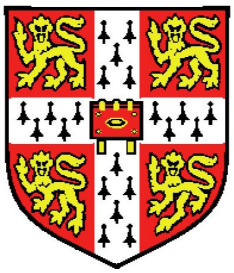
‘Powellsnakes’, a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets



Fernando Pedro Marinho Neto Pires de Carvalho

Cavendish Astrophysics

and Wolfson College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

December 2012

I would like to dedicate this work to all people that, in one way or another, helped to make it possible.

Acknowledgements

I acknowledge financial support from the Fundação Ciência e Tecnologia (Portugal) under the grant SFRH/BD/42366/2007 (four years).

I would like to thank both my PhD supervisors, Professor Michael Hobson (‘*I would sample from the full catalogue likelihood instead.*’) and Professor Anthony Lasenby (‘*Have you seen any relativist effect ?*’).
Thank you very much.

A warm and very special thanks to everybody at the Cavendish astrophysics group. You made me feel like being at home.
Thank you very much.

Another very special thanks goes to my office room mates and to all my colleagues. What a great time. If it did not happen more frequently, the only one to blame is myself.
Thank you very much.

I would like to thank the Planck Consortium, with whom I spent a significant fraction of my PhD time, for such a fantastic instrument, wonderful team and data. A very special mention to Planck HFI people and in particular to Cambridge Planck DPC (CPAC). Of course, I could not forget the people from IFCA (Santander-Planck LFI).
Thank you very much.

My other scientific family was the AMI+Planck team, from whom I learnt that having a lavish budget, in science as in life helps. Yet, excellence in science is still perfectly achievable, if only with a ‘piece of a wet string’ and a ‘sub-optimal correlator’, when one is not short in competence and ingenuity.
Thank you very much.

I would like to thank everybody at Wolfson College. You fed me, kept me warm and taught me my new country. From you I learnt about gowns, black ties, bow ties and that the best Port in the world can actually be found in Cambridge and not in Porto, where I am from!
Thank you very much.

Speaking of home...
A big, big kiss to my mother whose permanent love and support made it all possible. To my brothers, to my family,
thank you very much.

I could never forget my friends back home, whom have always been very supportive even financially when it came about.
Thank you very much.

Last, but not certainly not least, a very special thanks to the people that made direct contributions to this work and were not yet acknowledged in my previous publications or contributions:

- Graça Rocha, without whom I would never be here in first place. Thank you for believing me and bringing me into this new exciting world.
- Farhan Feroz, from whom I have learnt the 'dark arts' of inference, or I like to call it, 'how to make the fat lady sing'.
- Phil Graf, for the discussions about multiple hypothesis tests (classification) and how to evaluate the probability of being raped by a previously convicted person in cold weather, using a neural-network.
- '***CSI***', 'Cambridge, Santander & IPAC' and later 'Compact Source Investigation' group for the point source challenge graphics. Especially for plotting PwS after all other methods making it clear enough to be included in this thesis.
- Jean-Baptiste Melin, for the IDL scripts I have used to 'prèpárè' most of the SZ results.
- Paulo Marques, whom helped me port PwS's first version to Linux

and how to sue the Portuguese branch of a French multi-national company for copyright violation.

- Microsoft®, for letting me use their professional development system (Visual Studio®), for free, without making me sign a piece of paper acknowledging my work is worthless.

Thank you very much.

DECLARATION

This dissertation is the result of my own work, carried out in the Astrophysics Group of the Cavendish Laboratory, Cambridge, between May 2008 and December 2012, and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. Throughout the text, the plural pronoun '*we*' is used in the editorial sense, and should be taken to refer to the singular author, with the possible inclusion the reader. No part of this dissertation has been submitted for a degree, diploma or other qualification at this or any other university. The total length of this dissertation does not exceed sixty thousand words

Abstract

In this work we introduce a fast Bayesian algorithm designed for detecting compact objects immersed in a diffuse background.

A general methodology is presented in terms of formal correctness and optimal use of all the available information in a consistent unified framework, where no distinction is made between point sources (unresolved objects), SZ clusters, single or multi-channel detection. An emphasis is placed on the necessity of a multi-frequency, multi-model detection algorithm in order to achieve optimality.

We have chosen to use the Bayes/Laplace probability theory as it grants a fully consistent extension of formal deductive logic to a more general inferential system with optimal inclusion of all ancillary information [Jaynes, 2004].

Nonetheless, probability theory only informs us about the plausibility, a '*degree-of-belief*', of a proposition given the data, the model that describes it and all ancillary (prior) information. However, detection or classification is mostly about making educated choices and a wrong decision always carries a cost/loss. Only resorting to '*Decision Theory*', supported by probability theory, one can take the best decisions in terms of maximum yield at minimal cost.

Despite the rigorous and formal approach employed, practical efficiency and applicability have always been kept as primary design goals. We have attempted to select and employ the relevant tools to explore a likelihood form and its manifold symmetries to achieve the very high computational performance required not only by our '*decision machine*' but mostly to tackle large realistic contemporary cosmological data sets.

As an illustration, we successfully applied the methodology to ESA's (European Space Agency) *Planck* satellite data [Planck Collaboration et al., 2011d]. This data set is large, complex and typical of the contemporary precision observational cosmology state-of-the-art.

Two catalogue products are already released: (i) A point sources catalogue [Planck Collaboration et al., 2011e], (ii) A catalogue of galaxy clusters [Planck Collaboration et al., 2011f]. Many other contributions, in science products, as an estimation device, have recently been issued [Planck et al., 2012; Planck Collaboration et al., 2011g,i, 2012a,b,c].

This new method is called '***PowellSnakes***' (PwS).

Contents

List of Figures

Chapter 1

Introduction

The detection and characterisation of discrete objects is a common problem in many areas of astrophysics and cosmology. Indeed, every data reduction process must resort to some form of compact object detection, since either the objects themselves are the goal of the study or they act as contaminants and therefore must be removed. In such analyses, the key step usually involves the separation of a localised object signal from a diffuse background, defined as all contributions to the image aside from the objects of interest.

A well-established method to address this issue is to assume that most of the pixels are part of the background exclusively[1], the background is smoothly varying, i.e. has a characteristic length-scale much larger than that of the objects of interest, and the objects are bright compared with the background. A successful example of an object detection algorithm based on these assumptions is SExtractor [Bertin and Arnouts, 1996]. Its first step is to estimate the image background. The algorithm builds up an intensity histogram iteratively and clips it around its median. The resulting mesh (resembling a 'swiss-cheese') is then bilinearly interpolated to fill in the holes. After this background has been subtracted, the detection and characterisation of the objects is performed either by looking for sets of connected pixels above a given threshold or by boosting the image maxima with the help of an 'on-the-fly' convolution using a pre-defined kernel or the beam PSF.

Despite their general acceptance, such methods run into difficulties when the char-

[1] This is possible only if the fields are not very densely packed with objects.

acteristic extension of the fluctuations of the diffuse background match the size and the amplitudes of the objects of interest. Moreover, problems also arise when dealing with low or very low signal-to-noise-ratio (SNR) data, when the rms level of the background is comparable to, or even somewhat larger than, the amplitude of the localised objects of interest. A good example of this situation is the detection of the Sunyaev-Zel'dovich [Sunyaev and Zeldovich, 1972a] (SZ) effect in galaxy clusters, which have characteristic scales similar to that of the primordial CMB emission, and at the same time are very faint and extended. In such cases, traditional methods fail to provide a statistically supported prediction about the uncertainties on the parameter estimates.

The standard approach for dealing with such difficulties is to employ linear filtering, which is an extremely well-developed field, very firmly rooted in the principles of the orthodox school of statistics and signal processing [Trees, 2001]. These methods usually start by applying a linear filter $\psi(\boldsymbol{x})$ to the original image $d(\boldsymbol{x})$, and instead analyse the resulting filtered field. The filter is most often constructed by assuming a given (possibly parametrised) spatial template, $\tau(\boldsymbol{x})$, for the objects of interest. Depending on the application, this profile may contain parameters (to be estimated) and already include the the convolution with the instrumental beam. The common design goals for the filter follow the traditional, orthodox figures of merit: unbiasedness and efficiency. The optimal solution under these constraints is well-known to be the matched filter (MF) [North, 1963]. One may consider the filtering process as optimally boosting (in a linear sense) the signal from discrete objects, while simultaneously suppressing the emission from the background. The filtering methodology has yet another major advantage of being extremely fast and very simple to implement using 'off-the-shelf' routines (such as FFTs). The uncertainties in the parameter estimates are usually obtained from simulations. In practice, however, implementations of the filtering codes must be supported by ancillary steps in order to cope with the artefacts introduced as a consequence of the statistical description of the detection process being incomplete [López-Caniego et al., 2007; Melin et al., 2006]. A well known extension of the MF to spherical topologies was proposed by Tegmark and de Oliveira-Costa [1998]. A natural evolution of the MF, the matched multi-filter (MMF), follows exactly the same underlying principles and extends them to multi-channel data sets [Herranz et al., 2002b; Lanz et al., 2010]. The simultaneous multi-frequency analysis of a set of images has the immediate advantage of exploiting the objects' distinctive spectral signature, if any.

Two further advantages of this technique are: (i) it boosts the signal from the objects of interest simply by adding more data; and (ii) it improves the elimination of the background components by taking advantage of their correlation between channels. Once again, the thermal SZ effect embedded in primordial CMB emission provides a very good example. Owing to the well-defined and unique frequency dependence of the SZ effect, it is possible to design a filter that combines multi-frequency maps to make possible the extraction of deep catalogues even if the SZ component is sub-dominant in all the channels [Planck Collaboration et al., 2011f]. Herranz and Sanz [2008], suggested an extension to multi-channel analysis even when we are ignorant of the spectral signature of the targeted objects, the '*matched matrix filters*'. If the data can be assumed to be independent between channels then the matched matrix filters solution is just the co-addition of the matched filtered surfaces of each independent channel. Further development of traditional filtering techniques includes the 'Scale-Adaptive Filter' (SAF) [Herranz et al., 2002a; Sanz et al., 2001], in which the physical scale of the objects of interest is added as an extra degree of freedom and an additional condition for optimality is added in the derivation of the filter. Schäfer and Bartelmann [2007] generalised the scale-adaptive filter to the spherical topologies and added multi-channel support.

A very popular member of the filter family is the wavelets group, in particular the mexican-hat (MexHat) wavelet of order 2 and 1. Indeed, this MexHat wavelet is the MF or the SAF solution under particular assumptions about the statistical properties of the background and the objects profile [Sanz et al., 2001]. Since such conditions hold very well in modern cosmological data sets, such as those obtained from *WMAP* [Bennett et al., 2003] or *Planck* Planck Collaboration et al. [2011d], and the simplicity of the function allows easy and robust engineering, the MexHat wavelet has been a favourite detection tool of many authors [González-Nuevo et al., 2006; López-Caniego et al., 2006]. Nonetheless, obtaining good results with the MexHat filter is extremely dependent on the value of the acceptance/rejection threshold. The only way to ensure optimal performance is to run the code on realistic simulations and then assess the code's yield against the simulation's input catalogue, but a large number of runs is needed to fine-tune the threshold value. Exactly the same procedure must be followed to determine the uncertainties on the parameter estimates. This may not seem a severe limitation, since immense computing resources are now cheaply available. Given the increased level of accuracy and complexity of current cosmological data sets, however,

simulations must be rather sophisticated to provide a realistic test bed, and so even the enormous computational resources available are not sufficient to cope with the massive throughput demanded. For example, a single realistic *Planck* Full Focal Plane simulation (FFP) [Delabrouille et al., 2012; Reinecke et al., 2006] takes about one full week to run on a very large cluster and to have reasonable estimates of the parameter uncertainties and detection thresholds, at least several hundred independent simulations are needed.

To overcome these limitations of linear filtering methods, Hobson and McLachlan [2003] introduced a detection algorithm based on a Bayesian approach. As with the filtering techniques, the method assumed a parameterised form for the objects of interest, but the optimal values of these parameters, and their associated uncertainties, were obtained in a single step by evaluating their full posterior distribution. Another major advantage of this method is the consistent inclusion of physical priors on the parameters defining the objects and on the number of objects present, which improve the detection efficiency. Although this approach represented a further step in the direction of bringing a more solid statistical foundation to the object detection/characterization problem, its implementation was conducted using a Monte-Carlo Markov chain (MCMC) algorithm to sample from a very complex posterior distribution with variable dimensionality (dependent on the number of objects). This technique therefore proved extremely computational intensive. Despite the considerable progress that has recently been made towards increasing the efficiency of sampling-based Bayesian object detection methods [Feroz and Hobson, 2008], such algorithms are still substantially slower than simple linear filtering methods. In a recent work, Argüeso et al. [2011] suggested a semi-analytical hybrid Bayesian *maximum a posteriori* (MAP) scheme to overcome the complexity and the massive resources required the Hobson & McLachlan method. Its main advantage is a very simple and straightforward implementation. However, the method still relies on the MF to find the source's positions and, therefore, it does not embody the full Bayesian logic.

Meanwhile, Carvalho et al. [2009, 2012] and Feroz et al. [2009] have moved one step further towards the theoretically-optimal Bayesian solution by exploring the use of evidence ratio methods, which are the optimal decision-making tools (see section 2.2), rather than simply adopting the MAP solution.

Worthy of mention, although slightly off topic, is the literature on remote land

mines detection. Remote land mines detection principles follow the same type of statistical models, algorithms and solutions as the examples in this work. Svensson and Lundberg [2001] describes a Generalized Likelihood Ratio Test (GLRT) (see section 2.2.2) algorithm assuming a background and parametric object model along the lines of what we have used for point source detection (also see Lundberg [2001]). Anderson [2008] also suggests a GLRT-based procedure but using this time a non-parametric model and multi-wavelength observations. The images are modelled as Gaussian mixtures plus some 'anomalous' uniform pixel distribution representing the objects of interest. Finally, in Collins et al. [1999], the authors present a Bayesian algorithm, aiming to improve previous attempts based on the GLRT, but this time tackling known characteristic signal signatures in time-ordered data.

Our proposal here is to blend detection strategies, i.e. multi-channel filtering, Bayesian posterior sampling and evidence ratio evaluation, into a rigorous, hybrid, multi-model scheme (as opposed to traditional binary models). This novel methodology is simultaneously general, formally and statistically firmly grounded, and overcomes the computation inefficiencies of the pure sampling methodologies.

1.1 PowellSnakes (PwS) fundamentals

- ***Aim:***

  Detection and characterisation of localised objects embedded in a diffuse background.

- ***Concepts:***

  1. The only complete and coherent theory of inference in the presence of incomplete knowledge is Laplace-Bayes probability theory as an extension of logic [Jaynes, 2004].
  2. Detection is a decision theory problem; decisions ought to be taken. Owing to the presence of incomplete, undefined or noisy information the only possible way of supporting our decisions is through Bayesian probability theory.

3. Avoid heuristics whenever possible. Use only models based on trusted rigorous and well-established mathematical frameworks as this is the only possible way to achieve reliable control of results.
4. Detection and characterization of compact discrete objects is one of the most traditional and well-developed branches of observational astronomy. An immense repository of prior information is available. Therefore, in accordance with Laplace-Bayes probability theory, one should be allowed to use it in any eventuality.

- ***Methods:***

1. Always enforce statistical rigour. Never trade implementation simplicity for the fundamentals of probability theory.
2. Avoid as much as possible artificial parameter '*knobs*' in order to fine-tune the quality of the detection.
3. The algorithm must perform its tasks as fast as possible without compromising the quality of the detection.
4. The code must provide a full fledged professional solution to catalogue making, suitable to release science quality products, extracted from modern multi-channel multi-millon pixel cosmological data sets like those of the *Planck* or *WMAP* satellites. The implementation must be flexible enough to allow a gentle and stable integration into a 'pipeline' infrastructure.

Chapter 2

Bayesian inference and decision theory

2.1 Bayes theorem

The Bayesian system of inference is the only one that provides a consistent extension of deductive logic ($\{$ 0= false, 1= true$\}$) to a broader class of '*degrees-of-belief*' by mapping them into the real interval $[0, 1]$ [Jaynes, 2004]. Combining the multiplication rule together with the associativity and commutativity properties of the logical product, one may write the equation which will give us the posterior probability of a set of parameters $\boldsymbol{\Theta}$ taking into account the data $\boldsymbol{d}$, the underlying hypothesis H and the general 'background information' X[1]. This equation is also known as Bayes theorem

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, HX) = \frac{\Pr(\boldsymbol{d}|\,\boldsymbol{\Theta}, HX)\,\Pr(\boldsymbol{\Theta}|HX)}{\Pr(\boldsymbol{d}|HX)}, \tag{2.1}$$

where, for brevity, we denote $\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, HX) \equiv P(\boldsymbol{\Theta})$ as the posterior probability distribution of the parameters, $\Pr(\boldsymbol{d}|\,\boldsymbol{\Theta}, HX) \equiv \mathcal{L}(\boldsymbol{\Theta})$ as the likelihood, $\Pr(\boldsymbol{\Theta}|HX) \equiv \pi(\boldsymbol{\Theta})$ as the prior and $\Pr(\boldsymbol{d}|HX) \equiv \mathcal{Z}$ as the Bayesian evidence. The (unnormalised) posterior distribution is the complete inference of the parameter values $\boldsymbol{\Theta}$, and thus plays the central role in Bayesian parameter estimation.

The normalised posterior distribution may be easily obtained by integrating over all possible values of the parameters and equating the resulting expression to one, and

[1] The 'background information' comprises every assumption not explicitly stated, for instance the data model and the parametrisation of the data.

from this argument one can easily see that the evidence is given by

$$\mathcal{Z} \equiv \Pr(\boldsymbol{d}|HX) = \int \mathcal{L}(\boldsymbol{\Theta})\,\pi(\boldsymbol{\Theta})\,d^K\boldsymbol{\Theta}, \tag{2.2}$$

where K is the dimensionality of the parameter space. Inspecting this expression, one immediately recognizes that the evidence is the expectation of the likelihood over the prior, and hence is central to Bayesian model selection between different hypothesis H_i. We note that the evidence evaluation requires the prior to be properly normalised [Jaynes, 2004].

2.2 Decision theory

Probability theory defines only a state of knowledge: the posterior probabilities. There is nothing in probability theory *per se* that determines how to make decisions based on these probabilities. Indeed, a range of actions is always possible, even when using the same state of knowledge, because the cost of making a wrong decision usually changes according to the kind of problem under analysis. For example, in the case of object detection, one often considers each type of error, i.e. an undetected object or a spurious detection, as equally bad. For a moment, however, suppose we instead wished to determine whether or not a certain person was immune to a particular pathogen. Failing to detect a previously acquired immunity would only cost the price of an extra vaccine, but failing to determine that someone was not immune could seriously put her/his life at risk. Thus, even with the same degree of knowledge, the cost of choosing incorrectly is not the same in every case. To deal with such difficulties, one must apply Decision Theory (DT), which we now briefly summarise.

To apply decision theory, one must first define the *loss/cost function*

$$L(\boldsymbol{D}, \boldsymbol{E}), \tag{2.3}$$

for the problem at hand, where $\boldsymbol{D}$ is the set of possible decisions and $\boldsymbol{E}$ is the set of true values of the entities one is attempting to infer. In general, these entities can be either continuous parameters or discrete hypotheses, and so decision theory can be applied equally well both to parameter estimation and model selection. The loss

function simply maps the 'mistakes' in our estimations/selections, $\boldsymbol{D}$, into positive real values $L(\boldsymbol{D}, \boldsymbol{E})$, thereby defining the penalty one incurs when making wrong judgments. The Bayesian approach to decision theory is simply to minimise, with respect to $\boldsymbol{D}$, the expected loss

$$\langle L(\boldsymbol{D}, \boldsymbol{E}) \rangle = \iint L(\boldsymbol{D}, \boldsymbol{E}) \, \Pr(\boldsymbol{D}, \boldsymbol{E}) \, d\boldsymbol{D} \, d\boldsymbol{E}. \tag{2.4}$$

2.2.1 Parameter Inference (estimation)

In the estimation of a set of continuous parameters[1], the 'decisions' $\boldsymbol{D}$ are the parameter estimates $\widehat{\boldsymbol{\Theta}}$ and the 'entities' $\boldsymbol{E}$ are the true values $\boldsymbol{\Theta}^*$ of the parameters. Typically, the loss function is taken to be a function of the difference, or error, $\epsilon \equiv \widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*$. Some popular choices of loss functions are:

(i) the square error ϵ^2

(ii) the absolute error $|\epsilon|$

(iii) the uniform cost outside error bar, i.e. unity if $|\epsilon| > \Delta$ and zero if $|\epsilon| < \Delta$, where Δ is some pre-defined small quantity

In each case, one can easily find the optimal estimator by minimising the expected loss (2.4) with respect to $\widehat{\boldsymbol{\Theta}}$.

$$min_{|\widehat{\boldsymbol{\Theta}}} \, \langle L(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*) \rangle = min_{|\widehat{\boldsymbol{\Theta}}} \iint L(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*) \, \Pr(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}^*) \, d\widehat{\boldsymbol{\Theta}} \, d\boldsymbol{\Theta}^*. \tag{2.5}$$

The solutions are, respectively:

(i) the posterior mean[2].

(ii) the posterior median.

[1] In astronomical object detection, the majority of the interesting parameters (flux density, position, geometry, etc.) are continuous and real valued. Moreover, discrete parameters can always be handled within the same framework by resorting to Dirac delta functions.

[2] The posterior mean is the Bayesian optimal estimator under a very broad class of reasonable loss functions. When the posterior distribution is Gaussian all three common estimators match, and the posterior mode is often the simplest to compute.

(iii) the posterior mode[1].

The most popular choice of loss function among the astronomical community is the square error ϵ^2. When detecting astronomical objects, however, the requirements are usually not those of the square error function, which puts an extreme emphasis on values very far from the true ones. This extreme sensitivity to outliers[2] makes the posterior mean estimator less robust than, for example, the posterior median, which is much more resilient to deviations from the data model. An even better choice would be not to penalise the estimates at all if they fall within a small neighbourhood Δ around the true parameters values and prescribe a constant penalty otherwise. This is precisely the 'uniform cost inside error bar' loss function described above. This loss criterion closely matches what we would intuitively expect when assessing the quality of a detection algorithm. For example, if the estimated value of a source flux lies outside the allowed range it does not matter how far it lies from the true value, since it will always be counted as a spurious detection [Planck Collaboration et al., 2011e]. It is worth noting that, in contrast, the frequentist approach to statistics does not have any consistent and well-defined criterion for defining the optimality of an estimator. Instead, estimators are ranked in quality according to *ad hoc* criteria such as bias and efficiency. Indeed, it is well known that some unbiased estimators are sub-optimal in the decision-theory sense [Jaynes, 2004, ch. 17]. Their preferred tool is, most of the time, the MLE (Maximum Likelihood Estimator), which has good asymptotic properties [Trees, 2001, ch 2.4]. However, on real problems we are always dealing with finite data sets and sometimes even with problems where sufficient statistics do not exist. Even in these '*pathological cases*' cases, the Bayesian estimators keep their optimality untouched. For a classical example please refer to [Sivia and Skilling, 2006, ch. 2.4].

2.2.1.1 Interval estimation

In addition to an estimate $\hat{\boldsymbol{\Theta}}$, one typically summarises the inference implied by the full posterior distribution by quoting either joint (two-dimensional) or marginalised confidence intervals (or, more precisely, Bayesian credible intervals). One could, in

[1] This is only strictly true when $\Delta \longrightarrow 0$

[2] By *Outliers* we mean, values very far from the true ones, that do not follow the statistical data model we are employing.

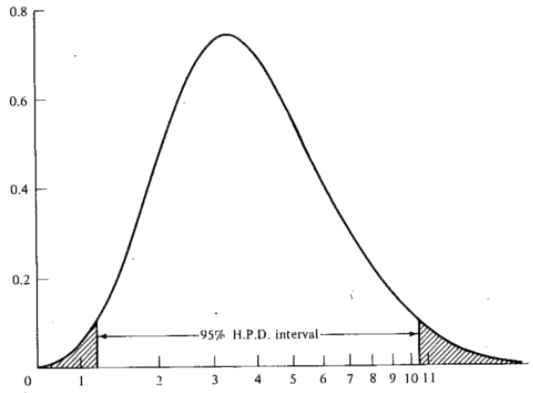


Figure 2.1: A 95% probability content HPD interval of a unimodal posterior distribution.

principle, obtain an optimal interval by employing an appropriate loss function, but a simpler approach is now widely accepted, namely the 'Highest Probability Density' interval HPD. The HPD interval containing the fraction α of the total probability is defined such that: $\Pr(\boldsymbol{\Theta} \in \text{HPD}|\boldsymbol{d}, HX) = \alpha$ and, if $\boldsymbol{\Theta}_1 \in \text{HPD}$ and $\boldsymbol{\Theta}_2 \notin \text{HPD}$, then $\Pr(\boldsymbol{\Theta}_1|\boldsymbol{d}, HX) \geq \Pr(\boldsymbol{\Theta}_2|\boldsymbol{d}, HX)$ (figure 2.1; see G. E. Box [1992, ch. 2]). The characterization of the HPD interval may be easily obtained by sampling from the posterior distribution or, when the dimensionality of the parameter space is low (≤ 4), by making a 'brute-force' evaluation over a grid spanning the parameter space where the posterior distribution is significantly different from zero. When the posterior distribution is known to be Gaussian or close to it, which is a very common case, the $\pm$rms interval is usually quoted instead. If the parameter space is discrete, as result of the intrinsically discrete nature of the data (e.g. pixelisation) or the processing pipeline, in practice the data quantum imposes a minimum limit on the accuracy of the more robust posterior mode or median estimators. But, since the Bayesian methodology gives us the full posterior distributions, a better recipe for hyper-resolution estimates (sub data quantum accuracy) would be to compute the expectation estimator over HPDs of different probability content (see section 3.4.2.2).

2.2.2 Model selection and catalogue making

In model selection, the decision theory 'entities' $\boldsymbol{E}$ are the hypotheses under consideration and the 'decisions' $\boldsymbol{D}$ are the chosen hypotheses, such that $L(D_i, H_j) \equiv L_{ij}$ is the loss associated with the decision $D_i \equiv$ choose H_i, when H_j is true;

$$L(D_i, H_i) \equiv \left\{ \begin{array}{c|cccc} & H_0 & H_1 & \dots & H_n \\ \hline D_0 & L_{00} & L_{01} & \dots & L_{0n} \\ D_1 & L_{10} & L_{11} & \dots & L_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_n & L_{n0} & L_{n1} & \dots & L_{nn}. \end{array} \right. \tag{2.6}$$

The main diagonal entries of the 'Loss' matrix ($L_{i,i}$) are always equal to zero because when a right decision is taken it does not incur any penalty. However, in general, $L_{i,j} \neq L_{i,j}$[1]. This kind of exercise is more generally known as a classification problem. Thus, inserting this form of the loss matrix into the right hand side of equation 2.4 and performing the integration using the delta Dirac functions to represent discrete values as infinite densities, the average loss reads

$$\langle L(\boldsymbol{D}, \boldsymbol{H}) \rangle = \sum_{ij} L_{ij} \ \Pr(D_i, H_j). \tag{2.7}$$

Minimizing 2.7 with respect to $\boldsymbol{D}$ is not a difficult task, but the general case leads to long and cumbersome expressions (see [Trees, 2001, ch. 2]).

Usually the constraints we use to construct a catalogue of objects in astronomy constitutes only a sub-set of the general classification problem with the following template

[1] see the example in the beginning of the current section.

loss matrix,

$$\begin{cases} L_{ii} & : (D_i \equiv \text{choose } H_i) \text{ when } H_i \text{ is true} \\ & = 0\text{; no penalty when you are right} \\ L_{1i} \quad i \neq 1 & : (D_1 \equiv \text{choose } H_1) \text{ when } H_i \text{ is true} \\ & \text{positive value; false positive or '}\textit{spurious}\text{'} \\ L_{i1} \quad i \neq 1 & : (D_i \equiv \text{choose } H_i) \text{ when } H_1 \text{ is true} \\ & \text{positive value; false negative '}\textit{miss}\text{'} \\ L_{ij} \quad i,j \neq 1 & : (D_i \equiv \text{choose } H_i) \text{ when } H_j \text{ is true} \\ & \text{usually = 0; alternative model selection error,} \end{cases} \tag{2.8}$$

where H_1, the '*target hypothesis*', is intentionally discriminated from the others. Later, H_0 will be taken, without any loss of generality, as the '*null/empty*' ($H_0 \equiv$ only background is present) or '*reference hypothesis*'.

Much simpler and enlightening, but still capable of a very broad and interesting range of practical applications, is the binary case consisting of just two hypotheses H_0 and H_1[1]. In this case, the decision criterion that minimises the expected loss is

$$\ln \left[\frac{\Pr(H_1|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})} \right] \underset{H_0}{\overset{H_1}{\gtrless}} \xi \tag{2.9}$$

where $\xi \equiv \ln \frac{L_{10}}{L_{01}}$. The ratio of the posterior distributions of the models given the data is a very popular quantity in the Bayesian probability and is usually known as the $odds \equiv \frac{\Pr(H_1|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}$[2]. Jaynes asserts that the best way to decide between two models is by computing the $odds$ and check it against a threshold. Using DT we could recover that result and at the same time give it a precise statistical sense and define a threshold for decision making based on the loss criterion:

- Choose H_1 if our decision criterion is larger than the threshold level ξ
- Choose H_0 if our decision criterion is smaller than the threshold level ξ
- Choose either at random, with equal probabilities, if the decision criterion is

[1] Later will become apparent why the simple binary decision model is not sufficient for optimal catalogue construction (please see 3.7).

[2] Although this definition of $odds$ is commonly found in literature, this is not the most general case (please refer to equation 2.13).

equal to the threshold.

Unfortunately, in astronomy many times it is not possible to assign meaningful values to the loss. Astronomers like to define the quality of the catalogues based on expected/maximum contamination (‘*false positive error*’) and expected/minimum completeness (1 - ‘*miss error*’). Of course, there is certainly a connection between these two methods, although quantifying it is not simple.

The techniques of the orthodox school match much closer these requirements, in particular the widely employed Neymann-Pearson likelihood ratio test [Neyman and Pearson, 1933]. The test prescribes the probability for a certain type of error, say the false positive ($\Pr(D_1|H_0)$) and then tries to find the best possible way of deciding between the two hypothesis by minimising the other error, ‘the miss’ ($\Pr(D_0|H_1)$). The solution is the very well known NP likelihood-ratio test:

$$\ln\left(\frac{\mathcal{L}_1}{\mathcal{L}_0}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \xi', \tag{2.10}$$

where $\mathcal{L}_{\{0,1\}}$ are the likelihoods of the hypotheses being tested and ξ' the threshold for acceptance/rejection. As a figure of merit for a certain test, they draw its Receiver Operating Curve (POC) and then they choose the threshold that best suits their purposes (see Trees [2001, ch. 2]). When the hypotheses do not depend on any parameter, the ‘*simple hypothesis*’, this is usually straightforward. However, in astronomy, most of our problems, if not all, depend on unknown parameters whose true values may take many different values inside a range: the ‘*intrinsic variability*’. These kind of problems are known as ‘*composite hypotheses*’. The best practice for this kind of problem is to derive sufficient statistics for the parameter estimates using a maximum-likelihood procedure first and then place them into the likelihood ratio and finally describe the ROC. Thus one considers

$$\ln\left(\frac{\mathcal{L}_{1max}}{\mathcal{L}_{0max}}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \xi'. \tag{2.11}$$

This is the *Generalized Likelihood Ratio Test* (GLRT) [Trees, 2001, ch. 2]. The problem is that drawing the ROC of the GLRT can be very difficult without resorting to MC simulations, except in very simple cases (see section 2.3).

Returning to the Bayesian solution and using Bayes theorem, expression 2.9 may

still be given a more familiar look:

$$\ln\left(\frac{\mathcal{Z}_1}{\mathcal{Z}_0}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \xi'' \tag{2.12}$$

where $\xi'' \equiv \ln\frac{L_{10}}{L_{01}} + \ln\left(\frac{\Pr(H_0)}{\Pr(H_1)}\right)$ is the detection level. When the hypothesis are simple then $\mathcal{Z}_i = \mathcal{L}_i$ and the selection expressions are '*almost*' equal. However, these small differences might play a very important role in detection.

- The Bayesian probability makes no distinction between composite or simple hypothesis, as both are uniformly included in the formulation. That is certainly not true for the NP test. The NP test can be extended to deal with composite hypothesis (a necessary condition for astronomy), but seldom a workable solution is feasible without resorting to simulations.
- Failing to realise the importance of the prior term, $\frac{\Pr(H_0)}{\Pr(H_1)}$, and relying only on the ROC to choose a detection threshold might lead to very wrong results. For an example, please consult [K. Riley, 2006, ch. 30, pag. 1132].
- Within the orthodox school there is no simple way to define the threshold level based on cost/loss.

2.2.2.1 Classification - extending to multiple hypotheses

Unfortunately, as we shall see (section 3.7) the binary decision case is not powerful enough to tackle realistic modern astronomical data sets as, most of the time, these cannot be accurately modelled by a simple mixture of the interesting signal embedded in a statistically uniform background. There are other localised 'background accidents', originating from different families of sources, which do follow neither the source nor the background model. The loss matrix template (equation 2.8) has a kind of symmetry that allows us to express the optimal decision rule as the ratio of

$$\ln\left[\frac{\Pr(H_1|\boldsymbol{d})}{\Pr(\widetilde{H_1}|\boldsymbol{d})}\right] \underset{\widetilde{H_1}}{\overset{H_1}{\gtrless}} \xi, \tag{2.13}$$

the ratio of the posterior probability of the target hypothesis against its complement ($\widetilde{H_1}$)[1]. If the same loss matrix template applies equally well to any source population loss, then by iterating through all object hypothesis, using formula 2.13, we are able to assign each individual source to a catalogue or 'classify' them. In case an object appears in more than one catalogue, choose that one where the left hand side of 2.13 is larger.

A practical way of tackling it, would start by evaluating the posterior ratio of each object model (H_j) against that of the background (H_0) and pick up the model with the largest ratio j^*.

$$\varrho_{j^*} \equiv \left[\frac{\Pr(H_{j^*}|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right], \quad j^* \neq 0 \tag{2.14}$$

Then following an argument of insufficient reason, we may state

> *Since the losses are the same regardless of the hypothesis ($j \neq 0$), why not choosing the one with the highest posterior ratio would help decrease the loss?* (2.15)

So, using equations 2.13 and 2.14 the classification condition reads

$$\frac{\varrho_{j^*}}{1+\sum_{i\neq j^*}\varrho_i} \underset{H_0}{\overset{H_{j^*}}{\gtrless}} \xi. \tag{2.16}$$

In the multiple option case, if the preferred option cannot reach the desired acceptance/rejection level, the putative detection will be attributed to the 'null', even if this was not the most favoured. Now imagine that the option with the highest ϱ_{j^*} has not passed the test in equation 2.16. So, we need to select a member of the complementary set of hypotheses. But which one? Individually, any of the hypothesis of the complement was already deprecated. Therefore the least risky is that with the lowest loss. Assigning to the 'null', you 'pay' for a 'miss' only. If you had instead assigned to any other hypothesis member of the complement set, you would be adding to the 'miss' loss that of a 'spurious'. So far, we have only been using our intuition guided by an educated application of the Bayesian logic. However, collecting the above considerations about reasonable loss choices, we have just defined the following loss matrix,

[1]This is the most general definition of 'odds' (see [Jaynes, 2004, ch. 4]). Choosing $\widetilde{H_1}$ means leaving the putative detection out of the targeted catalogue.

$$L_{ij} = \begin{cases} 0 & \text{if } i = j \text{ (no penalty if correct)} \\ L_{\text{miss}}(> 0) & \text{if } i = 0,\ j \neq 0 \text{ (miss)} \\ L_{\text{spurious}}(> 0) & \text{if } i \neq 0,\ j = 0 \text{ (spurious)} \\ L_{\text{miss}} + L_{\text{spurious}} & \text{if } i \neq 0,\ j \neq 0,\ i \neq j \text{ (spurious+miss)} \end{cases} \tag{2.17}$$

Placing definition 2.17 into equation 2.7 and minimizing with respect to the decisions, the result is a large and complex system of decision inequalities. Although, taking advantage of the above loss matrix symmetries, it is possible to reduce it to a much simpler form, without any loss of generality, with just 3 inequalities, of which only the following one is relevant.

$$\Upsilon \equiv \frac{\varrho_{(j\neq 0)^*}}{1 + \varrho_{\widetilde{(j\neq 0)^*\neq 0}}} \underset{H_{\widetilde{(j\neq 0)^*\neq 0}} \vee H_0}{\overset{\widetilde{H_0}}{\gtrless}} \xi, \tag{2.18}$$

where $H_{(j\neq 0)^*}$ is the hypothesis with the highest odds ratio ($\varrho_{(j\neq 0)^*}$) not including the 'null' (H_0), $H_{\widetilde{(j\neq 0)^*\neq 0}}$ is its complement, once again not including H_0, and $\xi = \frac{L_{\text{spurious}}}{L_{\text{miss}}}$. Two different situations may happen.

- $\Upsilon > \xi$
  Then we ought to choose something outside the 'null'. Using principle 2.15 only $(j \neq 0)^*$ is a possible choice.
- $\Upsilon \leq \xi$
  Of the two possible options only H_0 is allowed because principle 2.15 forbids any other possible choice.

The net result of the above conditions is inequality 2.16 as we intended to prove.

2.3 Catalogue figures of merit: Purity & Completeness

In astronomy it is often not possible to choose meaningful values for the loss, making the evaluation of the threshold ξ difficult. Astronomers like instead to measure the quality of a catalogue in terms of the expected/maximum contamination (false positive

rate) and the expected/minimum completeness (true positive rate). There is a connection between this approach and DT, but quantifying it is not trivial. Thus, in practice and from the point of view of detection performance, the two single most important figures of merit of a catalogue, expected to contain sources of a single family j, are

- *Purity*[1]:
  The percentage of real detections[2] over the total number of entries in the catalogue. For a Bayesian this is just another random parameter we can infer from the posterior distribution. In frequentist probability language it reads $\Pr(H_j|D_j)$. The definition may be complemented by considering only subsets of the catalogue entries, usually by thresholding the '*estimated*' object amplitude/flux/SNR ($\Pr(H_j|D_j, \widehat{S} > \xi)$).

- *Completeness*[3]:
  The percentage of objects, above a certain '*true*' amplitude/flux/SNR in our data, that we successfully included in j catalogue. In frequentist probability language, which we will follow in this case, it reads
  $\Pr(D_j|H_j, S > \xi)$[4].

2.3.1 Purity

From the basic axioms of probability theory [Jaynes, 2004, ch. 2,3], one may write

$$\Pr(H_j|\boldsymbol{d}) + \Pr(\widetilde{H_j}|\boldsymbol{d}) = 1,$$

$$\Pr(\widetilde{H_j}|\boldsymbol{d}) = \frac{\Pr(\widetilde{H_j}|\boldsymbol{d})}{\Pr(H_j|\boldsymbol{d}) + \Pr(\widetilde{H_j}|\boldsymbol{d})} = \frac{1}{1 + \frac{\Pr(H_j|\boldsymbol{d})}{\Pr(\widetilde{H_j}|\boldsymbol{d})}}. \tag{2.19}$$

[1] Some authors prefer 'Reliability'

[2] The definition of a '*real*' detection is not universally accepted. Often extra conditions on the quality of the estimation apply, for instance, that the flux estimate is inside a certain range of values around the true value.

[3] This probability in frequentist language is the '*Power of the test*'.

[4] The threshold is now a required part of the definition.

We then apply the above equation to each individual entry in a catalogue. So, the probability of the line i being a 'spurious' is given by,

$$p_i^{sp} = \frac{1}{1+\psi}, \quad \psi \equiv \frac{\Pr_i(H_j|\boldsymbol{d})}{\Pr_i(\widetilde{H_j}|\boldsymbol{d})} = \frac{\varrho_j}{1+\sum_{i\neq j}\varrho_i}. \tag{2.20}$$

We shall see (section 3.4) that the quantities necessary to evaluate expression 2.20 are all provided by our algorithm.

The number of false positives in a catalogue N_{sp}, may be represented as a sum of Bernoulli variables. Assuming all catalogue entries are statistically independent, then the sum of N of those variables is distributed as a Poisson–binomial distribution:

$$\mu = \sum_{i=1}^{n} p_i^{sp}, \quad \sigma^2 = \sum_{i=1}^{n} p_i^{sp}(1-p_i^{sp}), \tag{2.21}$$

where μ is the expected value and σ^2 the variance. To generate the most complete (largest) catalogue constraining the expected value of contamination $(= 1 - \text{Purity})$ to a prescribed value α, one way to proceed is as follows:

1. sort the list of candidate detections in p_i^{sp} ascending order (equation 2.20);
2. for each candidate, accumulate p_i^{sp} until μ (equation 2.21) exceeds the prescribed contamination α times the total number of lines already included
3. discard the last line.

Since μ is a sum of independent variables and N is usually a large number (hundreds), it is perfectly reasonable to assume the posterior distribution of N_{sp} converged to a Gaussian as result of the Central Limit Theorem (CLT)[1]. So, a good estimate of the number of spurious detections in the catalogue is

$$\Pr(N_{sp}|\boldsymbol{d}) \sim N\left(\sum_{i=1}^{N} p_i^{sp}, \sqrt{\sum_{i=1}^{N} p_i^{sp}(1-p_i^{sp})}\right), \tag{2.22}$$

[1]The approximation is perfectly justified as we are interested in the distribution mode neighbourhood only.

where $N(\mu, \sigma)$ is the Normal distribution centred in μ with variance equal to σ^2. An estimate of the fraction of spurious detections in the catalogue $\widehat{\alpha}$, reads

$$\widehat{\alpha} = \frac{\sum_{i=1}^{N} p_i^{sp}}{N} \pm \frac{\sqrt{\sum_{i=1}^{N} p_i^{sp}(1 - p_i^{sp})}}{N}. \tag{2.23}$$

A problem still remains since our calculation of $\Pr_i(\widetilde{H_j}|D_j)$ is only an approximation. Although, the algorithm also computes an estimate of the $\ln\left[\frac{\Pr(H_j|\boldsymbol{d})}{\Pr(\widetilde{H_j}|\boldsymbol{d})}\right]$ uncertainty (for a rigorous treatment see Keeton [2011]). Therefore, the above formulas ought to be corrected to account for the uncertainty on p_i^{sp}. It is easy to verify that, to a first approximation, the error on p_i^{sp}, reads

$$|\Delta p_i^{sp}| \simeq \gamma p_i^{sp}(1 - p_i^{sp}) \tag{2.24}$$

and γ is given by

$$\gamma = \sqrt{\sum_i \gamma_i^2}, \tag{2.25}$$

where γ_i is each hypothesis fractional uncertainty on the evaluation of the evidence ratio and the sum extends to all alternatives but the null. The corrected value of the catalogue's variance on the number of spurious, σ'^2, is always less than:

$$\sigma'^2 \lesssim (1 + \gamma) \sum_{i=1}^{n} p_i^{sp}(1 - p_i^{sp}), \tag{2.26}$$

and the variance on μ reads:

$$|\Delta\mu|^2 \simeq \gamma^2 \sum_{i=1}^{n} (p_i^{sp})^2(1 - p_i^{sp})^2 < \gamma \sum_{i=1}^{n} p_i^{sp}(1 - p_i^{sp}) \tag{2.27}$$

Thus, we get the final expression of predicted contamination of the catalogue by adding both contributions in quadrature:

$$\widehat{\alpha} = \frac{\sum_{i=1}^{N} p_i^{sp}}{N} \pm \frac{\sqrt{1 + 2\gamma}\sqrt{\sum_{i=1}^{N} p_i^{sp}(1 - p_i^{sp})}}{N}. \tag{2.28}$$

The uncertainty on the contamination of the catalogue for commonly accepted levels

($\sim 10\%$), catalogue size ($\gtrsim 1000$) and γ as large as 0.32[1], is always $\lesssim 1.2\%$.

