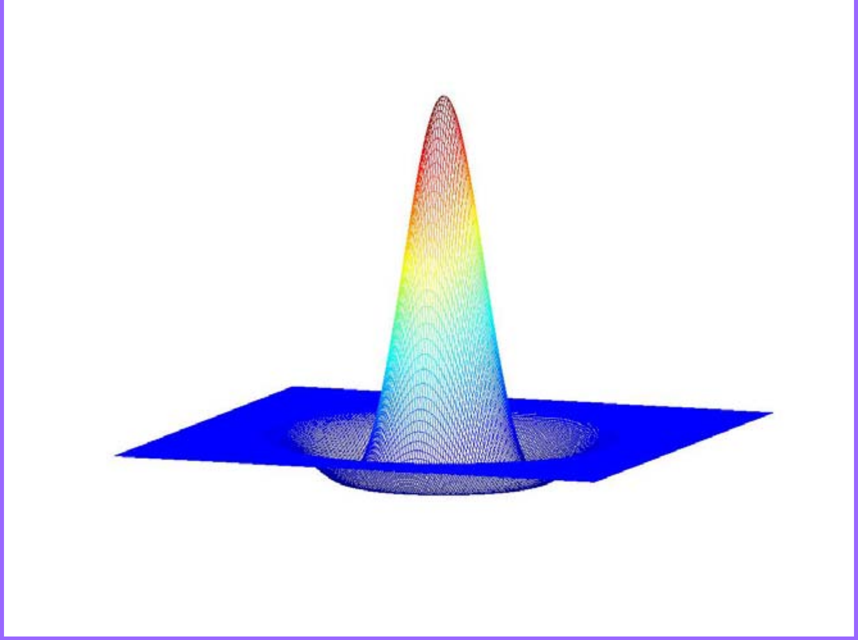
- Top Hat Filter in harmonic space
- The limits l_{min} and l_{max} are obtained from simulations and could be estimated iteratively from the data

The Mexican Hat Wavelet

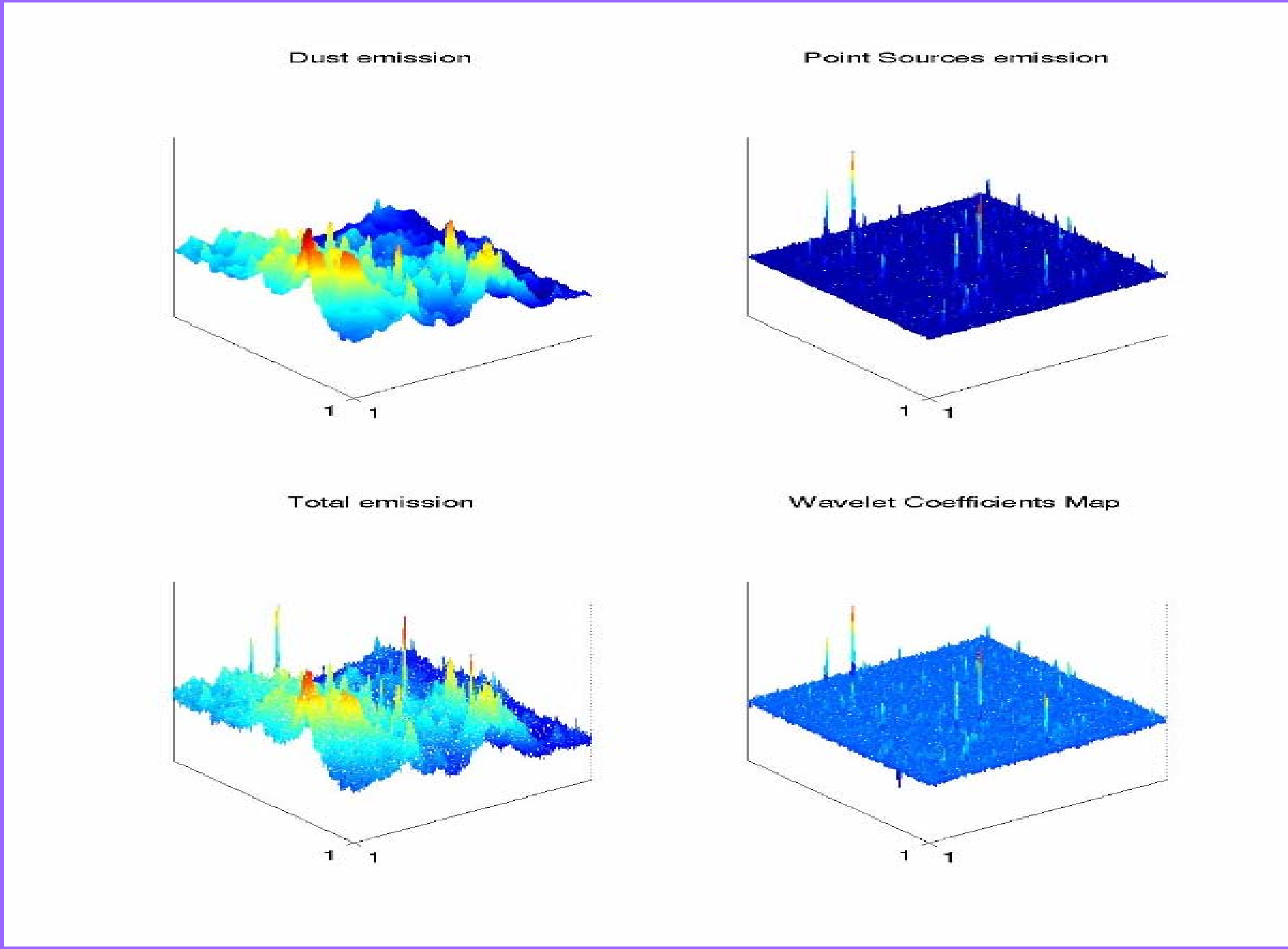
(Cayón et al. 2000, Vielva 2001a,2003)

