Universidad de Cantabria
Departamento de Física Moderna

CSIC - Universidad de Cantabria
Instituto de Física de Cantabria

Detection of Point Sources in Maps of the Cosmic Microwave Background Radiation by means of Optimal Filters

A dissertation submited in partial of the requirements
for the degree of Doctor of Philosophy in Physics

by

Marcos López-Caniego Alcarria

2006

CHAPTER 8

Conclusions

In this thesis we have presented a series of tools, linear filters and state-of-the-art detectors, that can be used to deal with the problem of component separation in CMB maps. In particular, with the detection of compact sources on a background.

In microwave astronomy, extragalactic objects, such as galaxies and most of the clusters of galaxies, appear as point objects in the sky, due to the present resolution of the experiments. If we assume that the instrument used for the observation has a known response, e.g. Gaussian, then, the compact source will appear as a Gaussian with a width equal to that of the response of the instrument.

In astronomy it is a common practice to linear filter the images in order to enhance very faint point sources, helping their detection. The best filter would be the one that makes it easier to distinguish between real sources present in the image and peaks coming from the background alone. In the case of simple thresholding, that considers only the amplitude of the peaks, the answer to the question of which one is the best linear filter is well known: the standard matched filter. But in the case of a Neyman-Pearson detector that considers other things apart from mere amplitudes, this is not longer true. Moreover, it has been shown that certain wavelets behave similarly to the matched filter and due to the practical problems in the implementation of this filter, they can certainly compete with it. Several filtering and detection techniques have been introduced in this thesis to detect point sources in one and two-dimensional images.

In **Chapter 2** we have introduced the problem of *detector design*, using a Neyman-Pearson rule that includes *a priori* information of the source distribution and the number densities of maxima (background and background plus source) to define the acceptance region. In this approach, based on maxima, we have included both the amplitude and the spatial information, i.e. the curvature. The curvature of the background is different from that of the sources, and we use this to improve our detection rule. Using this information, the chances of detection do not depend only on the amplification of the sources produced by the filtering but also on the filtered momenta up to the fourth order.

- We have applied our technique to a family of matched-type filters (MTF) by modifying the scale of the standard matched filter. We have considered the interesting case of color noise to represent the background, that can be modeled by an homogeneous and isotropic Gaussian random field, characterized by a scale-free power spectrum $P(q) \propto q^{-\gamma}$, $\gamma \geq 0$.
- In a practical example we have shown that the curvature plays an important role defining the acceptance region and we have proven that the number of detections in the case of a filter with a scale similar to the pixel size beats the number of detections in the case of the standard MF.
- This result has been tested with numerical simulations for a uniform distribution and white noise.

In **Chapter 3** we approach the problem of *filter design* in a different and interesting way. We design a new one-dimensional filter in such a way that optimizes the number of true detections for a fixed number of spurious sources, when used in conjunction with our improved detection rule based on amplification and curvature of the filtered field. The optimization of the number of true detections is performed by using the *a priori* pdf of the amplitudes of the sources. This filter depends on two free parameters and it is therefore called biparametric scale adaptive filter (BSAF).

We generalize the functional form of this filter, as well as other standard filters, and introduce another degree of freedom, α, that allow us to filter at any scale (αR), including that of the source R. We have shown the benefits of filtering at scales smaller than R, which significantly improves the number of detections.

- As an example, we have considered two different distributions of sources. A uniform distribution in the interval $\nu \in [0, 2]$ in the filtered field, and a scale-free power law distribution in the interval $\nu \in [0.5, 3]$, i.e. weak sources.
- The BSAF has proven to be significantly better than the standard MF, the scale-adaptive filter (SAF) and the MH wavelet in certain cases. In particular, the improvement in the number of detections of the BSAF at $\alpha = 0.3$ with respect to the standard MF is $\simeq 40\%$ for $\gamma = 0$, $n_b^* = 0.05$ and $R = 3$.

- We have also tested the performance of the filters for a mixture of weak, intermediate and bright sources. For a uniform distribution with $A \in [0, 5]\sigma_0$ and for a scale-free distribution with $A \in [0.5, 5]\sigma_0$, the BSAF also improves the MF. However, for a scale-free distribution with $A \in [3, 5]\sigma_0$, i.e., dominated by bright sources, we find that the optimal BSAF defaults to the standard MF, which gives the maximum number of detections in this case.

- By construction, the BSAF gives in any case the best performance among the considered filters: the SAF and the MF are particular cases of the BSAF and this guarantees that the parameters of the BSAF will default to the best possible filter in each case. Therefore, the number density of detections obtained with the BSAF will be at least equal to the best of the other two filters, and in some cases superior. We also remark that the BSAF is equivalent to the MH in some particular cases.

- Our results suggest that, from the practical point of view, one could use the BSAF when $0 \lesssim \gamma \lesssim 1$ since, in this range, clearly improves the number of detections with respect to the other filters. However, for $\gamma \gtrsim 1.0$ the usage of the MH is justified due to its robustness in its functional form and it gives approximately the number of detections obtained either with the BSAF or MF.