2.3.2 Completeness

Completeness is more closely related with the likelihood and a frequentist approach is somewhat less complicated. When the detections are very clear (high SNR), the uncertainties on the unknown parameters may be assumed small enough that one may accept that the estimates match the true parameter values with sufficient accuracy to consider them as 'previously known'. This case, without unknown parameters (unknown parameters are non-random variables in this context), is named a '*simple hypothesis test*' and applying the Neymann-Pearson likelihood ratio test (see section 2.2.2) simple and straightforward analytical solutions exist [Trees, 2001, ch. 2]. This is the best possible case scenario and it is usually employed to define an upper limit on the catalogue completeness[2]. Astronomers find that a good approximation for compact sources may be achieved considering the position and the geometry of the source as perfectly known and only allow the amplitude/flux/SNR to freely vary. In this case the NP test becomes:

(i) one-sided,

(ii) the test distribution belongs to the exponential family,

(iii) the likelihood ratio is monotonic.

This set of conditions guarantees an Uniformly Most Powerful (UMP) test exists with significance α and 'Power' P. Translating from the frequentist dialect into those of the astronomers, this means that one ***single*** amplitude/flux/SNR threshold is enough to secure a catalogue has a certain purity α and completeness P. These tests run into difficulties as soon as more realistic models, with a larger set of uncertainties considered, are essential to describe the data with the level of accuracy we want. Unfortunately, this is an important limitation. Even the crudest descriptions of clusters of galaxies cannot simply be modelled as point sources. A parameter describing the extent of the

[1] This value is an upper bound of the average posterior ratio uncertainty, computed by the method given in Sivia and Skilling [2006] and obtained on the extraction exercises we have performed with *Planck* data (please see 4.2).

[2] A realistic completeness prediction in more authentic scenarios, when the estimates uncertainties need to be considered, is usually significantly lower [Mooney et al., 1999].

cluster ought to be included[1]. But with extended objects, the conditions for a UMP do not hold anymore. The acceptance/rejection threshold now depends, at least, on the geometrical parameter that controls the extension of the objects and the completeness estimate of the sample is no longer trivial[2].

Recalling the equivalence between Bayesian posterior probabilities, when using non-informative priors, and the equivalent sampling distributions from the orthodox theory [G. E. Box, 1992, ch. 2], we define a new variable

$$\widehat{\zeta} = \frac{\widehat{A} - A}{\sigma_A} \sim N(0,1), \tag{2.29}$$

where A is the true value of the source amplitude, assumed non-random in this context, and σ_A is the variance of the random variable $\widehat{A}$ defined in (3.29). The $\widehat{\zeta}$ statistic is normally distributed (see equation 3.35). Now let us define $\gamma\,\sigma_A$ as the amplitude threshold for rejection/acceptance. So the completeness, $\Pr(D_j|H_j)$, reads

$$\int_{\gamma-\frac{A}{\sigma_A}}^{+\infty} N(0,1)\; d\zeta = \frac{1+\operatorname{erf}\left[\frac{\sqrt{2}}{2}\left(\frac{A}{\sigma_A}-\gamma\right)\right]}{2}, \tag{2.30}$$

where $\operatorname{erf}()$ is the Gaussian error function.

Apparently it seems PwS still relies on a SNR cut to predict the completeness of the catalogue. That means we would be throwing away most of the advantage of using a Bayesian detection method, as any catalogue is essentially useless unless it provides a measure of how representative it is of the population of the objects under scrutiny. This difficulty only emerged because of the simple Bayesian model we have used to describe the data. The way it is defined, completeness is a construct that allow us to assign probabilities to abundances of sources following an orthodox statistics methodology. However, the same source number counts can be directly inferred using a Bayesian framework. In this case we need to describe the data using a hierarchical model (see [Gelman et al., 2003, ch. 5] and [G. E. Box, 1992, ch. 5]). This will be the main focus of development in a forthcoming publication. Fortunately this apparent

[1] Please refer to chapter 5 for a full account.

[2] If the problem is a composite hypotheses linear on all unknown parameters and with measurement errors following a Gaussian law, then there are still full analytical GLRT solutions (see Mooney et al. [1999]).

limitation may be easily overcome by very realistic simulations that at the same time are simple and fast to construct: injecting mock source populations into the real maps and then recovering them. These simulations might not be optimal to assess purity but they provide the most realistic possible test bed for measuring completeness [Planck Collaboration et al., 2011e].

Chapter 3

Bayesian object detection in multi-frequency astronomical data sets

In the previous chapter we gave a brief summary of the probabilistic and logical Bayesian foundations will support PwS '*inference machine*'. This chapter will be devoted, first, to develop each element of the Bayesian inference framework in the context of the current work. Second, exploiting the symmetries and the particularities of this problem, suggest a fast algorithm, and code implementation, to successfully cope with a real, large and modern data set like that the *Planck* satellite will deliver. Finally, we will try to address how real data actually deviates from the assumed model, recommend an arrangement to tackle the additional difficulties and suggest a strategy for a successful, professional code to produce science quality catalogues.

3.1 Data model

The specification of the PwS statistical model for a single-frequency observation of localised objects embedded in a background is given in Carvalho et al. [2009]. This can be straightforwardly extended to accommodate multi-frequency observations. At each observing frequency ν, PwS treats the observed data $d_\nu(\boldsymbol{x})$, where $\boldsymbol{x}$ is the position vector in pixel space, as the superposition of a 'generalised' noise background $n'_\nu(\boldsymbol{x}) = b_\nu(\boldsymbol{x}) + n_\nu(\boldsymbol{x})$, consisting of background sky emission $b_\nu(\boldsymbol{x})$ and instrumental noise $n_\nu(\boldsymbol{x})$, plus a characteristic signal $s_\nu(\boldsymbol{x})$ coming from the sources. For ease

of notation, we will collect the fields at different frequencies into vectors. Moreover, the signal and background components in each frequency channel are assumed to have been smoothed with a known beam, which may differ between channels. The resulting model for the data vector $\boldsymbol{d}(\boldsymbol{x})$ reads

$$\boldsymbol{d}(\boldsymbol{x}) = \sum_{j=1}^{N_\mathrm{s}} \boldsymbol{s}_j(\boldsymbol{x}; \boldsymbol{\Theta}_j) + \boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{n}(\boldsymbol{x}), \tag{3.1}$$

where N_s is the number of sources, $\boldsymbol{s}_j(\boldsymbol{x}; \boldsymbol{\Theta}_j)$ is the signal vector due to the jth source, which depends on the parameter vector $\boldsymbol{\Theta}_j$ characterising the object, $\boldsymbol{b}(\boldsymbol{x})$ is the signal vector due to the diffuse astronomical backgrounds, and $\boldsymbol{n}(\boldsymbol{x})$ is the instrumental noise vector. The astronomical backgrounds denoted by $\boldsymbol{b}(\boldsymbol{x})$ are expected to exhibit strong correlations between different frequency channels, whereas the instrumental noise $\boldsymbol{n}(\boldsymbol{x})$ is expected to be uncorrelated between frequency channels, and also between pixels in the case of simple white noise[1].

We write the signal vector due to the jth source in 3.1 as

$$\boldsymbol{s}_j(\boldsymbol{x}; \boldsymbol{\Theta}_j) = A_j \boldsymbol{f}(\boldsymbol{\phi}_j) \boldsymbol{\tau}(\boldsymbol{x} - \boldsymbol{X}_j; \boldsymbol{a}_j), \tag{3.2}$$

where the vector $\tau(\boldsymbol{x} - \boldsymbol{X}_j; \boldsymbol{a}_j)$ denotes the convolved spatial template at each frequency of a source centred at the position $\boldsymbol{X}_j$ and characterised by the shape parameter vector $\boldsymbol{a}_j$, the vector $\boldsymbol{f}$ contains the emission coefficients at each frequency, which depend on the emission law parameter vector $\boldsymbol{\phi}_j$ of the source (see below), and A_j is an overall amplitude for the source at some chosen reference frequency. Thus, the parameters to be determined for the jth source are its overall amplitude, position, shape parameters and emission law parameters, which we denote collectively by $\boldsymbol{\Theta}_j = \{A_j, \boldsymbol{X}_j, \boldsymbol{a}_j, \boldsymbol{\phi}_j\}$. The totality of these parameters, for all the sources present, plus the number of sources N_s, are concatenated into the single parameter vector $\boldsymbol{\Theta}$.

[1] The condition of the instrumental noise being white is not necessary. The general case of correlated noise between pixels does not complicate the mathematical development, but can increase computational expense. In any case, the assumption of white noise applies extremely well to *Planck* data on the small scales of interest used for the identification of localised objects.

For convenience, we denote the signal vector generated by all the sources by

$$\boldsymbol{s}(\boldsymbol{x};\boldsymbol{\Theta}) \equiv \sum_{j=1}^{N_\mathrm{s}} \boldsymbol{s}_j(\boldsymbol{x};\boldsymbol{\Theta}_j). \tag{3.3}$$

The nature of the emission law parameter vector $\boldsymbol{\phi}$ depends on the class of object under consideration. PwS analyses the data assuming that all the objects belong to a single class, and repeats the analysis for each class of interest. The assignment of individual sources to a particular class is then performed via a model selection step (see section 3.7). The number and specification of classes can be arbitrary, including, for example, SZ clusters, point sources, etc. Previous multi-frequency versions of PwS have been limited to the case where all objects share the same, fixed emission law. SZ clusters fall exactly in this category as, ignoring the relativistic corrections, they all follow exactly the same spectral signature [Birkinshaw, 1999; Carlstrom et al., 2002], which does not depend on any parameters. For extragalactic point sources, however, the emission law is phenomenological and can vary between sources. Consequently, PwSII has been extended to accommodate such cases. For example, two important families of extragalactic point sources in *Planck* data are as follows.

- *Radio sources* are the dominant family of point sources for all *Planck* channels up to and including 217 GHz. Based on the work of Planck Collaboration et al. [2011c]; Waldram et al. [2007], we assume an emission law for such objects of the form

$$\ln f_\nu = \alpha \, \ln\left(\frac{\nu}{\nu_0}\right) + \beta \, \left[\ln\left(\frac{\nu}{\nu_0}\right)\right]^2, \tag{3.4}$$

  where $\boldsymbol{\phi} = \{\alpha, \beta\}$ are spectral parameters that can vary between sources, and ν_0 is the reference frequency (note that $f_\nu = 1$ at $\nu = \nu_0$). Setting $\beta = 0$, recovers the commonly-assumed power-law spectral behaviour with spectral index α. The more general form (equation 3.4) accommodates most of the common types of radio-source spectra, namely: flat, steep, and inverted.

- *Dusty galaxies* dominate the *Planck* highest frequency channels, starting at 217 GHz up to 857 GHz. Their spectral behaviour may be represented to very good

accuracy using the well-known greybody model

$$\ln f_\nu = \beta \ln\left(\frac{\nu}{\nu_0}\right) + \ln\left[\frac{B_\nu(T)}{B_{\nu_0}(T)}\right], \tag{3.5}$$

where the spectral parameters $\boldsymbol{\phi} = \{\beta, T\}$ are the dust emissivity and temperature respectively, $B_\nu(T)$ is the Planck law of blackbody radiation and ν_0 is once again the reference frequency [Serjeant and Harrison, 2005]. We have again normalised 3.5 such that $f_\nu = 1$ at $\nu = \nu_0$.

- *New/unexpected* sources can be easily accommodated inside the current framework by allowing the emission coefficient f_{ν_i} at each frequency to be a free parameter

$$\boldsymbol{f}_\nu = [f_{\nu_1}, \cdots, f_{\nu_n}]. \tag{3.6}$$

Such a SED model is certainly more generally applicable because it contains the other two models. This 'generic' model may be used with great advantage to prevent data artefacts being mistaken for a genuine source, although it may be disfavoured by the evidence because of its larger number of parameters (see section 3.7).

3.2 The likelihood

The likelihood is the probability of a data set conditional on the model that describes it H_j. Additionally, the model might be a function of the source parameters and the background properties.

$$\mathcal{L}\{H_j(\boldsymbol{s}(\boldsymbol{\Theta}), \boldsymbol{b})\} \equiv \Pr\{\boldsymbol{d}|H_j(\boldsymbol{s}(\boldsymbol{\Theta}), \boldsymbol{b})\}. \tag{3.7}$$

The form of the likelihood is determined by the statistical properties of the noise (background sky emission plus instrumental noise) in each frequency channel.

3.2.1 Single frequency

For clarity, before addressing the full multi-channel problem let us start, with the more straightforward single frequency case. Assuming the instrumental noise $\boldsymbol{n}(\boldsymbol{x})$, is a stationary Gaussian process then the likelihood reads:

$$\mathcal{L}\{H(\boldsymbol{s}(\boldsymbol{\Theta}),\boldsymbol{b})\} = \frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{d}-\boldsymbol{b}-\boldsymbol{s}(\boldsymbol{\Theta})\right]^t \boldsymbol{N}^{-1}\left[\boldsymbol{d}-\boldsymbol{b}-\boldsymbol{s}(\boldsymbol{\Theta})\right]\right\}}{(2\pi)^{N_{pix}/2}\left|\boldsymbol{N}\right|^{1/2}}. \tag{3.8}$$

where $\boldsymbol{N}$ is the covariance matrix $\boldsymbol{N} \equiv \langle \boldsymbol{n}\boldsymbol{n}^t \rangle$ and $N_{\mathrm{p}ix}$ the total number of pixels in the image. As we perform our analysis in small patches of sky, it is not unreasonable to assume the background is statistically homogeneous[1]. The $n \times n$ ($= \boldsymbol{N}_{\mathrm{p}ix}$) covariance matrix of a homogeneous background is of '*circulant*' type. A very important property of these matrices is that their eigenvectors, $\boldsymbol{\psi}_{u,v}$, are the Fourier modes

$$\begin{gathered}\boldsymbol{\psi}_{u,v} = \frac{1}{n}\left[\mathrm{e}^{\frac{-\mathrm{i}2\pi uvjk}{n^2}}, \ldots, \mathrm{e}^{\frac{-\mathrm{i}2\pi uv(n-1)(n-1)}{n^2}}\right], \\ \psi_{u,v} = \sum_{j,k=0}^{n-1} c_{(k-j)}\, \mathrm{e}^{\frac{-\mathrm{i}2\pi uvjk}{n}}\end{gathered} \tag{3.9}$$

where $j,k,u,v \in \{0,\ldots,n-1\}$, $c_{(k-j)}$ is the element $k-j$ of the covariance matrix $\langle \boldsymbol{n}\boldsymbol{n}^t \rangle$ and $\psi_{u,v}$, the eigenmodes set, is the well known '*power spectrum*'. So, rotating every quantity to the Fourier domain, even a non-diagonal covariance matrix (with spatial correlation), is transformed to diagonal, decoupling the Fourier modes and the full likelihood expression becomes the product of each Fourier mode individual likelihood.

$$\mathcal{L}(\boldsymbol{\Theta},\widetilde{\boldsymbol{b}}) = \prod_j \mathcal{L}_j(\boldsymbol{\Theta},\widetilde{b}_j), \tag{3.10}$$

where

$$\mathcal{L}_j(\boldsymbol{\Theta},\widetilde{b}_j) \propto \exp\left\{-\frac{W}{2}\left[\widetilde{d}_j - \widetilde{b}_j - \widetilde{s}_j(\boldsymbol{\Theta})\right]^2\right\}, \tag{3.11}$$

$W = \sigma^{-2}$, the symbols with a tilde on top are the Fourier transform of their respective real space entities and the index identifies the Fourier mode. Our likelihood is

[1] The *Planck* data set is obviously non-homogenous. Very large variations on the background statistics between high and low galactic latitudes are clear. To enforce the condition of statistical homogeneity, we must apply our algorithm to small patches, each one covering only very limited regions of the sphere, where we can safely assume the homogeneity condition applies.

still conditioned on the astronomical background component $\widetilde{b}_i$. These parameters are '*nuisance parameters*' as we are not interested in recovering them. So, they can be integrated out

$$\mathcal{L}_j(\mathbf{\Theta}) = \int \mathcal{L}_j(\mathbf{\Theta}, \widetilde{b}_j)\ \Pr(\widetilde{b}_j)\ \mathrm{d}\widetilde{b}_j, \tag{3.12}$$

where $\Pr(\widetilde{b}_j)$, is the prior probability of the $\widetilde{b}_j$ Fourier mode. This probability can be easily derived from the power spectrum statistical properties of a Gaussian process and it reads

$$\Pr(\widetilde{b}_j) \propto \exp\left[\frac{W_{\beta_j}}{2}\ \widetilde{b}_j^{\,2}\right], \tag{3.13}$$

where $W_{\beta_j} = \beta_j^{-2}$ is the inverse of the j mode of the power spectrum. Using the algebraic equality in appendix A-1 it is possible to rewrite equation 3.12 into the following form

$$\mathcal{L}_j(\mathbf{\Theta}) \propto \int \exp\left\{-\frac{1}{2}(W + W_{\beta_j})(\widetilde{b}_j - c)^2 + \frac{W\, W_{\beta_j}}{W + W_{\beta_j}}\widetilde{\vartheta}_j^2\right\}\ \mathrm{d}\widetilde{b}_j, \tag{3.14}$$

where $\widetilde{\vartheta}_j = \widetilde{d}_j - \widetilde{s}_j(\mathbf{\Theta})$. Evaluating the integral we finally get

$$\mathcal{L}_j(\mathbf{\Theta}) \propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2 + \beta_j^2}\right)\left[\widetilde{d}_j - \widetilde{s}_j(\mathbf{\Theta})\right]^2\right\}, \tag{3.15}$$

where we have moved into the proportionality constant an extra contribution coming from the integration. The expression $\left(\frac{1}{\sigma^2+\beta_j}\right)$ is the inverse of the total background power spectrum. One can hardly consider this result as unexpected. If the astronomical background is made of more than one component and these are independent then $\beta_j^2 = \sum_k \beta_{jk}^2$, is just the sum of the power spectrum of each one of these components, where β_{jk}^2 is the j Fourier component of the k component.

3.2.2 Extending to multi-frequency

Assuming the pixel noise is independent across channels, the joint likelihood for a single Fourier mode (j) is just the product of the individual likelihoods for each frequency

$$\mathcal{L}_j(\mathbf{\Theta}, \widetilde{b}_{jk}) \propto \prod_k^{N_f} \exp\left\{-\frac{W_k}{2}\left[\widetilde{d}_{jk} - \widetilde{b}_{jk} - \widetilde{s}_{jk}(\mathbf{\Theta})\right]^2\right\}, \tag{3.16}$$

where $W_k \equiv \sigma_k^{-2}$ and the subscript k indexes the channel. The above expression can be compactly represented by using matrix quantities

$$\mathcal{L}_j(\boldsymbol{\Theta}, \widetilde{\boldsymbol{b}}_j) \propto \exp\left\{-(\widetilde{\boldsymbol{b}}_j - \widetilde{\boldsymbol{\Delta}}_j)^t \, \boldsymbol{W_I} \, (\widetilde{\boldsymbol{b}}_j - \widetilde{\boldsymbol{\Delta}}_j)\right\}, \tag{3.17}$$

where $\boldsymbol{W_I} \equiv \boldsymbol{N_I}^{-1}$ is the $N_f \times N_f$ diagonal covariance matrix of the instrumental noise[1] and $\widetilde{\boldsymbol{\Delta}}_j = \widetilde{\boldsymbol{s}}_j(\boldsymbol{\Theta}) - \widetilde{\boldsymbol{d}}_j$. The prior on the astronomical component is assumed to be well described by a second order statistical model

$$\Pr(\widetilde{\boldsymbol{b}}_j) \propto \exp\left\{-\widetilde{\boldsymbol{b}}_j^t \, \boldsymbol{W}_{\beta_j} \, \widetilde{\boldsymbol{b}}_j\right\}, \tag{3.18}$$

where $\boldsymbol{W}_{\beta_j} \equiv \boldsymbol{N}_{\beta_j}^{-1}$ is the cross-channel covariance matrix of the j Fourier mode. For a multi-component background made of uncorrelated contributions, this matrix is just the sum of the individual covariance matrices. In a well designed, free of systematics and properly calibrated experiment this matrix should be close to singular. Each line, for each component matrix, can be obtained by multiplying any other by the spectral index relating these two frequencies. As we shall see, this very strong correlation among the same Fourier mode at different frequencies plays a very important role when the main contaminant of the signal we want to detect is the astronomical background and not the instrumental noise. This is exactly the case for *Planck* (see chapter 5). Taking advantage of the matrix formula in appendix A-1, the likelihood now reads

$$\begin{aligned}\mathcal{L}_j(\boldsymbol{\Theta}) \propto \int \exp\left\{-(\widetilde{\boldsymbol{b}}_j - \boldsymbol{c})^t \, (\boldsymbol{W_I} + \boldsymbol{W}_{\beta_j}) \, (\widetilde{\boldsymbol{b}}_j - \boldsymbol{c})\right\} \\ \exp\left\{-\widetilde{\boldsymbol{\Delta}}_j^t \, \boldsymbol{W_I}(\boldsymbol{W_I} + \boldsymbol{W}_{\beta_j})^{-1}\boldsymbol{W}_{\beta_j} \, \widetilde{\boldsymbol{\Delta}}_j\right\} \, \mathrm{d}^n\widetilde{b}_j.\end{aligned} \tag{3.19}$$

After performing the integration, we have finally derived a likelihood formula that embodies the assumed statistical assumptions

$$\mathcal{L}(\boldsymbol{\Theta}) = \Pr\left\{\boldsymbol{d}|H_{\boldsymbol{s}}(\boldsymbol{\Theta})\right\} \propto \prod_j \exp\left\{-\left[\widetilde{\boldsymbol{d}}_j - \widetilde{\boldsymbol{s}}_j(\boldsymbol{\Theta})\right]^t \, \mathfrak{N}_j^{-1} \, \left[\widetilde{\boldsymbol{d}}_j - \widetilde{\boldsymbol{s}}_j(\boldsymbol{\Theta})\right]\right\}, \tag{3.20}$$

where $\mathfrak{N}_j = \boldsymbol{N_I^{-1}} + \boldsymbol{N}_{\beta_j}^{-1}$ is the the generalised noise cross-power-spectra. The likelihood normalization constant could be evaluated by integrating the full likelihood over

[1] $\boldsymbol{N_I}$ is not a function of the Fourier mode j

the vectors $\widetilde{\boldsymbol{d}}_j$ and equating it to unity. We are, in fact, only interested in the likelihood ratio between the hypothesis H_s that objects (of a given source type s) are present and the null hypothesis H_0 that only background is present. The latter corresponds to setting the sources signal $\boldsymbol{s}(\boldsymbol{x};\boldsymbol{\Theta})$ to zero. Under our combined assumptions, the log-likelihood ratio has the form

$$\ln\left[\frac{\mathcal{L}_{H_\mathrm{s}}(\boldsymbol{\Theta})}{\mathcal{L}_{H_0}(\boldsymbol{\Theta})}\right] = \sum_{\boldsymbol{\eta}} \widetilde{\boldsymbol{d}}^t(\boldsymbol{\eta})\mathbf{N}^{-1}(\boldsymbol{\eta})\widetilde{\boldsymbol{s}}(\boldsymbol{\eta};\boldsymbol{\Theta}) - \tfrac{1}{2}\sum_{\boldsymbol{\eta}} \widetilde{\boldsymbol{s}}^t(\boldsymbol{\eta};\boldsymbol{\Theta})\mathbf{N}^{-1}(\boldsymbol{\eta})\widetilde{\boldsymbol{s}}(\boldsymbol{\eta};\boldsymbol{\Theta}). \quad (3.21)$$

We have changed to a more common notation where $\boldsymbol{\eta}$ is the bi-dimensional spatial frequency (the wavenumber $\boldsymbol{k} = 2\pi\boldsymbol{\eta}$). From 3.2 and 3.3, the Fourier transform of the signal due to all the sources may be written

$$\widetilde{\boldsymbol{s}}(\boldsymbol{\eta};\boldsymbol{\Theta}) = \widetilde{\boldsymbol{B}}(\boldsymbol{\eta}) \sum_{j=1}^{N_\mathrm{s}} A_j \boldsymbol{f}(\boldsymbol{\phi}_j)\widetilde{\tau}(-\boldsymbol{\eta};\boldsymbol{a}_j)\mathrm{e}^{\mathrm{i}2\pi\boldsymbol{\eta}\cdot\boldsymbol{X}_j}, \quad (3.22)$$

where the vector $\widetilde{\boldsymbol{B}}(\boldsymbol{\eta})$ contains the Fourier transform of the beam at each frequency and $\widetilde{\tau}(\boldsymbol{\eta};\boldsymbol{a})$ is the Fourier transform of the template for an unconvolved object at the origin, characterised by the shape parameters $\boldsymbol{a}$. Substituting 3.22 into 3.21 and re-arranging, one obtains the final form of the likelihood ratio we will use throughout

$$\begin{aligned}
&\ln\left[\tfrac{\mathcal{L}_{H_\mathrm{s}}(\boldsymbol{\Theta})}{\mathcal{L}_{H_0}(\boldsymbol{\Theta})}\right] = \\
&\textstyle\sum_j^{N_\mathrm{s}} \left\{ A_j \mathcal{F}^{-1}\left[\mathcal{P}_j(\boldsymbol{\eta};\boldsymbol{\phi})\widetilde{\tau}(-\boldsymbol{\eta};\boldsymbol{a}_j)\right]_{\boldsymbol{X}_j} - \tfrac{1}{2}A_j^2 \textstyle\sum_{\boldsymbol{\eta}} \mathcal{Q}_{jj}(\boldsymbol{\eta};\boldsymbol{\phi})|\widetilde{\tau}(\boldsymbol{\eta};\boldsymbol{a}_j)|^2 \right\} \\
&\textstyle- \sum_{i>j}^{N_\mathrm{s}} \left\{ A_i A_j \mathcal{F}^{-1}\left[\mathcal{Q}_{ij}(\boldsymbol{\eta};\boldsymbol{\phi})\widetilde{\tau}(\boldsymbol{\eta};\boldsymbol{a}_i)\widetilde{\tau}(-\boldsymbol{\eta};\boldsymbol{a}_j)\right]_{\boldsymbol{X}_i-\boldsymbol{X}_j} \right\},
\end{aligned} \quad (3.23)$$

where $\mathcal{F}^{-1}[\ldots]_{\boldsymbol{X}}$ denotes the inverse Fourier transform of the quantity in brackets, evaluated at the point $\boldsymbol{X}$, and the quantities $\mathcal{P}_j(\boldsymbol{\eta};\boldsymbol{\phi}) \equiv \widetilde{\boldsymbol{d}}^t(\boldsymbol{\eta})\mathbf{N}^{-1}(\boldsymbol{\eta})\boldsymbol{\psi}(\boldsymbol{\eta};\boldsymbol{\phi})$ the '*Data single equivalent channel*', and $\mathcal{Q}_{ij}(\boldsymbol{\eta};\boldsymbol{\phi}) \equiv \widetilde{\boldsymbol{\psi}}_i^t(\boldsymbol{\eta};\boldsymbol{\phi})\mathbf{N}^{-1}(\boldsymbol{\eta})\boldsymbol{\psi}_j(\boldsymbol{\eta};\boldsymbol{\phi})$ the '*Noise single equivalent channel*', in which the vector $\boldsymbol{\psi}_i(\boldsymbol{\eta};\boldsymbol{\phi})$ has the components $(\boldsymbol{\psi}(\boldsymbol{\phi})_i)_\nu = \widetilde{B}_\nu(\boldsymbol{\eta})(\boldsymbol{f}_i(\boldsymbol{\phi}))_\nu$, with ν labelling frequency channels.

We have written the likelihood ratio in this way since it combines multi-channel data into a single equivalent channel, but only if the source spectral signature does not

depend on any parameter $(\boldsymbol{f}_i(\boldsymbol{\phi}))_\nu \equiv (\boldsymbol{f}_i)_\nu$, as in the case of the SZ sources. This property can be exploited with great advantage in the construction of SZ catalogues. The '*single equivalent channel*' is by far the most resource demanding part of the likelihood evaluation and only needs to be evaluated once per patch[1]. As a side effect this likelihood form further highlights the importance of the final 'cross-term' on the right hand side of 3.23. Let us assume for a moment that this cross-term is negligible:

$$\sum_{i>j}^{N_\mathrm{s}} \left\{ A_i A_j \mathcal{F}^{-1} \left[\mathcal{Q}_{ij}(\boldsymbol{\eta}) \widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a}_i) \widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j) \right]_{\boldsymbol{X}_i - \boldsymbol{X}_j} \right\} \approx 0. \tag{3.24}$$

In this case, the parameters of each source enter the likelihood independently. This parameter independence allows us to perform our analysis one source at a time and forms the basis of the '*single source model*' discussed in section 3.4.1, which greatly simplifies the source detection problem. The physical meaning of the neglected cross-term is most easily understood by considering the simple, but important, example of point sources, for which $\tau(\boldsymbol{x}, \boldsymbol{a}) = \delta(\boldsymbol{x})$. In this case, equation 3.24 becomes

$$\sum_{i>j} A_i A_j \mathcal{F}^{-1} [\mathcal{Q}_{ij}(\boldsymbol{\eta})]_{\boldsymbol{X}_i - \boldsymbol{X}_j} \approx 0. \tag{3.25}$$

A sufficient condition for this expression to be small is that all sources are sufficiently well-separated that $\mathcal{F}^{-1}[\mathcal{Q}_{ij}(\boldsymbol{\eta})]_{\boldsymbol{X}}$ is close to zero for such distances. For simple, uncorrelated backgrounds, $\mathcal{Q}_{ij}(\boldsymbol{\eta})$ contains just linear combinations of the instrument beams in each frequency channel. Thus, the condition that equation 3.25 is small, it is just a generalization of the common assumption in astronomy that objects are well separated, or that object blending effects are negligible[2]. When detecting point sources, and assuming the blending is not severe, an efficient implementation of the full deblending term is possible, but this will be addressed in a forthcoming publication.

It is worth noting that maximising the likelihood ratio (equation 3.23) in the ab-

[1] In practical terms, the SZ '*single equivalent channel*' may account for a speed increase in the code execution of hundreds as in the case of our own implementation applied to *Planck* data.

[2] When the background is uncorrelated, this condition is immediately fulfilled if each pixel contains signal coming from one and only one source. However, this is not enough when there are strong correlations in the background as in the case for *Planck* data.

sence of the cross-term 3.25 with respect to the source amplitudes A_j,

$$\nabla_A \ln\left[\frac{\mathcal{L}_{H_s}}{\mathcal{L}_{H_0}}\right] = 0, \tag{3.26}$$

reads

$$\widehat{A}_j = \frac{\mathcal{F}^{-1}\left[\mathcal{P}_j(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta};\widehat{\boldsymbol{a}}_k)\right]_{\widehat{\boldsymbol{X}}_j}}{\sum_{\boldsymbol{\eta}} \mathcal{Q}_{jj}(\boldsymbol{\eta})|\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_j)|^2}. \tag{3.27}$$

We have recovered the expression for the Matched Multi-Filter (MMF) [Herranz et al., 2002b]. Thus, we see that the filtered field is merely the projection of the likelihood manifold onto the subspace of position parameters $\boldsymbol{X}_j$. Through Fisher matrix analysis and referring to the Cramér-Rao bound, the estimation of the MMF parameter uncertainties becomes straightforward (see chapter 5). This should be contrasted with traditional MMF approach in which the uncertainty on the source flux estimate is calculated assuming the values of all other parameters are fixed [Melin et al., 2006]. Substituting the maximum-likelihood estimate (equation 3.27) into the likelihood ratio (equation 3.23), one obtains for the jth object

$$\max\left[\ln\left(\frac{\mathcal{L}_{H_s}}{\mathcal{L}_{H_0}}\right)\right] = \tfrac{1}{2}\sum_{\boldsymbol{\eta}} \mathcal{Q}_{jj}(\boldsymbol{\eta})|\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_j)|^2\widehat{A}_j^2 = \tfrac{1}{2}\widehat{\mathrm{SNR}}_j^{\,2} \tag{3.28}$$

where $\widehat{\mathrm{SNR}}_j$ is the signal-to-noise ratio (at the peak) of the jth source, and the rms σ of the noise satisfies

$$\frac{1}{\sigma^2} = \sum_{\boldsymbol{\eta}} \mathcal{Q}_{jj}(\boldsymbol{\eta})|\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_j)|^2. \tag{3.29}$$

Thus, one recognises that the traditional approach to catalogue making, in which one compares the maximum SNR of the putative detections against a given threshold, is really nothing more than performing a Generalised Likelihood Ratio Test (GLRT) [Trees, 2001].

3.3 Priors

Prior probability distributions are part and parcel of Bayesian inference framework. If the data model provides a good description of the observed data and the signal-to-noise

ratio is high, then the likelihood will be very strongly peaked around the true parameter values and the prior will have little or no influence on the posterior distribution [Cam and Yang, 2000]. At the faint end of the source population, when we are getting close to the instrument sensitivity limit, however, priors will inevitably play an important role. Moreover, since for most cases in astronomy the faint tail overwhelmingly dominates the population, the selection of the priors becomes important and has to be addressed very carefully.

PwS separates the tasks of source detection and source estimation (see section 3.3.1). This separation has the advantage of allowing the use of different sets of priors at each stage. Typically, we first perform the source detection step using 'informative' priors, which encompass all the available information, since they provide the optimal selection criterion and the optimal estimators. After the set of detections has been decided, PwS proceeds to the estimation pass, in which 'non-informative' priors may be used instead.

- *Non-informative priors.*
  Non-informative priors are constructed such that a '*Maximum A Posteriori*' (MAP) estimator of any quantity should depend exclusively on the current data set[1]. One way of expressing this condition is that, when changing the data, the likelihood shape remains unchanged and only its location in the parameter space changes [G. E. Box, 1992, ch. 1]. Thus, the idea is to find an appropriate reparametrization of the likelihood that transforms the parameters into location parameters, for which the ignorance prior is locally uniform. Locally, in this sense, means the parameter range where the mass of the likelihood is concentrated. One then performs the inverse parameter transformation on the uniform prior to obtain the appropriate prior in the original parameter set. Finding such a transformation may, however, be very difficult for a general multi-dimensional prior. Nonetheless, in a large majority of applications, the parameters can be assumed independent, so that the prior factorises

$$\pi(\theta_1, \theta_2 \ldots, \theta_n) = \pi_1(\theta_1)\pi_2(\theta_2) \ldots \pi_n(\theta_n). \tag{3.30}$$

[1] These priors usually need not be properly normalised, since the normalisation constant does not depend on any parameters.

For one-dimensional distributions, Jeffreys devised a general way to derive the non-informative prior on a parameter based on invariance properties of the likelihood under a change of variable [G. E. Box, 1992; Jaynes, 2004]. The Jeffreys rule for constructing ignorance priors for the one-dimensional case reads

$$\pi(\theta) \propto \mathcal{J}^{1/2}(\theta), \tag{3.31}$$

where

$$\mathcal{J}(\theta) \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta^2} \right\rangle \tag{3.32}$$

is the Fisher information [Jeffreys, 1961]. We will adopt this approach and consider below the prior on each parameter of interest.

- *Informative priors.*
  Informative priors are those meant to add more information to the inference process than that carried by the data alone. The estimates are always the outcome of a trade-off between the data and prior information.

  - *Population related.*
    This type of probability distribution is something astronomers know very well. It describes how a certain physical quantity is distributed across a certain population of sources. An example is the distribution of flux densities for radio sources [de Zotti et al., 2005]. Usually, the general form of the distribution is known or at least can be constrained using physical or cosmological arguments only. But, many times, these laws depend on unknown or poorly known parameters. Bayesian hierarchical modeling (see [G. E. Box, 1992, ch. 5] and [Gelman et al., 2003, ch. 5]) is the perfect recipe for dealing with this kind of problems that pervade astronomy[1].

  - *Targeted.*
    This is commonly known as '*non-blind detection*'. These are the type of priors that apply to each object individually. Normally, they are assembled from previous observations of the same object and their main goal is to provide better constraints on the parameter estimates by combining, eventually

[1] The hierarchical modeling will be fully addressed in a future work.

heterogenous, data sets [Planck et al., 2012].

Some authors, such us Box and Tiao [G. E. Box, 1992], are clearly in favour of the use of *non-informative* priors, as these provide estimates based exclusively on the information contained in the data under study. Though, others prefer optimal estimators, based on informative priors, at the cost of some blending with external information [AMI Consortium et al., 2012].

3.3.1 Detection versus estimation (astrometry and photometry)

- *Detection*
  is defined as to ***decide*** whether a certain signal originates from one of the sources or not.

- *Estimation*
  is the process of ***assigning values*** and uncertainties to the physical parameters of interest. Estimation can be further sub-divided into:

  - *Photometry*,
    the estimation of the strength of the source emission.
  - *Astrometry*,
    the estimation of the object's position and geometry.

In PwS we do not make any distinction between the different parameters of our object model. Thus, astrometry and photometry are joined into a single entity '*parameter estimation*'. Bayesian probability allow us to define the full '*posterior distribution*' of the estimates unifying the concepts of '*point*' and '*interval estimation*'. Jaynes [Jaynes, 2004, ch. 6] asserts that, in the light of probability theory, there should not be any real difference between parameter estimation and model assessment at least in general. Moreover, by indexing the competing models and then finding the maximum of the posterior distribution of the index parameter given the data, we are after all selecting one of them (see section 3.4). Although, at the same time, he also claims [Jaynes, 2004, ch. 9] one can never say the data supports a certain hypothesis without stating a specific alternative. When we say we are testing for a family of objects H^j,

supported by a general source model (radio source, dusty source, cluster, etc.), we are at the same time assessing an entire set of sub-models, each one indexed on the number of sources present of that type, $H^j_{N_{obj}}$ against its complement $\widetilde{H^j_{N_{obj}}}$. Then by choosing the value of N_{obj} which maximises the evidence ratio, we are effectively selecting a sub-model, but at the same time bringing about an estimate of N_{obj} as well. We believe that is exactly what Jaynes had in mind.

Another, more subtle point, is about the choice of priors. Priors must bring into the inference arena the state of knowledge we possessed before start collecting the data. Often that prior knowledge is only vestigial or it is our desire to infer assuming that. These type of priors are commonly known as '*non-informative*'. In section 3.3 we define the conditions those priors need to meet, but only in the case of an estimation exercise. However, one should always remember that detection is about taking decisions by comparing hypotheses, which is fundamentally different. So are the underlining assumptions and consequently the non-informative priors as well. Making a catalogue always involves, at least, a detection plus an estimation step. We believe these operations must be kept independent as different priors may apply in each case.

3.3.2 Prior on positions

It is obvious that the distribution of sources is not uniform across the sky. The galactic regions (Milky Way and Magellanic Clouds) have a much higher density of detectable sources than the rest of the sky. Moreover, assuming extra-galactic sources to be uniformly distributed across the sky (no clustering) is not sufficient to ensure that the distribution of detectable sources is uniform, since the background/noise is itself inhomogeneous over the sky. Let us define a new entity which will soon prove very helpful, the '*homogeneous zone*'. A homogenous zone is an arbitrary ensemble of patches across which all statistical properties of the data and the object models may be assumed approximately invariant. The homogeneous zones should be appropriately sized to keep the expected number of sources λ on each mostly invariant[1]. PwS divides the sky into small patches, which may be further aggregated into homogeneous zones, and, in each such region, the assumptions of background homogeneity and a uniform

[1]Denser regions should be made smaller then sparsely populated ones. An homogeneous zone may contain from one single patch (the '*area atom*') to the full set.

source distribution are reasonable. Moreover, if the sky patches used are sufficiently small, our locally uniform model can easily cope with clustering when the gradient of the density of sources is small across the patch boundaries. The correctly normalised positions prior for the complete ensemble of sources in a patch is simply

$$\Pr(\boldsymbol{X}^{\boldsymbol{N_s}}|N_{\rm s},\ N_{\rm pix}) = \frac{1}{{N_{\rm pix}}^{N_{\rm s}}}, \tag{3.33}$$

where $N_{\rm pix}$ is the number of pixels in each homogeneous zone and $N_{\rm s}$ is the number of sources in that region.

3.3.3 Prior on the number of sources

Following the same rationale of local uniformity, i.e no clustering, the probability of finding $N_{\rm s}$ objects (above a given flux limit) in a sky homogeneous zone follows a Poisson distribution

$$\pi(N_s) = \Pr(N_s|\lambda) = e^{-\lambda}\,\frac{\lambda^{N_s}}{N_s!}, \tag{3.34}$$

where λ is the expected number of such objects in that region. Moreover, λ should be proportional to the region size $\lambda = \Lambda_s N_{\rm pix}\Delta_p$, where Λ_s is the number of sources per pixel and Δ_p is the pixel area. Note that Λ_s may change across the sky as we are only enforcing the uniformity locally within each homogeneous zone.

3.3.4 Prior on flux density

A good flux estimator should be unbiased, but this goal is often problematic. The optimal estimators in the sense of decision theory, i.e. those that minimise the expected loss/cost, are most often biased and they combine the data with external information from ancillary data sets. PwS thus includes two different sets of flux priors with distinct goals.

- *Non-informative*.
  Our data model depends linearly on the source fluxes A_j and is a particular case of the general linear model [G. E. Box, 1992]. Considering only a single source for simplicity (the solution for multiple sources is a mere repetition of

this simpler case), one may show that the likelihood can be written in a form that makes it clear that the flux is in fact a location parameter:

$$\mathcal{L}_{H_s}(A_j) \propto \exp\left[-\frac{\sum_{\boldsymbol{\eta}} \mathcal{Q}_{jj}(\boldsymbol{\eta})|\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_i)|^2}{2}(A_j - \widehat{A}_j)^2\right], \tag{3.35}$$

where $\widehat{A}_j$ is the MMF estimate of the flux (equation 3.27). The same result could have been obtained directly using formula 3.31. Thus, the prior on the flux must be locally uniform:

$$\pi(A_j) \propto c, \tag{3.36}$$

where j indexes the source. For a more general and rigorous treatment see [G. E. Box, 1992].

- *Informative*.
  Owing to the different statistical properties of point sources and SZ galaxy clusters, a different prior applies in each case. For point sources, we adopt the flux prior first suggested by Argüeso et al. [2011],

$$\pi(A_j) = \Pr(A_j|A_0\, p\, \gamma) \propto \left[1 + \left(\frac{A_j}{A_0}\right)^p\right]^{-\frac{\gamma}{p}}, \tag{3.37}$$

  where A_0 is the 'knee' flux, p is some positive number and γ is the exponent controlling the shape of the power law for fluxes much larger than the 'knee'. This provides a good model for the observed distribution of fluxes, fitting the de Zotti model almost perfectly [de Zotti et al., 2005]. Moreover, the distribution can be properly normalised as required for evidence evaluation. PwS sets a minimum flux and re-normalizes the remaining range, a practice the proponents of the distribution also followed. For galaxy clusters, the derivation of the prior follows a different approach. The *Planck* Sky Model (PSM v1.6) [Delabrouille et al., 2012] was used to draw realistic simulations of the cluster populations assuming a standard *WMAP* best-fit ΛCDM cosmology [Hinshaw et al., 2009] and the Jenkins mass function [Yoshida et al., 2001]. We found that the fluxes in

the sample cluster catalogues were quite well fitted by a power law

$$\pi(A_j) \propto A_j^{-\gamma}. \tag{3.38}$$

As we need to provide a properly normalised distribution, once again a minimum and, this time, a maximum flux also were assumed.

3.3.5 Prior on size

- *Point sources.*
  Point sources are best modelled by imposing the prior $\pi(r) = \delta(r)$ on the 'radius'. This condition might, however, be too restrictive, since to simplify the implementation of the code and to make it faster, PwS assumes the instrument beams are circularly symmetric, which is only an approximation to the true beam shapes. Thus, even for point sources, allowing the source radius to vary over a small range of values allows a better fit between the template and the pixel intensities and consequently a higher likelihood ratio/SNR value. Thus, in both the informative and non-informative case, our preferred radius prior for point sources is
  $$\pi(r_j) = \begin{cases} 1/\Delta & r_j \leq \Delta \\ 0 & r_j > \Delta \end{cases}, \tag{3.39}$$
  where $\Delta \ll$ FWHM (the full-width-half-maximum of the beam).