$$\psi(x) = \frac{1}{\sqrt{2\pi}}\left[2 - \left(\frac{x}{R}\right)^2\right] e^{-\frac{x^2}{2R^2}}$$

$$\hat{\psi}(k) \propto (kR)^2 e^{-\frac{1}{2}(kR)^2}$$



• Point sources are amplified in wavelet space



Results for Planck simulated data with the SMHW

Frequency (GHz)	#	Min Flux (Jy)	$\bar{E}(\%)$	$\bar{b}(\%)$	Galactic Cut(deg)	N_{R_o}	Completeness (%)
857	27257	0.48	17.7	-4.4	25	17	70
545	5201	0.49	18.7	4.0	15	15	75
353	4195	0.18	17.7	1.4	10	10	70
217	2935	0.12	17.0	-2.5	7.5	4	80
143	3444	0.13	17.5	-4.3	2.5	2	90
100(HFI)	3342	0.16	16.3	-7.0	0	4	85
100(LFI)	2728	0.19	17.0	-2.4	0	4	80
70	2172	0.24	17.1	-6.7	0	6	80
44	1987	0.25	16.4	-6.4	0	9	85
30	2907	0.21	18.7	1.2	0	7	85

- Using the recovered catalogue, mean spectral indices can be estimated with good accuracy
- The performance of the SMHW has been tested for realistic asymmetric beams (80 per cent of the detections are found in the worst Planck channel, 30 GHz)
- A joint MHW & MEM analysis has been performed on small patches of Planck simulated data, improving the results of each method on its own (Vielva et al. 2001b) ⇨ work in progress to extend it to the spherical case
- Work in progress ⇨ clustering of sources and polarization

The detector

- Filtering amplifies the signal over the background
- Whether we filter or not a detection criterion – the detector - is needed to decide if a given signal belongs to the background or to a true source
- The most common detector is thresholding, that uses only the intensity of the signal ($\nu > \nu_*$)
- More complex choices are possible, which include extra information about the field (intensity, eccentricity, curvature, multiscale information, multifrequency information…)
- The choice of the detector is a key issue for the performance of the filter

The Neyman-Pearson detector

$$L(\nu,\kappa) = \frac{n(\nu,\kappa)}{n_b(\nu,\kappa)} \geq L_*$$

ν: threshold, κ: curvature, L_*: constant

$n_b(\nu,\kappa)$: number density of maxima for background
$n(\nu,\kappa)$: number density of maxima for background + source

For a (1D) Gaussian background this is equivalent to (Barreiro et al. 2003, López-Caniego et al. 2004)

$$\varphi(\nu,\kappa) \equiv \frac{1-\rho y_s}{1-\rho^2}\nu + \frac{y_s-\rho}{1-\rho^2}\kappa \geq \varphi_*$$

ρ, y_s determined by background, filter and source profile
For a fixed n.d. of spurious sources (i.e. for a fixed φ_*), it gives the maximum n.d of detections ⇨filter comparison

The biparametric scale adaptive filter

López-Caniego et al., 2004 (see next talk)

$$\psi(q) \propto \frac{\tau(q)}{P(q)}\left[1 + c(qR)^2\right]$$

• c is a free parameter determined (numerically) by maximizing the number of detections

• the filter scale is allowed to vary

• Using a Neyman-Pearson criterion, the BSAF obtains up to ~40 per cent more detections than the MF in certain cases