- For the studied cases of sources distributions (except for the one dominated by bright sources) and fixing the values of γ, n_b^* and R, we find that the optimal parameters of the BSAF are only weakly dependent on the distribution of the sources.

- Regarding source estimation, we propose a linear estimator which is unbiased and of maximum efficiency, that we have also tested with simulations.

In **Chapter 4** we have considered the interesting case of *filter design + detector design* in two dimensions:

- We make a generalization to the two-dimensional case of the Neyman-Pearson detector, considering the number densities of peaks. This leads to a sufficient detector that, in the case of the spherically symmetric sources, is linear in the amplitude and curvature of the sources.

- We have also implemented the functional form of the MF, the MH and the Scale adaptive filter to two dimensions. Moreover, we have derived the BSAF in two dimensions to compare it with the previous filters in order to assess which one of them performs better when detecting sources.

- We have introduced an extra degree of freedom α in the previous filters that will allow us to filter at different scales αR, where R is the scale of the source.

  Note that by construction, the functional form of the BSAF includes the standard MF as a particular case and its performance in terms of number of true detections for a fixed number of spurious detections must be at least as good as the standard MF's one.

- As an application, we have considered an interesting case, a uniform distribution of weak sources with amplitudes $A \in [0, 2]\sigma_0$, where σ_0 is the dispersion of the field filtered with the standard matched filter, embedded in white noise ($\gamma = 0$).

- We have tested different values of the source size R and of the number density of spurious detections n_b^* that we fix. We find that the BSAF improves the number density of detections up to $\simeq$ 40% with respect to the standard MF ($\alpha = 1$) for certain cases.

  Note that since the Neyman-Pearson detector for the standard MF ($\alpha = 1$) defaults to the classic thresholding detector that is commonly used in astronomy, the results of this work imply that it is possible, under certain circumstances, to detect more point sources than in the classical approach.

In **Chapter 5**, we have presented a new method that combines images that contain localized sources in such a way that the output image has minimum variance and the image fusion gives at the position of the source an unbiased estimator of the amplitude. We studied the *linear* and *quadratic* fusion approach and tested it in the context of compact source detection in astronomical data. To do this, we have done one thousand simulations for each of the two cases we want to consider. First, white noise plus point sources, and second, color noise plus point sources for the 44 GHz channel of the Planck satellite specifications. In both cases we have compared the average number of real detections and false alarms for the *linear* and *quadratic* case with those obtained with the Mexican hat wavelet at the optimal scale, the best studied tool in CMB science to detect point sources.

- For the white noise case and 5σ threshold, if we compare the linear and MH case on the one hand, we find that the improvement is of the order of 23%. On the other hand, if we compare the quadratic fusion and the MH, the improvement is $\simeq$ 25%.

- When we consider the case of realistic simulations for the 44 GHz Planck channel at 5σ, comparing the MH with the *linear fusion*, the improvement of the latter is of the order of 12%.

- Furthermore, we find that using the *quadratic fusion* we detect $\simeq$ 35% more real sources than with the MH (with $r < 1\%$).
- Moreover, if we compare the quadratic fusion method with the results obtained when no filtering is done, we detect around four times more objects for the same ratio r.

  We remark that the parameter ϵ that appears in the *quadratic fusion* can be easily obtained. The implementation of the method is very straightforward and the CPU time necessary to test the method is very small, of the order of seconds for a 118x118 image.

In **Chapter 6** we have compared the performance of three filters when dealing with the detection of point sources in CMB flat sky maps. These filters are the well known MF, the Mexican Hat Wavelet (MH or MHW1) and the recently introduced Mexican Hat Wavelet 2 (MHW2). The latter can be obtained by applying twice the Laplacian operator to the Gaussian. This new wavelet is part of an effort to improve the MHW1, a tool that has proved very useful in the detection of compact sources and of non-Gaussianity in CMB maps.

- We have tested these filters in realistic two dimensional simulations of the microwave sky prepared in the framework of optimizing the efficiency in separating the various astrophysical components in the forthcoming ESA's Planck Satellite all-sky maps. As for the Galactic foregrounds and the SZ effect in clusters of galaxies, we have used the available *Planck Reference Sky Model*; we have adopted the standard *concordance model* for simulating CMB anisotropy maps and, as for extragalactic point sources we have used simulations based on the current number count models for sources. We then applied these filters to a sufficient number of flat patches to cover half of the sky ($2\pi \mathrm{sr}, b > |30^\circ|$).
- We have found that the MHW2 and the MF outperform the MHW1 in some aspects, specially at the lowest Planck frequencies. The three filters yield approximately the same number of real detections, down to the same flux detection limit, although the MHW1 yields a corresponding higher number of spurious detections, i.e. it is less reliable. Moreover, the MHW2 and the MF give comparable results for almost every one of the analyzed indicators.
- These results are very important because both wavelets, the MHW1 and the MHW2, are represented by a known analytical function. The only parameter that needs

to be obtained for each simulation is the "optimal scale", which is done locally in a very easy way.