- *Galaxy clusters.*
  Turning to galaxy clusters, a significant fraction of the clusters *Planck* will detect will be unresolved, and thus appear as point sources with a distinctive spectral signature. In many cases, however, galaxy clusters are large enough to be mapped as extended objects and a parameter controlling the scale of the cluster profile, the radius, needs to be included. The informative prior on the radius was derived using the same procedure as in section 3.3.4 and an exponential law
  $$\pi(r_j) \propto \exp\left(-\frac{r_j}{\ell}\right), \tag{3.40}$$
  was found to fit the simulated catalogues very well. We truncate the distribution

outside a minimum and maximum radius.

The non-informative prior follows a different law from that expected from the cosmological models. Our model for an individual source is the convolution of the source profile with the point spread function (PSF). The radius parameter r'_s that scales the resulting shape is a 'hybrid' parameter, as it shifts and scales the likelihood at the same time [Jaynes, 2004, ch. 12]. After applying the Jeffreys rule, the non-informative prior on r'_s reads:

$$\pi(r'_s) \propto \frac{1}{r'^2_s}. \tag{3.41}$$

Assuming that either the profile or the beam have centroids at the origin and the profile is a scaling profile $\tau(r/r_s)$ then,

$$r'_s = \sqrt{B^2 + \kappa^2\, r^2}, \tag{3.42}$$

where B^2 is a constant known as the function variance of the beam [Bracewell, 1999] and κ^2 is another dimensionless constant, the variance of the dimensionless variable r/r_s over the profile. The non-informative prior for the radius parameter then reads:

$$\pi(r) \propto \frac{r}{(\mathcal{B}^2 + r^2)^{\frac{3}{2}}}, \tag{3.43}$$

where $\mathcal{B} = B/\kappa$. For the general case B^2, the variance of the beam, should be replaced by the variance of $\sqrt{\mathcal{P}_j(\boldsymbol{\eta})}$. For unresolved objects, narrow clusters with radii smaller than the beam size, the prior grows linearly with r. For well resolved objects, $r \gg \mathcal{B}$, the prior decreases proportionally to r^{-2}.

3.3.6 Prior on spectral parameters

There is an extensive literature on the distribution laws of radio source spectral indexes [de Zotti et al., 2010; Planck Collaboration et al., 2011a,c, and references therein]. In general Gaussian distributions, or Gaussian mixtures with two modes, fit the available data reasonably well. However, the most interesting sources are exactly those that do not follow the canonical laws of emission. To avoid narrowing the range of possible alternatives too much, uniform priors are probably better choices unless we choose to

target a very specific family. The same holds for dusty galaxies.

By applying our standard procedure, the non-informative prior on the spectral parameters reads

$$\pi\left(\alpha_k\right) \propto \sqrt{\sum_{\nu_1\nu_2} \Theta_{\nu_1\nu_2} \left(\frac{\partial S}{\partial \alpha_k}\right)_{\nu_1} \left(\frac{\partial S}{\partial \alpha_k}\right)_{\nu_2}}, \tag{3.44}$$

where the sum extends over all frequency channels, $\left(\frac{\partial S}{\partial \alpha_k}\right)_{\nu_i}$ is the derivative of the SED of the source in order to parameter α_k evaluated at frequency ν_i and

$$\Theta_{\nu_1\nu_2} = \Sigma_{\boldsymbol{\eta}} \;\; \mathcal{N}^{-1}_{\nu_1\nu_2\boldsymbol{\eta}} \, \widetilde{B}_{\nu_1\boldsymbol{\eta}} \, \widetilde{B}_{\nu_2\boldsymbol{\eta}} \, \widetilde{\tau}_{\boldsymbol{\eta}}(\widehat{\boldsymbol{a}}),$$

where all symbols follow the same nomenclature as in 3.23.

3.3.7 Prior on the models

The prior ratio $\Pr(H_1)/\Pr(H_0)$ on the models is often neglected (i.e. assumed to equal unity), but plays a very important role in the PwS detection criterion. To give a proper account of its nature, let us imagine the simplest possible detection problem, where we know in advance all true values of the parameters that define an object, which translates into delta-functions priors. Substituting this condition into equation 2.9 and making use of 3.28, we obtain the following inequality:

$$\mathrm{SNR} \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{2 \left[\xi + \ln\left(\frac{\Pr(H_0)}{\Pr(H_1)}\right)\right]}. \tag{3.45}$$

One may interpret the term $\ln\left(\frac{\Pr(H_0)}{\Pr(H_1)}\right)$ as an extra 'barrier' added to the detection threshold because we are expecting more fake objects than the objects of interest, due to background fluctuations.

We saw earlier that, when an object is present, a local maximum in the likelihood is always present in the position parameter sub-space. This condition immediately implies that only likelihood maxima need be analysed. Nonetheless, one expects other likelihood maxima to occur as a result of background fluctuation 'conspiracies'. Assuming Poisson statistics for the number of sources and the number of likelihood maxima resulting from the background fluctuations, then the ratio of the probabilities is

given by (see A-2)

$$\frac{\Pr(H_1|N_s)}{\Pr(H_0|N_s)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{N_s} \tag{3.46}$$

where λ_0 is the expected number of maxima per unit area resulting from background fluctuations above the minimum limit of detection of the experiment, and λ_1 the expected number density of sources above the same limit.

If only background is present, the density of maxima, λ_0, resulting from the filtering procedure that creates the likelihood manifold can be estimated using the 2D Rice formula:

$$n_b(\nu,\kappa,\epsilon) = \frac{8\sqrt{3}\tilde{n}_b}{\pi\sqrt{1-\rho^2}}\,\epsilon(\kappa^2-4\epsilon^2)\,e^{-\frac{1}{2}\nu^2-4\epsilon^2-\frac{(\kappa-\rho\nu)^2}{2(1-\rho^2)}}, \tag{3.47}$$

where $\nu \equiv A/\sigma$ is the 'normalised peak amplitude', κ the 'normalised curvature', ϵ the 'normalised shear', and $\rho = \sigma_1^2/(\sigma_0\sigma_2)$, with $\sigma_n^2 = (2\pi)^{1+2n}\int_0^\infty \eta^{1+2n}|\mathcal{P}(\eta)|^2\,\mathrm{d}\eta$ [López-Caniego et al., 2005]. Marginalizing over all parameters we obtain the expected density of maxima of a Gaussian filtered field, which reads

$$\tilde{n}_b = \frac{\sigma_2^2}{8\pi\sqrt{3}\sigma_1^2}. \tag{3.48}$$

One is not interested, however, in all peaks, but only on those above a certain level ν_0, since PwS pre-selects the putative detections by imposing a minimum SNR level before attempting the evidence evaluation. The main reason for adopting this early selection is computational efficiency. The SNR alone provides a good proxy (see formula 3.28) for deciding whether a candidate peak is the result of the presence of a source or just a background fluctuation. Moreover, low SNR peaks tend to be 'badly-shaped' making the sampler very inefficient and resulting in a very large fraction of the samples being rejected. To make the things even worse, in most cases, these peaks themselves end up being rejected as objects. The applied flux cut must be taken into consideration to evaluate the correct expected number counts, which define the prior $\Pr(H_1)$ as well. Thus, λ_0 will read:

$$\lambda_0 = \int_{\nu_0}^{\infty} n_b(\nu)\mathrm{d}\nu, \tag{3.49}$$

where $n_b(\nu)$ is given by

$$n_b(\nu) = \frac{\tilde{n}_b\sqrt{6}}{2\sqrt{\pi}\rho_1}\left\{\left(1+\mathrm{erf}\left(\frac{\rho}{\rho_1\rho_2}\nu\right)\right)e^{-\nu^2\left(\frac{1}{2}+\left(\frac{\rho}{\rho_2}\right)^2\right)}\left(\frac{\rho}{\rho_2}\right) + \left(1+\mathrm{erf}\left(\frac{\rho}{\rho_1}\nu\right)\right)e^{-\frac{\nu^2}{2}}(\nu^2-1)\rho^2\rho_1 + \frac{\nu e^{-\nu^2\left(\frac{1}{2}+\left(\frac{\rho}{\rho_1}\right)^2\right)}}{\sqrt{\pi}}\rho\rho_1^2\right\}, \tag{3.50}$$

where $\rho_1 = \sqrt{2\left(1-\rho^2\right)}$ and $\rho_2 = \sqrt{2\left(\frac{3}{2}-\rho^2\right)}$. The expected number count of targeted objects above a certain flux threshold S, $\lambda_1 \equiv \langle N(>S)\rangle$, may be easily derived from their differential counts. The dominant type of extra-galactic point sources in *Planck* maps are galaxies which, in principle, do not follow the same statistics as the galaxy clusters. From general cosmological assumptions it is possible to derive that the expected differential counts for a certain population type of galaxies per flux interval at a certain frequency, always follow a power law $dN_\phi/dS = A_\phi\, S^{-b}$ [de Zotti et al., 2005]. For clusters of galaxies, however, we derived the differential number counts from a realistic set of simulations produced by the '*Planck Sky Model*' (PSM v1.6) [Delabrouille et al., 2012]. The package was parameterized to use a Jenkins mass-function [Yoshida et al., 2001] and a standard *WMAP* best-fit ΛCDM cosmology [Hinshaw et al., 2009] as before. We have found that a power law also fits quite well the expect number counts of clusters above a certain threshold. So, in either case, point sources or clusters, λ_1 may be written as

$$\lambda_1 = N(>S_0) = \int_{S_0}^{\infty}\frac{dN_\phi}{\mathrm{d}S}\mathrm{d}S = A_\phi\,(1-b)^{-1}\,S_0^{1-b},\ b\neq 1, \tag{3.51}$$

where we keep the parameters $\{A_\phi, b\}$ free. These parameters are usually provided by the user to target a specific type of object and/or instrumental setup.

3.4 Object detection strategy

So far we have only developed the logical and probabilistic underpinnings of PwS. It is now time to bring all the pieces together into a consistent strategy for the detection and characterisation of discrete objects. Our aim is to construct a robust, controlled, and

predictable algorithm. Some caveats will be identified and solutions suggested, always justified within the framework presented above.

3.4.1 Single object approach

Let us return to formula 2.9. At a first look, the evaluation of 2.9 seems quite a daunting task. In order to apply the full Bayesian approach, many complex integrals, over a very high dimensional volume (at least $4 \times N_s$), need to be evaluated[1]. Clearly a brute force method is not efficient and perhaps not possible, even with the massive computing resources generally available.

To find an effective solution, we begin by making two important assumptions: (i) the objects of interest are 'well separated', so that equation 3.24 holds; and (ii) all variables pertaining to each individual source are mutually independent, which has already been implicitly assumed throughout the exposition of our inferential infrastructure.

These conditions allow us to separate the integrals associated with each source. This is a very important simplification because it is now possible to deal with each source independently, one at a time. This is the 'single object approach' [Hobson and McLachlan, 2003] and replaces a single $N_{\rm param} \times N_s$-dimensional integral with a sequence of N_s integrals, each of dimension $N_{\rm param}$. The complete likelihood expression may now be replaced by the much simpler 'single source' form.

$$\frac{\Pr(H_s|\boldsymbol{d},N_s)}{\Pr(H_0|\boldsymbol{d},N_s)} =$$
$$\underbrace{\left\{\left[\int_{\alpha_{10}}\cdots\int_{\alpha_{1n}} \frac{\mathcal{L}_s(\alpha_{10},\cdots,\alpha_{1n})}{\mathcal{L}_0}\,\pi_{s1}(\alpha_{10})\cdots\pi_{sn}(\alpha_{1n})\,\mathrm{d}\alpha_{10}\cdots\mathrm{d}\alpha_{1n}\right]\right\}}_{z_{s1}}$$
$$\vdots$$
$$\underbrace{\left\{\left[\int_{\alpha_{N_s0}}\cdots\int_{\alpha_{N_sn}} \frac{\mathcal{L}_s(\alpha_{N_s0},\cdots,\alpha_{N_sn})}{\mathcal{L}_0}\,\pi_{s1}(\alpha_{N_s0})\cdots\pi_{sn}(\alpha_{N_sn})\,\mathrm{d}\alpha_{N_s0}\cdots\mathrm{d}\alpha_{N_sn}\right]\right\}}_{z_{sN_s}}$$
$$(N_{pix}\Delta_p)^{-N_s} \quad e^{-\lambda_s}\frac{\lambda_s^{N_s}}{N_s!} \quad \left(\frac{\lambda_s}{\lambda_0}\right)^{N_s}. \tag{3.52}$$

Although, one should exercise some care in defining the limits of integration in position

[1] Even when working with one small patch at a time, N_s is seldom smaller than 4.

space, since no significant likelihood mass can be shared among position integration domains. Apparently, this requirement creates such a wealth of complexity to the integral evaluation that the single source approach might at first be considered a poor choice. Fortunately, the method PwS uses to evaluate the evidence integrals automatically enforces this rule if the fields are not too crowded (see section 3.4.2). Under our assumptions, the odds of the model H_s (for a given source type), given N_s such sources, reads

$$\frac{\Pr(H_s|\boldsymbol{d}, N_s)}{\Pr(H_0|\boldsymbol{d}, N_s)} = (N_{\rm pix}\Delta_{\rm p})^{-N_s} e^{-\lambda_s} \frac{\lambda_1^{N_s}}{N_s!} \left(\frac{\lambda_s}{\lambda_0}\right)^{N_s} \prod_{j=1}^{N_s} \mathcal{Z}_{sj}, \tag{3.53}$$

Taking logarithms and rearranging, one finds

$$\ln\left[\frac{\Pr(H_s|\boldsymbol{d}, N_s)}{\Pr(H_0|\boldsymbol{d}, N_s)}\right] = \sum_{j=1}^{N_s} \ln(\mathcal{Z}_{sj}) - N_s P_s, \tag{3.54}$$

where we have defined the 'penalty per source' P_s as

$$P_s \equiv \ln \Lambda_s^{-1} + \ln\left(\frac{\lambda_0}{\lambda_s}\right) + \frac{1}{N_s}\left[\lambda_s + \ln N_s!\right]. \tag{3.55}$$

Thus, the total $\ln(\text{odds})$ for a single homogeneous zone is the sum of the partial $\ln(\text{evidence})$ for each source, plus an extra global penalty term that contributes, in the majority of the cases, negatively to the final balance and does not depend on any particular source, but exclusively on the ensemble properties.

The most robust source catalogue is that which maximises the $\ln(\text{odds})$ in 3.54, but we do not know the value N_s. Moreover, we have not yet addressed how many or which candidate detections will be finally selected for inclusion in the catalogue. Nonetheless, the expression 3.54 is a sum, so its maximum value is reached when only the positive terms are included. Thus, one possible procedure to select the optimal set of sources is as follows:

1. evaluate $\mathcal{Z}_j$ for each source;

2. partition the candidate detections into the pre-defined homogenous zones. For each zone:

(a) sort the candidate detections in descending order of $\mathcal{Z}$ and number them ($j = 1 \ldots$);

(b) with N_s, iterate down the list of catalogue entries evaluating formula 3.54;

(c) stop when moving from $N_s = k$ to its successor ($N_s = k + 1$) makes expression 3.54 decrease.

(d) This means, $\widehat{N_s}$ (the value of N_s that maximises the evidence ratio) has been found and the 'proto-catalogue' is formed from the first k entries.

This quantity, the ln(odds) for each object

$$\ln(\text{odds})_j \equiv \ln\left[\frac{\Pr(H_1|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right]_j = \ln(\mathcal{Z}_{1j}) - \widehat{P_s}, \tag{3.56}$$

($\widehat{P_s}$ is the penalty per source evaluated at $\widehat{N_s}$ or the catalogue penalty per source), has a pivotal role in catalogue making (see section 2.3).

We are not finished yet, however, because we have only selected the set of detections that maximises the odds. Other constraints may yet apply. For instance, we may impose a threshold per line different from zero as result of the loss criteria or, as we shall see, a prescribed contamination for the catalogue.

3.4.2 Evaluation of the odds ratio

Even using the simplified form of the likelihood assumed in the single-object approach, a 'brute force' evaluation of the resulting evidence integrals is still not feasible. One must instead use a Monte Carlo approach to the numerical integration. Evidence integrals are usually evaluated using Markov Chain Monte Carlo (MCMC) methods and thermodynamic integration. Such methods can fail, however, when the posterior distribution is very complex, possessing multiple narrow modes[1] that are widely separated. We therefore instead use 'nested sampling' [Sivia and Skilling, 2006], which is much more efficient, although not without its difficulties. Feroz et al. [2009] have developed a very efficient implementation of the nested sampling algorithm, called 'MultiNest', which is capable of exploring high-dimensional multimodal posteriors. Nonetheless,

[1] At least one central maximum per source plus other secondary maxima around the central higher peaks [Carvalho et al., 2009].

MultiNest is designed to be a general sampling and evidence evaluation tool and it is not particularly tuned for *Planck*.

In the interest of speed, PwS instead tries to take full advantage of the properties of the astronomical data sets. As already stated (see section 3.3), if our model explains the data well then the likelihood should peak steeply around the parameter true values, decay very rapidly to zero, and have most of its mass concentrated around the maxima vicinities. Thus, if one can first find the likelihood maxima, then one does not need a sophisticated multimodal sampling algorithm like MultiNest. A much simpler nested sampling scheme such as Mukherjee et al. [2006] would perform equally well. Moreover, reasonably high SNR maxima develop 'well-shaped' peaks, in the sense they are close to Gaussian, rendering the sampling highly efficient. Two other significant advantages are: (i) we can reduce our data set to a small neighbourhood enclosing the maxima, so that only a very small number of pixels close to the maxima contribute appreciably to the evidence value; and (ii) a much reduced parameter volume allows the same number of 'live points' to deliver a considerably higher accuracy on the evidence value, since they do not split among the several posterior peaks. This is the approach adopted in PwS, which we now outline in more detail.

3.4.2.1 Locating the likelihood maxima

Our first goal is to find the likelihood maxima. For illustration, let us focus on the example of galaxy clusters, each of which is described by 4 parameters: $\{X, Y, S, R\}$. An efficient 4-dimensional minimiser implementation is straightforward and immediately available [Press et al., 1992]. However, our manifold has many maxima and we need to check all of them, otherwise we might lose some sources.

One possibility would be to follow the approach used in Carvalho et al. [2009], where the Brent line minimiser was 'enhanced' with an ancillary step to allow it to 'tunnel' from one minimum to the next one using a scheme closely related with the equivalent quantum mechanical effect. To increase the effectiveness of the procedure, PwS I started a Powell minimization chain[1] in many different locations of the manifold in an attempt to find all the maxima. It should be remembered, however, that the likelihood only exhibits multiple maxima in the position sub-space; the other sub-spaces are

[1] 'Powell' = Powell minimiser + 'Snakes' = the path the solution follows before converging.

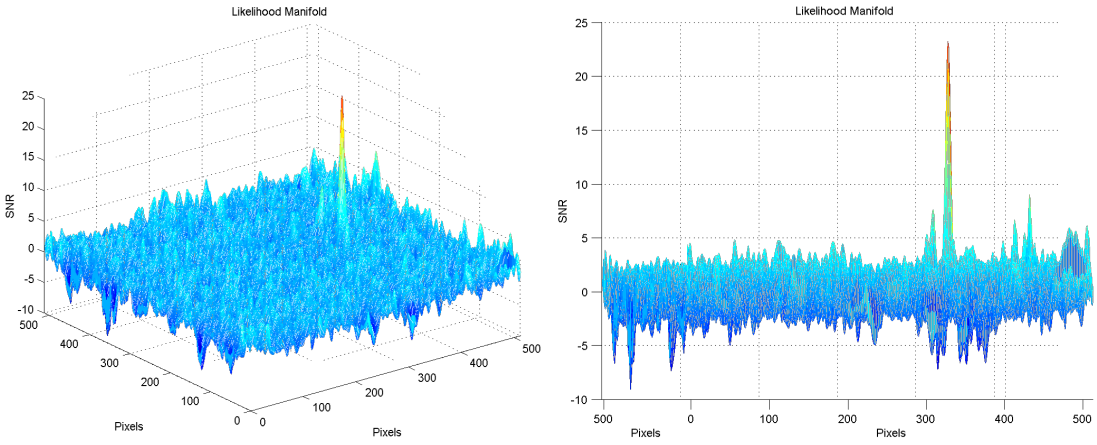


Figure 3.1: These two figures are views of a typical ln likelihood ratio manifold ($\ln\left[\mathcal{L}_{H_s}/\mathcal{L}_{H_0}\right]$) restricted to position space. The z axis was normalized using formula 3.28. This particular example was drawn from a SZ extraction exercise using the 'WG2' simulations (see section 4.3.1). The complexity of the manifold is overwhelming.

'well behaved'. Moreover, the likelihood in the position sub-space is merely the MMF filtered field. We therefore instead use a brute force peak finding algorithm that scans all pixels in this subspace, which is very easy to implement and almost instantaneous. Then, after collecting a list of peak positions, we start a 4-dimensional PowellSnakes optimisation at each such location to find the maximum-likelihood parameters for that particular peak.

A subtlety does arise in this approach, however, since to obtain the MMF filtered field, one needs to assume a size R for the objects to define the filter. Since we expect different clusters to have different radii, we might lose some peaks because of the mismatch between the true value of the cluster radius and that used in the filtering template. A simple solution would be that suggested by the MMF authors: apply the filter repeatedly using a different radius each time. Although practical, this is, however, not the most efficient approach. Fortunately, if the instrument beams and the sources possess reflection symmetries in both axes, then one can show that the Fisher matrix at each likelihood peak is block-diagonal (assuming the likelihood 3.23 and using the single-source approach assumption 3.25; see demonstration in appendix A-3), such that there is no correlation between the position subspace and the other parameters

(flux and size) of the cluster. This has two important consequences: (i) regardless of the radius used to construct the filter, a likelihood maximum will always appear at the location of the source and its position will not move as the filter scale varies; (ii) we do not need to perform a full 4-dimensional maximization but can (at least) separate the position variables from all others, which brings a tremendous simplification to the problem of finding the likelihood maxima. Thus, we can indeed start by finding the maxima in the position subspace using a brute force 'check-all-pixels' approach and then, after pinpointing the position of the source, search the remaining sub-spaces associated with the other variables.

A couple of final comments on this approach are worth making. First, it is well known that matched filters are excellent at finding and locating sources, but not as good at estimating fluxes. If the beam shape/size is not completely known but symmetric, even when building up a filter with the wrong beam geometry, the filter will correctly recover the positions of the objects. In general, however, the element in the Fisher matrix corresponding to the correlation between the radius and the flux of an object is non-zero. Therefore, if the filter is assembled using wrong beam parameters, bias in the flux estimates must be expected. Second, and perhaps more subtle, is that the symmetries of the Fisher matrix only hold on average. Thus, for each individual peak some residual correlation between the position and the other variables is expected. According to our current accumulated experience, however, this correlation is usually very small. Nonetheless, PwS still includes the option to use the peak positions obtained from the MMF filtered fields just as initial hints for a full N-dimensional Powell minimisation.

3.4.2.2 Exploring the posterior distribution

Our initial step provides the Maximum Likelihood (ML) estimates and the SNR of each detection candidates. This has a very useful side effect, since we do not need to explore the posterior distribution around all the maxima we find. Only a much smaller sub-set is chosen based on an SNR threshold. This SNR threshold should be low enough not to reject any substantial fraction of peaks associated with true detections and high enough to make the selected sample contain a large percentage of true sources and to include most 'well-shaped' maxima. This shorter list is then sorted in descending order of SNR

and one-by-one the maxima are sent to the nested sampler, which returns an evidence estimate and a set of weighted samples that we use to model the full joint posterior distribution[1]. The final catalogue is almost completely independent of SNR threshold if this is not too high[2]. From these posterior distribution samples one can compute any parameter estimate, draw joint distribution surfaces, predict HPD intervals of any content over the marginalized distributions to infer the parameter uncertainties, etc., as in the examples presented in [Planck Collaboration et al., 2011f, fig. 9] and Planck et al. [2012]. The current implementation of PwS (v3.6+) computes maximum-likelihood, posterior mode and posterior median estimates of the source parameters. However, as discussed in section 2.2.1.1, owing to the discrete nature of the parameter space or its processing, the uncertainties of these estimates will always be bounded from below by the 'atom' of the representation. Another difficulty happens when analysing low SNR detections. The likelihood manifold becomes almost flat and very 'noisy', introducing a very large dispersion on the estimates, especially on the MLE and the posterior mode. To overcome these difficulties, PwS uses the set of samples drawn from the parameter posterior distributions. To stabilise the estimators PwS employs the expected value estimator but over HPDs of different probability content. Table 3.1 shows a summary of how the PwS's estimators and uncertainties are implemented[3].

3.4.3 Gaussian approximation to posterior maxima

We generally expect from the Central Limit Theorem (CLT) and the likelihood asymptotic properties that, as we add more and more data, in the neighbourhood of the likelihood maxima, the parameters probability distribution should converge to a multivariate Gaussian [Gelman et al., 2003, ch. 4]. Therefore, a multivariate Gaussian approximation should yield a good description of the likelihood shape where its mass concentrates. So

$$\mathcal{L}_{s/0}(\boldsymbol{\Theta}) = \frac{\mathcal{L}_s(\boldsymbol{\Theta})}{\mathcal{L}_0} \approx \mathcal{L}_{s/0}(\widehat{\boldsymbol{\Theta}}) \exp\left(\frac{\boldsymbol{\delta\theta}^t \boldsymbol{F} \boldsymbol{\delta\theta}}{2}\right), \tag{3.57}$$

[1]The maximum-likelihood estimates from the maximization step are updated, if necessary, during the sampling phase.

[2]This initial selection by SNR thresholding is only applied on computation efficiency grounds. Very low SNR maxima take much longer to sample as the likelihood manifold peaks are not properly shaped and most of them turn out to be rejected.

[3]The expected value estimator, it is only evaluated over the 95% HPD to enforce its robustness against outliers.

PwS estimator	HPD content	Default priors	Comments
Posterior mode	10%	Non-informative	Default in non-blind detection
Expected value	95%	Informative	Evaluated over the HPD - default in blind detection
Median	100%	Informative	Currently not used
Non-symmetrical error bars ($2\sigma_{-}^{+}$)	95%	Informative	Limits of the HPD
Symmetrical error bars (σ)	100%	Informative	Square root of the HPD variance

Table 3.1: PwS Bayesian estimators implementation. The HPD probability content are default values and may be changed just like the default priors.

where $\widehat{\boldsymbol{\Theta}}$ are the coordinates of the likelihood maximum, $\mathcal{L}_{s/0}(\widehat{\boldsymbol{\Theta}})$ is the likelihood ratio at the maximum, $\boldsymbol{\delta\theta} = \left(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}\right)$ and $\boldsymbol{F} \equiv -\left[\nabla\nabla \ln(\mathcal{L}_{1/0})\right]_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}}$ is the curvature matrix in the neighbourhood of the maximum[1]. After equation 2.14 the decision rule reads

$$\varrho_s \equiv \ln\left[\frac{\Pr(H_s|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right] = \mathcal{R}_{s/0} + \mathcal{L}'_{s/0} + \ln(\mathcal{E}), \tag{3.58}$$

where $\mathcal{R}_{s/0} = \ln\left[\frac{\Pr(H_s)}{\Pr(H_0)}\right]$ is the ratio of the priors on the models, $\mathcal{L}'_{s/0} \equiv \ln\left[\mathcal{L}_{s/0}(\widehat{\boldsymbol{\Theta}})\right]$ and

$$\mathcal{E} \equiv \int \exp\left(-\frac{\boldsymbol{\delta\theta}^t\, \boldsymbol{F}\, \boldsymbol{\delta\theta}}{2}\right) \boldsymbol{\pi}(\boldsymbol{\Theta})\, \mathrm{d}^n\Theta, \tag{3.59}$$

where $\pi(\boldsymbol{\Theta})$ represents the set of priors and n the dimensionality of the parameter space. Considering all priors ought to be properly normalised then, by the average value theorem $\mathcal{E} \leq 1$. Accepting that previous to the experiment, no reason should lead us to prefer H_s over H_0, then $\mathcal{R}_{s/0} = 0$ and

$$\varrho_s \leq \mathcal{L}'_{s/0}. \tag{3.60}$$

The ratio of the posteriors is never larger than the ratio of the likelihoods evaluated at the maximum [Trotta, 2008].

If the $\boldsymbol{\pi}(\boldsymbol{\Theta})$ set only contains uniform priors and these do not truncate any signifi-

[1]The Fisher information matrix is the assemble average of $\boldsymbol{F}$.

cant amount of the likelihood mass, then $\mathcal{E}$ may be approximated by

$$\mathcal{E} \approx (2\pi)^{n/2} \frac{\sqrt{(\det \boldsymbol{F})^{-1}}}{\Delta V_p} \tag{3.61}$$

where $\Delta V_p \equiv \Delta\pi_1 \Delta\pi_2 \cdots \Delta\pi_n$ is the *prior* volume defined by the allowed ranges of the parameters. The Bayesian interpretation of $\boldsymbol{F}$ is such that

$$\boldsymbol{F}^{-1} = \left\langle \boldsymbol{\delta\theta}\boldsymbol{\delta\theta}^t \right\rangle \tag{3.62}$$

where $\left\langle \boldsymbol{\delta\theta}\boldsymbol{\delta\theta}^t \right\rangle$ is the covariance matrix of the parameter uncertainties [Sivia and Skilling, 2006, ch 3.5]. Therefore $\sqrt{(\det \boldsymbol{F})^{-1}} = \Delta V_{\mathcal{L}}$, where $\Delta V_{\mathcal{L}}$ is the *posterior* parameter uncertainty volume. Thus

$$\mathcal{E} \approx (2\pi)^{n/2} \frac{\Delta V_{\mathcal{L}}}{\Delta V_p} \tag{3.63}$$

may be interpreted as the '*parameter space volume compression*' ($\mathcal{E} \leq 1$) or the ratio between the volume of the initial (prior) parameter uncertainty and that after we have added and interpreted the data according to model H_s (posterior). Bringing everything together

$$\varrho_s \approx C + {\mathcal{L}'}_{s/0} - \frac{\Delta V_p}{\Delta V_{\mathcal{L}_{s/0}}} \tag{3.64}$$

where C is a constant that only depends on purely numerical values and the prior probabilities of the models $\mathcal{R}_{s/0}$. The decision value (ϱ_s) results from a balance between two opposite trends:

- *A positive term* ${\mathcal{L}'}_{s/0}$.
  Adding more adjustable parameters to the model it is always possible, in principle, to reduce the χ^2 of the residuals, thus increasing $\mathcal{L}_{s/0}$.

- *A negative term* $\frac{\Delta V_p}{\Delta V_{\mathcal{L}_{s/0}}}$.
  However, the more numerous the parameters are (a more complex model), the larger the compression becomes as $\Delta\pi_j/\sigma_j \leq 1$ by our assumptions.

The idea of simplicity, just as described by Occam's razor parable, that simpler more prescriptive models should be favoured, naturally occurs as another consequence of

the application of DT supported by Bayesian probability. The above results are much more general than this simple derivation might denote and may be derived under much more general conditions [Jaynes, 2004, ch. 20].

A case worthy of further study is when $\mathcal{E} \approx 1$. From formula 3.59, the exponential term is always ≤ 1. So, $\mathcal{E}$ is approximately 1 only if the priors $\boldsymbol{\pi}(\boldsymbol{\Theta})$, (i) enclose $\widehat{\boldsymbol{\Theta}}$, (ii) are narrow enough to consider $\exp(\cdots)$ is always ≈ 1 where they concentrate most of their mass. This only happens if the prior contains the best estimate and the data plus the model constrain the uncertainty no more than we already knew. A hardly useful case! This is yet another very interesting property of the Bayesian model selection: The parameters the likelihood cannot constrain will not play any role in model selection [Trotta, 2008].

Accepting that the beam Point Spread Function (PSF) and the sources themselves possess reflection symmetry on both axes, the extragalactic objects are uniformly distributed across the sky and their surface density λ is not too high[1], it is possible to reduce the dimensionality of the integral in expression 3.59. Using the F matrix symmetries (see section A-3), the integral may be split and the position variables part evaluated

$$\mathcal{E} \approx 2\pi\lambda\sqrt{F_{xx}^{-1}F_{yy}^{-1}}\,\overline{\mathcal{E}}, \tag{3.65}$$

where $\overline{\mathcal{E}}$ is expression 3.59 but only extended to the complement of the position parameter sub-space[2]. In any practical case the dimensionality of $\overline{\mathcal{E}}$ never exceeds 4 and the integral can be evaluated using a simple quadrature formula [Press et al., 1992]. Since we only need a very low numerical accuracy in the evaluation of the evidences to produce science quality catalogues (see section 2.3) the evaluation of this integral is usually a very fast operation.

In the particular case of point sources using a non-informative flux prior $\pi(S) = 1/\Delta S$, the decision variable reads

$$\varrho_s = \mathcal{L}'_{s/0} + \ln\left[\frac{(2\pi)^{3/2}\lambda(F_{xx}F_{yy}F_{ss})^{-1/2}}{\Delta S}\right] + \ln\left[\frac{\Pr(H_s)}{\Pr(H_0)}\right] \tag{3.66}$$

In the above expression the middle term can be made arbitrarily small just by increas-

[1] If centring a square whose side is $4 \times \mathrm{MAX}\{\sigma_x, \sigma_y\}$ on top of each source does not make them overlap, that ensures the source density is not too high.

[2] The units of area must be the same as those used to express λ.

ing ΔS. So, independently of how large $\mathcal{L}'_{s/0}$ is, if our data model cannot, a priori, limit the range of values allowed for S, H_s will never be accepted [Efstathiou, 2008; Linder and Miquel, 2008; Trotta, 2008] ! We have chosen the point sources example because it embodies a paradigmatic case of the above difficulties. One may argue source fluxes must be limited on physical grounds and ΔS cannot be made arbitrarily large. However the dynamic range of point source fluxes is so high that a flat prior that does not discard neither the faint nor the bright source populations would severely penalise the faint tail. That would make the decision engine so conservative that most of the low flux sources, the dominant population, would end up being rejected even if the detections were significant. However, the flat prior is the correct non-informative for this case (see section 3.3.4 and G. E. Box [1992]). This bring us, once again, to the problem that introduced this chapter: the difference between model selection and estimation. Indeed the flat prior is the non-informative prior for this case. However, the assumptions it was derived upon apply to estimation only and not to model selection.

3.4.4 A non-informative prior for model selection

In the previous sections we have made it clear that the assumptions for the derivation of non-informative priors were prompted by the need to create Bayesian data-only driven *estimators* and those were not appropriate for model assessment. Jeffreys [1961, ch. 5.2] suggested that when comparing a model with a free parameter say $\xi \geq 0$, against a simpler one $\xi = 0$, which exactly matches the detection cases, a better choice for a non-informative prior would be

$$\pi(\xi|\sigma_\xi) = \frac{2}{\pi\sigma_\xi(1+\xi^2/\sigma_\xi^2)}, \tag{3.67}$$

where σ_ξ is the standard deviation of *one single measurement* of ξ assuming the '*noise*' is Gaussian. The rationale behind his derivation is[1]:

- The error of the measuring device σ_ξ defines a natural scale of magnitude for the parameter.

[1]This author believes Jeffreys derivation applies equally well to any '*location*' parameter as those can be mapped into the present case by a simple coordinate translation.

- One single measurement is not enough to make us decide in favour or against any of the challenging hypothesis.

If the number of pixels enclosed by a point source is not small ($n \gtrsim 30$), a condition easily met by today's cosmological data sets, then some simplifying approximations can be introduced in the more general expression given in Jeffreys [1961, ch. 5.2] to obtain the following result for point sources

$$\varrho_s = \mathcal{L}'_{s/0} + \ln\left[\lambda_\Delta \frac{\sigma_s}{\sigma_\xi}\right] + C. \tag{3.68}$$

where $C = \ln\left[2\left(\frac{2}{\pi}\right)^{1/2}\frac{\Pr(H_s)}{\Pr(H_0)}\right]$ is just a constant, $\lambda_\Delta = \lambda\left(F_{xx}F_{yy}\right)^{-1/2}$ is the expected number of sources inside the position uncertainty area $= \left(F_{xx}F_{yy}\right)^{-1/2}$ and σ_s is given by 3.29. As we have demonstrated, this prior does not show any of the difficulties described in the previous section and actually resolves most if not all of Efstathiou's criticism. The lesson to learn is that there is a fundamental difference between detection (model assessment and decision) and characterization (estimation). This difference needs to be propagated into the construction of the non-informative priors with the risk of creating artificial paradoxes.

3.5 Statistical inhomogeneity of the background

Real observations will also inevitably exhibit some statistical inhomogeneity of the background, in contradiction to our assumed model. Consequently, the conditions of optimality derived therefrom no longer hold. This can lead to a number of difficulties in detecting and characterising discrete objects, particularly in regions of the sky that contain bright, very inhomogeneous and anisotropic backgrounds. Indeed, this general expectation has been borne out in applying earlier versions of PwS to detailed simulations of *Planck* observations [PSM 1.6, Delabrouille et al., 2012]. In particular, the presence of bright diffuse Galactic dust emission was found to lead to the PwS SZ catalogue (in common with catalogues produced by other methods, such as MMF) containing bright spurious detections. Hence one did not obtain a regular cumulative purity curve that slowly approaches unity as the ln(evidence), or the SNR, increases [Melin et al., 2012], in contradiction to what would be expected from theory if our

model explained the data properly.

Indeed, the detection of SZ galaxy clusters highlights further problems. Again in the analysis of *Planck* simulations using previous versions of PwS, one finds that bright spurious SZ signals are not only concentrated in complex background regions, with a fraction of the bright spurious detections spread all across the sky. By cross-correlating the resulting SZ catalogues with ancillary point source data sets, one finds that bright spurious cluster detections matched bright point source locations. In our preliminary attempts to address this problem, we therefore first performed a point source extraction step and subsequently subtracted/masked the best-fit point source profiles in the maps. This pre-processing step greatly helped in reducing the number of spurious detections, especially those with very high evidence values. Another approach has been suggested by the *Planck* WG5 team, namely the 'χ^2 test' [Planck Collaboration et al., 2011f]. This performed very well, although, once more, there is no easy way to choose a robust acceptance/rejection threshold for the test. Another difficulty occurs when extracting the SZ effect at each individual channel. The SNR was usually so low that the measurements ended up being quite noisy.

Can we do any better using Bayesian logic? The apparent failure of the 'best' test can be immediately explained using the main Bayesian decision equation 2.9. Our decision criterion is based on the $\ln(\text{odds})$, namely

$$\ln\left[\frac{\Pr(H_s|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right]. \tag{3.69}$$

The problem comes from the denominator $\Pr(H_0|\boldsymbol{d})$. When we find a point source, its probability of being a cluster, $\Pr(H_s|\boldsymbol{d})$, is very low, but the probability of those pixels being part of the background, $\Pr(H_0|\boldsymbol{d})$, is also very low, because point sources do not fit our model of the background either. We have already mentioned that the binary model is too simple to handle realistic astronomical situations. To secure the optimality of our methodology we must ensure that the data is well described by our model, and employ a multi-model approach, as described in sections 3.7 and 2.2.2.

3.6 Comments on non-Gaussianity

It is clear that our model of the observations, like any model, is only an approximation to the real data. This is true both for our model of the discrete objects and for our model of the background. For the latter, it is clear that the background emission in real observations is neither Gaussian nor statistically homogeneous. Regarding non-Gaussianity, we do not mean that of a primordial origin, which, if exists, would have an insignificant effect in our analysis. We are instead alluding to the non-Gaussianity induced by the Galactic emission components, the confusion noise created by the sources below the detection threshold, the instrumental noise artefacts coming from the incomplete removal of the cosmic rays glitches and, of course, a wealth of other possible sources.

Many authors simply ignore this issue and many others dismiss its importance. A very strong argument, used many times, is that despite the sky emission being admittedly non-Gaussian, the effect of the finite PSF of beams will combine many different sky locations into a single pixel. In addition, signal de-noising procedures further combine more samples. Some authors then appeal to the Central Limit Theorem (CLT) to claim that non-Gaussian effects in the final data must be completely negligible.

This argument seems particularly appealing, but a deeper analysis of the CLT shows that, in our particular problem, namely detection and separation of two signals, the effects of the CLT are not as important as those authors claim. Formally, the CLT only applies when $N \to \infty$, where N is the number of random deviates in the sum. For finite N, the CLT only guarantees the Gaussian approximation is good for '*a region around the mode*' [Bouchaud and Potters, 2004]. The size of this Gaussian region grows very slowly. In the worst case, the distributions of the individual deviates are skewed and have 'fat tails'. Let us focus on a real example: the Galactic emission. If the spectral brightness distribution follows a power law with a finite first moment, to guarantee the field has physical behaviour, the normalised central Gaussian region, $|u|$, only grows very slowly with N:

$$|u| \ll u_0 \propto \sqrt{\ln N} \tag{3.70}$$

where u_0 is the tail lower boundary. This means that the sum must have more than 1000 terms to make the Gaussian approximation acceptable up to about $|u| \sim 2.6$.

In detection problems, where we want to separate the maxima created by the sources from the background fluctuations, we are dealing all the time with the background distribution upper tail:

$$\mathcal{P}_{\underset{u_0}{>}} \equiv \int_{u_0}^{\infty} \Pr(u)\, \mathrm{d}u. \tag{3.71}$$

If the background field intensity distribution follows a power law: $\Pr(I_\nu) \propto {I_\nu}^{-\mu}$, with $\mu > 2$, to guarantee its energy is finite, then the probability that a sum of N deviates falls into the upper tail region of the sum normalised distribution is:

$$\mathcal{P}_{\underset{u_0}{>}} \propto \frac{1}{N^{\mu/2-1}\left[\ln N\right]^{\mu/2}}. \tag{3.72}$$

This is a very serious problem. Object detection methodologies are designed typically to suppress the background and amplify what does not fit its model. The non-gaussianity component is not part of our background model, so its effect on the detection process is doubly pernicious: not only it is not removed, it is amplified.

There seem to be only two ways of circumventing this problem: (i) to include the non-Gaussian effects in the statistical models; and (ii) to manipulate and add as much data as possible to make it more Gaussian. Owing to the complexity of *Planck* data it is almost impossible to give a proper account of the non-Gaussian effects without making the problem unsolvable. So, a workable solution must necessarily combine as much data as possible, and then analyse the outcome. The only possible way of doing this is to use multi-channel analysis all the time.

Our own experience corroborates this view. The SNR values of the PwS selected detections and the thresholds the frequentist methods normally employed ($\gtrsim 4.0$), are much higher than what would be expected according to the purity levels of the catalogues if the statistics were purely Gaussian. Although, the channels with the largest beams, where each pixel is the result of a much higher number of different contributions, do indeed have detection thresholds lower and closer to those expected from the Gaussian theory. A good practical example of how the multi-channel processing can help the reduction of the impact of the non-Gaussian distributions on the detection process is the recovery of the SZ signal [Melin et al., 2012].

Owing to the residual non-Gaussianity left in the background, especially close to the Galactic plane, we should now expect a higher number of background fluctua-

tions reaching above the evidence threshold level than those predicted by the Gaussian model. So, eventually, we need to correct the prior on the models: $\frac{\Pr(H_1)}{\Pr(H_0)}$, as this prior was derived assuming that the background had purely Gaussian statistics. The simplest way, we believe, is just to count the total number of fluctuations above the SNR threshold adopted, before embarking on the evaluation of the evidence. In particular, one should compare this number with what would be expected from the Gaussian model plus the predicted source counts above the SNR threshold, then take the larger quantity. Denoting this value by T, a corrected estimate of λ_0 (see formula 3.46) would read

$$\lambda_0 \simeq T - \lambda_1. \tag{3.73}$$

This very simple 'trick' provides a first-order correction to the effects of background non-Gaussianity.