Bayesian approach to discrete object detection

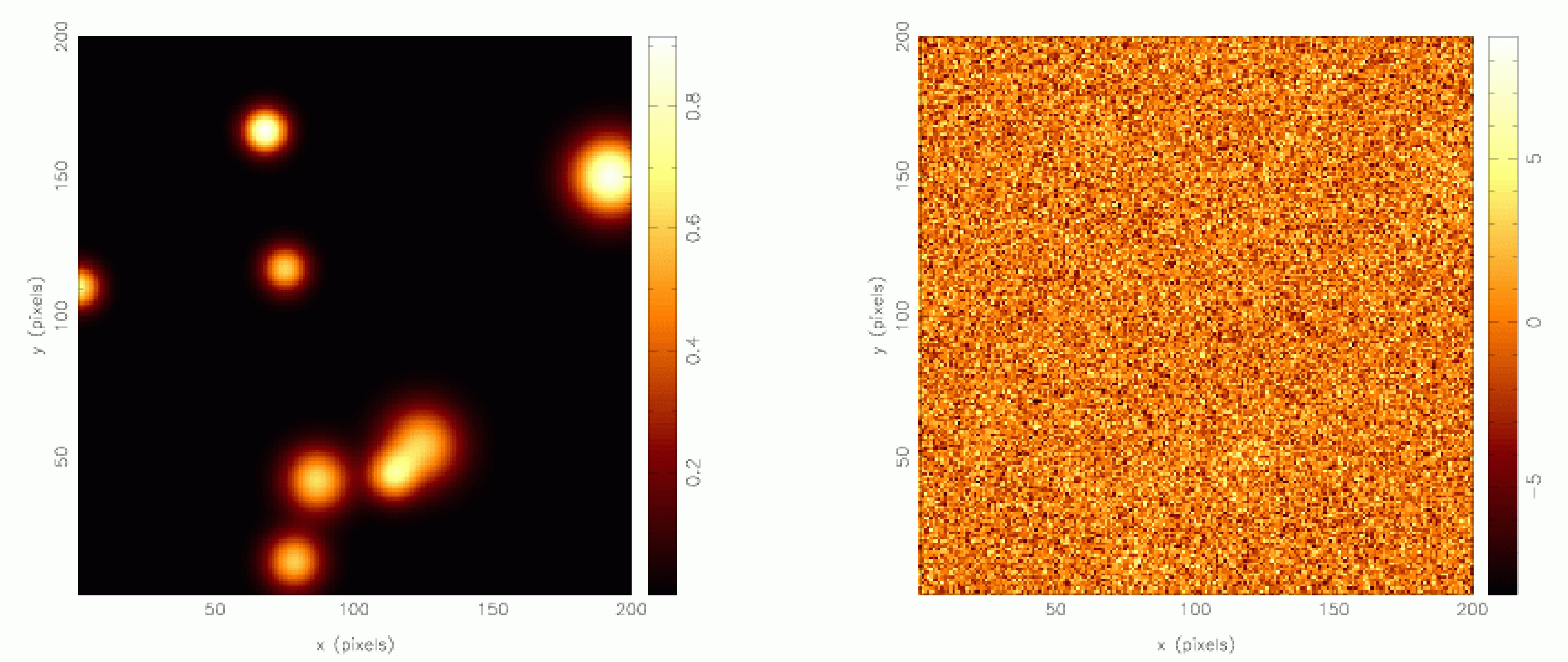
(Hobson & McLachlan 2003)

The optimal values of the object parameters θ (position, scale, amplitude…) and their associated errors are obtained evaluating their (unnormalized) posterior distribution $\Pr(\theta|D)$ using a Markov-chain Monte Carlo method

