Detecting Point Sources in Planck CMB maps with needlets and multiple testing

Javier Carrón Duque

javier.carron@uniroma1.it

Collaborators: Nicola Vittorio, Domenico Marinucci, Alessandro Buzzelli

Department of Physics
Università Tor Vergata

28 May 2019

Outline

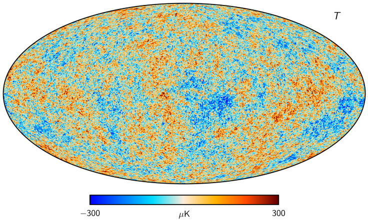
1. Introduction
   - Cosmic Microwave Background
   - Needlets and multiple testing
2. Methodology
   - STEM algorithm
   - Implementation
3. Results
   - Simulations
   - Planck Data
4. Conclusions and future work

Outline

1. Introduction
   - Cosmic Microwave Background
   - Needlets and multiple testing
2. Methodology
   - STEM algorithm
   - Implementation
3. Results
   - Simulations
   - Planck Data
4. Conclusions and future work

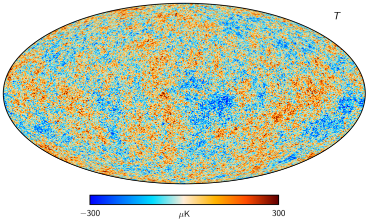
Cosmic Microwave Background (CMB)

- A lot of information can be extracted from a good observation of the CMB.
- There have been 3 multifrequency satellite missions: COBE (1989), WMAP (2001), and Planck (2009).



Cosmic Microwave Background (CMB)

- A lot of information can be extracted from a good observation of the CMB.
- There have been 3 multifrequency satellite missions: COBE (1989), WMAP (2001), and Planck (2009).
- Furthest observation with photons ($\sim 377\,000$ years after initial singularity).



Everything else is a foreground.

Foregrounds

Two types:

- Diffuse galactic radiation: free-free, synchrotron, thermal dust, spinning dust...
- Point sources: radio galaxies, dusty galaxies.

Foregrounds

Two types:

- Diffuse galactic radiation: free-free, synchrotron, thermal dust, spinning dust...
- Point sources: radio galaxies, dusty galaxies.

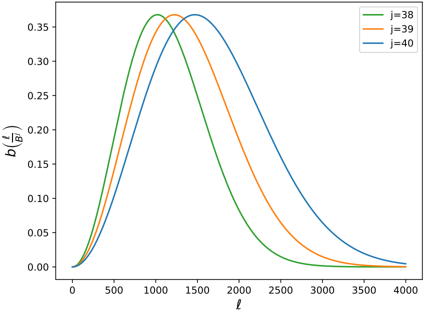
- Planck observed in 9 different bands from 30 GHz to 857 GHz.
- They are combined to extract the CMB temperature maps.
- Four algorithms to clean the maps:

▷ `Commander` ▷ `NILC` ▷ `SMICA` ▷ `SEVEM`

Needlets I

- Functions defined over the sphere, used to filter the map.

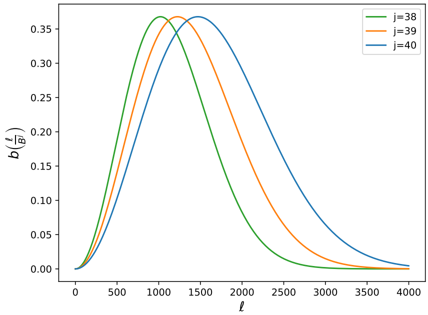
$$\psi_j(x,\xi) := \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} Y_{\ell m}(x)\overline{Y}_{\ell m}(\xi)$$



Needlets I

- Functions defined over the sphere, used to filter the map.

$$\psi_j(x,\xi) := \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} Y_{\ell m}(x) \overline{Y}_{\ell m}(\xi)$$