3.7 Multi-frequency, multi-model catalogue making

In general detection methods always exploit some particular trait of dissimilarity between the targeted objects and the background they are embedded in. These might be different spatial properties and/or a contrasting emission SED. For the reasons outlined above, we believe that a deeper and purer catalogue can only be obtained through multi-frequency analysis. An excellent example of the power of such an approach is provided by the detection of SZ clusters. Despite the SZ signal being sub-dominant on all *Planck* channels (the signal level is below that of the background), an optimal combination of the different frequencies can boost these extremely faint signals to the point where one can now build reliable catalogues of many hundreds of such objects [Planck Collaboration et al., 2011f; Reichardt et al., 2012]

We have also demonstrated above that our simple binary decision making approach is too naïve to handle 'real-life' situations. The introduction of a multi-model (more than two models) decision rule cannot, however, be achieved simply by extending the binary case ([Jaynes, 2004, ch. 3]). Although, in most of the situations of interest in astronomy, it is always possible to implement a general multi-model decision rule evaluating in succession each one of the possible source models against that of the background 2.2.2.1. So, if the implementation of the algorithm is versatile enough

to support different object models then a code version designed with just the binary evidence comparison in mind can be easily extended to multi-model. Some authors recognise the power of the Bayesian methods 'if' the data only includes exactly what we anticipated would be there. But they are critical of the possible consequences some very interesting unexpected new entity would have in the inference machine. Observing the posterior distributions might provide an early warning. Confirming whether the data has moved the posterior mass against one of the 'prior walls' should raise some alarm. Or, if replacing the priors by non-informative ones and repeating the exercise considerably changes the evidence, that might be another symptom of model mismatch. Generally, it means our data has enough power to constrain the estimates but somehow the prior is getting in the way. Or, a bit more technically, it has a significant impact on the estimation. In general, one should never expect this to happen. But, real data is always full of (mostly unwanted) surprises (see 3.5). That is when our generic model (see 3.1) unleashes its full power. Not only it is there for the most exciting sources, those not falling in any of the known SED models, as for least interesting as well, the localised artefacts. Some might question whether this 'safety net' might on the contrary mislead us. As the physical SEDs models are just a special case of the generic, what prevents our 'decision machine' not picking it even when one of the other models is present. The generic SED is necessarily more complex (has more adjustable parameters) than any of the physical ones[1]. But this is now the case when Bayesian evidence really excels as a model selection tool. Either the generic model fits the data much better (higher SNR) or owing to the enlarged prior volume of the parameter space, the generic model will always be deprecated in favour of the others. This is an interesting example of Occam's razor in full swing. For a very detailed and elegant exposition why this happens please refer to Jaynes [2004, ch. 20] and Jeffreys [1961, ch. 5.2].

The last step of PwS is to assemble the final catalogue from a list of candidates. During this stage, PwS performs the following steps:

1. maps flat sky patches back onto the sphere at the positions of the putative detections;

2. applies a detection mask, if any;

[1] Otherwise the physical SED models would only be a re-parametrization of the generic one.

3. merges multiple detections of the same source obtained in different patches into a single candidate detection; and

4. makes the final catalogue by rejecting those entries that do not meet the pre-established criterion of purity or loss.

The last step is critical to the success of our methodology. If the selection criterion is based on losses, then we just need to trim the 'proto-catalogue' further by applying the decision rule 2.16. But, as we mentioned previously, it is much more common in astronomy to require a catalogue should be as complete as possible (contain the largest possible number of detections) but with a predicted contamination ratio (purity). However completeness is as important as purity. Completeness is the percentage of objects, above a certain amplitude/flux, present in our data that we successfully included in our catalogue, or, in probability language $\Pr(D_s|H_s, S_s > \xi)$, where S_s stands for any quantity we are imposing the threshold on and ξ is the threshold (see section 2.3). To give a proper assessment of the purity and completeness, except on the simplest, though not realistic, cases frequentist methods ought to resort to simulations. The characterization (purity and completeness) of the catalogue is entirely defined by imposing a single SNR cut[1]. Obviously, the simulations need to emulate the most realistic astronomical backgrounds at the frequencies of interest and the instrumental effects as closely as possible. Modern cosmological data sets are increasingly larger and more complex, rendering this task immensely resource consuming.

If we trust the statistical model we employ to describe the data then decision theory, supported by Bayesian logic, supplies the optimal tools to construct a catalogue with a predicted purity. This is exactly what we have shown in section 2.3. However, completeness needs to be addressed using a different Bayesian probabilistic model, a Bayesian hierarchical or random effect model (see section 2.3.2). Very fortunately this apparent limitation of our simplified approach may be easily overcome by very realistic simulations and at the same time of simple and fast construction: by injecting, and later recover, mock source populations into the real maps. These simulations might not be optimal to assess purity but they provide the most realistic possible test bed for measuring completeness [Planck Collaboration et al., 2011e].

[1] If the catalogue is extracted from a homogenous region, both in terms of instrumental noise and astronomical background, then a flux/amplitude cut may be used instead.

In some extreme cases, 'complex regions' (SMC, LMC, Orion, etc.) or the Galactic plane for instance, the splitting into small patches is not enough to satisfy neither the isotropy nor the homogeneity of the background or the source number counts. Moreover, the background statistics deviate so much from the assumed model or source blending is so acute that the operation of the code may break and the results stop being reliable. In this case we actually apply a mask to those regions (see A-4). Although, in the case of point sources, moderate source blending (no more than 3 sources per beam) is not out of reach for PwS. In this case the 'one source at a time' approximation no longer holds. Thus, we have now to employ the full likelihood expression and expand the single source parameter space to the Cartesian product of the spaces of the individual sources we pretend to de-blend. This extension of PwS will be addressed in a forthcoming publication.

Chapter 4

Results

In the two previous chapters we described, firstly, PwS's probabilistic and logic foundations. Then, we suggested an algorithm, founded on these principles and capable of producing catalogues of compact objects. We have always tried to keep a strong focus on the actual practicalities of the implementation of a truly functional solution. The current chapter will be devoted to presenting the results of the application of the developed code to a wide range of compact object detection scenarios. These scenarios include not only state-of-the-art realistic simulations, but the actual data derived from the last generation of spaceborne microwave probing instrumentation: ESA's *Planck* satellite[1].

A wealth of PwS results have already been published. These results were either obtained from simulations or actual *Planck early release* data [Planck Collaboration et al., 2011d]. In this work we aim to explore, with more emphasis, those aspects that received less attention from the community, as a complement to those previously released.

[1] *Planck* (http://www.esa.int/Planck) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states (in particular the lead countries France and Italy), with contributions form NASA (USA) and telescope reflectors provided by a collaboration between ESA and a scientific consortium led and funded by Denmark.

	HFI						LFI		
Channel [GHz]	100	143	217	353	545	857	30	44	70
$\lambda[\mu m]$	3000	2098	1382	850	550	350	10000	6818	4286
Detector technology	Spider-web and polarization bolometers						InP HEMT		
Detector temperature [K]	0.1						20.0		
Number of detectors$^{(a)}$	0 + 8	4 + 8	4 + 8	4 + 8	4 + 0	4 + 0	4	6	12
Angular resolution (FWHM) [arcmin]	9.88	7.18	4.87	4.65	4.72	4.39	32.89	30.23	12.97
Small scale noise $[K^2\text{rad}^2] \times 10^{-15}$	0.78	0.25	0.59	7.61	1492	423847	11.0	18.0	11.0

Table 4.1: $^{(a)}$(HFI) spider-web + polarisation sensitive bolometers; (LFI) Feeds.

4.1 *Planck* satellite - the data

4.1.1 Instrument and data description

European Space Agency (ESA) *Planck* satellite is the latest generation of space missions aimed to measure the anisotropies of the cosmic microwave background (CMB) [Planck Collaboration, 1999; Planck Collaboration et al., 2011d]. *Planck* observes the entire sky in nine channels spanning a wide frequency range starting at 30 GHz (microwave) up to 857 GHz (sub-mm). The scientific payload is made of two instruments the Low Frequency Instrument (LFI) [Mennella et al., 2011; Zacchei et al., 2011] and the High Frequency Instrument (HFI) [Planck HFI Core Team et al., 2011]. Table 5.1 summarises the main characteristics of both instruments. *Planck* orbits Lagrange L2 point [Lagrange, 1892] and scans the sky the way as shown in figure 4.1. *Planck* scans the sky in rings as described in Tauber et al. [2010]. The scanning strategy essentially follows a cycloidal path (normally 1 deg day^{-1}), with each individual ring, about 2 arcminutes wide, taking approximately 50 minutes to acquire. During that time, the satellite spins multiple times around its axis, allowing many high precision readings of the same sky location.

Figure 4.2 shows a typical set of HFI '*time-lines*'. The very thin vertical lines, shown in all panels, are glitches induced by cosmic-ray hits on the telescope focal plane[1]. These glitches are modelled, characterised and then subtracted from the data stream. Only the most intense are flagged for posterior removal. The upper panel shows a typical '*time-line*' from a '*cosmological*' channel. The periodic modulation is the result of the CMB dipole. Some very prominent and periodic peaks on the 545 GHz channel are the read outs of the Galactic emission particularly strong at this frequency.

[1]The LFI data is not affected by these glitches.

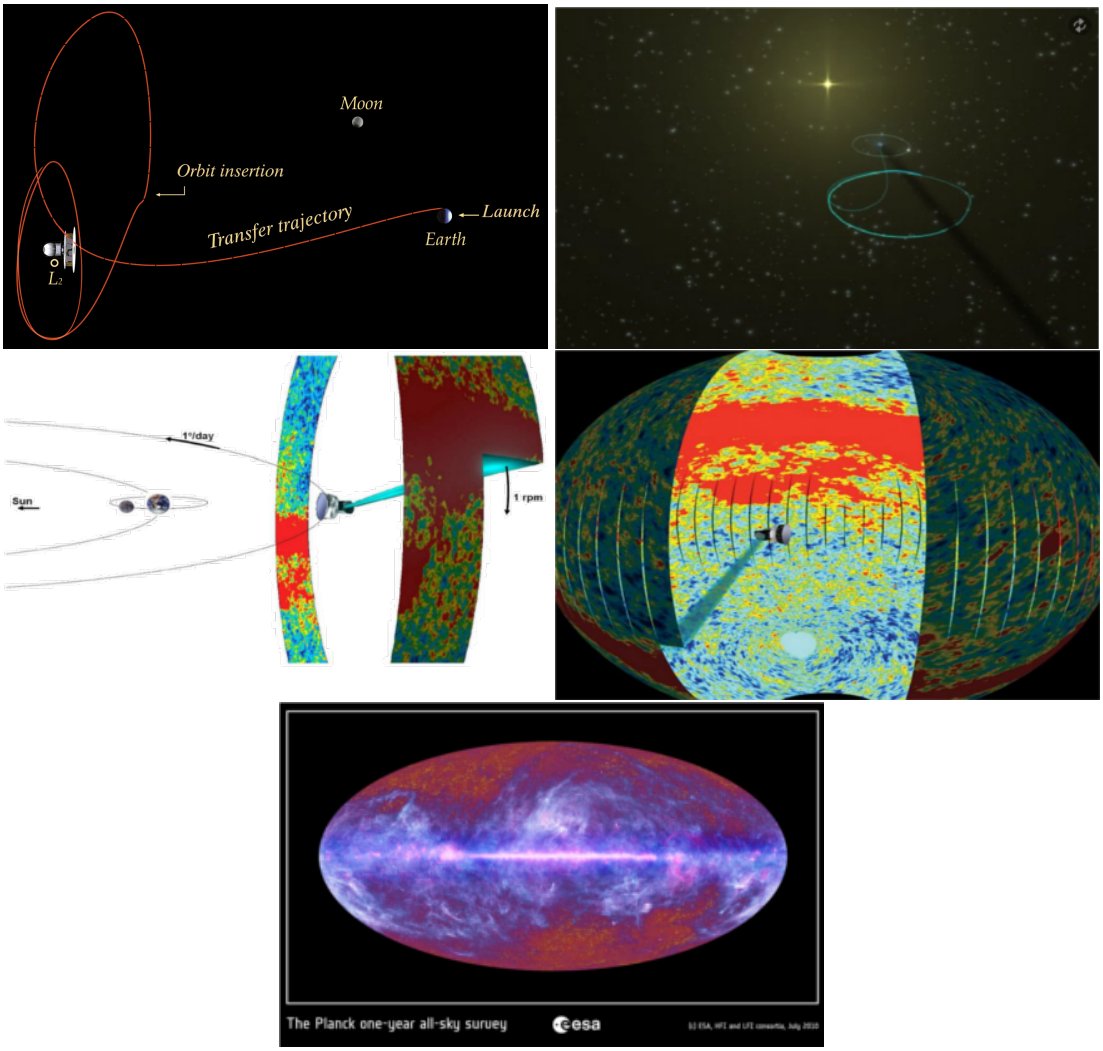


Figure 4.1: *Upper panels*: *Planck*'s flight and orbit around L2. *Middle panels*: *Planck*'s scanning strategy: *Planck* completes about one full rotation around its spin axis each minute. At the same time it precesses around a fixed direction, perpendicular to the symmetry axis, at the rate of 1° per day, completing a full sky survey in slightly less than one year. *Lower panel*: Composite image (false colours) of several *Planck* channels after the first year survey (images provided by the European Space Agency (ESA) website ESA [2009]).

The same features can be seen in the 143 GHz channel but attenuated. The '*dark*' bolometer (lower panel), a bolometer that does not receive any light, measures the signal drift caused by tiny variations in the instrument temperature and other transient

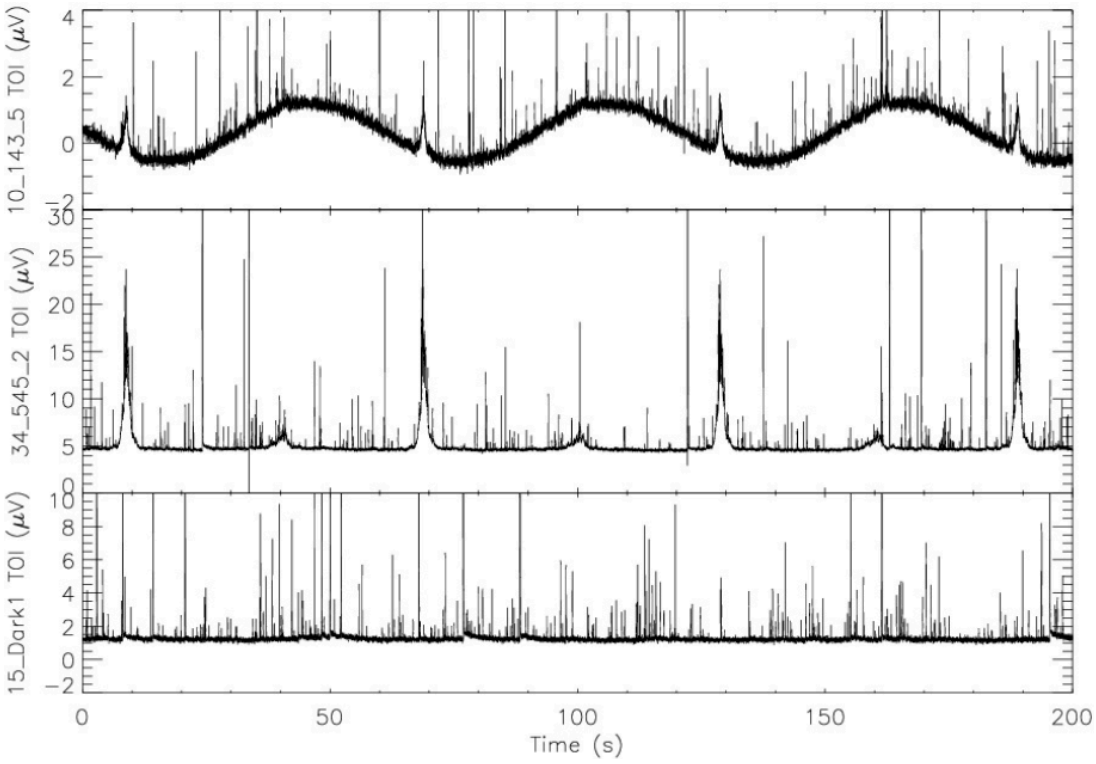


Figure 4.2: The *upper panel* shows a typical raw 'cosmological' TOI (143 GHz). The *middle panel* shows another raw TOI but this time from one of the sub-mm channels (545 GHz). The *lower panel* is the time domain representation of a '*dark*' bolometer, i.e. a bolometer does not receive any light (plot from Planck HFI Core Team et al. [2011]).

effects.

Figure 4.3 shows the same data as in 4.2, but after being pre-processed by the initial *Level 2* pipeline tasks (see section 4.2.1; for a complete description refer to Planck HFI Core Team et al. [2011]).

As result of the scanning strategy, there are regions observed many thousands of times ($\sim$ 10000), close to the ecliptic poles, and others observed just a few times around the ecliptic equator (see figure 4.4). This effect makes the instrumental noise very anisotropic, adding extra complexity to the statistical modeling of the background. The time ordered data are then put through a *map-making* pipeline package that assembles the pixel maps and the covariance (noise) information at each channel.

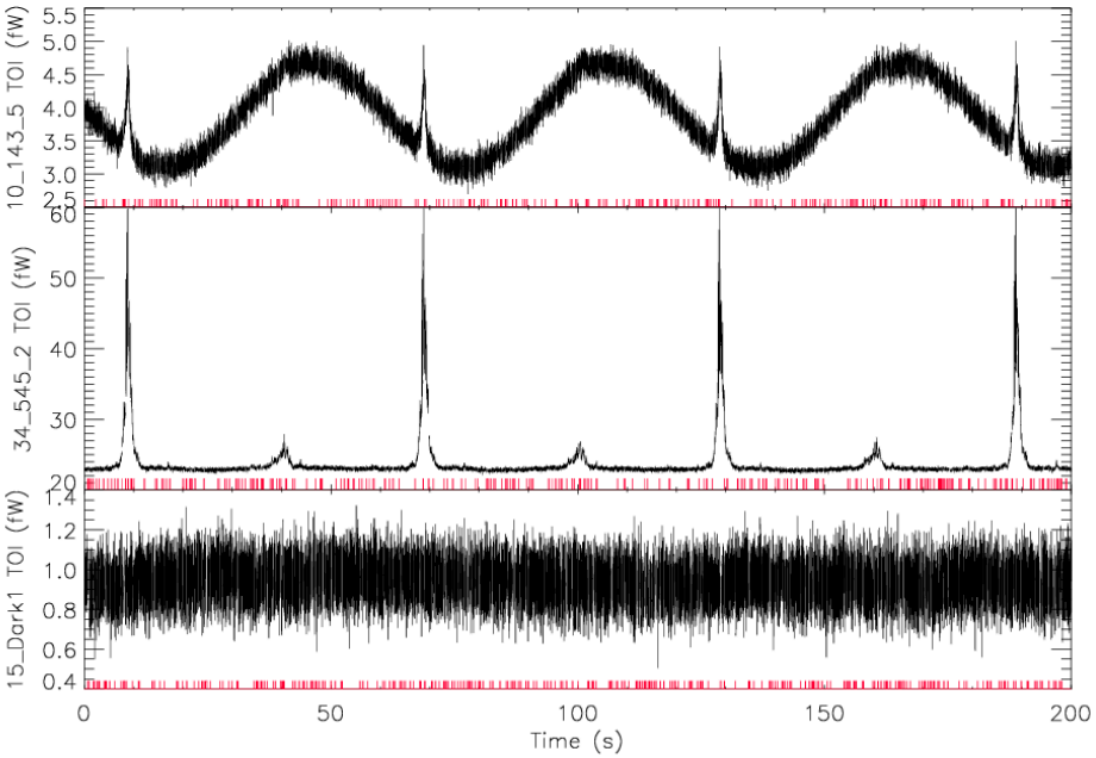


Figure 4.3: The panels show the same data as in figure 4.2 after detrending and glitch removal. The vertical red bars at the bottom of each panel indicate which portions of data are flagged (plot from Planck HFI Core Team et al. [2011]).

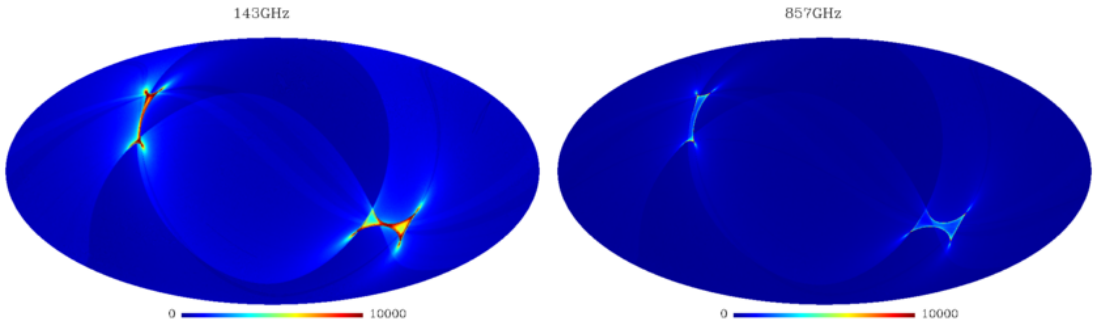


Figure 4.4: Full sky maps of two HFI channels, 143 and 857 GHz, showing the number of times each zone of the sky is observed ('*hit counts*') (plots from Planck HFI Core Team et al. [2011]).

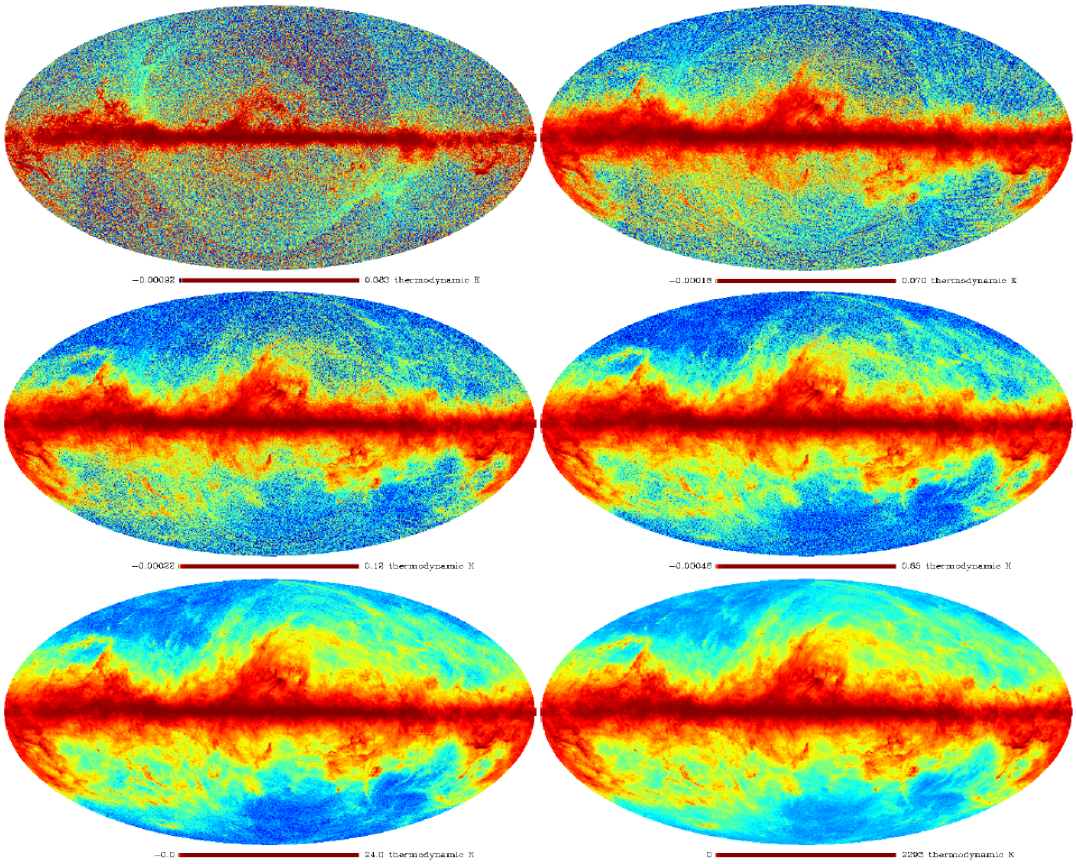


Figure 4.5: The six HFI channel pixel maps with the CMB component subtracted. Left to right and downwards: $100, 143, 217, 353, 545, 857$ GHz. The map units are K thermodynamic (plots from Planck HFI Core Team et al. [2011]).

Figure 4.5 shows the data from the 6 HFI channels with the CMB component removed. In the *cosmological* channels there is a significant amount of background noise clearly visible. This happens because the Galactic components away from the Galactic plane are very faint below the instrumental noise levels. As one goes up in frequency, the contribution from the CMB component fades away at the same time the Galactic dust and the extra-galactic dusty point sources emission becomes dominant.

Exploiting the high precision pointings in the time ordered data, accurate noise pixel maps can be easily constructed. Firstly, the scanning rings are split into two exclusive subsets, the *null test* maps. Then the map-making code is run over each set of *null test* maps individually, and finally a half-difference map is computed from

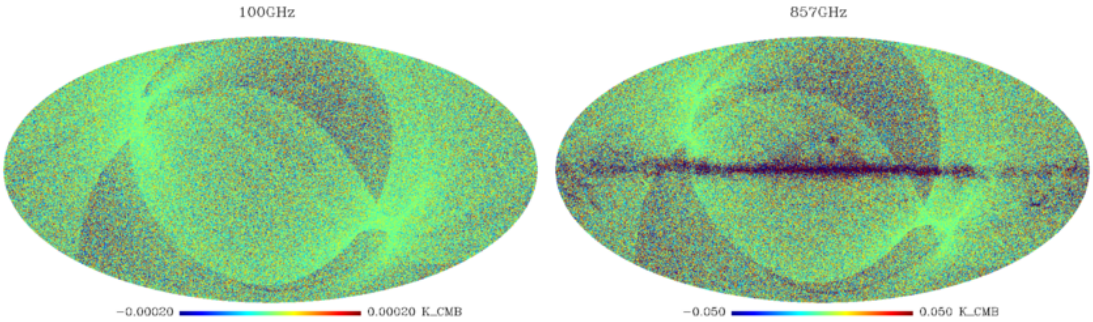


Figure 4.6: Two *null test* half-difference noise maps. The map units are K thermodynamic (plots from Planck HFI Core Team et al. [2011]).

these. Figure 4.6 shows two *null test* half-difference noise maps: 100 and 857 GHz. It is very clear that noise pattern correlates extremely well with what would be expected from the *hit count* maps. Although, for the 857 GHz channel some very strong noise can be seen around the Galactic plane as well. This is a spurious effect produced by strong gradients, on sub-pixel scales, occurring in that zone.

4.1.2 Data processing and analysis (pipeline)

Figure 4.7 outlines the basic building blocks of *Planck*'s data processing 'pipeline'.

- **Level 1 (L1):**
  contains the most basic low level data retrieval activities like reading out the bolometers. It is its goal to form a steady stream of read-out values, time stamps and pointings for each one of the focal plane bolometers. It delivers a time stream of data to *level 2*.

- **Level 2 (L2):**
  aggregates the data processing tools that operate at the time ordered data (TOI) level. This set of tools makes the *back-bone* of all analysis and data processing activities. Its goal is to deliver a well characterised set of pixel maps, temperature, polarization and noise, at each of the of *Planck* channels. At the same time it provides the *instrument model* (IMO), which includes the beam, calibration and noise definition. At its last stage it runs the destriping package to remove

the ring offsets, which happen as result of the $1/f$ noise and assemble the pixel and noise maps.

- **Level 3 (*L3*):**
  composes the assemble of data processing tools that operate directly on data and/or pixel noise maps. PwS is a *L3* package fully integrated in the HFI *Planck* pipeline DMC infrastructure.

The integration into the HFI DMC infrastructure posed one more challenge. The code ought to run multiple instances in parallel, reliably and in the shortest possible amount of time. An extensive and thorough validation of the data statistical modelling, implementation design and product quality (detection and characterisation) is a basic requirement of any scientific quality catalogue. In order to comply with the necessary requirements to undergo such stress tests, a complex code structure was developed, integrated in the DMC pipeline infrastructure and finally tested. Although fast and parallel execution have always been the main driving vectors of development, flexibility was paramount as well. Owing to its sound foundations and malleable implementation, one single piece of code is capable of handling almost, if not all the requirements to deliver all *Planck* compact source products.

4.2 Detection on real data

PwS has already been extensively used inside the *Planck consortium*, both in point source (point like shapeless objects) [Planck Collaboration et al., 2011e] and SZ cluster extraction [Planck Collaboration et al., 2011f] and characterization [Planck et al., 2012; Planck Collaboration et al., 2011g,i, 2012a,b,c].

In its current incarnation, the PwS code package is made of three different programs.

- Two natively compiled codes: (i) the pre+post processor utility (PPP) and (ii) the detection '*core*' (PwSC). The PPP utility handles all tasks non-associated with detection, like: masking, patch creation, pipeline interfacing, catalogue conversion, etc. The PwSC implements the actual detection and estimation tasks. These codes are written in C++.

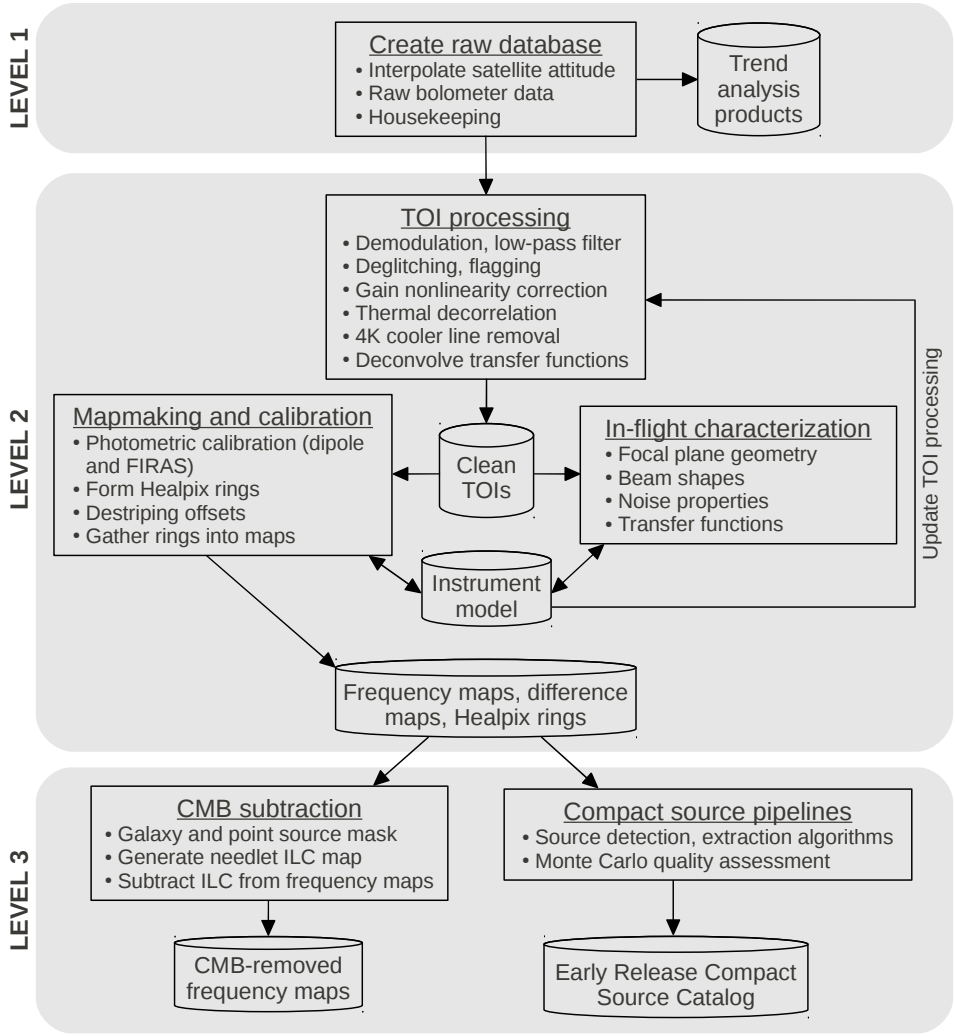


Figure 4.7: Outline of the HFI pipeline data flow and hierarchy of the processing tasks for the *Planck* early data release in 2011. PwS is a *Level 3* (L3) job (plot from Planck HFI Core Team et al. [2011]).

- A Python script that handles all HFI DMC pipeline activities. Activities like creating a stream of control and data parameters across the multiple serial and parallel tasks, spawning and synchronizing all jobs, etc.

4.2.1 Masks and ill-behaved pixel handling

Real data can deviate strongly from the models we use to describe it. In the *Planck* case, the highly inhomogeneous fields and the large frequency range of the observation channels adds an extra layer of complexity. In order to obtain reliable results, some regions of the sky ought to be masked and removed. Those regions include zones of very high noise, like the ill-observed pixels and others where the background model assumptions severely break. For instance, the Galactic plane or, in case of the SZ detection, the bright point sources (see section 3.5). Figure 4.8 shows the total combined mask: ill-observed pixels + Galactic dust emission + bright point sources, after being apodised. The pixels in the rejected areas can be further filled in to preserve the statistical homogeneity of the background. Another important role of the masks is to restrict the analysis to sub-sets of the data where their statistical properties are more homogenous and/or carry added scientific value. For instance, the background on high Galactic latitude regions are dominated by the very homogeneous CMB emission and it is the target of the extra-Galactic science. The opposite can be said about the low Galactic latitude regions.

4.2.2 Patch projection

PwS operates on small flat square patches (see appendix A-4). These flat patches are obtained using a gnomonic projection. PwS can split the sky is three different geometric arrangements:

- Healpix aligned with a spin (see upper panel of figure 4.9); this type of patch layout aligns the patch X and Y axis with the pixel dividing lines of the Healpix pixelixation, at each patch location, by orientating the patch appropriately. Its goal was to keep the same pixel density at each location and reduce possible aliasing effects (Moiré fringes) [Figueiredo et al., 2001].

- Constant latitude (Galactic or Ecliptic); the patches are projected by arranging the patch centres on top of lines of constant latitude and the patch X axis aligned with them.

- '*Non-blind*'; the patch X axis is oriented the same way as in the *constant latitude*

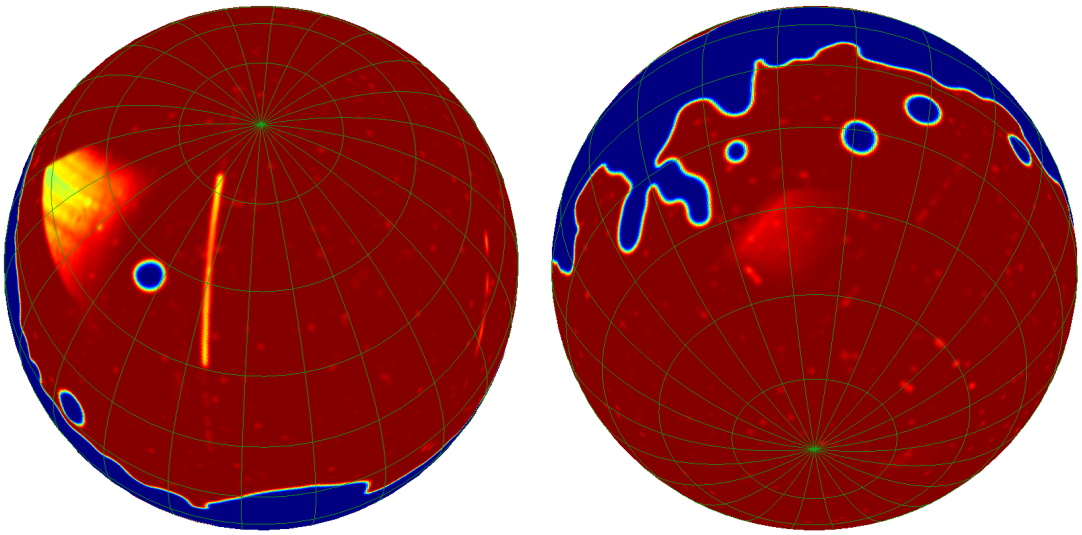


Figure 4.8: These two pictures show the combination of the ill-observed pixel, Galactic dust emission and bright point source mask apodised with 1° FWHM Gaussian kernel. Blue regions will be rejected and red kept. Left panel is the north pole and right panels the south. Coordinates are 'Galactic'.

case, but the centres are located on positions derived from a catalogue of sky coordinates[1].

The background and foreground statistical properties are assumed homogeneous inside each patch. Pixels sitting inside the masked areas are then replaced using some form of '*in-painting*' that tries to preserve the background traits across those areas [Bajkova, 2005][2].

Each patch is 512×512 pixels (SZ) or 256×256 (point sources) with respectively $14.66^\circ \times 14.66^\circ$ and $7.33^\circ \times 7.33^\circ$ areas (each pixel is $\sim 1.718' \times 1.718'$). In turn each patch is further sub-divided in three different zones (see figure 4.9, lower panel). All likelihood peaks sitting on top of the '*rejection*' zone are immediately discarded. Those on the '*detection*' region are assessed using the GLRT and then subtracted to avoid contaminating the '*core*' area. Only those likelihood maxima lying in the '*core*'

[1] The exact locations may be either the source positions or, if information from a '*blind*' extraction is available, the centre of the patch with the lowest noise and a matching detection (a detection is matched if it is inside a 5 arcmin radius).

[2] Simply ignoring the flagged pixels in our analysis would prevent the use of the FFT. The FFT can only be used if the samples are equally spaced and there are no 'holes'.

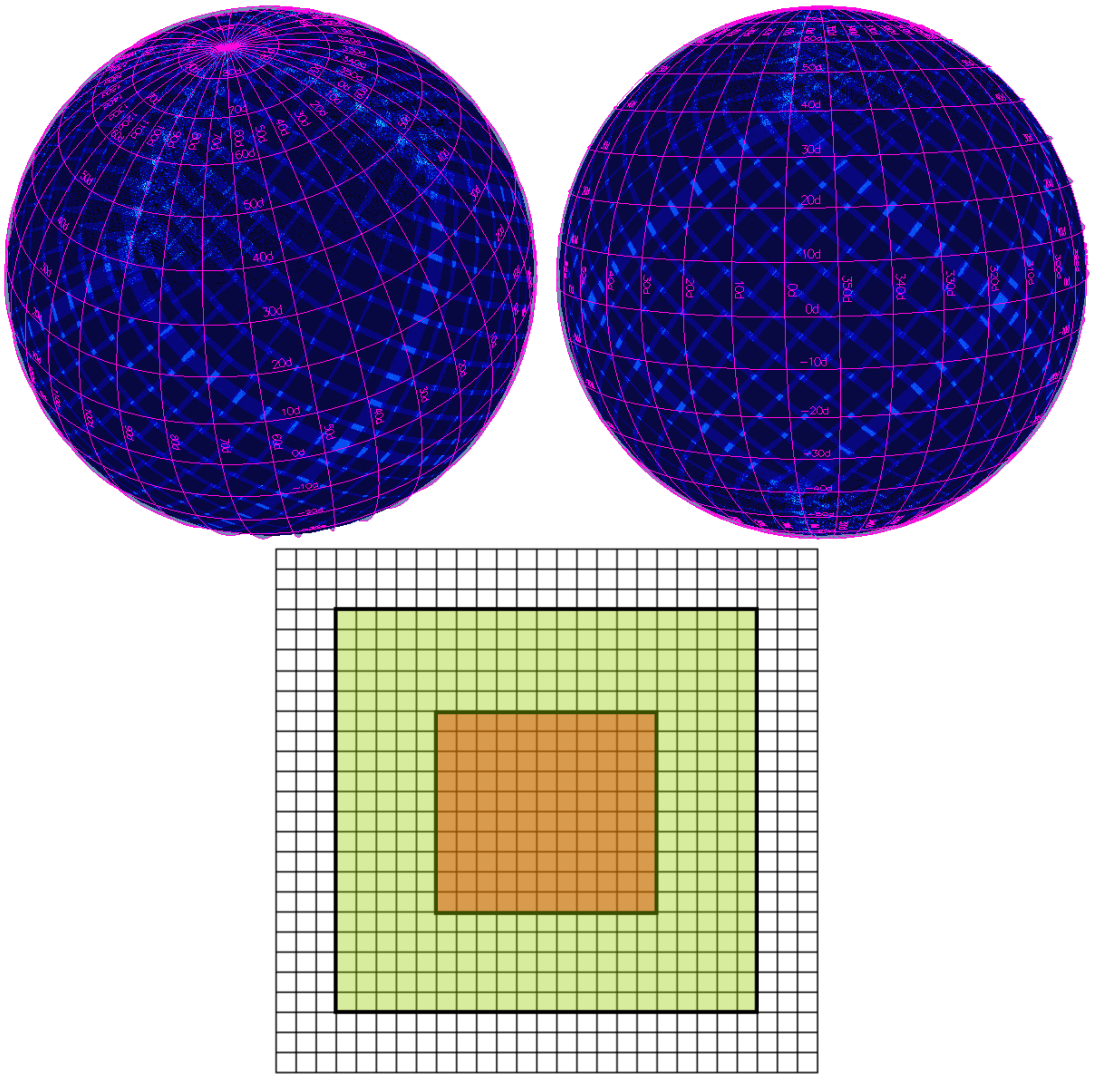


Figure 4.9: *Upper Panel*: The pictures show one possible distribution of patches PwS splits the sky: '*Healpix aligned with a spin*'. Lighter regions are those where the patch overlap is higher. *Lower Panel*: Zones of a single patch. The unfilled region is the '*rejection*' zone, the green the '*detection*' and the light brown the '*core*' regions.

zone are put through the Bayesian sampler and then added to an intermediate catalogue containing possible candidates to inclusion in the final catalogue. For a 85% mask, PwS cuts the sky in about 3100 patches when performing a SZ extraction and about four times more when doing point source detection. This large number of patches creates a very large inter-patch overlap with each cluster detected 4.7 times on average.

Later, when assembling the final catalogue, PwS rejects all multiple detections of a single object only keeping the highest SNR in GLRT mode or the highest $\ln(odds)$ in Bayesian operation.

4.2.3 Detection of point sources

Planck's Early Release Compact Source Catalogue (ERCSC) was the first official *Planck* product released and it is fully described in Planck Collaboration et al. [2011e] and references therein. The ERCSC was meant to be a single frequency extraction '10σ' product with the following goals:

- Performance
  - 90% reliability overall, $> 95\%$ reliability at high galactic latitude.
  - flux density cutoff: SNR of 10 or better.
  - flux density accuracy: better the 30%.
  - positional accuracy: better than beam FWHM $/5$ (1 sigma radial).
- The flux density cutoff will vary across the sky as the variance of background changes.

The goals of this early product were modest because it would only include a single full sky scan with many regions so far poorly observed. That gave rise to strong inhomogeneities in the instrumental noise properties, leaving many pixels still undefined and an overall sub-optimal map. PwS was selected, amongst several other algorithms (see Planck Collaboration et al. [2011e] and section 4.3.3), to deliver the point source catalogues up to and including 143 GHz channel.

Figure 4.10 shows the full sky spatial distribution of *Planck*'s ERCSC point sources catalogues extracted by PwS. Table 4.2 gives a brief summary of most important extra-Galactic ($|b|$[1]$> 30°$) traits of the ERCSC PwS catalogue. PwS faintest '*reliable source*' (see Planck Collaboration et al. [2011e]) flux density on every channel is well inside that of the goals. PwS operation on small patches allows adjusting the detection sensitivity threshold on a local basis according to the local background fluctuations. Owing

[1] b is the Galactic latitude cut.