$$\Pr(\theta \mid D) \equiv \Pr(D \mid \theta)\Pr(\theta)$$

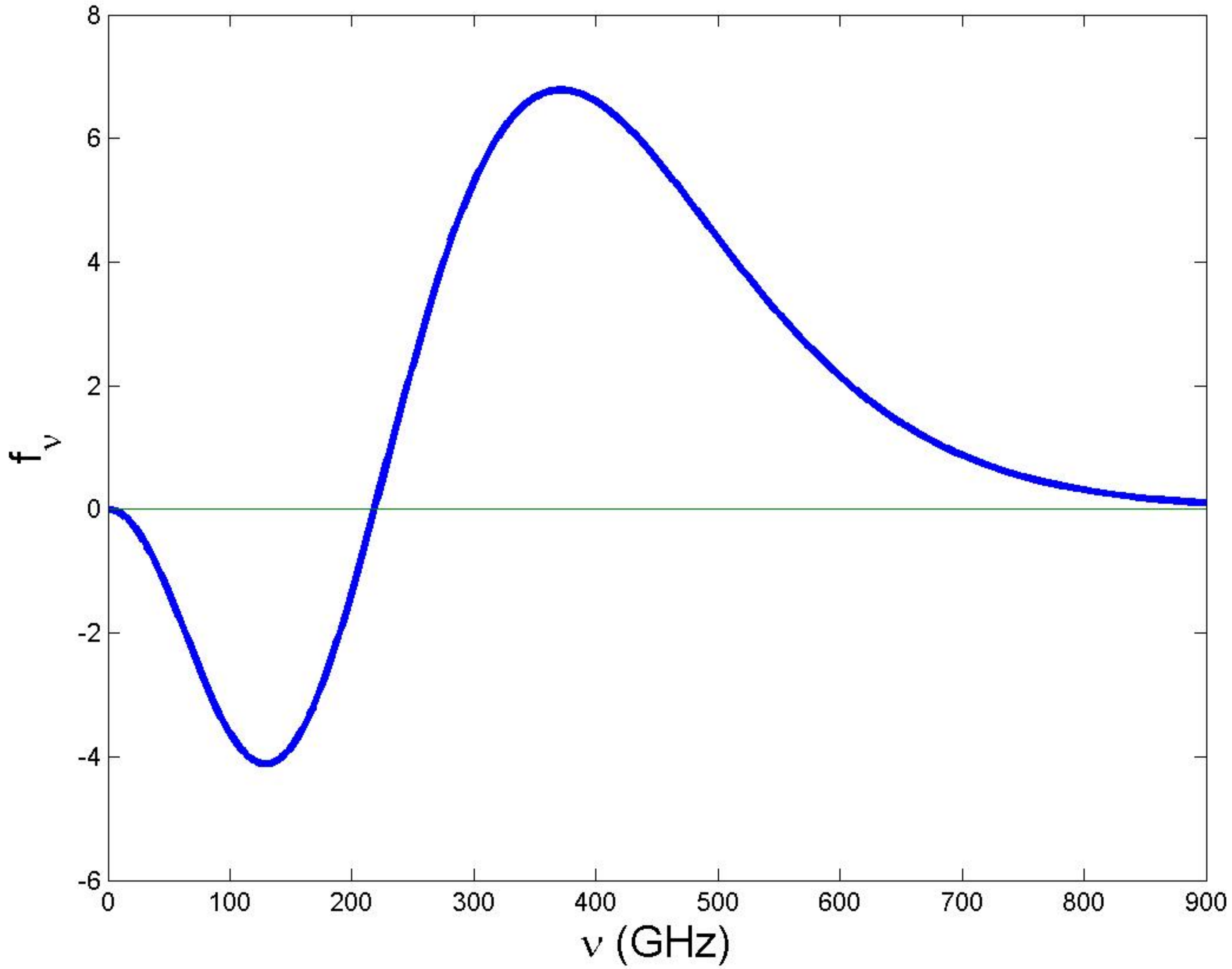
where $\Pr(D|\theta)$ is the likelihood and $\Pr(\theta)$ is the prior

An exact method and a much faster iterative approach (McCLEAN algorithm) are proposed



Toy example: Gaussian objects embedded in white noise with signal-to-noise 0.25-0.5

The McCLEAN algorithm identifies 7 out of the 8 objects (two of them combined into a single object) with no spurious detections



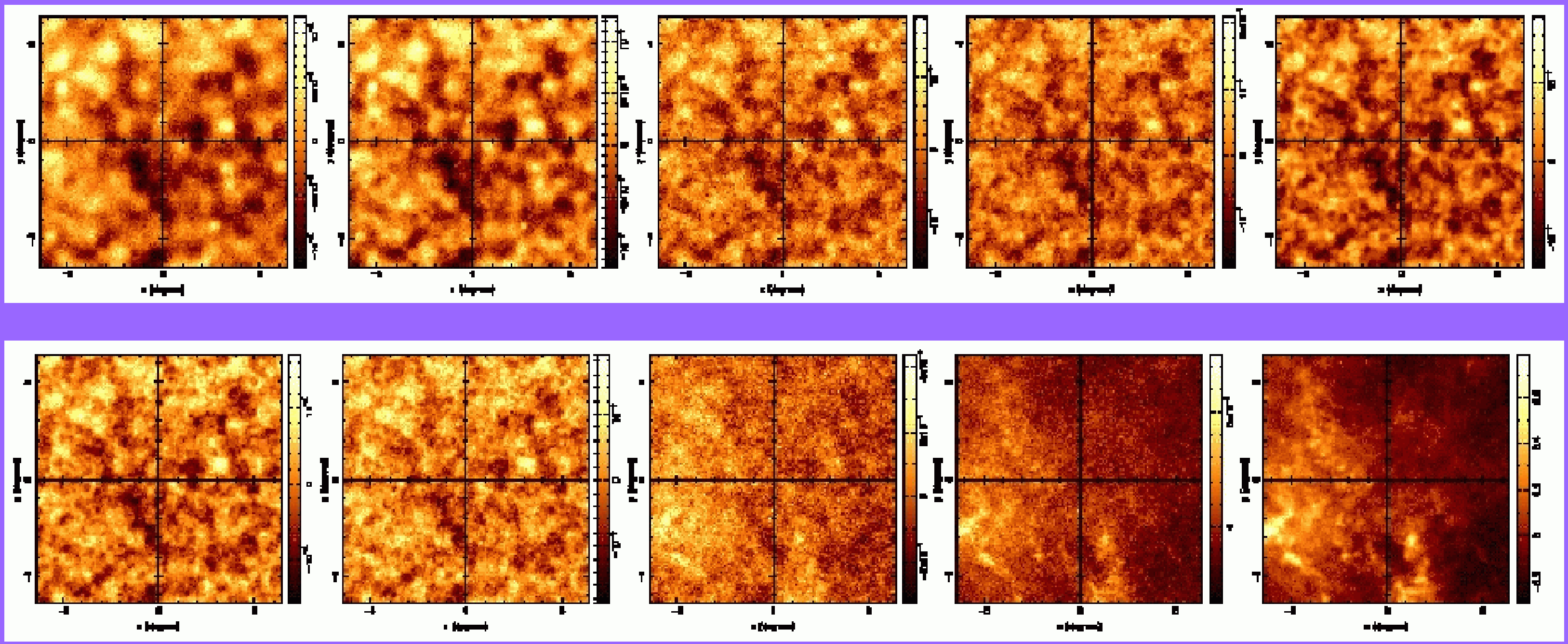
8
6
4
2
0
-2
-4
-6
0
100
200
300
400
500
600
700
800
900
f_ν
ν (GHz)