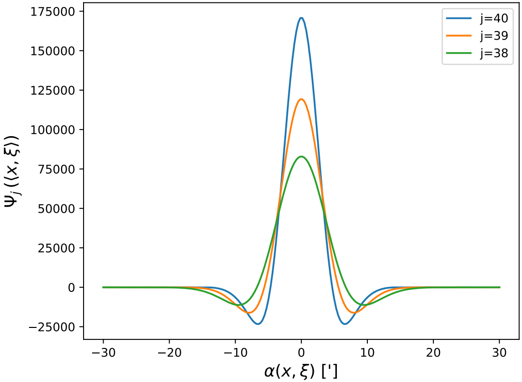
Six advantages

1. Do not rely on tangent plane approximation.
2. Calculated from spherical harmonics.
3. Simple reconstruction formula.
4. Different scales are uncorrelated.

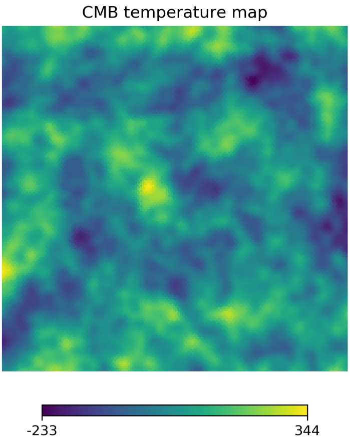
Needlets II

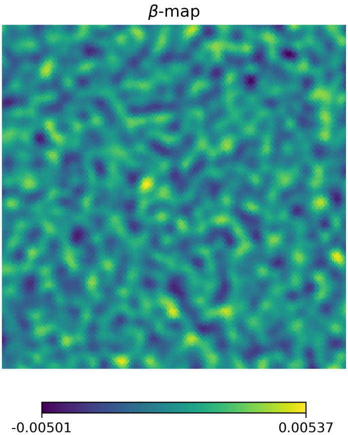
Six advantages

5. Very concentrated in pixel space.
6. Localised in multipole space.



Needlet filtering





Multiple testing

- Let us imagine two scenarios:

- We have a map with a maximum with a high intensity x.
- Assuming a gaussian map, we expect to find 1 maximum at intensity x in the entire map.
- The maximum x is compatible with gaussianity, it may be a random fluctuation.

Multiple testing

- Let us imagine two scenarios:

- We have a map with a maximum with a high intensity x.
- Assuming a gaussian map, we expect to find 1 maximum at intensity x in the entire map.
- The maximum x is compatible with gaussianity, it may be a random fluctuation.

- Assuming a gaussian map, we expect to find 1 maximum at intensity x in the entire map.

Multiple testing

- Let us imagine two scenarios:

- We have a map with a maximum with a high intensity x.
- Assuming a gaussian map, we expect to find 1 maximum at intensity x in the entire map.
- The maximum x is compatible with gaussianity, it may be a random fluctuation.

- Assuming a gaussian map, we expect to find 1 maximum at intensity x in the entire map.
- But we have now 1000 maxima at intensity x.
- Each maximum x is compatible with gaussianity, but the maxima population is not.

Benjamini-Hochberg approach

1. We obtain the values of the maxima.
2. For each value x, we count the number of maxima in the map at intensity x or higher: $n(x)$.
3. For each value x, we calculate the expected number of maxima at intensity x or higher, assuming gaussianity: $N(x)$.
4. If there is x_0 such that $N(x_0) < \quad n(x_0)$, we report the first $n(x_0)$ maxima.

Benjamini-Hochberg approach

1. We obtain the values of the maxima.
2. For each value x, we count the number of maxima in the map at intensity x or higher: $n(x)$.
3. For each value x, we calculate the expected number of maxima at intensity x or higher, assuming gaussianity: $N(x)$.
4. If there is x_0 such that $N(x_0) < \alpha n(x_0)$, we report the first $n(x_0)$ maxima.

- $\alpha \sim 0.01$ is a confidence factor.