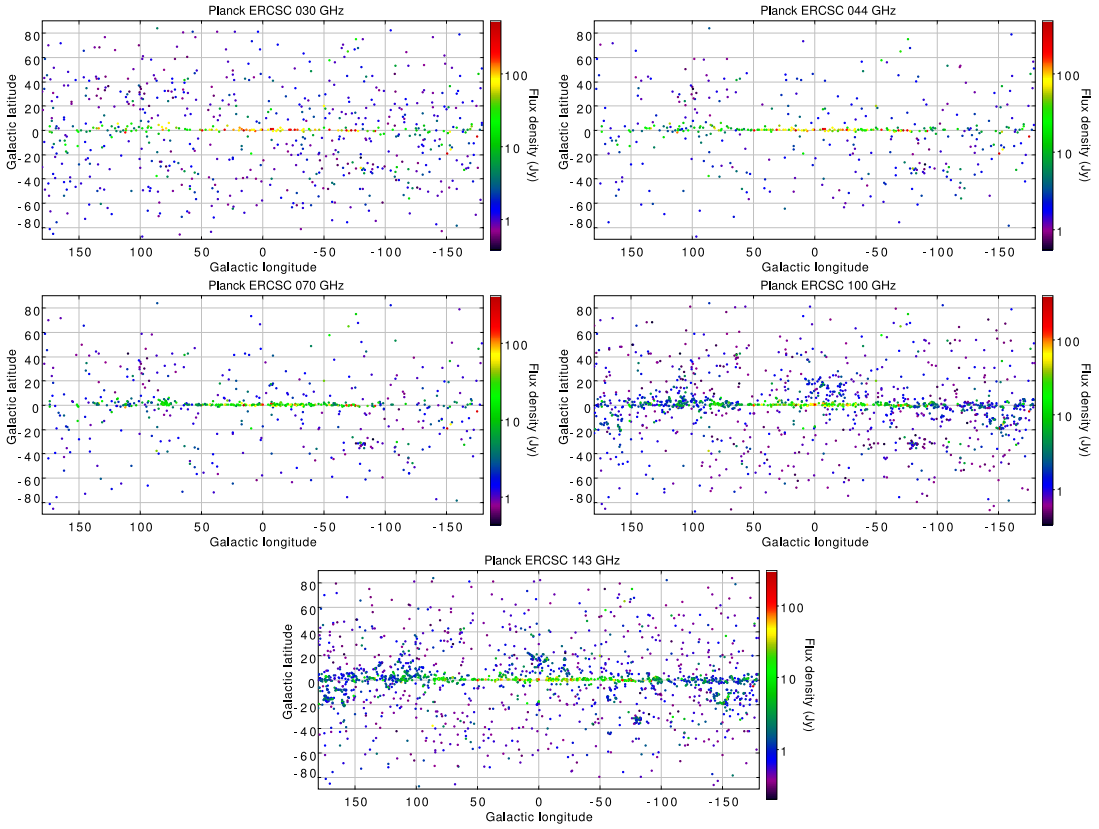


Figure 4.10: Sky distribution of the point source catalogues extracted by PwS at the lowest *Planck* five channels . The colour of each source shows the estimated flux density in [Jy] (plots from Planck Collaboration et al. [2011e]).

to *Planck* very inhomogeneous backgrounds this is a major advantage as it allows us to go '*deeper*' in '*cleaner*' regions[1].

4.2.4 Detection of galaxy clusters

The intra cluster medium gas (ICM) is rich in hot electrons which, through inverse-Compton scattering, disturb the CMB photons energy distribution to create localized secondary anisotropies. This effect is known as the Sunyaev-Zeldovich effect (SZ) [Birkinshaw, 1999; Carlstrom et al., 2002; Sunyaev and Zeldovich, 1972b]. The kinetic energy of the scatterers (hot electrons) is essentially the combined effect of two

[1]This is the reason why the '*local* 10σ' can be significantly different from the global average.

Frequency [GHz]	30	44	70	100	143
$\lambda[\mu\text{m}]$	10000	6818	4286	3000	2098
Resolution (FWHM) [arcmin]	32.89	30.23	12.97	9.88	7.18
$10\sigma^{(c)}$ goal [mJy]	790	820	660	650	580
# Sources (total)	705	452	599	1381	1764
# Sources $\wedge\ \lvert b\rvert^{(a)} > 30°$	307	143	157	332	420
$S^{(b)} > 10\sigma^{(d)} \wedge \lvert b\rvert > 30°$ (median) [mJy]	1173	2286	2250	1061	750
$S > 10\sigma^{(d)} \wedge \lvert b\rvert > 30°$ (faintest) [mJy]	487	1023	673	500	328
$S \wedge \lvert b\rvert > 30°$ (faintest) [mJy]	480	585	481	344	206

Table 4.2: $^{(a)}$ Galactic latitude cut; $^{(b)}$ Source flux density; $^{(c)}$ overall sensitivity goal; $^{(d)}$ local background σ.

distinct components. One, as result of the peculiar velocity of the cluster, manifests as the bulk movement of the electron gas in the comoving frame (CMB rest frame; Hubble flow) and it is named '*kinetic SZ*' (ΔT_{SZ_k}). Another type of SZ anisotropy is the '*thermal SZ*' (ΔT_{SZ}) and it originates from the random local component of the electrons velocity as result of their temperature[1]. The ratio of the temperature perturbation amplitude as result of the kinetic and thermal effects is given by Birkinshaw [1999],

$$\frac{\Delta T_{SZ_k}}{\Delta T_{SZ}} = \frac{1}{2}\frac{v_z}{c}\left(\frac{k_B T_e}{m_e c^2}\right)^{-1} = 0.085\left(\frac{v_z}{1000\,\text{kms}^{-1}}\right)\left(\frac{k_B T_e}{10\,\text{keV}}\right)^{-1} \lesssim 5\%, \quad (4.1)$$

where v_z is the velocity of the cluster projected in the *line of sight* ($l.o.s$) (positive when it moves away from the observer) and T_e the electrons temperature. Observational data shows that for cluster typical values of the peculiar velocity $(v_z/1000\,\text{kms}^{-1}) \lesssim 0.5$ and electron temperatures $(k_B T_e/10\,\text{keV})^{-1} \sim 1$, the amplitude of the kinetic effect is no more than 5% of the thermal component. Regardless of the interesting possibilities that might arise from the study of the kinetic effect, its detection posses a significant challenge with *Planck*'s instrumental setup. The kinetic effect is superimposed and shares the same geometry of the much stronger thermal signal, severely limiting their separation by spatial filtering. On the other hand its frequency signature follows that of the primordial CMB fluctuations rendering a successful isolation very difficult.

The thermal SZ effect moves some of the low-energy CMB photons to higher energies. The transferring process leaves a deficit in the low energy bands and a excess

[1]The same hot electrons also radiate through thermal free-free in the X-ray wave band.

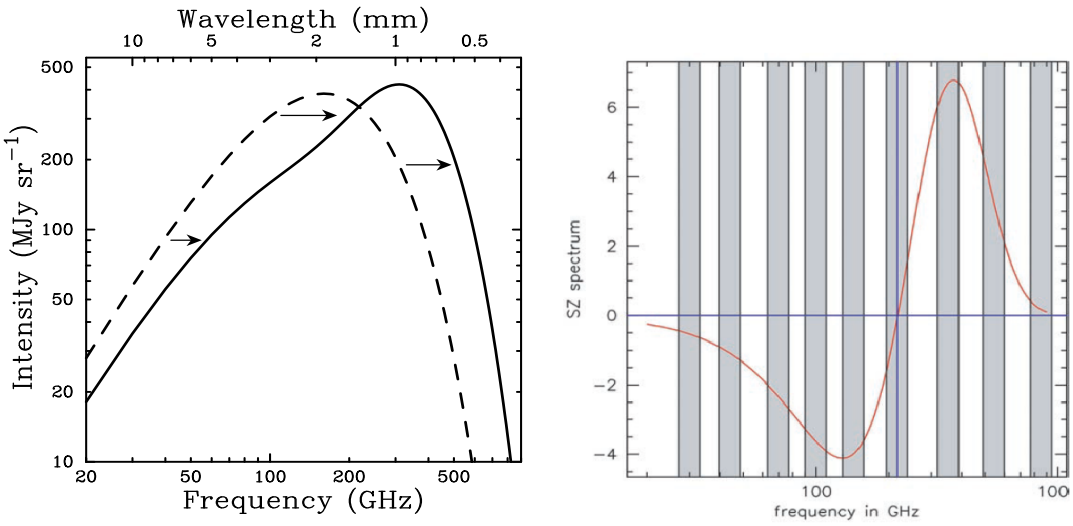


Figure 4.11: The *left panel* shows the shift on the CMB black-body emission spectrum induced through thermal inverse-Compton scattering; The *right panel* shows the distortion on the background CMB signal as predicted by equation 4.6. The nine *Planck* channel bands are superimposed.

in the upper bands (see figure 4.11; Carlstrom et al. [2002] and Planck Collaboration [1999]). The original and distorted populations cross at $\sim$ 217 GHz, leaving at that particular frequency the background photon population undisturbed, thus not generating any observable secondary anisotropies of thermal origin. This aspect of the thermal SZ effect might be explored to great advantage allowing us to successfully isolate the kinetic effect. However, it should be noted that despite of the fact *Planck* observes at that precise frequency, after integrating the thermal SZ spectrum over the finite bandwidth, the residual thermal signal is strong enough to severely limit the detection and characterization of the kinetic in a per cluster basis. However a statistical detection might still be possible but not in the scope of the current study. In this work our attention is on individual object detection and owing to the above difficulties, we will only focus on the extraction of the thermal signal.

The thermal SZ anisotropy (SZ), may be characterized by the product of two functions

$$\frac{\Delta T_{SZ}}{T_{CMB}} = f(\zeta)\, y(\overrightarrow{\theta}), \tag{4.2}$$

where $f(\zeta)$ represents the spectral behaviour and $y(\overrightarrow{\theta})$ the projected spatial geometric profile. The SZ temperature variation spectral dependence $f(\zeta)$, is given by [Carlstrom et al., 2002]

$$f(\zeta) = \left(\zeta \frac{\mathrm{e}^{\zeta}+1}{\mathrm{e}^{\zeta}-1} - 4\right)\left(1+\delta_{SZ}(\zeta, T_e)\right), \tag{4.3}$$

where ζ is the dimensionless frequency

$$\zeta = \frac{h\nu}{k_B T_{CMB}}. \tag{4.4}$$

The extra term, $\delta_{SZ}(\zeta, T_e)$, is an additional SED correction factor as result of hot electrons energy distribution distortion when taking into account the relativistic effects [Birkinshaw, 1999; Challinor and Lasenby, 1998; Itoh et al., 1998]. Throughout this work we shall ignore this term, as for the majority of the predicted cluster population, its contribution is always very small. Although, owing to *Planck*'s high sensitivity to the SZ signal (see chapter 5), a statistical detection of the relativistic correction should be possible and will be attempted in a forthcoming publication. We have found it more convenient to express the SZ perturbation in surface brightness units

$$\frac{\Delta I_{SZ}}{I_0} = g(\zeta) y(\overrightarrow{\theta}), \tag{4.5}$$

where $I_0 = 2(K_B T_{CMB})^3/(hc)^2$ and the SZ dimensionless SED

$$g(\zeta) = \frac{\zeta^4 \mathrm{e}^{\zeta}}{(\mathrm{e}^{\zeta}-1)^2}\left(\zeta \frac{\mathrm{e}^{\zeta}+1}{\mathrm{e}^{\zeta}-1} - 4\right)\left(1+\delta_{SZ}(\zeta, T_e)\right). \tag{4.6}$$

The intensity of signal measured at each one of the *Planck* channels (ν_j) is the SZ SED integrated over the channel bandpass

$$g_{\nu_j} = \int_{\zeta} g(\zeta)\, \Gamma_{\nu_j}(\zeta) \mathrm{d}\zeta, \tag{4.7}$$

where $\Gamma_{\nu_j}(\zeta)$ is the channel normalised bandpass [Planck HFI Core Team et al., 2011; Schäfer et al., 2006]. The SZ signal, as it shows in *Planck* data, can be modelled by

$$\Delta \boldsymbol{I}_{SZ} = y(\overrightarrow{\theta})\boldsymbol{g'}(\zeta), \tag{4.8}$$

where the symbols in bold represent vector quantities with each element representing one of the *Planck* channels and $\boldsymbol{g'}(\zeta) = I_0 \boldsymbol{g}(\zeta)$. A very helpful property of the SZ signal is that the frequency dependence of the signal, $\boldsymbol{g'}(\zeta)$, is the same and parameterless across the cluster population (see equations 3.23 and 3.22).

Let us now focus on the spatial geometry of the SZ anisotropy. The surface brightness perturbation is proportional to $y(\overrightarrow{\boldsymbol{\theta}})$, the Comptonisation (see 4.8). The Comptonisation is a dimensionless quantity defined by

$$y(\boldsymbol{r}) = \frac{\sigma_T}{m_e c^2} \int_{l.o.s.} P_e(\boldsymbol{r}) \mathrm{d}l \tag{4.9}$$

[Birkinshaw, 1999]. Assuming the electron pressure profile is scaled by the parameter r_s, then a more convenient parametrisation $P_e(\boldsymbol{r}) = P_0 f(\boldsymbol{r}/r_s)$, expresses the observable

$$\begin{cases} y(\overrightarrow{\theta}) & = y_0 \, \Gamma(\overrightarrow{\theta}/\theta_s) \\ \Gamma(\overrightarrow{\theta}/\theta_s) = \Gamma(\boldsymbol{\rho}/r_s) & = \int_{-\infty}^{+\infty} f(\boldsymbol{r}/r_s) \, (\mathrm{d}l/r_s) \\ y_0 & = \frac{P_0 \sigma_T r_s}{m_e c^2} \end{cases} \tag{4.10}$$

in such a way that allows us to take advantage of one of the most important aspects of the SZ effect, its independence of redshift ($\boldsymbol{\rho}$ is the transverse position vector). As the *angular diameter distance* (see Hogg [1999]) is the scaling factor that transforms $\boldsymbol{\rho}$ into $\overrightarrow{\theta}$ and r_s into θ_s, this implies $\Gamma(\boldsymbol{\rho}/r_s) = \Gamma(\overrightarrow{\theta}/\theta_s)$. Current descriptions of the cluster physical properties, such as electron density, pressure, temperature or entropy (see Arnaud et al. [2010]; Nagai et al. [2007]; Planck Collaboration et al. [2011f] and references therein) all assume spherical symmetric

$$\Gamma(\overrightarrow{\theta}/\theta_s) = \Gamma(\theta/\theta_s). \tag{4.11}$$

The first PwS code implementations were based on an isothermal beta model

$$f\left(\frac{r}{r_s}\right) = K^{-1} \left(1 + \frac{r^2}{r_s^2}\right)^{-\frac{3}{2}\beta}, \tag{4.12}$$

where $K = \sqrt{\pi} \frac{\Gamma\left(\frac{3}{2}\beta - \frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}\beta\right)}$ is a convenient dimensionless normalization constant to make

the profile equal to 1 at the origin (the origin matches the profile's maximum value) [Birkinshaw, 1999]. The β profile then reads

$$y(\theta) = y_{max}\left(1+\frac{\theta^2}{\theta_s^2}\right)^{\frac{1}{2}-\frac{3}{2}\beta} \tag{4.13}$$

where $y_{max} = y_0/K$. Unfortunately the β profile only converges to a finite total flux if $\beta > 2/3$. However there are physical reasons to believe that values of β smaller than that limit actually provide a good description of the cluster pressure making necessary to impose a cut-off radius. To overcome this complication Herranz et al. [2002b] suggested an alternative profile, the multi-quadratic

$$y(\theta) = y_{max}\frac{\gamma}{\gamma-1}\left[\left(1+\left(\frac{\theta^2}{\theta_s^2}\right)\right)^{-1/2} - \left(\gamma+\left(\frac{\theta^2}{\theta_s^2}\right)\right)^{-1/2}\right], \tag{4.14}$$

where $\gamma = r_v/r_s > 1$ and r_v is the virial radius of the cluster[1]. This profile behaves much like the β and has nice asymptotic properties like finite fluxes regardless of γ value. PwS code implemented the multi-quadratic profile in its early incarnations (before 2010). Later on we moved to a Generalised Navarro-Frenk-and-White (GNFW) pressure profile [Arnaud et al., 2010; Nagai et al., 2007]

$$f\left(\frac{r}{r_s}\right) = K^{-1}\left(\frac{r}{r_s}\right)^{-\gamma}\left[1+\left(\frac{r}{r_s}\right)^{\alpha}\right]^{\frac{\gamma-\beta}{\alpha}}, \tag{4.15}$$

where α, β, γ are parameters setting the shape of the profile and $K = \frac{2}{\alpha}\,B\left(\frac{1-\gamma}{\alpha}, \frac{\beta-1}{\alpha}\right)$ ($B(\ldots)$ is the complete Beta function) is the normalisation constant (see figure 4.12).

Eventually the single most interesting quantity the SZ signal can provide is the cluster total integrated flux[2]

$$Y \equiv \int_s y(\overrightarrow{\theta}/\theta_s)\,d\Omega = y_0\,r_s^2\int_{-\infty}^{+\infty}\!\!\!\!\!\!\!\!\!\!\int P_e(\boldsymbol{r}/r_s)\,(\mathrm{d}l/r_s)^3\,. \tag{4.16}$$

No unique definition of the boundary of a cluster is universally used. The current

[1] The virial radius is the radius within virial equilibrium holds.

[2] Y has the dimensions of a solid angle, usually arcmin^2.

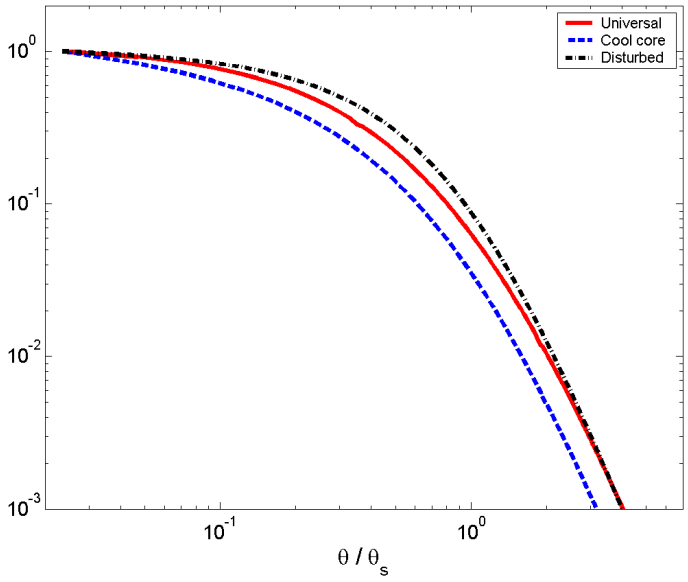


Figure 4.12: Normalised GNFW profiles as described in Planck Collaboration et al. [2011h]: Universal (black, solid), Morphologically Disturbed (magenta, dotted) and Cool core (blue, dashed).

most favoured is a spherical surface, r_Δ, enclosing an average over-density, at the cluster redshift, Δ times larger than the critical density, or a multiple of it[1]. Common values of r_Δ in the literature are $\{r_{200}, r_{500}, r_{1500}\}$. Therefore the integral in formula 4.16 should only extend to that limit in the transverse direction. Something that, in principle, should not raise any difficulties in a numerical implementation. However, as a matter of convenience, the integral can be expanded to infinity if the profile decays to zero quickly enough to enclose a finite flux. The predicted amount of flux lying outside the cluster boundary must be kept to a small fraction of the total flux. The Y value may be easily evaluated once we know the projected profile parameters y_0 or y_{max} and r_s,

$$Y = \begin{cases} \frac{8\pi\, y_0\, r_s^2}{(3\beta-2)(3\beta-4)} & \beta \text{ profile} \\ 2\pi\gamma\, y_{max}\, r_s^2 & \text{multi-quadratic} \\ \frac{4\pi y_0\, r_s^2}{\alpha} B\left(\frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right) & \text{GNFW.} \end{cases} \tag{4.17}$$

[1] $\rho_c = 3H_0^2/8\pi G$.

where $B(a,b)$ is the complete *Beta* function. Another very important parameter $c_\Delta = r_\Delta/r_s = \theta_\Delta/\theta_s$, the is '*concentration parameter*'. The concentration parameter relates the two scaling radius (or angles): pressure and over-density[1]. From now on we shall exclusively focus on the GNFW profile.

Most of the cluster observations made in the X-ray band [Arnaud et al., 2010; Böhringer et al., 2001, 2007; Nagai et al., 2007], are sensitive to over-densities not lower than $\Delta = 500$. Hence, a significant amount of the literature in the field uses r_{500} as the 'natural' scale to measure cluster sizes. Owing to the exquisite *Planck* sensitivity to faint and smooth SZ perturbations (see chapter 5) our measurements can reach well beyond the r_{500} limit, being accepted that $5 \times r_{500}$ provides a much better estimate of the limits of the cluster as 'seen' by *Planck*. The total flux inside a spherical shell of radius $\delta \times r_{500} = \delta \times \theta_{500}$ is given by

$$Y_{\delta\times\theta_{500}} = \frac{4\pi\, y_0 \left(\frac{\theta_{500}}{c_{500}}\right)^2}{\alpha} B\left(\frac{(\delta c_{500})^\alpha}{1+(\delta c_{500})^\alpha}; \frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right), \tag{4.18}$$

or as a fraction of the total flux Y

$$\frac{Y_{\delta\times r_{500}}}{Y} = I_{\left\{\frac{(\delta c_{500})^\alpha}{1+(\delta c_{500})^\alpha}\right\}}\left(\frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right). \tag{4.19}$$

where $B(x;a,b)$ is the incomplete *Beta* function and $I(x;a,b)$ is the fractional incomplete *Beta* function. Using the '*Universal*' GNFW pressure profile of Arnaud et al. [2010] (A10) as reference one could check that $Y_{5r500}/Y \sim 96\%$. This error falls, most of the time, well below the statistical uncertainty on the measurement of Y_{5r500} [Planck Collaboration et al., 2011f]. When the uncertainty on the measurement of Y is small enough, as for certain high SNR detections more rigorous formulas need to be

[1] Actually, r_s is introduced as a scaling parameter for the dark matter halo. It then propagates to the other quantities as the dark matter distribution completely dominates, according to these models, the overall geometry of all physical quantities.

employed[1]. The exact flux inside a cylinder of radius equal to $\delta \times \theta_{500}$ is given by:

$$Y_{(cyl)\,\delta\times\theta_{500}} = \frac{4\pi\, y_0 \left(\frac{\theta_{500}}{c_{500}}\right)^2}{\alpha} \int_0^{\pi/2} B\left(\frac{(\delta c_{500})^\alpha}{\cos(t)^\alpha + (\delta c_{500})^\alpha}; \frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right) \cos(t)\mathrm{d}t, \tag{4.20}$$

or as a percentage of the total flux

$$\frac{Y_{(cyl)\,\delta\times\theta_{500}}}{Y} = \int_0^{\pi/2} I_{\left\{\frac{(\delta c_{500})^\alpha}{\cos(t)^\alpha + (\delta c_{500})^\alpha}\right\}} \left(\frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right) \cos(t)\mathrm{d}t. \tag{4.21}$$

Combining equations 4.21 and 4.19 we get the following expression

$$\frac{Y_{(cyl)\,\delta\times\theta_{500}}}{Y_{\delta\times\theta_{500}}} = \frac{\int_0^{\pi/2} I_{\left\{\frac{(\delta c_{500})^\alpha}{\cos(t)^\alpha + (\delta c_{500})^\alpha}\right\}} \left(\frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right) \cos(t)\mathrm{d}t}{I_{\left\{\frac{(\delta c_{500})^\alpha}{1 + (\delta c_{500})^\alpha}\right\}} \left(\frac{3-\gamma}{\alpha}, \frac{\beta-3}{\alpha}\right)}, \tag{4.22}$$

which relates the spherical $\delta \times \theta_{500}$ quantities to their cylindrical counterparts. The interest in developing these formulae clearly exceeds the academic. An accurate numerical implementation of $I(x; a, b)$ exists (see Press et al. [1992, ch. 6]) and for the range of expected values of the profile parameters they are very fast. It is worth mentioning that if the β parameter is actually lower than the universal values reported in A10, as recently described in Planck Collaboration et al. [2012a], then the spherical approximation no longer holds and the the full cylindrical expressions must be used instead (see 4.13). As for any 'physically viable' set of GNFW parameters ($\beta > 3, \gamma < 3$; parameters that integrate to a finite Y value) our implementation of the above formulae is very efficient, we have chosen to parameterise the detection profile using the total flux Y instead. This parameterisation is very advantageous in the sense that it provides directly the full joint posterior distribution of the cluster most interesting parameters, Y and θ_s. In practical terms we just need to make sure that the area below the projected profile is always equal to unity. To allow a direct comparison with other extraction methods, namely the Matched Multi Filters [(MMF) Herranz et al., 2002b; Melin et al., 2006], PwS implements a cut-off radius of $5 \times \theta_{500} \Rightarrow \delta = 5$. Although, using formula 4.22 one can recover the spherical quantities even in the case of profiles

[1] Our catalogues include a small number of detections with SNR $\gtrsim 20$. Although, the actual uncertainty in Y_{5r500} is much closer to 10% owing to the $Y - \theta$ degeneracy.

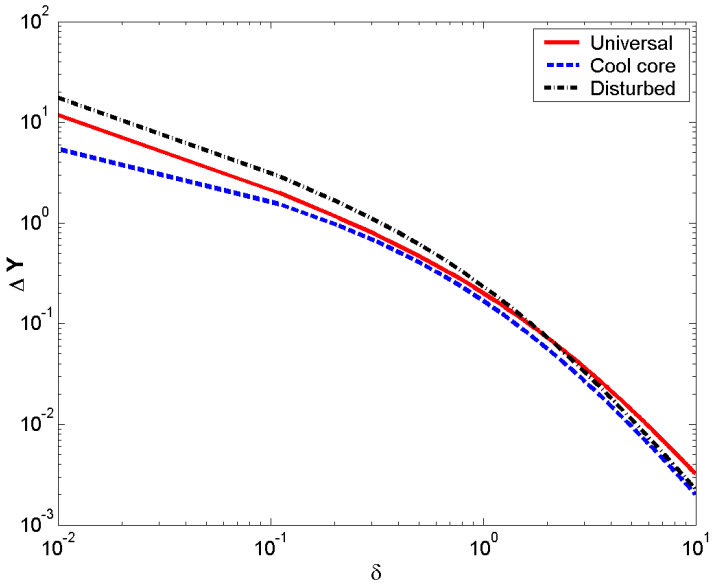


Figure 4.13: $\Delta Y = \frac{Y_{cyl}-Y_{sph}}{Y_{sph}}$ is the fractional difference between Y_{cyl} and Y_{sph} inside a radius $\delta \times \theta_{500}$.

with smaller β (shallower in the outskirts) like the best fit profile in Planck Collaboration et al. [2012a].

4.2.4.1 Early SZ catalogue

Planck's '*Early Sunyaev-Zeldovich cluster catalogue*' (ESZ) is an integral part of ERCSC. It comprises 189 galaxy clusters detected via their thermal SZ effect. The ESZ was extracted from *Planck*'s early data and all 6 HFI channels were employed in a blind multi-frequency extraction exercise. A Matched Multi-Filter code, MMF3 [Melin et al., 2006], was used as the reference extraction tool. PwS, together with another MMF implementation, MMF1, acted as an ancillary detection validation tool. All three methods ran independent blind extraction tasks. Then only those putative detections above a SNR cut of 6 found by at least 2 methods were selected for inclusion. Figure 4.14[1] shows the locations of the ESZ catalogue entries on the sky. This catalogue was still quite 'shallow' as the survey data only contained one full sky scan.

[1]Meanwhile, some of the newly discovered clusters were verified [Planck et al., 2012; Planck Collaboration et al., 2012c; Sayers et al., 2012].

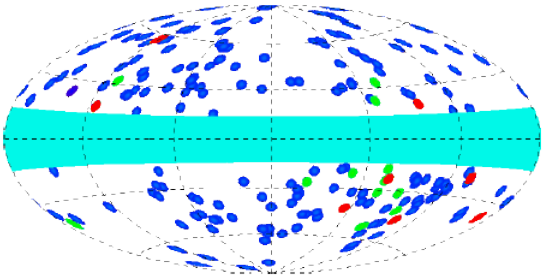


Figure 4.14: Distribution of cluster detections on the sphere (size of the clusters not shown at scale). The blue circles are previously known clusters, in green are clusters that were confirmed as part of the validation task and the red circles show clusters not yet confirmed. The light blue region is the employed Galactic mask ($14°$) (plot from Planck Collaboration et al. [2011f]).

However, not only the sheer size of the catalogue, $\sim 5\times$ larger than all previously SZ observed cluster, as its volume and high SNR, were already a promise of a very interesting scientific product for the coming releases. A full and detailed description of the ESZ contents can be found in Planck Collaboration et al. [2011f] and Planck Collaboration et al. [2011e].

4.2.4.2 $Y - \theta$ degeneracy

The SZ signal is very faint and sub-dominant across all *Planck* observing frequencies. Thus, a deep and complete understanding of the statistical uncertainty, clearly predominant in this analysis, is foremost. As we already stressed in Chapter 1 this is eventually the single major advantage of a Bayesian methodology over the more empirical and frequentist solutions: a full characterisation of the errors, and their correlations, by means of the joint posterior distributions. In contrast to point sources, the cluster signal cannot be appropriately modelled without introducing an extra parameter, θ_s (or any proportional quantity), to give account of the cluster spatial extension. We shall see in chapter 5, that by virtue of *Planck* modest resolution, the cluster size estimates cannot be properly constrained. As consequence of the likelihood 3.23, there is a large correlation between the cluster total flux estimate (Y) and its size (θ). Unfortunately, this effect propagates the size uncertainty into the Y estimate, increasing it dramatically (see Planck Collaboration et al. [2011f]). Another, more subtle problem, is that

the distribution of the parameters uncertainties is not close to Gaussian, especially if the detection SNR is low. However, using the Bayesian posterior distributions one can harness all possible information about the joint parameter constraints giving the data and all ancillary information.

Figure 4.15 shows several examples of posterior probability contours obtained by PwS with *Planck* data. The two plots on the top row plus that on the bottom row left were taken from Planck et al. [2012]. By combining the posterior probability distributions of multiple observations made by complementary heterogeneous instrumental setups, it is possible to 'shrink' the parameter uncertainty volume. AMI with its higher resolution ($\sim 1'$) but lower sensitivity[1] creates 'more vertical' posterior probability shapes in the $\theta - Y$ plane. By combining both posteriors, one is able to provide much tighter constraints not only on Y but on θ as well (black curve). The top contours were drawn from observations of two well known clusters: *A773* and *MACS J1149+2223*, that were observed by X-ray telescopes as well. The red arrow marks the X-ray θ_{500} best fit estimate, which is consistent with the joint SZ constraints. The plot on the lower right corner of the figure (taken from Planck Collaboration et al. [2011f]) exposes how severe the increase in the Y uncertainty is as result of the $Y - \theta$ degeneracy. The image shows two typical cases of *Planck* ESZ: A large cluster ($\theta_s \gtrsim 20'$) with high SNR that *Planck* can resolve (blue contours) and an unresolved cluster ($\theta_s \lesssim 5'$) with low SNR. These probability contours are in good agreement with what one would expect from the theory: when the data is informative, the likelihood dominates and tends to a multidimensional Gaussian [Cam and Yang, 2000][2]. The low SNR probability contours also show that for less informative data sets, the posterior distributions may deviate strongly from Gaussian. However, this has never being a problem for a Bayesian method.

4.2.4.3 SZ photometry constrained by X-Rays measurements

To decrease the dispersion on the estimates of Y, it is important to break up the important $Y - \theta$ degeneracy [Planck Collaboration et al., 2011f]. Very high resolution

[1] Sensitivity in this context means '*sensitivity to the Y signal*'. AMI's instrumental noise is much lower than that of *Planck*, but AMI's backgrounds are dominated by CMB and its detections are strongly contaminated by radio point sources.

[2] The other three examples displayed are high SNR detections in the ESZ catalogue as well.

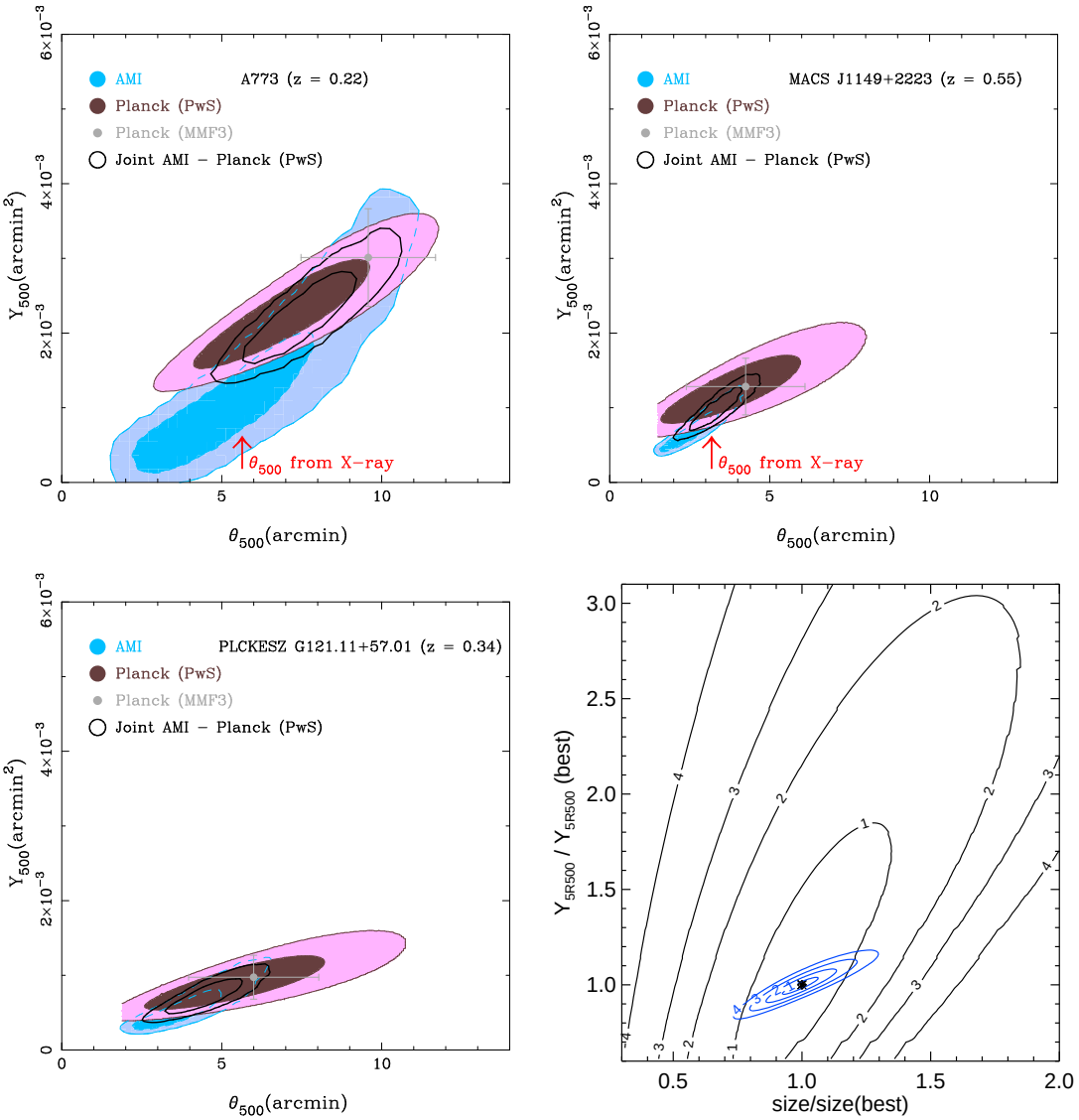


Figure 4.15: The plots on the *upper row* and *lower row left*, show the superposition of two joint $Y_{500} - \theta_{500}$ probability contours. The probability mass of the contours is 0.466 and 0.911 respectively. The *Planck* ones are shown in magenta, *AMI* in blue and the common in black. The cross is MMF3 best estimates and symmetrical error bars. The *red upper arrow* (on the top plots only) is the best X-ray estimate of θ_{500}. The graph on the *lower row right* depicts the joint $Y_{500} - \theta_{500}$ probability contours of two typical clusters detected by *Planck*: well resolved in blue and unresolved in black. (plots from Planck et al. [2012]; Planck Collaboration et al. [2011f])

X-rays observations of the gas in the intra-cluster medium supply accurate and very well constrained measurements on θ_{500} (see Böhringer et al. [2001, 2007]; Piffaretti et al. [2011] and references therein). Employing to great advantage '*targeted*' priors (see section 3.3) derived from the constraints on position and size deduced from the X-rays data we were able to refine the '*blind*' Y estimates as in the cases described in Planck Collaboration et al. [2012c], Planck Collaboration et al. [2012b], Planck Collaboration et al. [2011i], Planck Collaboration et al. [2011g] and Planck Collaboration et al. [2011f].

4.2.5 Validation

No science quality catalogue is complete without a proper validation. Validation tests the assumptions, physical and statistical, and whether our models are actually a truthful representation of reality. Only a thorough validation can actually provide a sensible way to assess how systematics, mis-modeling and statistical bias impair the properties of the catalogue and its estimates. The first 3 ERCSC PwS channel catalogues were fully and thoroughly validated (see Planck Collaboration et al. [2011e], Planck Collaboration et al. [2011c] and Planck Collaboration et al. [2011b]). Table 4.3 contains a summary of the validation results. Despite of the strong gradients close to the Galactic plane and very non-uniform background, PwS performed clearly inside the imposed constraints even for Galactic latitudes as low as 5°. However, the coarse resolution of *Planck* in addition to the above mentioned factors brought about a severe break in the underlying assumptions of the algorithm, rendering it sub-optimal for regions of the Galactic plane. Nevertheless, the overall performance was still inside the catalogue goals.

Further analysis of formula 3.56 shows that the penalty per source, $\widehat{P_s}$, concentrates into a single number the least well defined prior quantities, namely, the expected number counts of a population above a certain flux (λ_1) the expected number of background fluctuations above the same threshold (λ_0), etc. The non-Gaussianity of the background, the instrumental or map-making artefacts and the uncertainty on the priors, most make this value deviate from the fiducial prediction. A properly calibrated $\widehat{P_s}$ may be easily found, initially by using simulations and later consolidated through validation. PwS catalogues and estimates have been extensively validated in a broad

Frequency [GHz]	30	44	70
# $\lvert b\rvert^{(a)} > 5^\circ$	563	278	320
# Identified	547 (97%)	265 (95%)	289 (90%)
# $\lvert b\rvert < 5^\circ$	142	176	280
# Identified	95 (67%)	144 (82%)	...
# Total	705	454	600
# Identified	642 (91%)	409 (90%)	...

Table 4.3: $^{(a)}$ Galactic latitude cut.

variety of different astronomical scenarios (for examples please refer to Planck et al. [2012]; Planck Collaboration et al. [2011f]).

4.3 Simulations

Before being selected as part of the official *Planck* compact source catalogue pipeline (ERCSC + ESZ), PwS underwent many different quality assessment (QA) tests based on simulated data. There were several challenges [Melin et al., 2012; Rocha, G. et al., 2013] and many *Planck* internal QA actions. These QA exercises were designed not only to compare the relative effectiveness of the different algorithms involved but to check whether their performance met the required goals. In a initial phase, in either case, point sources or cluster SZ extraction, many (> 10) different algorithms and implementations competed. Later, the initial bundle of catalogue extraction methods (detection + characterisation) was reduced to a smaller set:

- On point source extraction;

  SExtractor (SEx). The most widely used 'classical' extraction package [Bertin and Arnouts, 1996].

  IFCA Mexican Hat Wavelet (IFCAMex). An implementation of the Mexican Hat Wavelet, specifically tuned for cosmological microwave surveys developed at IFCA [González-Nuevo et al., 2006; López-Caniego et al., 2006].

  Paris Matched Filter (PMF). A single frequency derivative of Melin et al. [2006].

  PowellSnakes (PwS). This work [Carvalho et al., 2009, 2012].

- On SZ cluster extraction;
  after a numerous initial collection of participants [Melin et al., 2012] only 4 codes/implementaions were finally selected to participate in the making of the different extraction/estimation products: (i) Two different implementations of the Matched Multi-Filter (MMF1, MMF3) [Herranz et al., 2002b; Melin et al., 2006]; (ii) PowellSnakes (PwS).

The simulated *Planck* data sets that provided the underlying framework for the QA exercises were, in both cases, the current state-of-the-art in realism and accuracy.

4.3.1 The simulated maps

The sky emission at *Planck* frequencies was generated by the pre-launch 'Planck Sky Model' package [Delabrouille et al., 2012, (PSM)]. This package is currently the most sophisticated and complete simulation of the sky as 'seen' by *Planck*. It contains a very large collection of flexible and highly parameterizable codes that simulate the diffuse and compact emission components across the extended range of frequencies *Planck* observes. Concurrently, it produces a full sky pixelised template of the predicted *Planck* instrumental noise. Finally, it assembles a set of mock full sky maps, in temperature and polarization, together with the predicted noise covariance matrix at each pixel. A detailed description of this package can be found in Delabrouille et al. [2012] and a distinct application, in the context of diffuse component separation, may also be found in here Leach et al. [2008]. The SZ extraction exercises were based on these simulations, the '*WG2*' simulations.

The point source QA tests and challenges were performed using another set of simulations even more accurate from the point of view of the instrumental effects. These are the '*Full Focal Plane*' (FFP) simulations. *Planck*'s pipeline has yet another component, not described in diagram 4.7: the '*Level-S*'. '*Level-S*' is a set of numeric tools aimed at modelling the instrumental output of *Planck*'s instruments [Reinecke et al., 2006]. Roughly speaking, it takes the simulated sky emission from the PSM and the scanning strategy as input and produces clean and calibrated time-lines, detector by detector as read in *Planck*'s focal plane. Essentially, it replaces L1 and feeds L2 with the necessary inputs. Then, the map-making pipeline is run over these mock data and the maps are produced. So, not only the full instrumental setup is reproduced in

these simulations, but all possible pipeline induced effects are also included.

4.3.2 Catalogue figures of merit

The figures of merit of a catalogue are clearly non-unique and depend heavily on the scientific goals of the catalogue. In chapter 2 some general criteria like '*Reliability/Purity*', '*Completeness*' or '*Estimation accuracy*' that apply to every catalogue were defined based on probability grounds and decision theory (see section 2.3). The '*sample*' versions of these quantities that we employed along the several QA exercises and challenges were the following:

- *Reliability/purity*,
  is defined as:

$$\frac{\text{Number of detections correctly included} \mid \widehat{\Theta} > \xi}{\text{Total number in the output catalogue} \mid \widehat{\Theta} > \xi}, \tag{4.23}$$

  where the '$\mid \cdots$' means 'given that', $\widehat{\Theta}$ is the estimate of a certain parameter (value in the output catalogue; usually SNR or flux density) and ξ is a threshold that will parameterize the reliability. '*Correctly included*' might have very different meanings that depend on the goals of the catalogue (see section 4.2.5)[1].

- *Completeness*
  of a catalogue is:

$$\frac{\text{Number of detections correctly included} \mid \Theta_0 > \xi}{\text{Total number in the input catalogue} \mid \Theta_0 > \xi}, \tag{4.24}$$

  where Θ_0 is the *true* value of the parameter (value in the input catalogue; usually SNR or flux density) and ξ has exactly the same sense as in the reliability definition. A close derivative of completeness is '*total completeness*' where we drop the threshold condition in the denominator. 'Total completeness' gives us the percentage of objects over the complete targeted population we detect when

[1] This definition of reliability is also mentioned as 'integrated reliability/purity'. Another very common definition of reliability is 'binned reliability' defined by binning parameter ξ; $\xi_i < \widehat{\Theta} < \xi_{i+1}$.

we threshold on ξ[1].

- *Estimation accuracy* (astrometry + photometry),
  is some kind of loss measurement the parameter estimates in the catalogue incur as consequence of an imperfect estimation (see 2.2.1). The smaller the loss the highest the estimation accuracy. However, many times the frequentist estimators figures of merit, unbiasedness and minimum variance are instead employed[2].