Multifrequency filters

Herranz et al. 2002c

- Tested in Planck simulated data (small patches) containing SZ, CMB, Galactic emission, point sources and white noise

- Spatial profile of clusters assumed to be known



Two different approaches

- Combination method

The individual frequency maps are optimally combined (the weights give maximum amplification of objects with the desired spatial profile and frequency dependence and can be determined from the data) and filtered with a MF (constructed for the new combined map)

- Matched multifilter (MMF)

The n maps are filtered with n filters and then added together

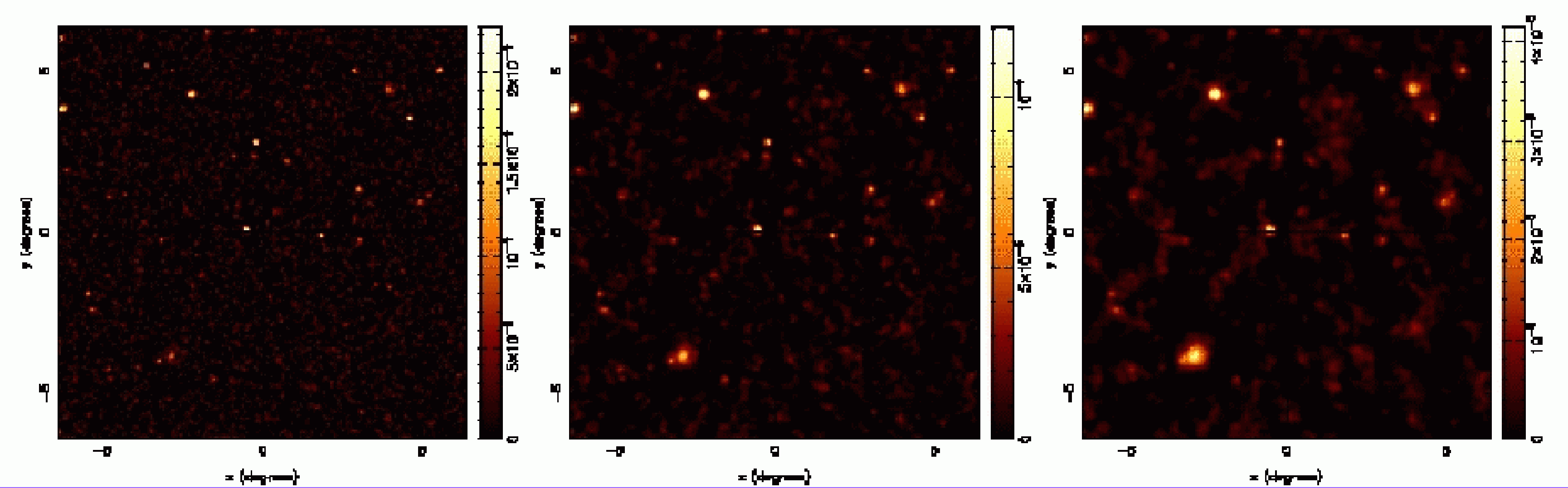
$$\psi(q) = \alpha P^{-1} F$$

P: cross-power spectrum matrix, $F=[f_\nu \tau_\nu]$, $\alpha=f(P,F)$

The filters are constructed such that $\langle w(x_0) \rangle = A$

⇨ estimation of the source amplitude

- The combination method is faster but the multifilter approach detect ~ few per cent more clusters
- The cluster size r_c will not be known a priori ⇨ the maps are (multi)filtered with different scales ⇨ a cluster will have maximum amplification when filtered with the correct scale ⇨ r_c can be estimated