Outline

1. Introduction
   - Cosmic Microwave Background
   - Needlets and multiple testing
2. Methodology
   - STEM algorithm
   - Implementation
3. Results
   - Simulations
   - Planck Data
4. Conclusions and future work

STEM algorithm

1. The map is filtered with a needlet.
2. All the maxima are extracted from the map (intensity and location).
3. The expected distribution of maxima is computed.
4. The Benjamini-Hochberg approach is applied to extract the list of candidates.

- Cheng et al. (2016)

False Discovery Rate

- The False Discovery Rate for the STEM algorithm is defined as:

$$FDR := \frac{\#(\text{False detections})}{\#(\text{Total detections})}$$

- It depends on the parameters of the algorithm B, j, and α.

False Discovery Rate

- The False Discovery Rate for the STEM algorithm is defined as:

$$FDR := \frac{\#(\text{False detections})}{\#(\text{Total detections})}$$

- It depends on the parameters of the algorithm B, j, and α.
- It can be proven that, for high enough j,

$$\boxed{FDR \lesssim \alpha}$$

Code

- Development of a small package in `Python 3`:

  `MTNeedlet`: https://javicarron.github.io/mtneedlet/

- It contains all functions needed for needlet filtering and applying the STEM algorithm.
- Fully documented and flexible in order to apply it to other problems.

Code

- Development of a small package in `Python 3`:

  `MTNeedlet`: https://javicarron.github.io/mtneedlet/

- It contains all functions needed for needlet filtering and applying the STEM algorithm.
- Fully documented and flexible in order to apply it to other problems.
- Using this package, we test the algorithm on simulations and apply it to real Planck data.

Simulations

- We obtain 200 CMB realizations.
- Realistic noise is added.
- Artificial point sources may be added: 200 points with intensities from 0 to 7σ.
- We apply the STEM algorithm with $B = 1.2$, $j = 38, 39, 40$, and $\alpha = 0.05, 0.01, 0.002$.

Simulations

- We obtain 200 CMB realizations.
- Realistic noise is added.
- Artificial point sources may be added: 200 points with intensities from 0 to 7σ.
- We apply the STEM algorithm with $B = 1.2$, $j = 38, 39, 40$, and $\alpha = 0.05, 0.01, 0.002$.

Goal:

Maximize discovery power while keeping $FDR \leq 0.01$.

Real data

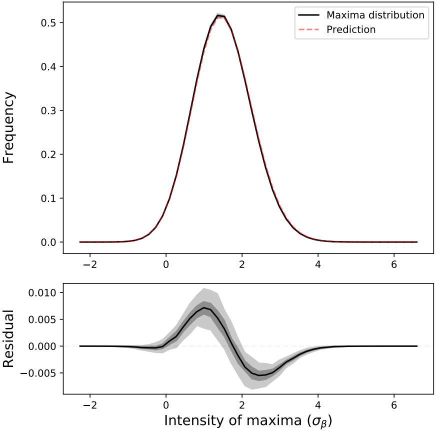
- The algorithm is applied to Planck CMB temperature maps, version 2 (2015) and 3 (2018).
- For version 3, inpainted and not inpainted maps are used.
- Number of detections is obtained for the whole map, outside inpainting mask ($\sim 98\%$ of the sky), and outside common confidence mask ($\sim 78\%$ of the sky).
- Detections are compared to the Planck Source Catalogues.

Outline

1. Introduction
   - Cosmic Microwave Background
   - Needlets and multiple testing
2. Methodology
   - STEM algorithm
   - Implementation
3. Results
   - Simulations
   - Planck Data
4. Conclusions and future work

No sources

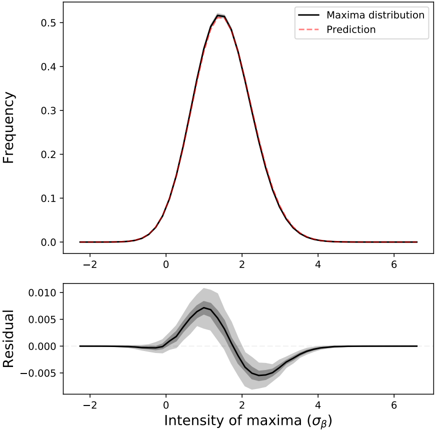
Only CMB+noise, no sources:



- The maxima distribution follows the theoretical one.