4.3.3 Detection of point sources

The point source extraction challenges were part of the preparatory effort to characterise the, at the time, future *Planck* ERCSC (see [Rocha, G. et al., 2013]). We will focus on the last of these challenges where FFP release 3 version 4 simulations (FFP3v4) were used. Figure 4.16 shows only 3 of the 9 temperature *Planck* simulated data channels side-by-side with the Galactic mask employed. In each case the Galactic mask left 25% of the sky uncovered. The choice of these particular 3 channels, 30, 143, and 857 GHz, was not arbitrary. The 30 GHz channel is representative of the LFI, with typical LFI values of sensitivity and beam size. The 143 GHz is the central HFI cosmological channel with very low noise and Galactic contamination and the 857 GHz is the strongest sub-mm channel designed to measure the CIB and dust emission. The point source catalogues were drawn from two different families of sources: radio and infrared (dusty). Each one of these point sources components were further sub-divided into faint and strong. Strong sources are those with fluxes higher than 200 mJy. The faint population was only intended to add to the background a realistic confusion noise component. The two populations were treated differently in FFP simulations (see Reinecke et al. [2006]). The strong point sources were convolved with the beam in real space taking into account the beam's spatial variation and the scanning strategy. The faint population was smoothed in harmonic space using an 'average' beam. Only the strong population was targeted by the algorithms.

Figure 4.17 shows the spatial distribution of PwS detections. The Galactic, LMC and SMC regions show an above average density of sources and higher fluxes as ex-

[1] There is as well an equivalent definition of 'binned completeness' the goes along the same lines as those of reliability.

[2] See Jaynes [2004, ch. 17.2] for a very insightful discussion about the merits (or lack of them !) of the frequentist estimators figures of merit.

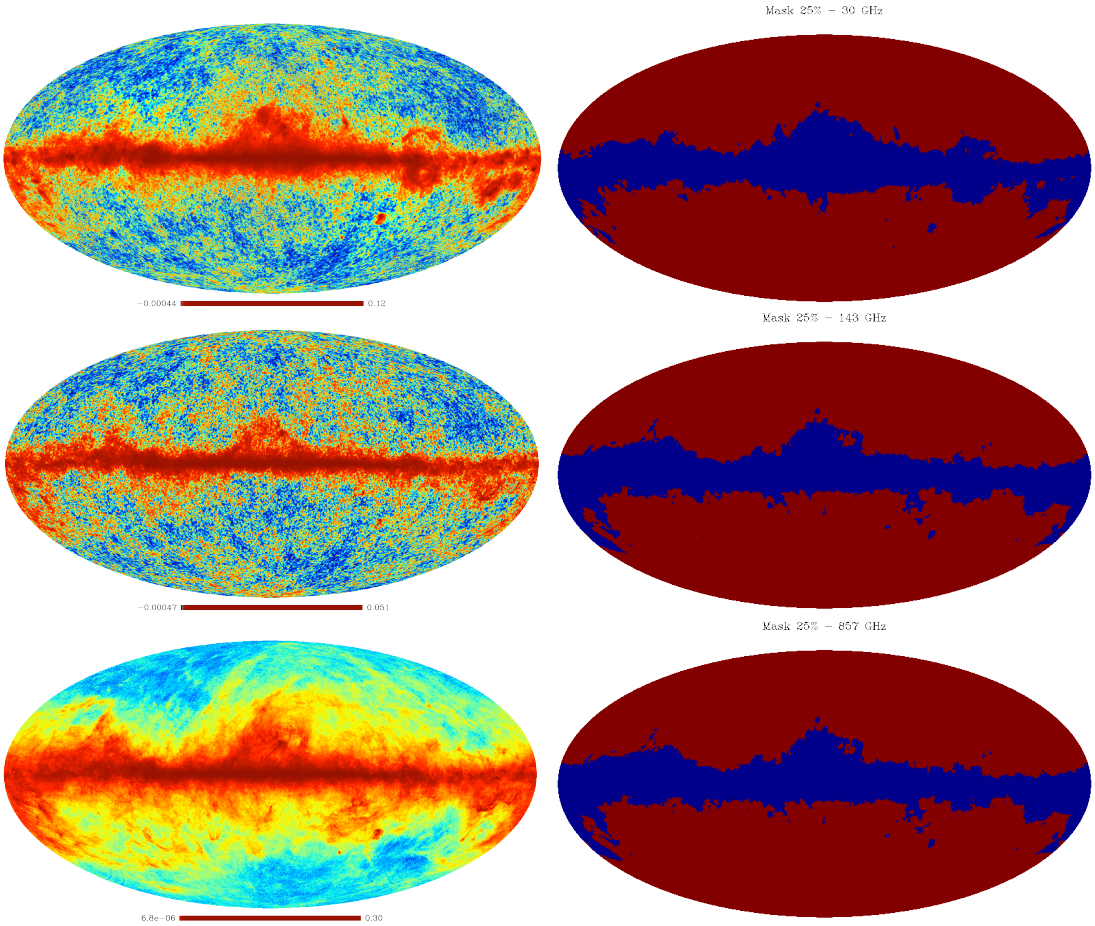


Figure 4.16: Each row refers to a different observation channel: $30, 143$ and 857 GHz in top-down order. The figures on the left column show the simulated temperature data [K CMB] and on the right one the respective Galactic mask. Each Galactic mask removes 25% of the sky.

pected. The density contrast is more easily perceived on the 143 channel. Owing to the very strong background on the 857 channel, very few sources below 1 Jy were detected. These fainter sources are all very close to the Ecliptic poles where the noise levels are considerably lower as result of the much higher 'hit counts' (see figures 4.4 and 4.6).

Figure 4.18 shows two 143 GHz small patches ($10° \times 10°$) cut from background regions with distinct statistical traits: (i) 'Deep' (very homogenous, isotropic and Gaussian) and (ii) 'shallow' (complex structure and less Gaussian). It is perceived from this

figures how challenging the task of detection and parameter estimation can be with these data sets. The SNRs of the objects are so small that one can hardly have any glimpse of them at all. Yet, PwS's catalogues reached more than 95% purity (see table 4.3)!

The next series of graphics (figure 4.19, 4.20 and 4.21) portrays the comparative performance of the extraction algorithms that competed in the Compact Source Investigation group challenge [Rocha, G. et al., 2013, (CSI)]. The CSI was assembled with the purpose of developing and characterize potential point sources extraction codes for the coming *Planck* ERCSC. The next example specifically targets extra-galactic compact objects. As we were not concerned with the Galactic region, a 25% Galactic mask was used to remove the complex low Galactic latitude regions (see figure 4.16 right column). The challenge consisted of 9 point source single-frequency extraction exercises (one per *Planck* channel). However, in the interest of compactness, we will only focus on 3 channels. The channel selection criterion was that previously mentioned. The different panels in figures 4.19, 4.20 and 4.21 represent the following:

- The graph in *first row from the top*,
  shows the outcome of the integrated completeness comparative analysis as a function of the flux density (see formula 4.24).

- The graph in *second row from the top*,
  shows the fraction of non-matched sources ($1-$ 'integrated purity') as a function of the flux density (see formula 4.23). So, in this case, lower figures mean better performance.

- The graphics in *third row from the top and fourth on the right*,
  are about the quality of the algorithm's photometry. The first is a direct comparison of the flux estimates: simulated fluxes versus estimated. The second plot is the relative estimation error $\frac{S_{est}-S_{sim}}{S_{sim}}$ in percentage.

- The graph in *fourth panel (from the top) on the left*,
  is the most important of all. We have tried to combine the two catalogue detection figures of merit into a single plot to produce a unique criterion that would allow us to identify which code was performing better in terms of detection. The x axis is the 'integrated reliability', the y axis is 'global completeness' and the

common parameter that relates both quantities is the flux threshold. Drawing a vertical line through a certain reliability value, it intercepts the performance curves at some point. The point of interception gives the fraction of sources each algorithm can detect at that reliability and the flux (SNR) cut to achieve it. The curve that intersects the line further apart from the x axis immediately identifies the best performer.

- The *bottom panel*
  is about positional accuracy. The *right plot* shows the position residuals. The *plot on the left* shows un-normalized histograms of the residuals depicted on the right.

4.3.3.1 CSI point sources challenge - 30 GHz channel

It terms of completeness PwS and the IFCAMex clearly outperform the other two codes on the faint tail. They are both more than 90% complete at 400 mJy, with PwS already achieving a purity in excess of 90% as well. On the '*reliability-completeness*' (RC) plot, PwS's curve is always above that of the IFCAMex, meaning that for any level of reliability PwS is always more complete. The extraction performance of SExtractor is far from optimal on this channel. This does not come as a surprise (see chapter 1). In terms of photometry all the codes perform equally well with no significant bias beside the expected Eddington bias [Eddington, 1913] on the faint tail. However there are some notorious outliers considerably more numerous on SExtractor estimates. In terms of astrometry (positional accuracy) PwS once again outperforms all other but now by a narrow margin (see figure 4.19).

4.3.3.2 CSI point sources challenge - 143 GHz channel

The 143 channels is the 'cleanest', with very low instrumental noise and Galactic foreground emission, therefore, a close to optimal performance from all codes is to be expected (figure 4.20). Indeed this was the case. All codes, but SExtractor, were more than 90% complete at 200 mJy (the challenge minimum flux limit). Purity was above 95% (ERCSC requirement) at this flux limit as well. Looking at the 'RC' plot, the detection performance of all algorithms was almost undistinguished besides SExtrac-

tor[1]. The same can be said about photometry. Even though SExtractor flux estimation looks unbiased, the estimates dispersion is much higher than that of the other three. On positional accuracy, PwS estimates show no bias, like all other codes, but clearly much smaller dispersion.

4.3.3.3 CSI point sources challenge - 857 GHz channel

The 857 GHz channel presented, without any doubt, the most difficult scenario. The Galactic foregrounds, that in this case extend well beyond the mask, are extremely complex with structure and strong gradients on sub-pixel scales. That severely breaks most of PwS statistical assumptions about the background (see 3.1). Thus, a sub-optimal performance should now be expected. The completeness range of values required by the *Planck*'s ERCSC starts at much higher flux values in this case, typically between 1 and 2 Jy. Once again the detection performance of the algorithms, PwS, IFCAMex, and PMF is mostly equivalent above 2 Jy. The same cannot be said about SExtractor[2]. However, despite of its potential non-optimal design for this type of data, PwS still shows a slightly better completeness at the same level of reliability. However, in the estimation exercises the differences were more obvious. If in photometry terms the differences seem not to be significant, perhaps with the exception of SExtrator that, once again exhibits a larger dispersion, with regard to positional accuracy PwS clearly shows a much better performance than all other competitors.

4.3.4 Detection of galaxy clusters

Before being selected by *Planck*'s consortia as one of the official catalogue extraction tools, PwS underwent numerous QA tests and two public challenges as described in Melin et al. [2012]. The data employed throughout this section were the same as in the '*open challenge*' of Melin et al. [2012]. In this section we will only try to highlight some topics that were addressed with less detail in the cited publication.

The detection of the secondary CMB anisotropies induced by the SZ effect is so faint that even the brightest SZ sources create a signal that is below the background at

[1] PwS's curve stops at a very high reliability. This happened because PwS made a catalogue 'hard' cut at 200 mJy and at this flux the reliability was still very high.

[2] This poor performance of SExtractor on this channel was unexpected. We believe that SExtractor's sub-optimal performance might have been caused, at least partially, by the choice of the kernel.

	Distribution	Parameters	Inferior limit	Superior limit
Position (x, y)	$1/\Delta$ (Uniform)		patch	patch
Angular size (θ_s)	$\mathrm{e}^{-\frac{\theta_s}{\theta_0}}$	$\theta_0 = 5.0$	1.30	40.0
Flux (Y)	Y^{-a}	$a = 1.6$	$5\ 10^{-4}$	0.2

Table 4.4: PwS's default priors when operating in 'blind-mode'. The position prior limits are fixed by the patch size and the expected number of sources in that patch (see section 3.3.2). The units of θ_s are [arcmin] and of Y are [arcmin^2].

any of the *Planck*'s channels. Figure 4.22 shows two patches cut from the simulations centred on the Coma cluster ($l = 57.86, b = +88.01$), eventually the brightest in the sky. The 143 GHz channel is very close to the minimum of equation 4.6 and 353 to the maximum (see figure 4.11 right panel). Not even on these channels, where the SZ signal reaches its maxima, negative and positive, is one able to see any hint of the presence of the brightest cluster in the sky! This is the main reason why 'classical' detection packages like SExtractor are unable to produce SZ catalogues of scientific quality. Although, this cluster is detected by PwS with an SNR in excess of 30 (see figure 4.23)! This is only possible because of the very special instrumental setup of *Planck* (see chapter 5).

Figure 4.24 displays three graphics taken from Planck et al. [2012]. These plots show PwS's joint $Y_{500} - \theta_{500}$ posterior probability densities drawn from 10 different realizations of simulations developed as described in that paper. The cluster's (Y_{500}, θ_{500}) true values where always recovered inside the $2D$, 2σ contour ($\sim 91.1\%$). Unfortunately, the degradation in the Y constraints as consequence of the $Y - \theta$ degeneracy is evident. These are all high SNR clusters.

However, the large majority of *Planck*'s potential detections are much fainter, close to the instrument sensitivity limit. As we already mentioned (section 3.3), in the later case a careful understanding of the priors is crucial. Table 4.4 contains prior's default parameter values (these can be changed through the parameter file). More information about the use of priors by PwS may be found in table 3.1. Figure 4.25 shows the effect of the priors (table 4.4) on the joint $Y - \theta$ posterior probabilities. The red curves were obtained using 'flat' priors on both Y_{5r500} and θ_s and the black curves using PwS's default priors. The systematic effect of the prior on high SNR detections is completely negligible when compared with the statistical uncertainty and the shape is very close

to Gaussian near the maximum (figure 4.25 top). Although, the effect of the prior is clearly more pronounced in the case of low SNR as in the plots on the bottom row. PwS's default priors not only slightly shift the contours towards lower values as shrink them considerably as well. The effect is even stronger on low SNR large clusters where the Y density, y, is very small. Yet, we could not see any significant systematic bias in the analysis of the simulated clusters as in Planck et al. [2012] (see figure 4.24) or any of the public challenges [Melin et al., 2012].

We have tried to understand the effect of the priors, part and parcel of any Bayesian method, on the estimates Y_{5r500} and θ_S, to give a proper characterization of the catalogue estimates. The plots in figures 4.26 and 4.27 were based on PwS's extraction exercises based on the same simulation set as that used on the '*open*' challenge (see Melin et al. [2012]). The most prominent feature in the Y plots is the large Eddington bias that affects almost 90% of the detected cluster population as may be verified by looking at the bottom plot. The max-likelihood estimate (left column) produces no bias for bright clusters ($Y_{5r500} \gtrsim 0.002$ arcmin^2). However, in the faint tail, which clearly dominates (see bottom plot; 90% of the population is faint), exhibits a considerably larger dispersion than the 'expected value over the posterior' on the right plot. Although, a very small negative bias can be seen for the 10% bright population. The 'posterior mode' estimator (upper rows central plots) does not show any significant improvement with relation to the max-likelihood on the faint tail, but it still displays the small negative bias present on the expected value estimator as well. This result was somewhat unexpected. According to the theory (see 2.2.1) the mode of the posterior estimator should have yielded the best performance. After the second challenge we changed the implementation of the mode and expected value using HPD of different content estimators as described in table 3.1. That made both estimators much more robust, particularly on the faint end. Results obtained with the new implementation of the estimators were not ready in time to be included in this work. They can be found in *Planck*'s intermediate papers and forthcoming *Planck* releases.

The unresolved/barely resolved population of clusters ($\theta_s \lesssim 6'$) stands for more than 90% of the total sample (figure 4.27 bottom plot). Owing to *Planck*'s large beams, sub-optimal constraints on the cluster radii should not come as a surprise. Indeed, figure 4.27 supports exactly that. In particular the max-likelihood estimator (on the left) shows a particularly large dispersion over that population of clusters. However, ex-

actly as in the case of flux (figure 4.26), the application of PwS's default prior and the expected value over the posterior estimator considerably reduced the dispersion of the radius estimator (right column). Still, a non-negligible negative bias that affects the resolved population ($\sim 10\%$) was the 'penalty' for better constraints on the unresolved/barely resolved population that accounts for more than $\sim 90\%$ of the sample. Despite of the non-optimal implementation of the mode estimator, still a non-negligible improvement could be seen in the dispersion of the cluster size estimates.

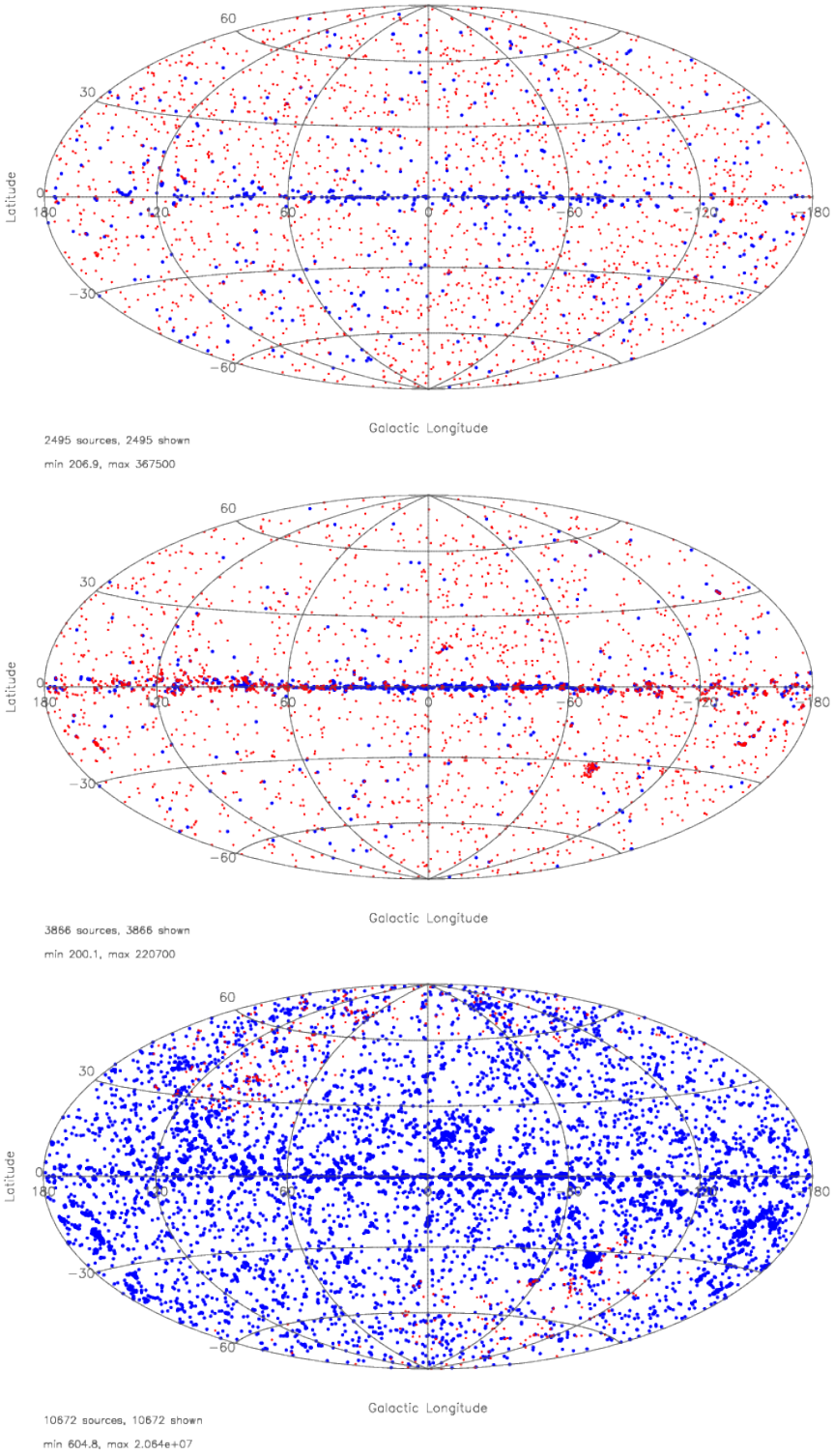


Figure 4.17: Spatial distribution of PwS detections on 30, 143 and 857 GHz channels. The sources sitting on top of the Galactic mask were left in the plots. Source fluxes are in [mJy]. Sources painted in blue have fluxes larger than 1 Jy (plots from Rocha, G. et al. [2013]).

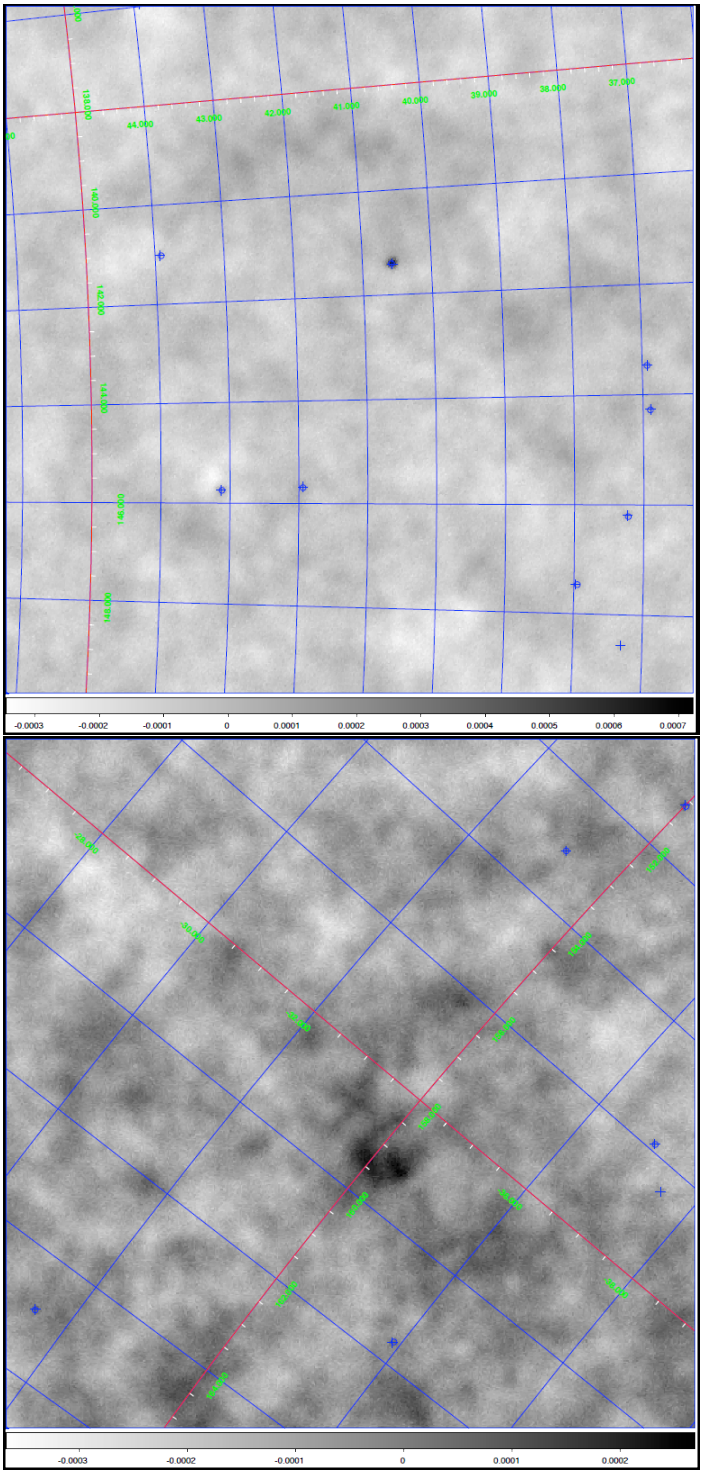


Figure 4.18: Small patches ($10^\circ \times 10^\circ$) of two distinct regions in the 143 GHz temperature map. The simulated sources are marked with blue crosses and those that were detected are marked with blue circles. The *top panel* shows a 'deep' sky zone with a very homogeneous and isotropic background. The *bottom panel* depicts a noisier ('shallower') patch with a more complex background. The temperature units in the grey scale are [K RJ] and the coordinates are Galactic [degrees].

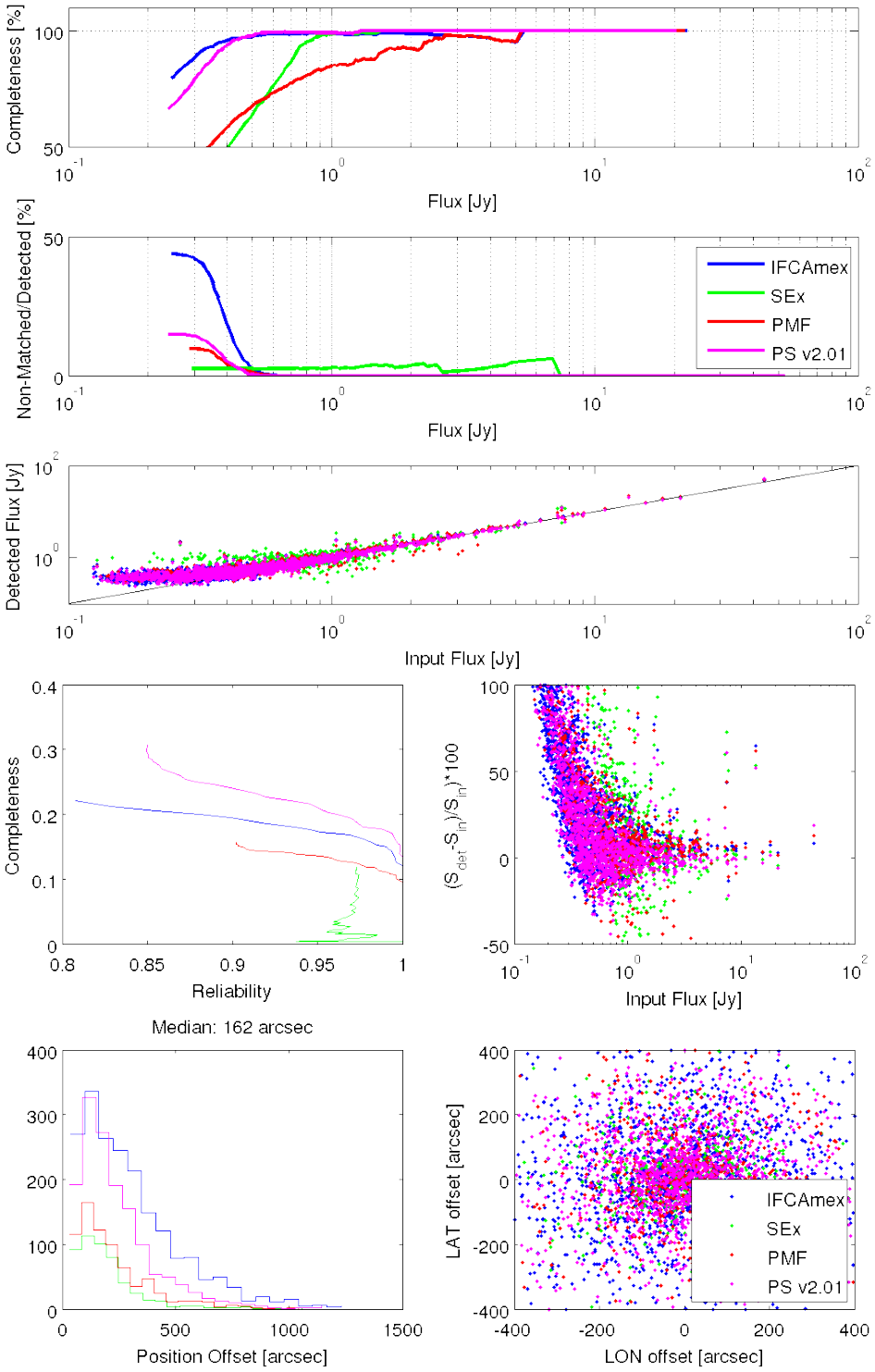


Figure 4.19: CSI point source challenge summary of results for the 30 GHz channel (PwS labelled PS in the figure; plots from Rocha, G. et al. [2013]).

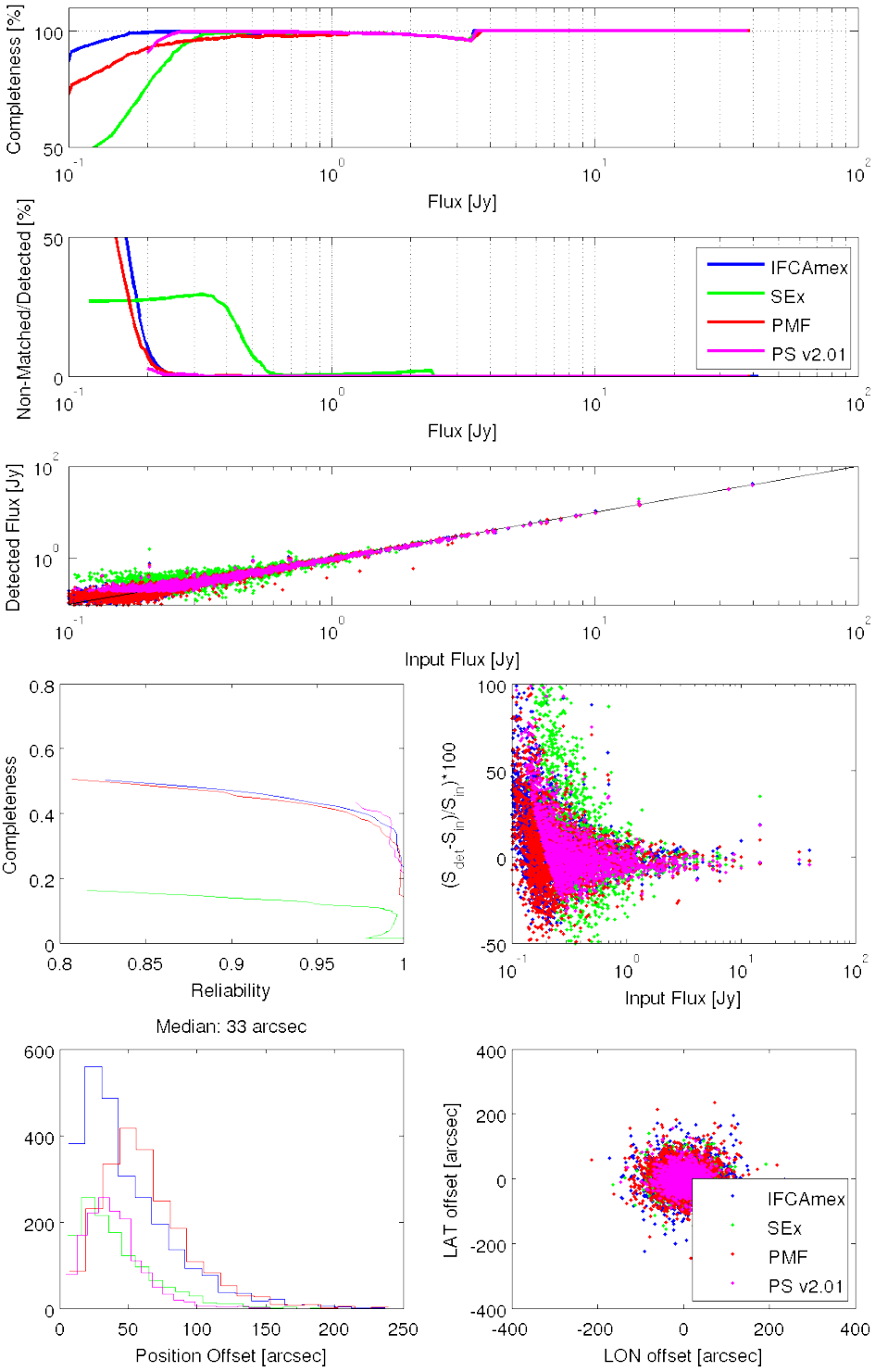


Figure 4.20: CSI point source challenge summary of results for the 143 GHz channel (PwS labelled PS in the figure;plots from Rocha, G. et al. [2013]).

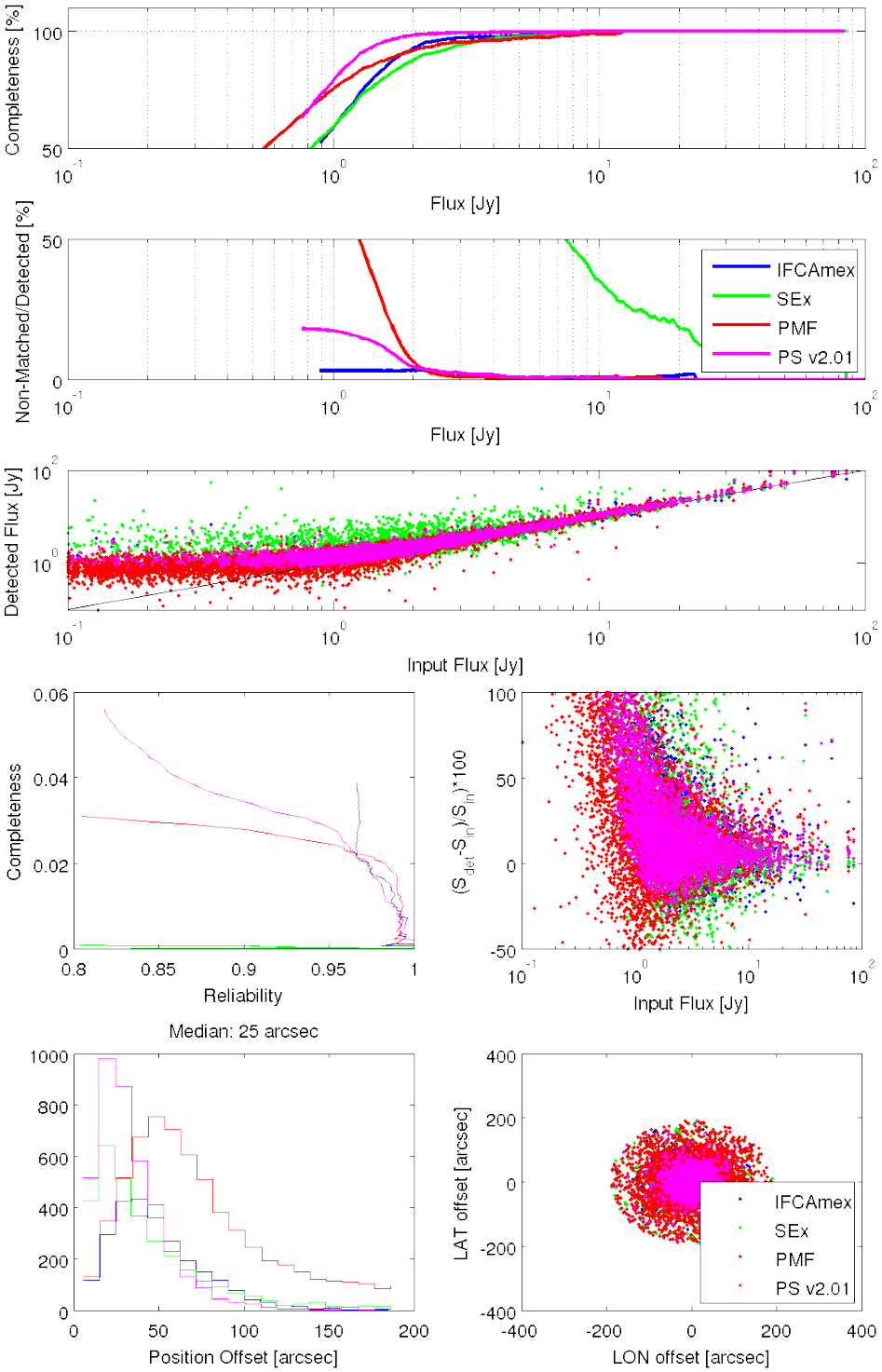


Figure 4.21: CSI point source challenge summary of results for the 857 GHz channel (PwS labelled PS in the figure; plots from Rocha, G. et al. [2013]).

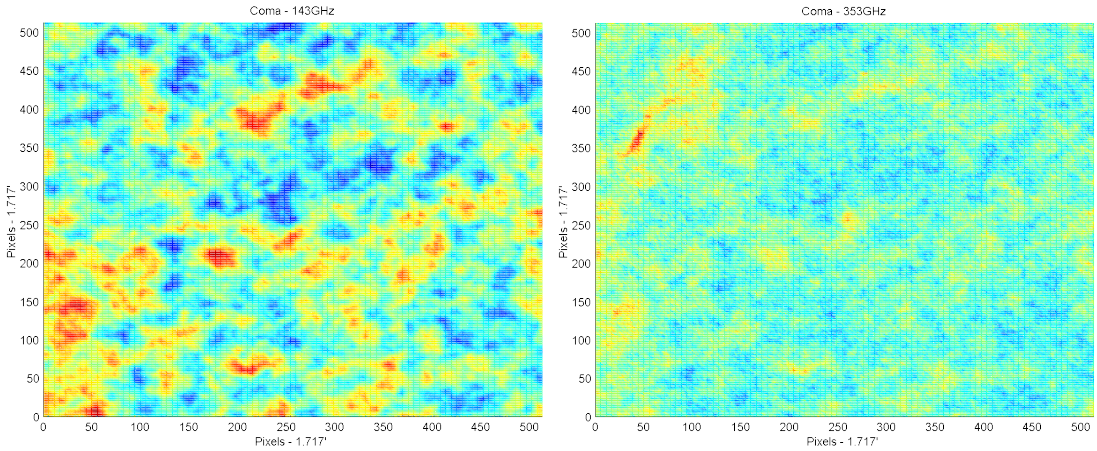


Figure 4.22: These patches are 512×512 pixels wide ($\sim 14.66\times \sim 14.66$) and show the north polar region (Galactic coordinates) in two *Planck* channels: 143 and 353 GHz. They were cut from '*WG2*' simulations as described in section 4.2.2. The colour scale is not the same on both the plots.

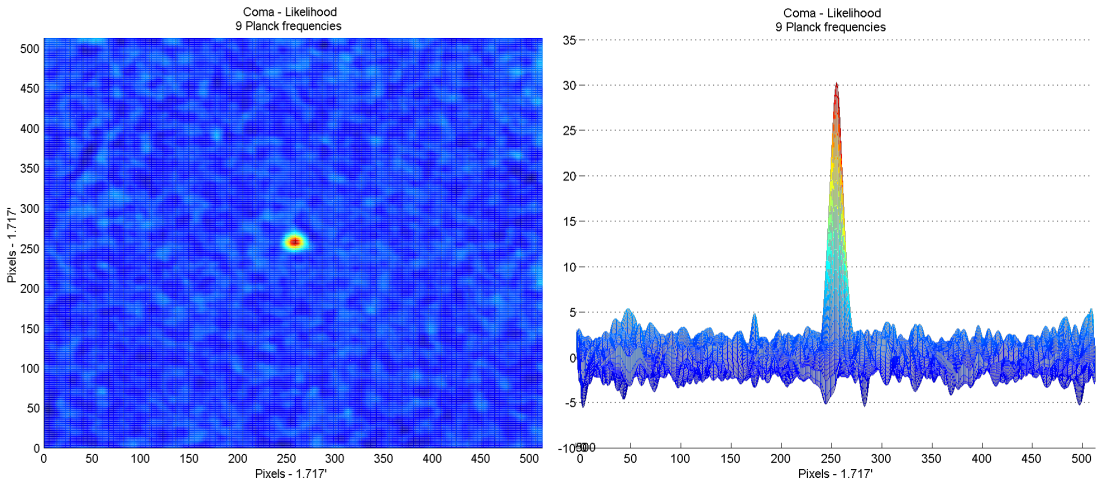


Figure 4.23: Two views (*left*: top view; *right*: side view) of a cut ($\theta_s = c$) through the SZ multi-channel likelihood manifold projected on position space. The likelihood manifold was drawn from the same data as in figure 4.22 but now using all (9) *Planck* channels. The z axis scale was normalised to SNR units. The central very large maximum is the detection corresponding to a simulation of the Coma cluster. The likelihood manifold is profusely multi-modal.

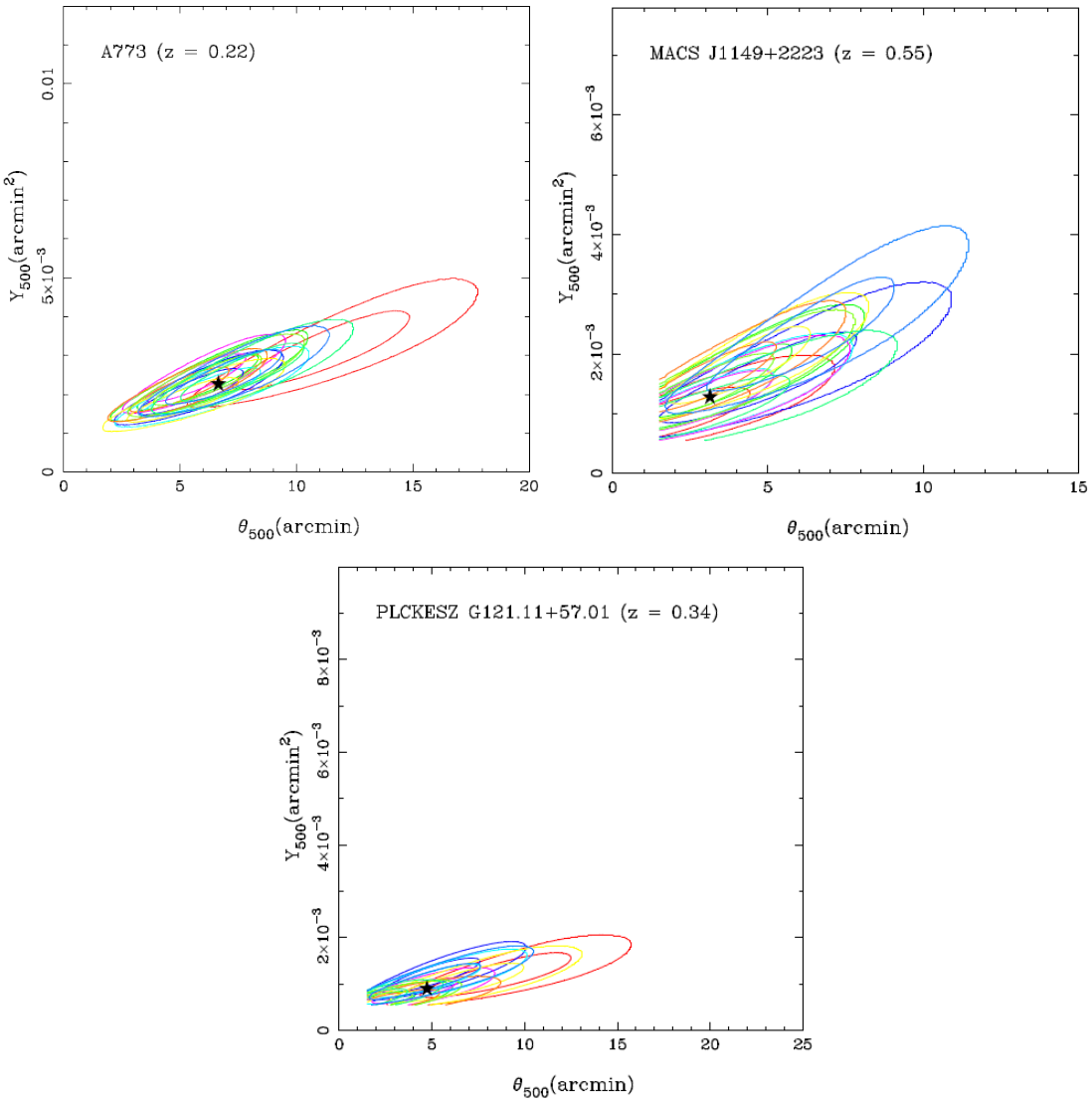


Figure 4.24: These plots show joint probability contours of clusters simulated following a GNFW 'universal pressure profile' [Arnaud et al., 2010] with fluxes and sizes derived from estimates obtained from *Planck*'s early data (MMF3). The clusters are the same as in figure 4.15. Each contour refers to a different simulation of the same three clusters as described in Planck et al. [2012]. The 'star' shows the input (Y_{500}, θ_{500}) pair (plots from Planck et al. [2012]).