Planck would provide a catalogue of ~10000 clusters in 2/3 of the sky

Bayesian non-parametric method for Planck

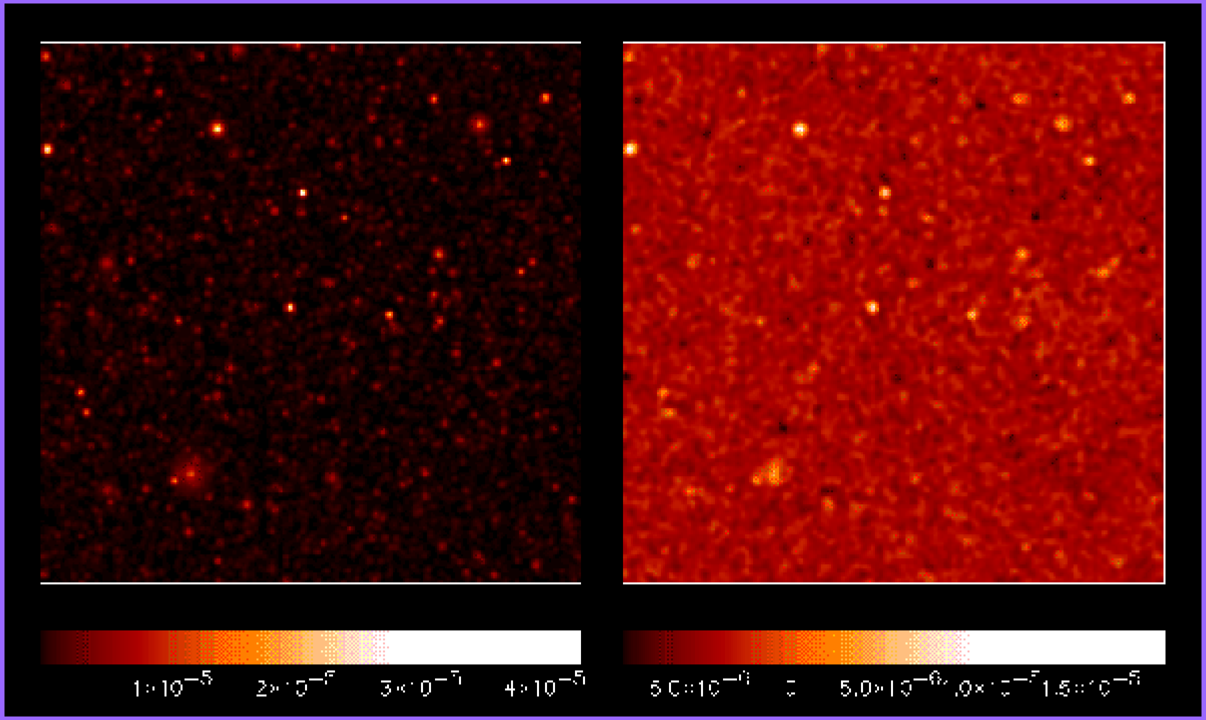
Diego et al. 2002

- Tested in Planck simulated data (small patches) containing SZ, CMB, Galactic emission, point sources and white noise
- No assumptions needed about the profile of the cluster

Two-step method

1. Cleaning of the maps: point sources (removed with the MHW), dust (using the 857 GHz channel) and CMB (using the 217 GHz channel)
2. The posterior probability $P(y_c|d)$ is maximized to obtain a map of the Compton parameter y_c. A power spectrum for the tSZ must be assumed (but the results are weakly dependent on the chosen prior)

Input versus reconstructed y_c map



Recovered catalogue of ~ 7500 clusters in 2/3 of the sky for Planck

Extraction of the kinematic SZ effect

The determination of peculiar velocities from the kSZ in individual clusters is a very difficult task

- Very weak signal (one order of magnitude below tSZ)
- Same spectral dependence as the CMB
- tSZ, foregrounds, radio sources, noise...