No sources

Only CMB+noise, no sources:

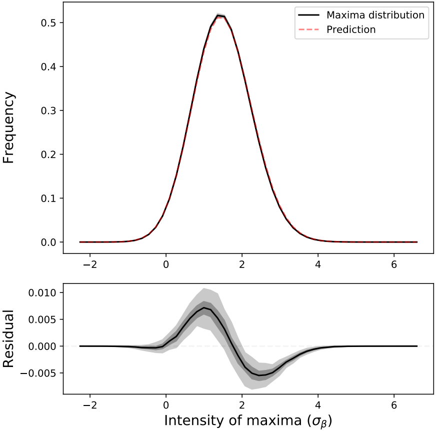


- The maxima distribution follows the theoretical one.
- Low amount of point sources are reported.

α	$j = 38$	$j = 39$	$j = 40$
0.05	4.0 %	6.5 %	99.5 %
0.01	1.5 %	2.5 %	98.5 %
0.002	0.5 %	0.5 %	93.5 %

No sources

Only CMB+noise, no sources:



- The maxima distribution follows the theoretical one.
- Low amount of point sources are reported.

α	$j = 38$	$j = 39$	$j = 40$
0.05	4.0 %	6.5 %	99.5 %
0.01	1.5 %	2.5 %	98.5 %
0.002	0.5 %	0.5 %	93.5 %
0.05	0.045	0.075	4.595
0.01	0.015	0.030	3.215
0.002	0.005	0.005	2.270

With sources

- The needlet filtering boosts the signal to noise ratio, up to a factor $\lesssim 2$.

With sources

- The needlet filtering boosts the signal to noise ratio, up to a factor $\lesssim 2$.
- False Discovery Rate:

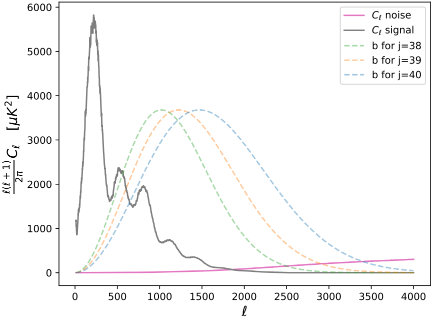
α	$j = 38$	$j = 39$	$j = 40$
0.05	0.042	0.040	0.082
0.01	0.009	0.009	0.046
0.002	0.0013	0.0018	0.0325

With sources

- The needlet filtering boosts the signal to noise ratio, up to a factor $\lesssim 2$.
- False Discovery Rate:

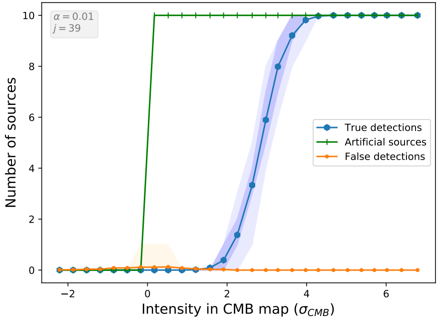
α	$j = 38$	$j = 39$	$j = 40$
0.05	0.042	0.040	0.082
0.01	0.009	0.009	0.046
0.002	0.0013	0.0018	0.0325

- $FDR_{j=40} > \alpha$, it is contaminated by noise.



Sensitivity

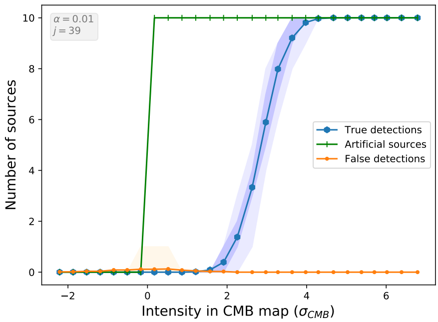
- We obtain the sensitivity curve.