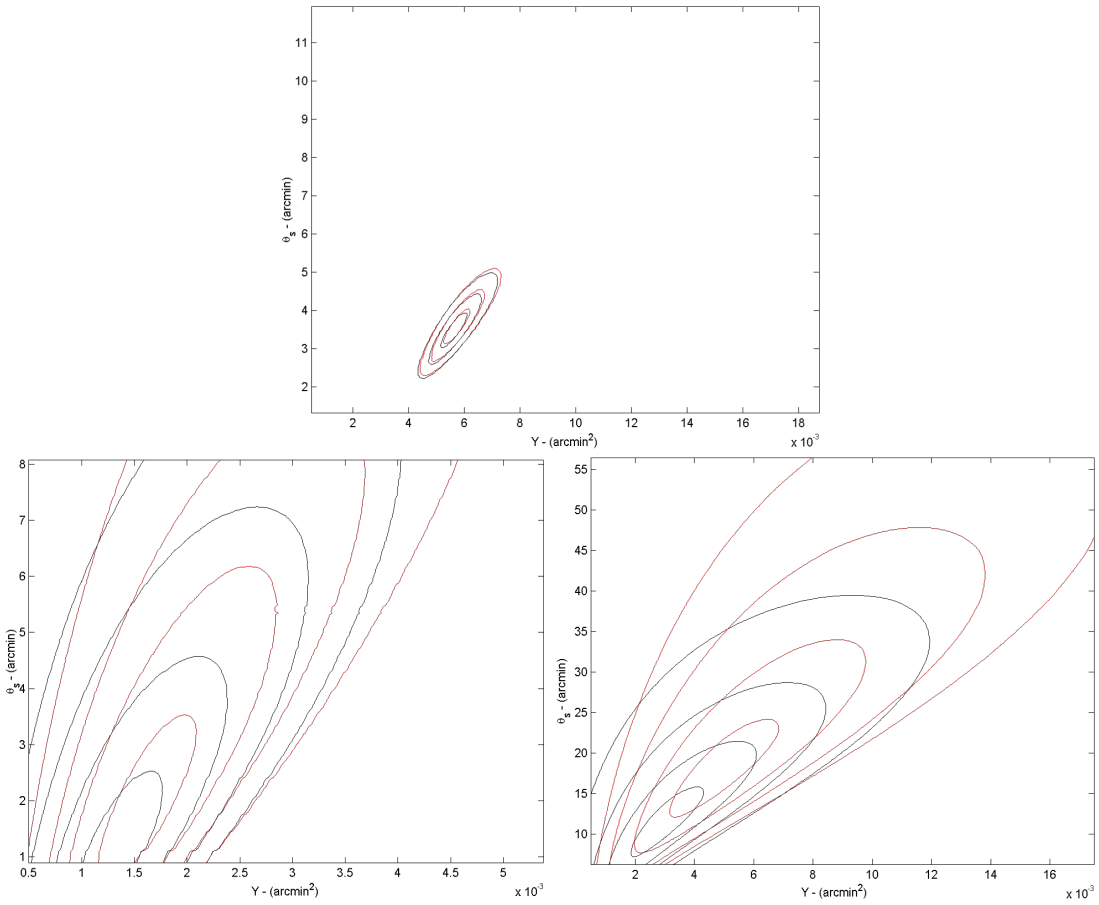


Figure 4.25: The red contours show the joint posterior distributions assuming all priors are 'flat' (uniform) where the likelihood mass concentrates. The black contours were made from the same data but using the priors described in table 4.4. The plot in the *upper row* shows a typical high SNR detection. The *lower row* depicts two examples of low SNR detections: on the *left* a small point like cluster and on the *right* an extended and well resolved. The plot x axis shows $Y \equiv Y_{5r500}$ [arcmin^2]. The y axis is θ_s [arcmin].

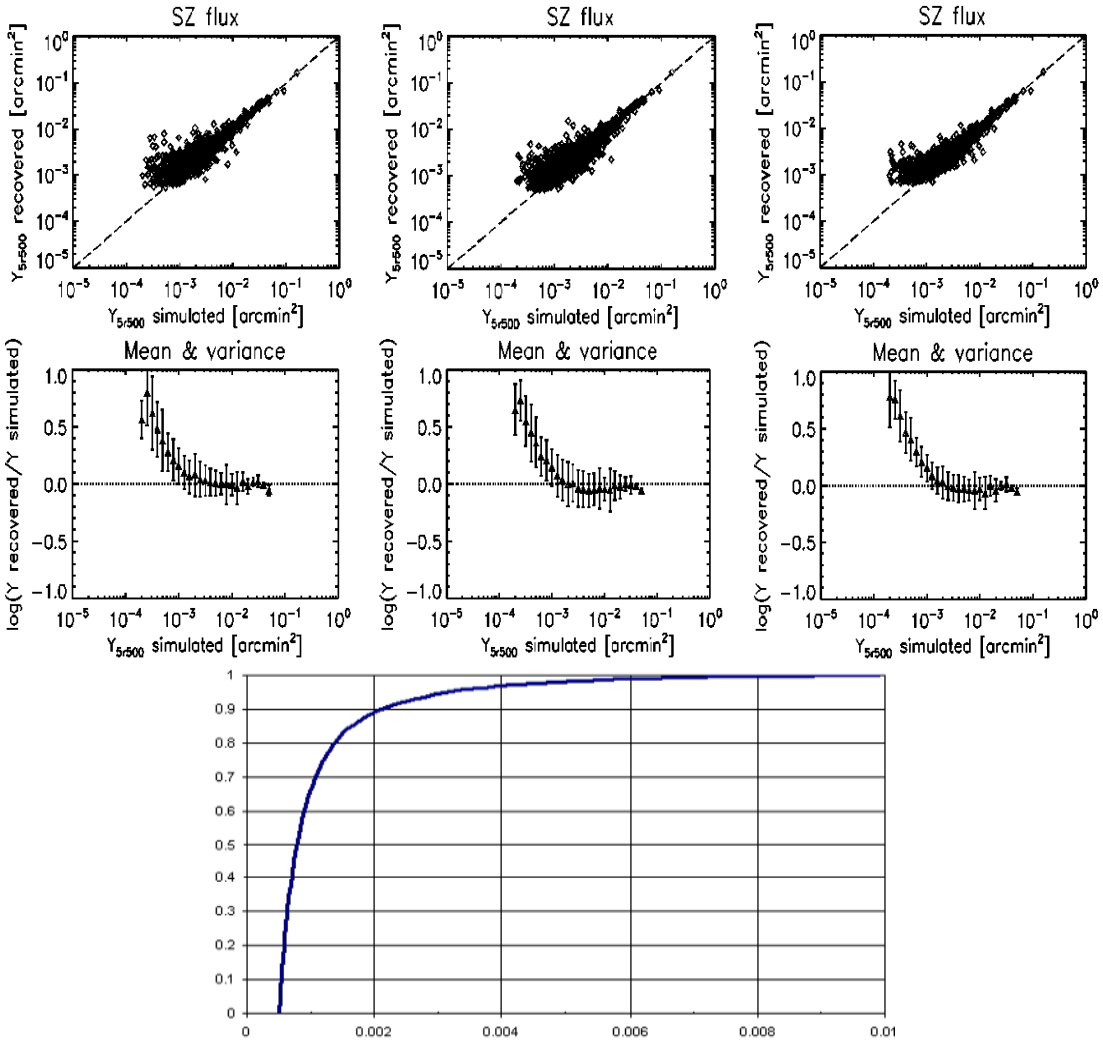


Figure 4.26: The *upper row* and *middle row* compare PwS's recovered Y_{5r500} with the simulated input value. On the *left column*, the best fit estimates were obtained through a mode estimator using 'flat' priors on both parameters, Y_{5r500} and θ_s, the equivalent of a maximum likelihood estimator. On the *middle and right column* the flat priors were replaced by those described in table 4.4. The *middle column* shows the posterior mode estimate and the column on the *right* the posterior expected value. The Y_{5r500} units are [arcmin^2]. The *lowest row* shows an normalised integrated histogram (CDF) of the Y_{5r500} injected distribution (x axis units in [arcmin^2]).

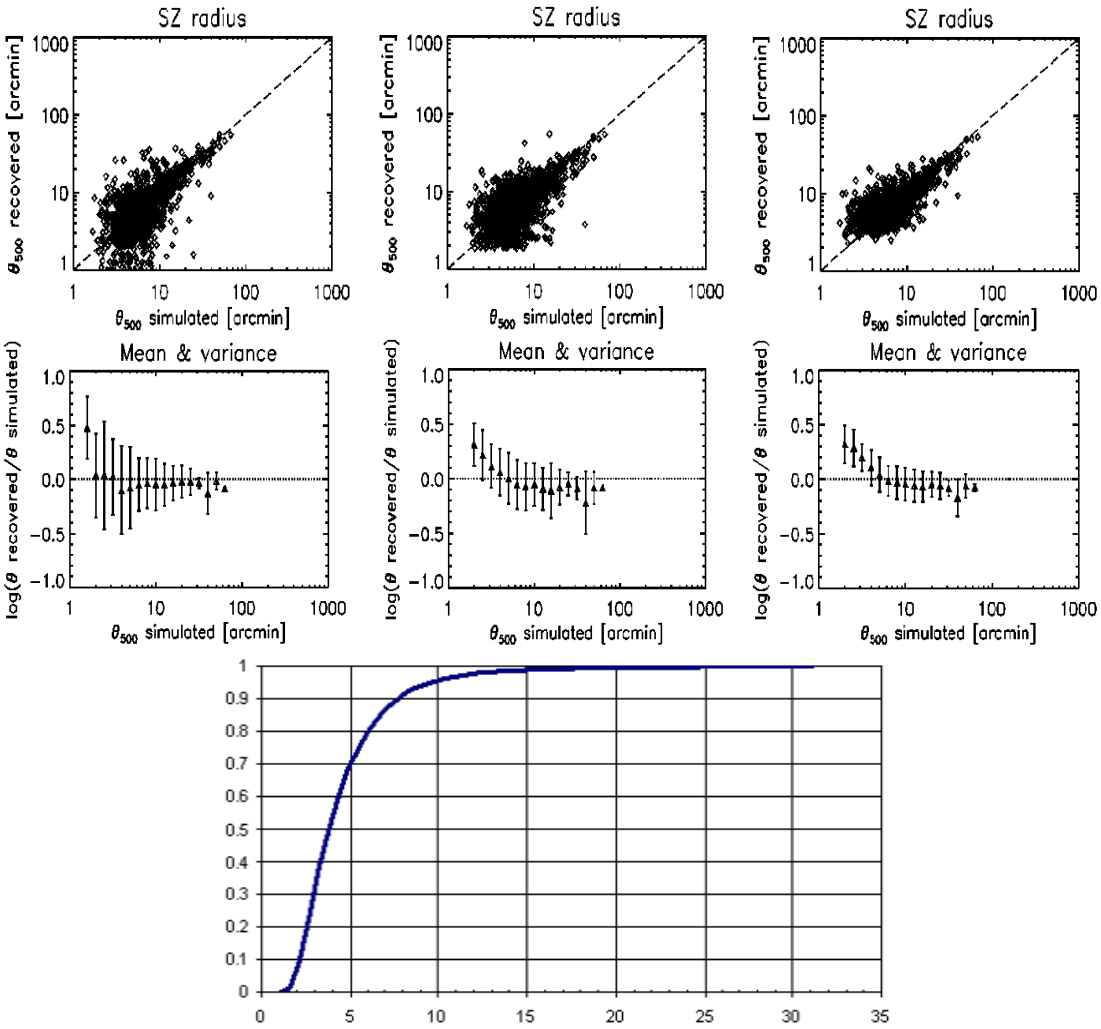


Figure 4.27: These plots were obtained the same way as those in figure 4.26, but this time using parameter θ_s. The x axis on the *last row* plot shows θ_s plotted in units of [arcmin].

Chapter 5

'*Planck*' detection capabilities - a simplified model

For the sake of completeness, in this last chapter we have attempted to approach the same detection formalism but now employing a somewhat different technique. This technique, the Fisher information matrix analysis, is more frequently found in the realm of the frequentist school of probability [Trees, 2001, ch. 2]. In fact even in Bayesian statistics the asymptotic distribution of the posterior mode depends on the Fisher information matrix and not on the prior [Cam and Yang, 2000][1]. Up to now, we have always explored our Bayesian detection formalism to make inferences about parameter estimation, or decision making, giving the current data set. We now approach this from a different angle, where we assume the parameters of our test hypothesis are non-random, though unknown, and the data are realisation(s) of a well characterised random process. Even though the Bayesian formalism is the only fully consistent and optimal way of interpreting the data accounting for all ancillary information, the suggested method (frequentist) might, in this particular case, provide a shorter and perhaps simpler way to draw inferences about our data collecting systems. Moreover, owing to the asymptotic symmetry in the sense of large data sets of both approaches[2], in the end one should expect to arrive at the same conclusions. Even though this chapter aims at exploring a different angle of inference, with a distinct foundation, all common

[1] This result was first anticipated by Laplace for the exponential family of probability distributions.

[2] Modern cosmological data sets like those produced by spaceborne instruments like the *WMAP* or *Planck* satellites are sufficiently large to make the asymptotic properties hold with good accuracy.

symbols to earlier chapters keep their meanings unchanged.

5.1 Fisher analysis

The '*Fisher information matrix*' is defined as

$$\mathcal{I}_{ij} = -\left\langle \left(\frac{\partial^2 \ln \mathcal{L}}{\partial\alpha_i \partial\alpha_j} \right)_{\boldsymbol{\alpha}} \right\rangle_H , \tag{5.1}$$

where $\mathcal{L}$ is the likelihood[1], $\boldsymbol{\alpha} \equiv [\alpha_0 \cdots \alpha_n]$ is a vector representing the parameter set and the ensemble average is taken assuming hypothesis H. Using the Cramér-Rao bound and supposing that a unbiased estimator exists, then a lower bound on the parameter covariance matrix $C_{ij} \equiv \langle \delta\alpha_i \delta\alpha_j \rangle$ reads

$$C_{ij} \geq (\mathcal{I}^{-1})_{ij}. \tag{5.2}$$

Equality holds only when the estimator is minimum variance ('*efficient*'). The likelihood mode is an asymptotically unbiased and efficient estimator, therefore reaching the Cramér-Rao bound [Trees, 2001][2].

The likelihood expression derived in the context of the detection of a single isolated source (equation 3.23 with restriction 3.24) and tailored with the SZ effect detection symbols reads

$$\ln\left(\mathcal{L}_{H_s}\right) \propto Y\mathcal{F}^{-1}\left[\mathcal{P}(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta};\theta_s)\right]_{\boldsymbol{X}} - \tfrac{1}{2}Y^2 \sum_{\boldsymbol{\eta}} \mathcal{Q}(\boldsymbol{\eta})|\widetilde{\tau}(\boldsymbol{\eta};\theta_s)|^2. \tag{5.3}$$

Owing to the symmetry properties of the Fisher matrix (see A-3), it is possible to isolate the sub-space of the position parameters from the remaining set $\{Y, \theta_s\}$, where we shall now focus our attention. The evaluation of the Fisher matrix coefficients for

[1] This definition of the Fisher information matrix assumes the likelihood is twice differentiable and certain regularity conditions apply (see [Trees, 2001]).

[2] In some cases the convergence of the likelihood mode estimates can be slow. A simple example is the maximum-likelihood estimate of a Gaussian population variance with known mean.

the $\{Y, \theta_s\}$ parameter subset can be obtained with the following expression

$$-\left\langle \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_i \partial \alpha_j} \right)_\alpha \right\rangle_{H_s} = -\int_{\boldsymbol{\eta}} \mathcal{Q}(\boldsymbol{\eta}) \, \frac{\partial \, \tilde{h}(Y, \theta_s, \boldsymbol{\eta})}{\partial \alpha_i} \cdot \frac{\partial \, \tilde{h}(Y, \theta_s, \boldsymbol{\eta})}{\partial \alpha_j} \, \mathrm{d}\boldsymbol{\eta}, \tag{5.4}$$

where $\tilde{h}(Y, \theta_s, \boldsymbol{\eta}) = Y \, \tilde{\tau}(\boldsymbol{\eta}; \theta_s)$. Performing the operations above depicted and after some trivial algebraic manipulations, we obtain the following expressions for the Fisher matrix coefficients

$$\begin{cases} \mathcal{I}_{YY} = \frac{2\pi}{\sigma_{Y_0}^2} & \int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right) \, f_{yy}(\psi) \, \psi \, \mathrm{d}\psi, & f_{yy}(\psi) = \tilde{\tau}^2(\psi) \\ \mathcal{I}_{Y\theta_s} = \frac{2\pi Y}{\theta_s \sigma_{Y_0}^2} & \int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right) \, f_{y\theta_s}(\psi) \, \psi \, \mathrm{d}\psi, & f_{y\theta_s}(\psi) = \frac{d\tilde{\tau}^2(\psi)}{d\psi} \, \frac{\psi}{2} \\ \mathcal{I}_{\theta_s\theta_s} = \frac{2\pi Y^2}{\theta_s^2 \sigma_{Y_0}^2} & \int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right) \, f_{\theta_s\theta_s}(\psi) \, \psi \, \mathrm{d}\psi, & f_{\theta_s\theta_s}(\psi) = \left[\frac{d\tilde{\tau}(\psi)}{d\psi}\right]^2 \psi^2 \end{cases}, \tag{5.5}$$

where $\psi = \eta \, \theta_s$, $\mathcal{Q}'\left(\theta_b \boldsymbol{\eta}\right) = \sigma_{Y_0}^2 \mathcal{Q}(\boldsymbol{\eta})$, $\sigma_{Y_0} = \left(\int \mathcal{Q}(\boldsymbol{\eta}) \, \mathrm{d}\boldsymbol{\eta} \right)^{-1/2}$, $\mathcal{R} = \theta_b / \theta_s$ and θ_b is the '*natural*' scale of the instrument 'equivalent' PSF[1]. We have tried to express all entities using dimensionless quantities to make them as general as possible. The parameter $\mathcal{R}$ is a proxy for how well the instrument is able to resolve targets. The function $\mathcal{Q}'(\phi)$ reaches its maximum at the origin and then quickly decreases as ϕ increases[2].

- Small $\mathcal{R} \Rightarrow \theta_b \ll \theta_s$. The objects are much larger than the PSF scale. In this case the beam spatial transfer function may be assumed approximately constant, equal to the zero frequency sensitivity, across the range of values the functions $f_{yy}(\psi), f_{y\theta_s}(\psi), f_{\theta_s\theta_s}(\psi)$ are significantly different from zero, implying $\mathcal{Q}'(\phi) \sim \mathcal{Q}'(0)$. This means the object is fully resolved and the solution acts as if the PSF were not present.

- Large $\mathcal{R} \Rightarrow \theta_b \gg \theta_s$. The object transverse sizes are much smaller than the PSF scale. In this case the beam transfer function drops so quickly that $\tilde{\tau}(\psi)$, may be approximated by its value at the origin ($\tilde{\tau}(\psi) \sim \tilde{\tau}(0)$) and $f_{y\theta_s}(\psi), f_{\theta_s\theta_s}(\psi) \sim$

[1] We say 'equivalent' because generally the instrumental setup, like all modern cosmological probes, may observe in multiple channels, each one with a different PSF.

[2] This is not always true for interferometers. The beam transfer function drops off very quickly for low spatial frequencies (large scales) as result of the minimum baseline and/or under-sampled uv plane. That usually means that when $\mathcal{R}$ is small (object much larger than the beam 'natural' scale) the interferometers are mostly insensitive to those objects.

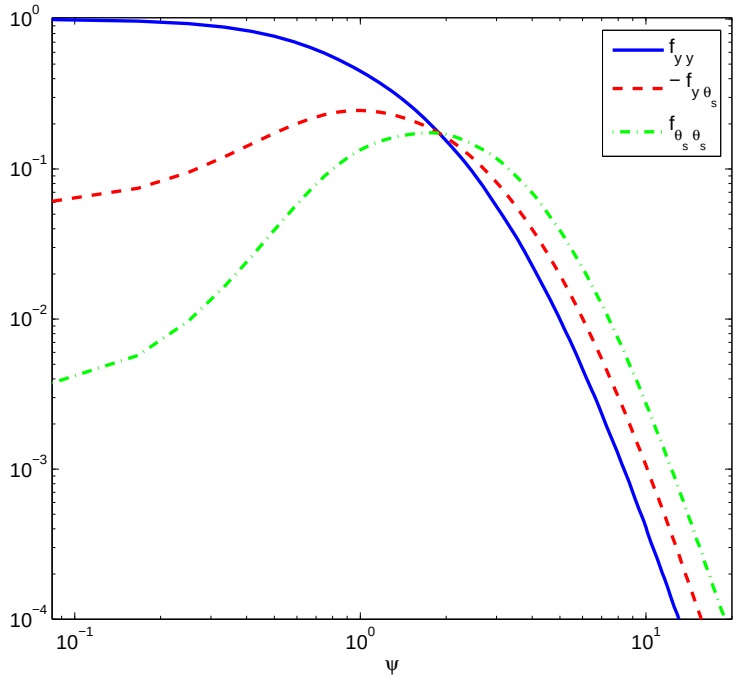


Figure 5.1: $f_{yy}(\psi)$, $f_{y\theta_s}(\psi)$, $f_{\theta_s\theta_s}(\psi)$ functions from equation 5.5. $\tau(\boldsymbol{x})$ is the universal GNFW profile integrated in the $l.o.s$ (see equation 4.15 and Arnaud et al. [2010]). The profile was normalised to enclose a unit volume.

$f_{y\theta_s}(0), f_{\theta_s\theta_s}(0) \sim 0$, rendering the actual profile of the object completely irrelevant. This case is usually known as unresolved or point like objects.

The cluster template profile we have been using, a GNFW profile integrated in-the-line-of-sight (4.15), does not possess a simple closed analytical formula except at the origin. So, in order to proceed, one needs to resort to numerical evaluations. Figure 5.1 depicts the functions $f_{yy}(\psi)$, $f_{y\theta_s}(\psi)$, $f_{\theta_s\theta_s}(\psi)$. Its worth noting that the dependence of the Fisher coefficients on the extension of the source is fully contained in these functions which are not random and becomes completely defined after choosing the detection profile.

5.2 *Planck* SZ sensitivity (σ_Y)

From equations 3.35 and 3.29 we know that the flux (Y) posterior distribution given θ_s is a Gaussian with dispersion given by

$$\sigma_Y = \delta Y_{(\theta_s)_{known}} = \frac{1}{\sqrt{\mathfrak{I}_{YY}}} = \frac{\sigma_{Y_0}}{\sqrt{2\pi}\sqrt{\int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right)\ f_{yy}(\psi)\ \psi\ \mathrm{d}\psi}}. \tag{5.6}$$

Then, by definition, the SNR of a detection is

$$\mathrm{SNR} \equiv \frac{\widehat{Y}}{\delta Y_{(\theta_s)_{known}}} = \widehat{Y}\ \sqrt{\mathfrak{I}_{YY}}, \tag{5.7}$$

where $\widehat{Y}$ is the the maximum likelihood estimate of Y. Equation 2.30 describes the completeness of the catalogue as function of Y and depends only on σ_Y and the actual acceptance/rejection SNR threshold γ. In the current context, the completeness function is usually known as the '*selection function*'. The selection function attempts to characterise the population of clusters by defining a pure 'Y cut sample'. This kind of sample is a very powerful tool to constrain cosmology via the mass function and the distribution of Y, a well known low dispersion proxy for the cluster mass (see Allen et al. [2011] and references therein). Mass function derived number counts are very steep functions of the cosmological parameters, particularly σ_8 and Ω_m, placing therefore strong constraints on cluster abundance and evolution. So, understanding the limits our instrument places on σ_Y, the 'Y *sensitivity*', is of pivotal importance in the specification of the instrument capability for constraining cosmology.

5.2.1 Background model

In section 3.1 (equation 3.1), we separated the background into two components, regarding their origin: instrumental (pixel) noise and astronomical background emission. *Planck* pixel noise may be very safely assumed as completely uncorrelated not only spatially but across observation channels as well. On the other hand, the astronomical background is highly, though not perfectly, correlated spatially and across channels. In the current analysis it is more convenient to decompose the background in channel 'i' using its correlation properties instead. So, the background in channel 'i' may be

defined as

$$\boldsymbol{b}_i = \boldsymbol{c}_i + \boldsymbol{n}_i, \tag{5.8}$$

where $\boldsymbol{c}_i$ is assumed as being perfectly correlated across channels and $\boldsymbol{n}_i$ completely uncorrelated. By *perfectly correlated* we mean in this context

$$\text{if} \quad \boldsymbol{c}_j = k_j \boldsymbol{c} \quad \text{then} \quad \langle \boldsymbol{c}_i^t \boldsymbol{c}_j \rangle = k_i^* k_j \, \boldsymbol{c}^t \boldsymbol{c}, \tag{5.9}$$

where $\boldsymbol{c}$ is some arbitrary background reference and k_j is a constant (eventually complex to accommodate phase differences). The *completely uncorrelated* component $\boldsymbol{n}_i$ is such $\langle \boldsymbol{n}_i^t \boldsymbol{n}_j \rangle = \sigma_i^2$ if $i = j$ and 0 otherwise. This simplistic background model will include one single correlated component only. Therefore, without any loss of generality, one may always choose appropriated units so that $k_i = k_j = 1$. Accepting the background is homogeneous, its cross-covariance matrix is then circulant and taking advantage of the linearity of the Fourier transform, the following properties apply:

$$\begin{aligned} \langle n_{i\alpha} n_{j\beta} \rangle &= \begin{cases} \sigma_i^2 & \text{only if } i = j \text{ and } \alpha = \beta \\ 0 & \text{otherwise} \end{cases} \\ &, \\ \langle c_{i\alpha} c_{j\beta} \rangle &= \begin{cases} c^2 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5.10}$$

where the Greek symbols identify the spatial Fourier modes.

The simplest possible emission model we may define possesses a single background component, in this case CMB[1], plus the foreground signal we are searching for, in this particular example the SZ emission. Of course, we need to consider the antenna beam PSFs and the instrumental noise (assumed white) as well. So, in this stripped down model, the '*correlated*' component is represented by

$$\langle \boldsymbol{c}_i^t \boldsymbol{c}_j \rangle = \alpha_{ij} \, \text{CMB}. \tag{5.11}$$

[1] At all *Planck* channels that will be used in this analysis ($\{30, 44, 70, 100, 143, 217, 353\}$ GHz), CMB is the dominant background component at high galactic latitudes

where α_{ij} is the cross channel degree of correlation of the CMB emission defined by

$$\alpha_{ij} = \frac{\langle \boldsymbol{c}_i^t \boldsymbol{c}_j \rangle}{k_i^* k_j \; \boldsymbol{c}^t \boldsymbol{c}}, \tag{5.12}$$

An α_{ij} close to 0 means that the cross-power spectrum of the CMB is close to 0. On the opposite side of the range, ($\alpha_{ij} \approx 1$), means the cross-power spectrum of the CMB perfectly matches the power spectrum. For simplicity we will accept α as independent of the channel ($\alpha = \alpha_{ij}$). Likewise, we will model the CMB as simple scale-free power spectrum with an index $= 2$

$$\langle \boldsymbol{c}_i^t \boldsymbol{c}_j \rangle = \alpha \; \mathrm{CMB}_0 \; l^{-2}, \tag{5.13}$$

where CMB_0 is a constant. The '*uncorrelated*' component is represented as the sum of two contributions

$$\sigma_i^2 = \varepsilon_i^2 + \; (1 - \alpha) \; \mathrm{CMB}_0 \; l^{-2}, \tag{5.14}$$

where ε_i^2 is the pixel noise and the second term the leakage from the correlated component. This exceptionally stripped-down model lacks most of the complexity of the real data, namely:

- No provision for other astronomical components: free-free, dust, confusion noise from undetected unresolved sources, etc.
- The α parameter is assumed constant for all Fourier modes.
- The relative balance of all background components is assumed constant across all sky (statistical homogeneity).

Despite its extreme simplicity we could reproduce the results of the *Planck* ESZ catalogue with reasonable accuracy (see section 5.2.3)

5.2.2 The evaluation of $\mathcal{Q}(\eta)$

From equation 5.7 one can see that if the source extension approaches 0 then $f_{yy} \to 1$. Therefore for a certain fixed flux Y, the SNR is proportional to $\sqrt{\int \mathcal{Q}(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta}}$. As $\mathcal{Q}(\boldsymbol{\eta}) \geq 0$[1], the higher and the wider $\mathcal{Q}(\boldsymbol{\eta})$ is, the better the cluster can be detected. We

[1] $\mathcal{Q}(\boldsymbol{\eta})$ is a quadratic form.

shall now proceed to examine the dominant factors controlling the shape of $\mathcal{Q}(\boldsymbol{\eta})$. One single Fourier mode of $\mathcal{Q}_\eta$, reads

$$[a_0 z_0, \ldots, a_n z_n] \left(c^2 \begin{bmatrix} a_0^2 + X_0, & a_0 a_1, & \ldots & , a_0 a_n \\ a_1 a_0, & a_1^2 + X_1, & \ldots & , a_1 a_n \\ \vdots, & \vdots, & \vdots & , \vdots \\ a_n a_0, & a_n a_1, & \ldots & , a_n^2 + X_n \end{bmatrix} \right)^{-1} \begin{bmatrix} a_0 z_0 \\ \vdots \\ a_n z_n \end{bmatrix}, \tag{5.15}$$

where a_i is the beam transfer function at channel '*i*', z_i is the SZ SED at channel's '*i*' effective frequency and $X_i = \sigma_i^2/c^2$ (we have omitted the subscript η for clarity). A closed form for $\mathcal{Q}_\eta$ may be obtained, after a long but otherwise uninteresting algebraic manipulation (see Henderson and Searle [1981] for a detailed proof), and it reads

$$\mathcal{Q}_\eta = \sum_i \Theta_i z_i^2 - \frac{c^2}{1 + c^2 \sum_i \Theta_i} \left(\sum_i \Theta_i z_i \right)^2, \tag{5.16}$$

where $\Theta_i = \frac{a_i^2}{\sigma_i^2}$. A much more insightful form to present this result is the following

$$\mathcal{Q}_\eta = \frac{\sum_i \Theta_i z_i^2 \; + \; c^2 \left[\sum_i \sum_{j>i} \Theta_i \Theta_j (z_j - z_i)^2 \right]}{1 + c^2 \sum_i \Theta_i}, \tag{5.17}$$

because it allows us to identify two different regimes where some approximations can be introduced to produce simplified, more predictive, expressions:

- **Uncorrelated component dominates** ($c^2 \sum_i \Theta_i \ll 1 \Rightarrow \sigma_i^2 \gg c^2 a_i^2$)

$$\mathcal{Q}(\boldsymbol{\eta}) \approx \sum_i \frac{a_i^2(\boldsymbol{\eta}) z_i^2}{\sigma_i^2}. \tag{5.18}$$

  $\mathcal{Q}_{\boldsymbol{\eta}}$ is the co-addition of the different channels. This result is exactly what would be expected when optimally (in the sense of minimizing the noise) combining completely uncorrelated channels. This regime happens for the high '*l*' Fourier modes.

- **Correlated component dominates** ($c^2 \sum_i \Theta_i \gg 1 \Rightarrow \sigma_i^2 \ll c^2 a_i^2$)

$$\mathcal{Q}(\boldsymbol{\eta}) \approx \frac{\sum_i \left\{ \Theta_i(\boldsymbol{\eta}) \sum_{j>i} \Theta_j(\boldsymbol{\eta}) (z_j - z_i)^2 \right\}}{\sum_i \Theta_i(\boldsymbol{\eta})}. \tag{5.19}$$

  Within the range the correlated component dominates, $\mathcal{Q}(\boldsymbol{\eta})$ is a weighted combination of the square of the differences between the channels. This is the normal case for the low '*l*' region.

The most interesting result is that in neither case $\mathcal{Q}(\boldsymbol{\eta})$ depends on the correlated component! So even for low '*l*' where the main contaminant of the signal is the astronomical background (mainly CMB), an appropriately tuned multi-frequency setup (see 5.2.4) might eliminate it to the point where the only noise source left is pixel noise. This situation can be exploited with great advantage when pixel noise is decidedly subdominant in comparison with CMB as in the case of *Planck*. However, the amplitude of $\mathcal{Q}(\boldsymbol{\eta})$, when the correlated component dominates, is proportional to the differences of the SZ spectral signature at the observation channels. If the observation frequencies are closely packed either in the positive or negative side of the SZ spectrum, this might bring $\mathcal{Q}(\boldsymbol{\eta})$ to very low values, making the instrument very insensitive to those scales. On the other hand *Planck's* HFI instrument frequency distribution, with channels on both sides (positive and negative) of the SZ spectrum, is such that it optimally exploits this feature. The width of $\mathcal{Q}(\boldsymbol{\eta})$ is mainly dictated by the high '*l*' regime when the uncorrelated component is dominant. Analysing this expression it becomes apparent, with no surprise, that the narrower the beam (higher resolution) the better.

5.2.3 Comparison with the ESZ

To assess how representative this detection model is of a modern cosmological data set, we have tried to predict the *Planck* Early SZ catalogue of clusters of galaxies (ESZ) [Planck Collaboration et al., 2011f]. We have employed all entries in the catalogue (189). We have only kept the detection SNR estimates though. The large beams and high instrumental noise levels in *Planck* data at the time of publication of the catalogue[1] severely limited the cluster size constraining capability of the instrument.

[1] This catalogue was an '*early release*'.

	LFI			HFI			
Channel [GHz]	30	44	70	100	143	217	353
ΔT_{SZ} $[\Delta T/T]$	-1.954	-1.901	-1.753	-1.509	-1.042	-0.011	2.235
Noise $[K^2\text{rad}^2]10^{-15}$ ..	11.0	18.0	11.0	0.78	0.25	0.59	7.61
Beam FWHM [arcmin]	32.65	27.92	13.01	9.53	7.08	4.71	4.50

Table 5.1: Instrumental setup. No colour corrections were employed.

$\langle\Delta\rangle$	σ_Δ	$\langle\Delta\rangle/\sigma_\Delta$	$\widehat{\beta}\,(y_0=0)$	ρ
-0.2	2.26	-8.8%	0.98	0.851

Table 5.2: $\Delta = \text{SNR}_{pred} - \text{SNR}_{esz}$, $\widehat{\beta}(y_0 = 0)$ is the best fit slope of a line that includes the origin and ρ is the correlation coefficient.

Unfortunately, owing to the important Y, θ_s correlation (see sections 5.3, 4.2.4.2 and Planck Collaboration et al. [2011f]) the large uncertainty in the cluster extension propagates through to the Y estimates. To limit the uncertainty on the SZ flux estimate, the cluster radius estimates in the catalogues were derived from high resolution X-ray observations [Planck Collaboration et al., 2011f]. In this work we have used the original SZ estimates extracted directly from the *Planck* data exclusively using PwS. Table 5.1 shows the values of the parameters we employed in our model to try to predict the ESZ catalogue SNRs using formula 5.7. This set of values was collected from the *Planck* early high and low frequency instrument performance data [Mennella et al., 2011; Planck HFI Core Team et al., 2011; Zacchei et al., 2011]. We took $\text{CMB}_0 = 2\pi\ 2000.0\ (\mu K)^2$ from Hinshaw et al. [2009] and assumed a cross channel degree of correlation $\alpha = 0.99$. In figure 5.2 we plot the SNRs from the ESZ catalogue (SNR_{esz}) versus the prediction (SNR_{pred}) and a histogram of the residuals ($\Delta \equiv \text{SNR}_{pred} - \text{SNR}_{esz}$). Despite its very basic design, this model predicts with reasonable accuracy the ESZ SNRs (see table 5.2). It is mostly unbiased, with low dispersion and good correlation. It is remarkable that such a simple and stripped-down model can actually make such close prediction of the catalogue content.

5.2.4 σ_Y predictions and selection function

We can use to great advantage this model to predict σ_Y, the instrument sensitivity (see equation 5.6), for a variety of different instrumental setups and get a first approxi-

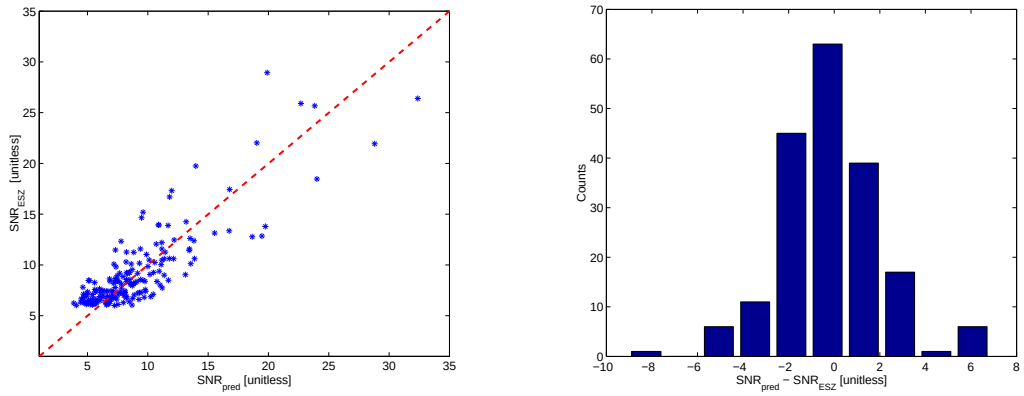


Figure 5.2: *Left panel*: SNR prediction from the current model SNR_{pred} versus the SNR from the ESZ catalogue. The red (dashed) line is the $1:1$ line. The plot's statistical data may be found in table 5.2. *Right panel*: Histogram of the residuals ($\Delta = \mathrm{SNR}_{pred} - \mathrm{SNR}_{esz}$).

mation of its selection function. Another possibility is to suggest new instrumental designs given survey goals. We set off by predicting how the spatial extension of the clusters, as seen by *Planck* HFI and LFI instruments[1] constrain the instrument sensitivity to the different scales and therefore the sample selection. Both sensitivity curves, HFI and LFI (see figures 5.3 and 5.4) were drawn using the data in table 5.1. Only four channels were used in contrast to the ESZ catalogue where all HFI frequencies (6) were used[2]. The 100 GHz channel was added to the LFI setup to make the cases more comparable (same amount of data). Five different values of α (cross-channel correlation), $\{0.0, 0.9, 0.98, 0.99, 1.0\}$, were tested. The cluster extension range goes from 0.0 arcmin, unresolved point like object, to 30.0 arcmin, a well resolved object. The dashed curves were plotted assuming the instrumental noise was $1/2$ of the level present in the early release data, or close to the level one should expect in the full mission data. As expected, there is a general trend in that both instruments are less sensitive to more extended clusters with lower flux densities (y). But this where the affinity ends. As

[1] After adding redshift information the observed spatial scale θ_s can be converted into the actual cluster physical size.

[2] The non inclusion of the two highest *Planck* channels was dictated mainly by the fact the dominant background component in these frequencies is dust, which clearly breaks some assumptions of our model.

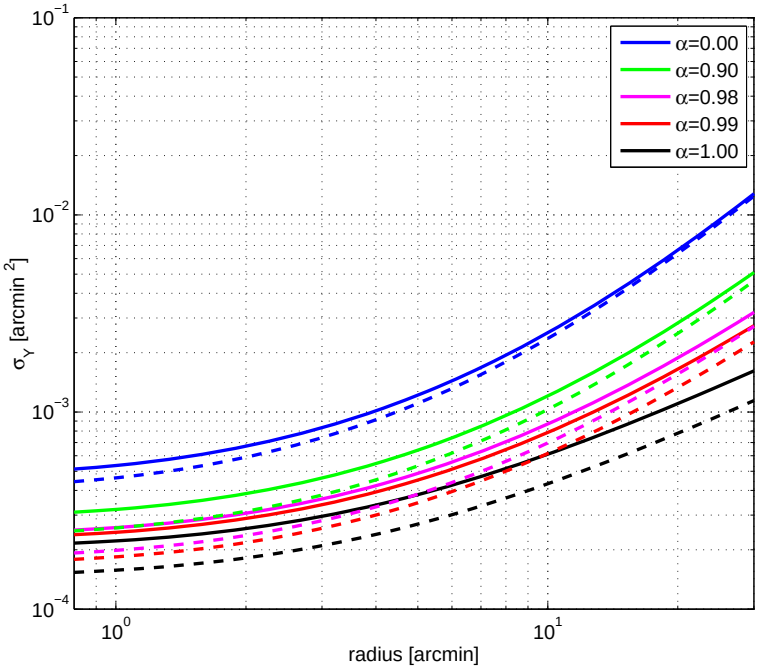


Figure 5.3: Predicted *Planck* HFI sensitivity as a function of the cluster radius. Solid lines were drawn assuming the noise values of the early release papers (see table 5.1) corresponding to 1 full sky scan and the dashed lines halving the instrumental noise values (~ 4 full sky scans).

result of the much wider beams and higher noise levels[1], the LFI is much less sensitive to the SZ emission, effectively making it almost blind to any cluster except the brightest ones[2]. Something more interesting is that, despite of the higher resolution and much lower noise, if the cross-channel correlation $\alpha = 0$ (the channels are simply co-added), the HFI sensitivity becomes comparable to that of the LFI. Another aspect of the problem is the impact on the sensitivity when reducing the instrumental noise. The pixel noise is, roughly speaking, inversely proportional to the square root of the number of times a pixel is observed ('*hit count*'). The dashed curve shows the sensitivity if the pixel noise is reduced to half of the ESZ values, which accounts for four

[1] One order of magnitude for the cosmological channels.

[2] No more than a dozen clusters are likely to be detected with this configuration. Some examples are Coma, $A2319$ and $A2256$.

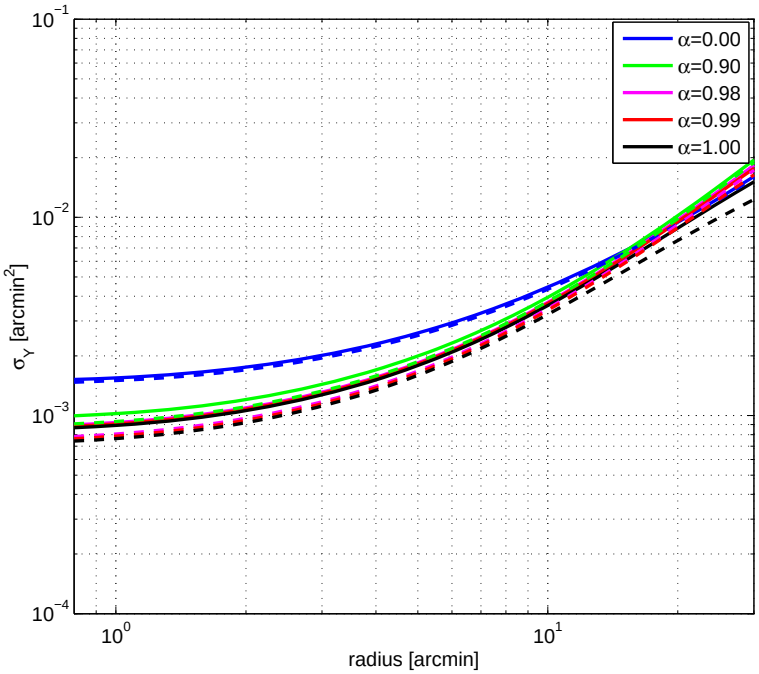


Figure 5.4: Predicted *Planck* LFI sensitivity as a function of the cluster radius. Follows the same conventions as figure 5.3 (the LFI setup includes the HFI 100 GHz channel).

full sky surveys instead of a single one. The difference is quite significant for the HFI and mostly irrelevant for the LFI. The extreme sensitivity of the HFI to the SZ signal comes to too high a cost. Unless a very accurate characterization of the beams PSF and very effective elimination of the $1/f$ noise (map-making) is achieved, the α values will not get close to 1 and the sensitivity will significantly drop as figure 5.3 shows. *Planck* was the first instrument to release a SZ cluster catalogue constructed with data from both sides of the SZ spectrum [1]. The atmospheric window starts to close very quickly at frequencies above $\sim$ 150 GHz even on the driest and highest places. Observing in those channels always means expensive spaceborne missions. What would then be the impact of using an instrument with the HFI specifications but at the LFI frequencies?