Make use of:

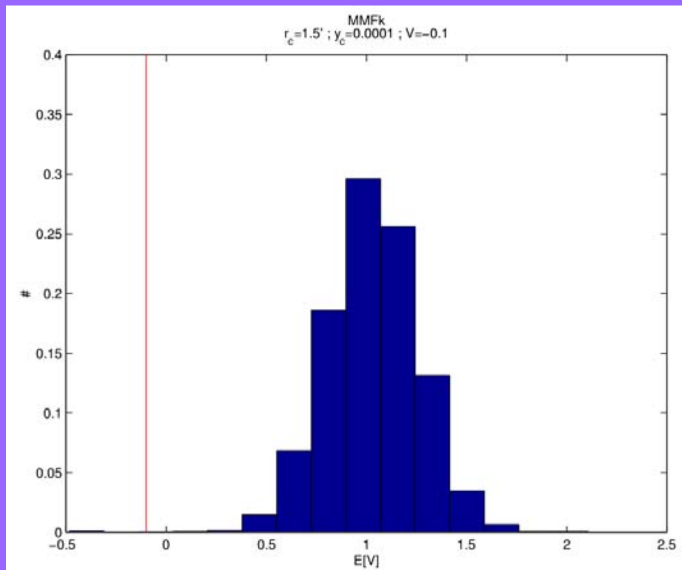
- Shape ⇨ cluster profile convolved with beam
- Cross-correlations with tSZ
- Multifrequency observations ⇨ separation from tSZ and foregrounds
- Non-Gaussian statistics ⇨ separation from CMB

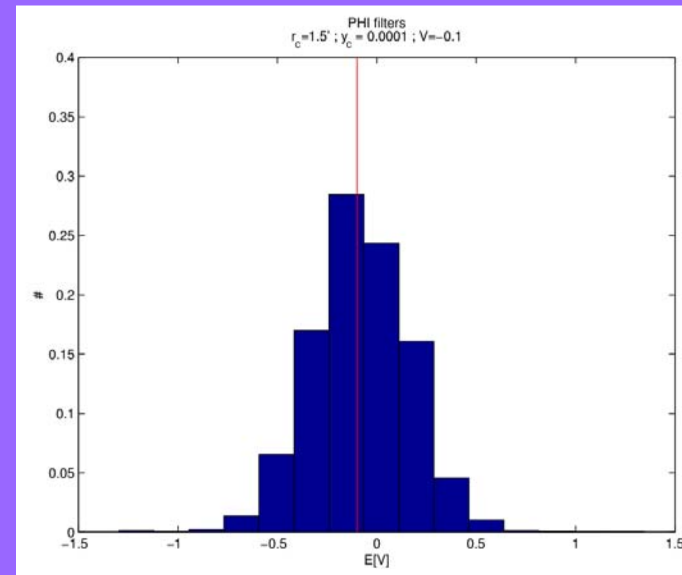
Multifrequency filters

Herranz et al. 2004a

- Two signals with the same spatial profile (tSZ and kSZ) ⇨ introduces a systematic bias in the estimation of the amplitude
- Negligible correction for tSZ but very important for the kSZ (~1 order of magnitude)
- Use instead unbiased multifrequency filters that cancel the tSZ
- Tested in Planck simulated data (small patches) containing SZ, CMB, Galactic emission, point sources and white noise
- Position and spatial profile of clusters assumed to be known