Sensitivity

- We obtain the sensitivity curve.
- We calculate the 90%-complete limit, in units of σ_{CMB}:

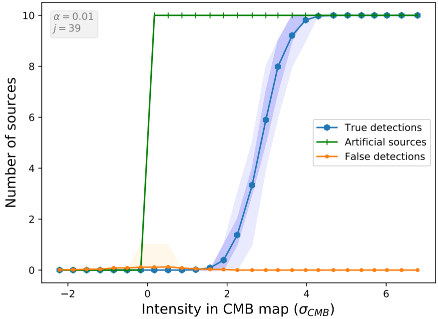
α	$j = 38$	$j = 39$	$j = 40$
0.05	5.08	3.50	2.50
0.01	5.36	3.66	2.66
0.002	5.74	3.86	2.74



Sensitivity

- We obtain the sensitivity curve.
- We calculate the 90%-complete limit, in units of σ_{CMB}:

α	$j = 38$	$j = 39$	$j = 40$
0.05	5.08	3.50	2.50
0.01	5.36	3.66	2.66
0.002	5.74	3.86	2.74



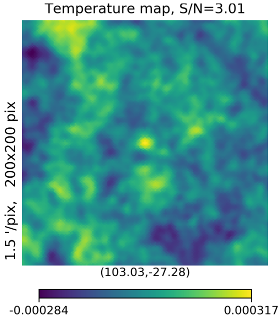
The choice is $j = 39$, $\alpha = 0.01$

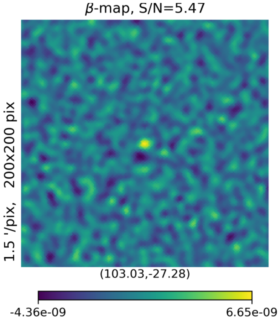
Example: SMICA

Map	version	inpainted	detections	mask I	mask C	cat	cat C
smica	2.01	False	558	10	2	69	0
smica	3.00	False	318	16	2	197	0
smica	3.00	True	11	3	1	7	0

Example: SMICA

Map	version	inpainted	detections	mask I	mask C	cat	cat C
smica	2.01	False	558	10	2	69	0
smica	3.00	False	318	16	2	197	0
smica	3.00	True	11	3	1	7	0





Planck cleaning algorithms

All algorithms:

- Candidates to PS reported in most maps.
- Few candidates outside the confidence mask.
- Around half of the candidates in Planck PS catalogues.
- Last release mostly compatible with non-contaminated Gaussian maps.

Planck cleaning algorithms

All algorithms:

- Candidates to PS reported in most maps.
- Few candidates outside the confidence mask.
- Around half of the candidates in Planck PS catalogues.
- Last release mostly compatible with non-contaminated Gaussian maps.

SEVEM: Large amount of sources.

NILC: No candidates for the inpainted version.

SMICA: Lowest number of candidates, one candidate outside confidence mask.

Commander: Largest amount of candidates because it does not preprocess the map to remove sources.

Outline

1. Introduction
   - Cosmic Microwave Background
   - Needlets and multiple testing
2. Methodology
   - STEM algorithm
   - Implementation
3. Results
   - Simulations
   - Planck Data
4. Conclusions and future work

Conclusions

- We have tested the STEM algorithm through **simulations** under different conditions in order to find the optimal choice of parameters and keep the *FDR* within limits.
- The last release of **Planck data** is mostly compatible with non-contaminated maps, although a few point sources are reported.
- A **software** has been developed to work with *needlets* and *multiple testing*. It is now publicly available.

Future work

- Only temperature maps have been studied, the algorithm can be also applied to **polarization** maps: E and B modes.
- Needlets are very **flexible**, these techniques could be applied with other goals, such as extracting a lensing map.

Detecting Point Sources in Planck CMB maps with needlets and multiple testing

Javier Carrón Duque

javier.carron@uniroma1.it

Collaborators: Nicola Vittorio, Domenico Marinucci, Alessandro Buzzelli

Department of Physics
Università Tor Vergata

28 May 2019

Thank you!!