Moreover, the correct definition of the Matched Filter implies

- Firstly, it is necessary to estimate the value of the power spectrum for all the Fourier modes present in the image, specially for the low modes where the power spectrum is *noisy*.
- Secondly, the use of such a noisy power spectrum to construct the MF often yields a filter with many discontinuities in Fourier space which, in turn, produces ringing effects in the filtered image. Therefore some smoothing in the spectra needs to be done before constructing the filter, which introduces further arbitrariness.
- Thirdly, sometimes it will not be possible to properly estimate some Fourier modes, for example when using masks with missing data, and these modes will have to be guessed.
- Fourthly, further problematics when implemented on the sphere.

  Therefore, the most relevant conclusion of this analysis is that the MHW2 can be surely a better choice for the definition of a blind source catalogue because it gives numbers of detected and of spurious sources comparable to the ones obtained with the MF but it easier to implement, more robust and has much lesser CPU requirements.

In **Chapter 7** we have used the MHW2 filter to obtain estimates of (or upper limits on) the flux densities at the WMAP frequencies of a complete all-sky sample of 2491 sources at $|b| > 5^\circ$, brighter than 500 mJy at 5 GHz (or at 1.4 or 0.84 GHz, in regions not covered by 5 GHz surveys but covered by either the NVSS or the SUMSS).

- We have obtained flux density estimates for the 939 sources detected at $\geq 3\sigma$ and 3σ upper limits for the others.
- At the 5σ level (and above) we have detected 383 sources, from them, 101 (i.e. 26%) are "new", in the sense that they are not present in the WMAP catalogue; 43 of them are above the estimated completeness limit of the WMAP survey: $\simeq 1.1$ Jy at 23 GHz. This illustrates how the prior knowledge of source positions can help their flux measurements.
- We have detected 37 WMAP extragalactic sources with low-frequency counterparts included in our sample at $< 5\sigma$. This is probably due to the fact that our

error estimates exceed substantially those given by the WMAP team, particularly at 23 GHz, where we have the highest detection rate. In fact, the WMAP errors do not correspond to the rms fluctuations in the source neighborhood but to the uncertainties in the amplitude of the Gaussian fit. As a consequence, the WMAP catalogue starts being incomplete at flux densities more than 10 times higher than their typical formal errors.

- Our flux density estimates for sources detected at $\geq 5\sigma$ are generally in very good agreement with the WMAP ones at 23 GHz. At higher frequencies WMAP fluxes tend to be slightly but systematically higher than ours, probably due to having ignored the deviations, increasing with frequency, of the point spread function from a Gaussian shape. Our estimates use the real beam shape at every frequency.
- We have worked out and applied a method to correct flux estimates for the Eddington bias, without the need of an *a-priori* knowledge of the slope of source counts below the detection limit.
- Our selection criterion leaves out 25 WMAP sources, only 12 of which however turn out to be $\geq 5\sigma$ detections after our analysis, and only 3 have 23 GHz fluxes $\gtrsim 1.1\,\mathrm{Jy}$, the estimated completeness limit of the survey. Thus, our approach has proven to be competitive with, and complementary to the blind technique adopted by the WMAP team. If fact we missed 3 sources brighter than the estimated completeness limit ($S_{23\mathrm{GHz}} = 1.1\,\mathrm{Jy}$), but we detected 43 new ones. Our completeness level is thus $\simeq 98.7\%$.
- On the whole, 26% of sources we have detected at $\geq 5\sigma$ are not present in the WMAP catalogue. On the other hand, the efficiency of the process is low. Only 383 of the 2491 sources in our input sample were detected at $\geq 5\sigma$ in at least one WMAP channel.

Future work

In this thesis we have developed some new techniques that deal with the detection of point sources in one dimensional data arrays and two dimensional maps. Moreover, we have developed a linear detector that also takes into account the spatial information, not only amplification, and can help in the detection. These techniques have been successfully tested in realistic simulations of the Planck satellite and on real data from the WMAP space mission.

- In a future work, we will exploit the non-blind catalogue of point sources at the frequencies of WMAP to investigate the high-frequency properties of sources selected at low frequencies.

- We will extend some of these tools to the sphere, complementing other existing techniques already developed. Filtering the data directly on the sphere allows one to take full advantage of the properties of wavelets, in particular the MHW1 and MHW2 in the detection of compact sources as compared with the Matched filter.

- We will complement the non-blind detection of sources in patches of the WMAP data with a blind approach on the sphere.

- Last, but not least, we plan to use these techniques to detect point sources on Planck as soon as the first maps are available after it is launched in the spring of 2008. We hope to contribute with our algorithms to the production of the Early Release Point Source Catalog (ERPSC) and the final catalogue.