Recovered V for (biased) MMF and unbiased MMF



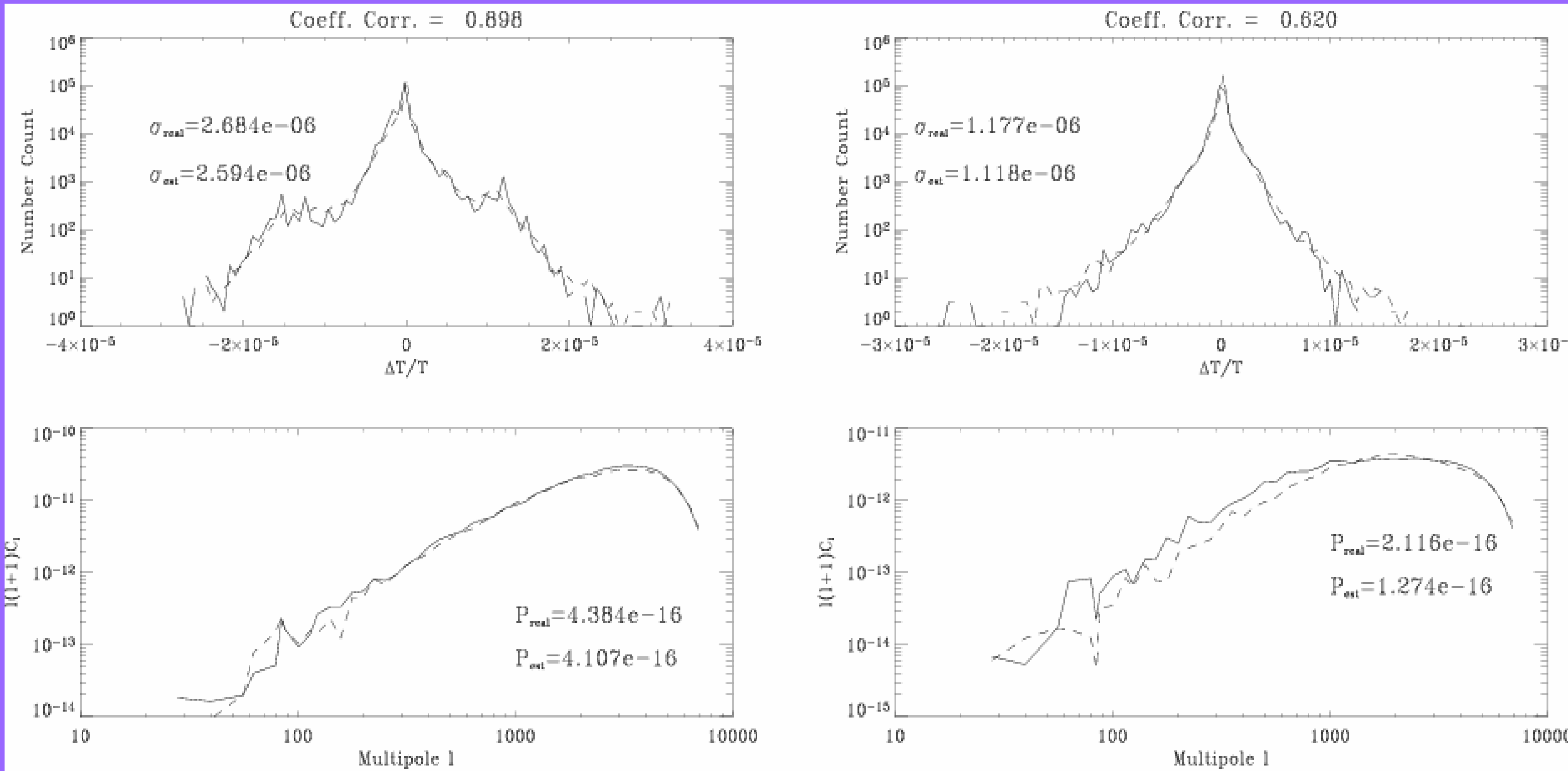


- Simulated clusters with y_c=0.0001, r_c=1.5 arcmin, V=-0.1 (which corresponds to v~300 Km/s for T_e=5 keV)
- For the new multifilter, the results are unbiased (tested also for smaller velocities and y_c parameter)
- 1-sigma error in estimated velocities:
  ~0.26 (~800 km/s) for y_c=0.0001
  ~0.65 (~2000 km/s) for y_c=0.00004

“Mask + reconstruct” method

Forni & Aghanim 2004

- Test the intrinsic power of the method (ideal conditions)
- Starting point: tSZ and CMB+kSZ maps
- The tSZ map is used to construct a mask with those pixels with tSZ emission (above a threshold ν)
- The mask is applied on the CMB+kSZ map and the masked pixels are reconstructed using interpolation
- The kSZ map is obtained as the difference between the CMB+kSZ and the interpolated maps
- This process is repeated for several thresholds
- The kSZ maps are combined with weights that take into account the non-Gaussian character of this emission to obtain a final kSZ map



Figures show results for the best and worst cases out of 15 simulations

Average correlation between input and rec. ~ 0.78

Mean error on σ of the kSZ map ~ 5 %

Extraction of statistical information from undetected sources

- Parameters of point source counts $n(S) = kS^{-\eta}$

  Pierpaoli 2002 ⇨ uses high order moments of the underlying distribution (CMB + undetected sources) to fit η

  Herranz et al. 2004b ⇨ obtain k and η fitting the characteristic function of the point sources distribution with an α-stable model

- Bispectrum from undetected point sources

  Komatsu & Spergel 2001, Argüeso et al. 2003

  Values from WMAP (Komatsu et al. 2003)

  $$(9.5 \pm 4.4) \times 10^{-5}\, \mu K^3 sr^2$$ at 41 GHz

  $$(1.1 \pm 1.6) \times 10^{-5}\, \mu K^3 sr^2$$ at 61 GHz

- Properties of the underlying distribution function to discriminate between undetected extragalactic point sources and SZ emission

  Rubiño-Martín & Sunyaev 2003

- Measurement of bulk flows from the kSZ effect

  Kashlinsky & Atrio-Barandela 2000

  Aghanim et al. 2001

  Atrio-Barandela et al. 2003

Final remarks

- Multifrequency information (if available) should be included in the methods to detect point sources
- Some methods need to be tested in more realistic conditions (cluster profiles, relativistic effects, foregrounds, asymmetric beams, etc)
- Assessment of the impact of source extraction in the CMB
- They should be extended to deal with polarization data
- Looking forward to see the methods applied on real data !!!