Figure 5.5 shows the predicted sensitivity of an instrument with the HFI specifications, beam widths and noise levels, but operating at LFI frequencies. A quick

[1] SPT (South Pole Telescope Consortium) last catalogue of about 220 SZ detected clusters was extracted from the 95 and 150 GHz channels only [Reichardt et al., 2012].

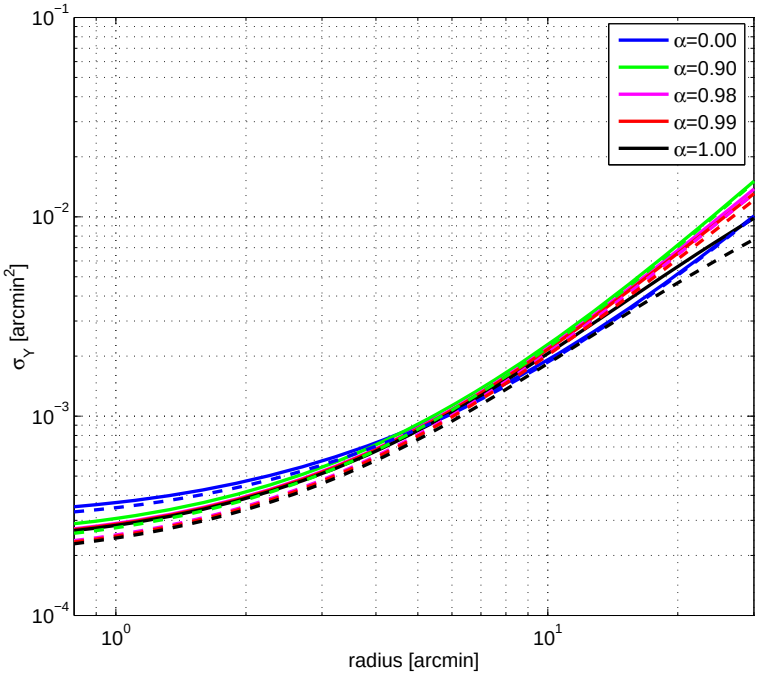


Figure 5.5: Sensitivity of a virtual experiment keeping the HFI performance traits unchanged (noise + beam PSF) and shifting the channel central frequencies to LFI values. Lines colour and shape follow previous cases (see figure 5.3).

inspection of the SZ signal strength at the LFI channels would have anticipated a different result. The average SZ signal strength for the HFI frequency set is ~ 1.20 and for the LFI ~ 1.78 or about 48% higher[1]. Although, the HFI observing frequency set has clearly higher sensitivity throughout all cluster scales except for the $\alpha = 0$ case[2]! And the sensitivity degradation with the cluster extension is quicker, almost a decade when it reaches ~ 12 arcmin. Just co-adding channels ($\alpha = 0$) is not as penalizing as in the HFI case. Indeed clustering the observation frequencies into a narrow band effectively reproduces a similar effect as removing the correlation (see 5.17). σ_Y is not as dependent on the instrumental noise levels as in the HFI case. So, deeper integration times might not help increasing the size of the sample as the sensitivity is dominated

[1]For instance the 217 GHz channels carries almost no signal.

[2]The better sensitivity of the $\alpha = 0$ case should not come as a surprise given the fact that in fact the LFI frequencies carry more SZ signal.

not only by the instrumental noise but mainly by the astronomical background. This situation leaves almost no hope for ground based instruments to tackle large faint SZ clusters. However, large ground based dishes or interferometers, with much improved resolution and very low noise detectors, exploiting with great advantage the difference in scales (spatial filtering) can successfully tackle narrow clusters which make up the more interesting population of massive high redshift clusters (see AMI Consortium et al. [2008]; Marriage et al. [2011]; Reichardt et al. [2012]).

5.3 The Y, θ_s correlation

As previously mentioned (see 4.2.4.2) our Bayesian analysis of the joint probability $\Pr(Y, \theta_s|d)$ showed a strong degeneracy between the total cluster flux (Y) and size estimates (θ_s). The large *Planck* beams can only provide weak constraints on the cluster extension. Unfortunately, the cross-correlation between Y and θ_s, propagates the large extension uncertainty into the Y estimation. So, using once again the Cramér-Rao bound (equation 5.2) and the Fisher information matrix (5.5) we were able to predict the actual Y error bar (δY_{eff}) when considering the effect of the uncertainty in θ_s as well. The '*effective*' SNR then reads:

$$\mathrm{SNR}_{eff} = \frac{\widehat{Y}}{\delta Y_{eff}} = \mathrm{SNR}\ \sqrt{1-\rho^2}, \tag{5.20}$$

where ρ is the correlation coefficient given by

$$\rho = -\frac{\mathcal{I}_{Y\theta_s}}{\sqrt{\mathcal{I}_{YY}}\sqrt{\mathcal{I}_{\theta_s\theta_s}}} = \frac{-\int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right)\ f_{y\theta_s}(\psi)\ \psi\ \mathrm{d}\psi}{\sqrt{\int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right)\ f_{yy}(\psi)\ \psi\ \mathrm{d}\psi}\sqrt{\int_0^\infty \mathcal{Q}'\left(\mathcal{R}\psi\right)\ f_{\theta_s\theta_s}(\psi)\ \psi\ \mathrm{d}\psi}}. \tag{5.21}$$

Figure 5.6 depicts the variation of ρ with the cluster radius for the HFI setup (table 5.1). The prediction just confirmed what one would expect from the joint $\{Y, \theta_s\}$ posterior distributions (see section 4.2.4.2): that there exists an important correlation between Y and θ_s. There is a good match between these model predictions ($\rho \sim 0.90$) and the actual ESZ results ($\rho \sim 0.85$) obtained from *Planck* early data [Planck Collaboration et al., 2011f]. The prediction seems to be slightly biased high ($\sim 5\%$). However, as one could conclude from the plots, lower noise estimates tend to produce higher cor-

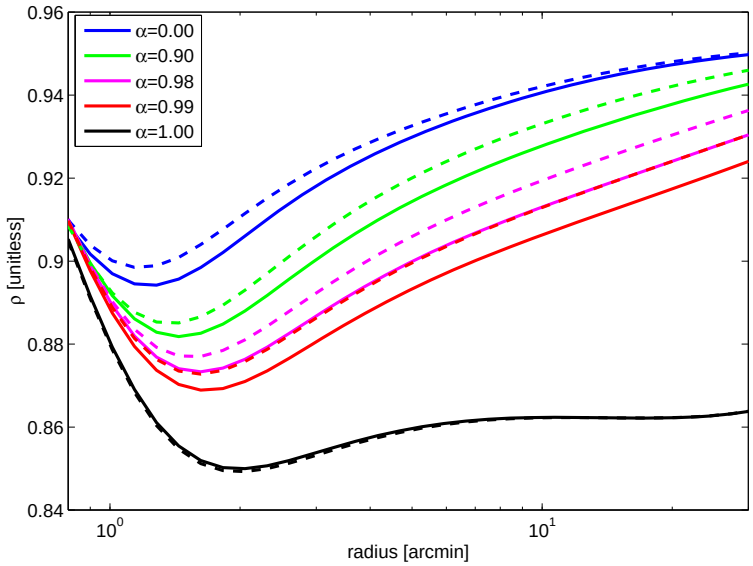


Figure 5.6: HFI configuration correlation coefficient ρ evolution with the cluster extension (equation 5.21). Line colour and shape follow previous cases (see figure 5.3).

relation coefficient values. Such a simple model does not include several sources of contamination such as dust, confusion, etc. We believe that these extra noise components, left out of the model, would bring the prediction in complete agreement with the ESZ results. Equation 5.20 relates ρ the $\{Y, \theta_s\}$ correlation coefficient to the effective uncertainty on the Y estimate. Regrettably, ρ is, throughout the full cluster size range, very high, imposing a severe detrimental impact on the Y constraint. In figure 5.7 we have tried to predict the impact of the strong $\{Y, \theta_s\}$ correlation in the ESZ catalogue Y estimates. The correlation coefficients are tightly distributed around 0.90 and with good accuracy one may say that the increase in the Y uncertainty (δY_{eff}) as result of the θ_s error propagation into the Y estimate is very severe $\delta Y_{eff}/\delta Y_{(\theta_s)_{known}} \sim 2.36$. The agreement between these predictions and the results in section 4.2.4.2 obtained by sampling the Y, θ_s joint posterior distributions is very good.

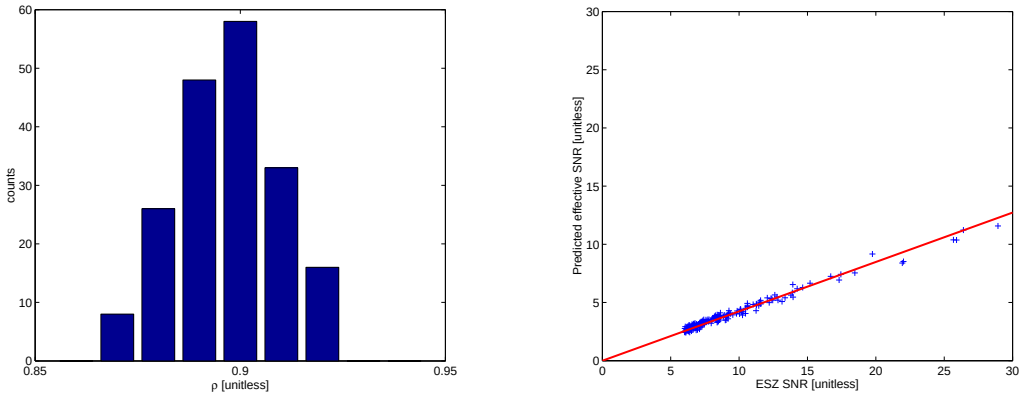


Figure 5.7: *Left panel*: Histogram of the correlation coefficient distribution in the ESZ catalogue. *Right panel*: ESZ catalogue SNR versus predicted 'effective' SNR (see equation 5.20). The red line is the best fit line ($y_0 = 0$) with a slope of 0.424.

5.4 Comments

- *Planck* HFI, owing to its unique instrumental setup and observing conditions is particularly well equipped to find faint extended clusters. This capability makes it extremely sensitive to very smooth and feeble gas pressure in the outskirts of the cluster, making it an excellent complement to X-ray measurements.

- However, as result of its large beams mainly, its capability to constrain the extension of cluster is very limited. Moreover, the large correlation between the cluster extension estimate θ_s and its total flux Y propagates the large size uncertainty into the flux estimate rendering it very '*noisy*'.

- By combining estimate constraints from X-ray observations with those from *Planck* SZ measurements, it is possible to provide low dispersion unbiased proxies of the cluster total mass, a valuable cosmological datum (see Planck Collaboration et al. [2011f] and references therein).

Chapter 6

Conclusions

The *Planck* satellite, and many other modern cosmological data sets, present completely new challenges for the detection and description of compact objects. With the immense accessible computing power now available, the need for more sophisticated algorithms, capable of extracting from the data a 'deeper' content, demands an improved understanding of the data as an information repository one wants to unlock. These new more ambitious goals impose that one should aim at (i) low or very low SNRs; (ii) strongly correlated backgrounds with typical scales similar to those of the objects being sought; (iii) full, consistent and probability/decision theory based understanding of the acceptance/rejection criterion and (iv) full joint uncertainty characterization of the estimates. These attributes render traditional object detection methods sub-optimal, since: (i) it is difficult to perform an optimal separation of the sources from the background fluctuations; and (ii) the complete characterisation of the parameter uncertainties together with their correlations is decisive and traditional methods do not provide it. Moreover, with the increased sensitivity of the contemporary instruments, the faint tail of the source populations, with more important and frequently more complex uncertainties, frequently dominates the sample.

A better strategy is to develop an object detection methodology from a strong statistical foundation first. The linear filtering family of tools is the attempt by the orthodox frequentist school of probability to overcome these limitations. The matched filter and all its derivatives are based on the Neymann-Pearson likelihood ratio, and their optimal performance is still extremely dependent on the choice of the acceptance/rejection threshold and on implementation details. Despite their widespread use, the actual prac-

tical designs of these tools do not yet implement a sound framework to handle the parameter estimates uncertainty manifold and in particular its complex degeneracies and correlations.

Bayesian methods have the great advantage of providing a coherent probability methodology with the option to include, in a completely consistent way, all ancillary information. This is really a major advantage since a wealth of different heterogeneous cosmological and astrophysical data sets are now widely available. Only through a consistent joint analysis, one is able to get the most of them.

But probability theory by itself only gives us a degree of belief. In order to produce a catalogue, decisions must be made as well. Decision Theory is unambiguous: $\ln\left[\frac{\Pr(H_1|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right]$ is the optimal decision tool (in the binary case), although the binary model is manifestly not powerful enough to handle a real data set. The necessary extension to a multi-model, multi-channel foundation is mandatory for an operational and viable solution. The simplest extension to a multi-model decision rule, under common loss functions normally employed in astronomy, would read instead: $\ln\left[\frac{\Pr(H|\boldsymbol{d})}{\Pr(\widetilde{H}|\boldsymbol{d})}\right]$ or as it is more commonly known $\ln(\text{odds})$

The evaluation of the evidence ($\Pr(H|\boldsymbol{d})$) poses several challenges: (i) high dimensionality integrals with complex boundaries ; (ii) non-Gaussian distributions; (iii) very large volumes of data; and (iv) fast operation. To achieve our goal we focused on taking advantage of the symmetries of the multi-channel likelihood manifold to design an efficient, though rigorous, exploration tool, mostly independent of the manifold geometry. After finding the likelihood maxima using a very fast algorithm that follows the implementation of the matched filter, we set off using a single-nested sampling procedure adopting the previously found values as an initial guide.

Owing to its full consistent probability foundation, PwS can provide a sound, generally applicable, and complete statistical characterization of its results. Simultaneously, we can offer effective solutions for the difficulties accompanying real data, in a real modern environment, without compromising any of our goals.

6.1 Future directions of development and products.

6.1.1 Directions of development

PwS Bayesian modeling is, however, still incomplete. Some might question that conditioning our reasoning by priors impairs or even misleads the inference, especially if these are neither complete nor accurate. We are partially sympathetic with this view, though: (i) if the data contain enough information and can actually constrain the object parameters estimates, the priors are mostly not influential (see introduction of chapter 5); (ii) PwS capability of using informative and non-informative priors as well, allows the user to select the appropriate prior for the required goal; and (iii) by exploring one of the most powerful Bayesian analysis tool, the '*hierarchical*' or '*random effect*' data models (see G. E. Box [1992, ch. 5], Jaynes [2004, ch. 6] and Gelman et al. [2003, ch. 5]), one can fully explore the optimal decision and estimation devices, only provided by informative priors, without limiting or biasing the results. Fundamental physics can most of the time constrain the general shape of the priors but not the actual parameter values they depend upon. This is much like how the laws of physics, usually through integral/differential equations, define the general behavior of physical systems but leave some undefined constants/conditions that can only be derived from the intial/boundary conditions. Bayesian hierarchical model methods allow the joint and simultaneous estimation of the object parameters and the parameters in the priors, the hyper-parameters, that provide the population laws. This will constitute the main subject of future PwS developments.

6.1.2 Future products

PwS is already commissioned to deliver several new *Planck nominal and full mission* products and science results, namely:

- *Nominal mission*
  - as one of the extraction tools to construct the SZ cluster catalogue of the forthcoming '*Planck Catalogue of Compact Sources*' (PCCS);

  - to create the cosmological sample of SZ clusters, among two other algorithms, and the '*noise maps*' aiming at the characterization of '*Planck's selection function*', with the aim of setting constraints on Ω_m and σ_8;

- *Full mission products*

  - as a multi-spectral extraction tool of the full mission complete compact sources catalogue: point sources and SZ clusters;
  - to improve the characterization of the cluster sample through joint estimation using ancillary data sets (SDSS);
  - and to attempt the detection of the relativistic correction to the non-relativistic thermal SZ effect.

Appendices

A-1 Quadratic Identities

- Let A, B, z, a and b be real scalars. Then

$$A(z-a)^2 + B(z-b)^2 = (A+B)(z-c)^2 + \frac{AB}{A+B}(a-b)^2, \tag{A-1}$$

with

$$c = \frac{1}{A+B}(Aa + Bb).$$

- Let x, a and b be k x 1 vectors and A and B be k x k symmetric matrices such that $(A+B)^{-1}$ exists. Then,

$$(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{a}) + (\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{B}(\boldsymbol{x}-\boldsymbol{b}) = (\boldsymbol{x}-\boldsymbol{c})'\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}(\boldsymbol{x}-\boldsymbol{c}) + (\boldsymbol{a}-\boldsymbol{b})'\boldsymbol{A}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\boldsymbol{B}(\boldsymbol{a}-\boldsymbol{b}) \tag{A-2}$$

where

$$\boldsymbol{c} = \left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\left(\boldsymbol{A}\boldsymbol{a}+\boldsymbol{B}\boldsymbol{b}\right).$$

- Assuming that $\boldsymbol{A}$ and $\boldsymbol{B}$ have inverses then,

$$\boldsymbol{A}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\boldsymbol{B} = \left(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\right)^{-1}. \tag{A-3}$$

For proofs please refer to G. E. Box [1992]

A-2 The ratio $\frac{\Pr(H_1|N_s)}{\Pr(H_0|N_s)}$

If X_0 is Poisson distributed with parameter λ_0,

$$\Pr(X_0 = n|\lambda_0) = \mathrm{e}^{-\lambda_0}\frac{\lambda_0^n}{n!},$$

X_1 with parameter λ_1 and $Y = X_1 + X_0$, then

$$\Pr(X_1 = n|Y = y) = \binom{y}{n}p^n(1-p)^{y-n}, \quad p = \frac{\lambda_1}{\lambda_1 + \lambda_0}.$$

If we draw N_s samples from the mixture Y and all of them must match the hypothesis H_k, then

$$\Pr(H_k|N_s) = \Pr(X_k = N_s|Y = N_s) = \left(\frac{\lambda_k}{\lambda_k + \lambda_0}\right)^{N_s}.$$

Using the above expression twice we can now write the expression for the prior on the models:

$$\frac{\Pr(H_1|N_s)}{\Pr(H_0|N_s)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{N_s}.$$

A result that could hardly surprise us.

A-3 Fisher information matrix symmetries

If the condition for well separated objects holds (condition 3.25), then the parameters referring to each source individually can be separated in the likelihood (equation 3.23). So, if ζ_i represents an arbitrary parameter of source i and ζ_k of source k, then

$$\frac{\partial^2 \ln \mathcal{L}_{H_1}}{\partial\zeta_i\partial\zeta_k} = 0,$$

holds. Being H_0 not dependent on any parameter, we subtract $\ln \mathcal{L}_{H_0}$ from $\ln \mathcal{L}_{H_1}$, just to make the algebra simpler. To prove the property that allow us to split the maxima scrutiny into, at least, two separated searches, one in position space and the other in its

complement, we only need to show that

$$\left\langle \frac{\partial^2 \ln \mathcal{L}_{H_1}}{\partial x \partial \widetilde{x}} \right\rangle_{H_1} = 0,$$

where both variables refer to the same object, hence the reason why we have dropped the subscript, x is one of the position variables and $\widetilde{x}$ is an arbitrary variable belonging to the complement of the position sub-space of the same source. Let us define

$$\widetilde{\psi}(\boldsymbol{a}, -\boldsymbol{\eta}) \equiv A\, \widetilde{\tau}(\boldsymbol{a}, -\boldsymbol{\eta}),$$

where $\widetilde{\tau}(\boldsymbol{a}, -\boldsymbol{\eta})$ is defined in equation 3.22. Replacing the above definition in the likelihood and developing the expression, one gets

$$\left\langle \frac{\partial^2 \ln \mathcal{L}_{H_1}}{\partial x \partial \alpha} \right\rangle_{H_1} = -\mathrm{i}2\pi \sum_{\boldsymbol{\eta}} \eta_x \Omega(\boldsymbol{\eta})\ \widetilde{\psi}(\boldsymbol{a}, \boldsymbol{\eta})\ \frac{\partial \widetilde{\psi}(\boldsymbol{a}, -\boldsymbol{\eta})}{\partial \widetilde{x}}.$$

Explicitly bringing the negative part of the Fourier spectrum into the sum and remembering that if $f(\boldsymbol{x})$ is real then $\widetilde{f}(\boldsymbol{\eta}) = \widetilde{f}^*(-\boldsymbol{\eta})$ (* is the complex conjugate operation), then the above expression now reads

$$= -4\pi \sum_{\boldsymbol{\eta} \geq 0} \eta_x\ \mathrm{Im}\left\{ \Omega(\boldsymbol{\eta})\ \widetilde{\psi}(\boldsymbol{a}, \boldsymbol{\eta})\ \left(\frac{\partial \widetilde{\psi}(\boldsymbol{a}, -\boldsymbol{\eta})}{\partial \widetilde{x}} \right)^* \right\}.$$

$\Omega(\boldsymbol{\eta})$ is by definition a real quantity. Imposing $\tau(\boldsymbol{a}, \boldsymbol{x})$ has symmetry of reflection on both axes is a sufficient condition to guarantee $\widetilde{\tau}(\boldsymbol{a}, \boldsymbol{\eta})$ is real, then all terms of the above sum are identically null and

$$\left\langle \frac{\partial^2 \ln \mathcal{L}_{H_1}}{\partial x \partial \widetilde{x}} \right\rangle_{H_1} = 0,$$

as we were about to demonstrate.

This in turn implies that if

$$\overrightarrow{r} \in \boldsymbol{X} \quad \text{and} \quad \overrightarrow{\tilde{r}} \in \widetilde{\boldsymbol{X}} \quad \text{then} \quad \overrightarrow{r}.\mathcal{I}.\overrightarrow{\tilde{r}} = 0. \tag{A-4}$$

where $\boldsymbol{X}$ is the likelihood position sub-space, $\widetilde{\boldsymbol{X}}$ its complement and $\mathcal{I}$ is the Fisher information matrix. Equation (A-4) states that any direction belonging to position sub-space is conjugated, in the sense of conjugated gradients, with any other outside of that sub-space. This result shows that we do not need to perform a full N-dimensional maximization but at least we can separate the position variables from the rest. For an exhaustive discussion please refer to [Press et al., 1992, ch. 10.6]

A-4 PwS algorithm implementation

PwS operation may be broadly divided in three main steps:

- **Pre-processing**.
  Makes the masks, creates the flat patches geometry and computes the pixel values.
- **Detection/estimation**.
  Performs the detection by creating a list of candidates and respective parameter estimates.
- **Post-processing**.
  Removes repeated detections of the same source across different patches, selects the candidates to be included in the catalogue, formats, and writes it to the media.

A-4.1 PwS step-by-step

A-4.1.1 Pre-processing

- Reads in the data channel maps together with those defining the masks, Galactic and point sources (SZ). Renders auxiliary maps to flag heavily contaminated areas and possible ill-defined pixels.

- Defines the patches' geometry using a gnomonic projection. Patches are squares, usually 256 or 512 pixels wide and the pixel area is $\sim 1.718' \times 1.718'$. The pixel values are computed using a bilinear interpolation. With the mask information from the first step, the pixels flagged as '*not usable*' are filled with values that preserve the statistical properties of the background ('*in-painting*').

A-4.1.2 Detection/estimation

- For each one of the patches repeat the next steps at least twice:
  - Estimate the cross-power spectrum matrix and invert each Fourier mode.
  - Create the likelihood manifold.
  - Scan the likelihood manifold searching for its maxima using a grid. Store the maxima coordinates as initial hints.
  - Sort the putative detections in descending order of likelihood ratio.
  - For each one of the candidates do:
    - Use a Powell minimizer to estimate the optimal values (likelihood maximum estimate) of the source's parameters: position, flux and radius[1].
    - If this is **not** the last iteration:
      - If the SNR of the candidate is above a certain threshold (high) mask it. Go back to the beginning of the processing of this patch and start all over again[2].
      - Subtract the object from the maps. Continue with the next iteration (candidate source).
      - If running in *The Generalized Likelihood Ratio Test* (GLRT) mode. Test the computed SNR ratio value against the threshold for acceptance/rejection. If the detection is accepted then add it to the intermediate catalogue.

[1] Even when doing point source extraction PwS *never* assumes the objects as point like and shapeless. A narrow, but with finite dimensions, gaussian is employed instead.

[2] This operation is restricted to point source detection only.

· If running in $ln(odds)$ (Bayesian) mode.
With the estimates from the previous step predict an initial bounding parameter volume and explore the posterior distribution using a simple nested sampling algorithm. Evaluate the $\ln(odds) \equiv \left[\ln\left(\frac{\Pr(H_1|\boldsymbol{d})}{\Pr(H_0|\boldsymbol{d})}\right)\right]$.
Several sets of priors can be used. With the samples drawn from the posterior compute the best parameter estimates, mode or expected value estimators and the uncertainties on the parameters and the $\ln(odds)$. Always add the detection to the intermediate catalogue. If necessary it will be removed during the post-processing stage.

∗ Subtract the object from the maps and continue with the next candidate.

A-4.1.3 Post-processing

- Map the position coordinates of the detections from the patch pixels back onto sphere coordinates.
- Using the auxiliary mask maps filter the intermediate catalogue removing those detections laying on top of the flagged areas.
- Remove repeated detections from the intermediate catalogue choosing always that one with the largest SNR (GLRT) or $\ln(odds)$ (Bayesian). Generate the catalogue using a SNR cut (GLRT) or probability of a false positive (Bayesian).

A-5 Implementation history

The data analysis philosophy and set of algorithms described in this thesis have not so far been fully implemented in a coded version of PwS. We are working towards this aim, and the release corresponding to the full set of features described here will be PwS v4.0. The versions that have been used in published data analyses so far are v1.5 and v3.1 for the SZ Challenge [Melin et al., 2012], v2.01 for the lower frequency

point sources in the *Planck* ERCSC [Planck Collaboration et al., 2011e] and for all frequency channels in the Compact Source Investigation workshop (CSI) [Rocha, G. et al., 2013], v3.6 in application to SZ cluster detection in the *Planck* ESZ sample [Planck Collaboration et al., 2011f] and to characterise a single cluster parameters in a non-blind exercise [Planck et al., 2012; Planck Collaboration et al., 2011g,i, 2012a,b,c]. It is worth noting that these versions include a pre-processing tool specifically designed to convert data sets distributed within the *Planck* collaboration into the format required by PwS. The main tasks performed by this tool are:

- taking account of the masking and/or flagging of ill-observed pixels and contaminated regions;
- projecting the spherical maps into flat patches[1];
- mapping of coordinates from the sphere into the patches and back;
- removal of multiple detections of the same source in different patches;
- assembly of the output catalogues into the required format.

The existing released versions of PwS differ from what will be available in v4.0 mainly in the limitation to a binary model selection step in determining when to accept a putative source detection and to a non-parametrised frequency spectrum in multi-frequency detection. The latter restriction meant that, while SZ cluster detection could be carried out using all *Planck* frequencies simultaneously, point source detections, in common with the other methods available, were carried out for each frequency channel separately. PwS v4.0 will aim at genuine multi-frequency and indeed multi-model detection, using all the available data simultaneously.

[1]The patch set usually contains about $12,000$ $7.33° \times 7.33°$ flat patches or $3,000$ $14.66° \times 14.66°$.

References

S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. *ARAA*, 49:409–470, September 2011. doi: 10.1146/annurev-astro-081710-102514. 116

AMI Consortium, J. T. L. Zwart, R. W. Barker, P. Biddulph, D. Bly, R. C. Boysen, A. R. Brown, C. Clementson, M. Crofts, T. L. Culverhouse, J. Czeres, R. J. Dace, M. L. Davies, R. D'Alessandro, P. Doherty, and et al. The Arcminute Microkelvin Imager. *MNRAS*, 391:1545–1558, December 2008. doi: 10.1111/j.1365-2966.2008.13953.x. 126

AMI Consortium, C. Rodríguez-Gonzálvez, T. W. Shimwell, M. L. Davies, F. Feroz, T. M. O. Franzen, K. J. B. Grainge, M. P. Hobson, N. Hurley-Walker, A. N. Lasenby, M. Olamaie, and et al. Detailed Sunyaev-Zel'dovich study with AMI of 19 LoCuSS galaxy clusters: masses and temperatures out to the virial radius. *MNRAS*, 425:162–203, September 2012. doi: 10.1111/j.1365-2966.2012.21419.x. 36

J. Anderson. A generalized likelihood ratio test for detecting land mines using multispectral images. *Geoscience and Remote Sensing Letters, IEEE*, 5(3):547 –551, july 2008. ISSN 1545-598X. doi: 10.1109/LGRS.2008.922316. 5

F. Argüeso, E. Salerno, D. Herranz, J. L. Sanz, E. E. Kuruoğlu, and K. Kayabol. A Bayesian technique for the detection of point sources in cosmic microwave background maps. *MNRAS*, 414:410–417, June 2011. doi: 10.1111/j.1365-2966.2011.18398.x. 4, 39

M. Arnaud, G. W. Pratt, R. Piffaretti, H. Böhringer, J. H. Croston, and E. Pointecouteau. The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation. *A&A*, 517:A92, July 2010. doi: 10.1051/0004-6361/200913416. xv, xvi, 81, 82, 84, 108, 115

A. T. Bajkova. A High-Accuracy Method for the Removal of Point Sources from Maps of the Cosmic Microwave Background. *Astronomy Reports*, 49:947–957, December 2005. doi: 10.1134/1.2139810. 74

C. L. Bennett, R. S. Hill, G. Hinshaw, M. R. Nolta, N. Odegard, L. Page, D. N. Spergel, J. L. Weiland, E. L. Wright, M. Halpern, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. Wollack. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Foreground Emission. *ApJS*, 148:97–117, September 2003. doi: 10.1086/377252. 3

E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *A&AS*, 117: 393–404, June 1996. 1, 91

M. Birkinshaw. The Sunyaev-Zel'dovich effect. *Physics Reports*, 310:97–195, March 1999. doi: 10.1016/S0370-1573(98)00080-5. 26, 77, 78, 80, 81, 82

H. Böhringer, P. Schuecker, L. Guzzo, C. A. Collins, W. Voges, S. Schindler, D. M. Neumann, R. G. Cruddace, S. De Grandi, G. Chincarini, A. C. Edge, H. T. MacGillivray, and P. Shaver. The ROSAT-ESO flux limited X-ray (REFLEX) galaxy cluster survey. I. The construction of the cluster sample. *A&A*, 369:826–850, April 2001. doi: 10.1051/0004-6361:20010240. 84, 90

H. Böhringer, P. Schuecker, G. W. Pratt, M. Arnaud, T. J. Ponman, J. H. Croston, S. Borgani, R. G. Bower, U. G. Briel, C. A. Collins, M. Donahue, and et al. The representative XMM-Newton cluster structure survey (REXCESS) of an X-ray luminosity selected galaxy cluster sample. *A&A*, 469:363–377, July 2007. doi: 10.1051/0004-6361:20066740. 84, 90

Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2 edition, February 2004. ISBN 0521819164. 58

Ronald Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill Science/Engineering/Math, 3 edition, June 1999. ISBN 0073039381. 41

L.L. Cam and G.L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer, 2000. ISBN 9780387950365. 34, 88, 112

J. E. Carlstrom, G. P. Holder, and E. D. Reese. Cosmology with the Sunyaev-Zel'dovich Effect. *ARAA*, 40:643–680, 2002. doi: 10.1146/annurev.astro.40.060401.093803. 26, 77, 79, 80

P. Carvalho, G. Rocha, and M. P. Hobson. A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I. *MNRAS*, 393:681–702, March 2009. doi: 10.1111/j.1365-2966.2008.14016.x. 4, 24, 47, 48, 91

P. Carvalho, G. Rocha, M. P. Hobson, and A. Lasenby. Powellsnakes II: a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets. *MNRAS*, 427:13841400, November 2012. doi: doi:10.1111/j.1365-2966.2012.22033.x. 4, 91

A. Challinor and A. Lasenby. Relativistic Corrections to the Sunyaev-Zeldovich Effect. *ApJ*, 499:1, May 1998. doi: 10.1086/305623. 80

Leslie Collins, Ping Gao, Student Member, Lawrence Carin, and Senior Member. An improved bayesian decision theoretic approach for land mine detection. *IEEE Trans. Geosci. Remote Sensing*, 37:811–819, 1999. 5

G. de Zotti, R. Ricci, D. Mesa, L. Silva, P. Mazzotta, L. Toffolatti, and J. González-Nuevo. Predictions for high-frequency radio surveys of extragalactic sources. *A&A*, 431:893–903, March 2005. doi: 10.1051/0004-6361:20042108. 35, 39, 44

G. de Zotti, M. Massardi, M. Negrello, and J. Wall. Radio and millimeter continuum surveys and their astrophysical implications. *A&AR*, 18:1–65, February 2010. doi: 10.1007/s00159-009-0026-0. 41

J. Delabrouille, M. Betoule, J.-B. Melin, M.-A. Miville-Deschênes, J. Gonzalez-Nuevo, M. Le Jeune, G. Castex, G. de Zotti, and et al. The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths. *ArXiv e-prints*, July 2012. 4, 39, 44, 56, 92

A. S. Eddington. On a formula for correcting statistics for the effects of a known error of observation. *MNRAS*, 73:359–360, March 1913. 97

G. Efstathiou. Limitations of Bayesian Evidence applied to cosmology. *MNRAS*, 388: 1314–1320, August 2008. doi: 10.1111/j.1365-2966.2008.13498.x. 55

European Space Agency ESA. PLANCK - MULTIMEDIA GALLERY. http://sci.esa.int/planck/31075-images-and-videos/?farchive_objecttypeid=18,19,22&farchive_objectid=30912, 2009. [Online; accessed 25-September-2013]. xii, 66

F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384:449–463, February 2008. doi: 10.1111/j.1365-2966.2007.12353.x. 4

F. Feroz, M. P. Hobson, and M. Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398:1601–1614, October 2009. doi: 10.1111/j.1365-2966.2009.14548.x. 4, 47

M. Figueiredo, J. Zerubia, and A.K. Jain. *Energy Minimization Methods in Computer Vision and Pattern Recognition: Third International Workshop, EMMCVPR 2001, Sophia Antipolis France, September 3-5, 2001. Proceedings*. Number v. 3 in Lecture Notes in Computer Science. Springer, 2001. ISBN 9783540425236. URL http://books.google.co.uk/books?id=yb8otde21fcC. 73

G. C. Tiao G. E. Box. *Bayesian Inference in Statistical Analisys*. John Wiley & Sons, Wiley-Interscience, 1992. 11, 22, 34, 35, 36, 38, 39, 55, 131, 133

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, 2 edition, July 2003. ISBN 158488388X. 22, 35, 51, 131

J. González-Nuevo, F. Argüeso, M. López-Caniego, L. Toffolatti, J. L. Sanz, P. Vielva, and D. Herranz. The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps. *MNRAS*, 369:1603–1610, July 2006. doi: 10.1111/j.1365-2966.2006.10442.x. 3, 91

H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):pp. 53–60, 1981. ISSN 00361445. URL http://www.jstor.org/stable/2029838. 119

D. Herranz and J. L. Sanz. Matrix Filters for the Detection of Extragalactic Point Sources in Cosmic Microwave Background Images. *IEEE Journal of Selected Topics in Signal Processing*, 2:727–734, November 2008. doi: 10.1109/JSTSP.2008.2005339. 3

D. Herranz, J. L. Sanz, R. B. Barreiro, and E. Martínez-González. Scale-adaptive Filters for the Detection/Separation of Compact Sources. *ApJ*, 580:610–625, November 2002a. doi: 10.1086/342651. 3

D. Herranz, J. L. Sanz, M. P. Hobson, R. B. Barreiro, J. M. Diego, E. Martínez-González, and A. N. Lasenby. Filtering techniques for the detection of Sunyaev-Zel'dovich clusters in multifrequency maps. *MNRAS*, 336:1057–1068, November 2002b. doi: 10.1046/j.1365-8711.2002.05704.x. 2, 33, 82, 85, 92

G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *ApJS*, 180:225–245, February 2009. doi: 10.1088/0067-0049/180/2/225. 39, 44, 121

M. P. Hobson and C. McLachlan. A Bayesian approach to discrete object detection in astronomical data sets. *MNRAS*, 338:765–784, January 2003. doi: 10.1046/j.1365-8711.2003.06094.x. 4, 45

D. W. Hogg. Distance measures in cosmology. *ArXiv Astrophysics e-prints*, May 1999. 81

N. Itoh, Y. Kohyama, and S. Nozawa. Relativistic Corrections to the Sunyaev-Zeldovich Effect for Clusters of Galaxies. *ApJ*, 502:7, July 1998. doi: 10.1086/305876. 80

E. T. Jaynes. *Probability Theory, The logic of science*. Cambridge University Press, 2004. vi, 5, 7, 8, 10, 16, 18, 35, 36, 41, 54, 60, 61, 94, 131

H. Jeffreys. *Theory of Probability (Third Edition)*. Oxford University Press, 1961. 35, 55, 56, 61

S. Bence K. Riley, M. Hobson. *Mathematical Methods for Physics and Engineering (3rd edition): A Comprehensive Guide*. Cambridge University Press, 2006. 15

C. R. Keeton. On statistical uncertainty in nested sampling. *MNRAS*, 414:1418–1426, June 2011. doi: 10.1111/j.1365-2966.2011.18474.x. 20

Joseph-Louis Lagrange. *Oeuvres de Lagrange. Tome 6*. Gauthier-Villars (Paris), 1892. 65

L. F. Lanz, D. Herranz, J. L. Sanz, J. González-Nuevo, and M. López-Caniego. A multifrequency method based on the matched multifilter for the detection of point sources in CMB maps. *MNRAS*, 403:2120–2130, April 2010. doi: 10.1111/j.1365-2966.2010.16260.x. 2

S. M. Leach, J.-F. Cardoso, C. Baccigalupi, R. B. Barreiro, M. Betoule, J. Bobin, A. Bonaldi, J. Delabrouille, G. de Zotti, C. Dickinson, H. K. Eriksen, J. González-Nuevo, F. K. Hansen, D. Herranz, M. Le Jeune, M. López-Caniego, E. Martínez-González, M. Massardi, J.-B. Melin, M.-A. Miville-Deschênes, G. Patanchon, S. Prunet, S. Ricciardi, E. Salerno, J. L. Sanz, J.-L. Starck, F. Stivoli, V. Stolyarov, R. Stompor, and P. Vielva. Component separation methods for the PLANCK mission. *A&A*, 491:597–615, November 2008. doi: 10.1051/0004-6361:200810116. 92

E. V. Linder and R. Miquel. Cosmological Model Selection:. Statistics and Physics. *International Journal of Modern Physics D*, 17:2315–2324, 2008. doi: 10.1142/S0218271808013881. 55

M. López-Caniego, D. Herranz, J. L. Sanz, and R. B. Barreiro. Detection of Point Sources on Two-Dimensional Images Based on Peaks. *EURASIP Journal on Applied Signal Processing*, 2005:2426–2436, August 2005. doi: 10.1155/ASP.2005.2426. 43

M. López-Caniego, D. Herranz, J. González-Nuevo, J. L. Sanz, R. B. Barreiro, P. Vielva, F. Argüeso, and L. Toffolatti. Comparison of filters for the detection of point sources in Planck simulations. *MNRAS*, 370:2047–2063, August 2006. doi: 10.1111/j.1365-2966.2006.10639.x. 3, 91

M. López-Caniego, J. González-Nuevo, D. Herranz, M. Massardi, J. L. Sanz, G. De Zotti, L. Toffolatti, and F. Argüeso. Nonblind Catalog of Extragalactic Point Sources from the Wilkinson Microwave Anisotropy Probe (WMAP) First 3 Year Survey Data. *ApJS*, 170:108–125, May 2007. doi: 10.1086/512678. 2

M. Lundberg. Infrared land mine detection by parametric modeling. In *Proceedings of the Acoustics, Speech, and Signal Processing, 2001. on IEEE International Conference - Volume 05*, ICASSP '01, pages 3157–3160, Washington, DC, USA, 2001. IEEE Computer Society. ISBN 0-7803-7041-4. doi: 10.1109/ICASSP.2001.940328. 5

T. A. Marriage, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, E. S. Battistelli, J. R. Bond, B. Brown, B. Burger, J. Chervenak, S. Das, M. J. Devlin, S. R. Dicker, W. Bertrand Doriese, Dunkley, and et al. The Atacama Cosmology Telescope: Sunyaev-Zel'dovich-Selected Galaxy Clusters at 148 GHz in the 2008 Survey. *ApJ*, 737:61, August 2011. doi: 10.1088/0004-637X/737/2/61. 126

J.-B. Melin, J. G. Bartlett, and J. Delabrouille. Catalog extraction in SZ cluster surveys: a matched filter approach. *A&A*, 459:341–352, November 2006. doi: 10.1051/0004-6361:20065034. 2, 33, 85, 86, 91, 92

J.-B. Melin, N. Aghanim, M. Bartelmann, J. G. Bartlett, M. Betoule, J. Bobin, P. Carvalho, G. Chon, J. Delabrouille, J. M. Diego, D. L. Harrison, D. Herranz, M. Hobson, R. Kneissl, A. N. Lasenby, M. Le Jeune, M. Lopez-Caniego, P. Mazzotta, G. M. Rocha, B. M. Schaefer, J.-L. Starck, J.-C. Waizmann, and D. Yvon. A Comparison of Algorithms for the Construction of SZ Cluster Catalogues. *ArXiv e-prints*, October 2012. 56, 59, 91, 92, 98, 100, 138

A. Mennella, R. C. Butler, A. Curto, F. Cuttaia, R. J. Davis, J. Dick, M. Frailis, S. Galeotta, A. Gregorio, H. Kurki-Suonio, C. R. Lawrence, and et al. Planck early results.

III. First assessment of the Low Frequency Instrument in-flight performance. *A&A*, 536:A3, December 2011. doi: 10.1051/0004-6361/201116480. 65, 121

J.E. Mooney, Z. Ding, and L.S. Riggs. Performance analysis of a glrt in late-time radar target detection - abstract. *Journal of Electromagnetic Waves and Applications*, 13 (10):1339–1341, 1999. doi: doi:10.1163/156939399X00673. 21, 22

P. Mukherjee, D. Parkinson, and A. R. Liddle. A Nested Sampling Algorithm for Cosmological Model Selection. *ApJL*, 638:L51–L54, February 2006. doi: 10.1086/501068. 48

D. Nagai, A. V. Kravtsov, and A. Vikhlinin. Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. *ApJ*, 668:1–14, October 2007. doi: 10.1086/521328. 81, 82, 84

J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:pp. 289–337, 1933. ISSN 02643952. 14

D. O. North. Analisys of the factors which determine signal/noise discrimination in radar (1943). *Proc IRE*, 51:1016–1028, 1963. 2

R. Piffaretti, M. Arnaud, G. W. Pratt, E. Pointecouteau, and J.-B. Melin. The MCXC: a meta-catalogue of x-ray detected clusters of galaxies. *A&A*, 534:A109, October 2011. doi: 10.1051/0004-6361/201015377. 90

Planck, AMI Collaborations, :, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, E. Battaner, R. Battye, K. Benabed, A. Benoît, J.-P. Bernard, and et al. Planck Intermediate Results II: Comparison of Sunyaev-Zeldovich measurements from Planck and from the Arcminute Microkelvin Imager for 11 galaxy clusters. *ArXiv e-prints*, April 2012. vii, xiv, xv, 36, 51, 71, 86, 88, 89, 91, 99, 100, 108, 139

Planck Collaboration. Planck blue book, 1999. http://www.rssd.esa.int/SA/PLANCK/docs/Bluebook-ESA-SCI(2005)1_V2.pdf. 65, 79

Planck Collaboration, J. Aatrokoski, P. A. R. Ade, N. Aghanim, H. D. Aller, M. F. Aller, E. Angelakis, M. Arnaud, M. Ashdown, J. Aumont, and et al. Planck early results. XV. Spectral energy distributions and radio continuum spectra of northern extragalactic radio sources. *A&A*, 536:A15, December 2011a. doi: 10.1051/0004-6361/201116466. 41

Planck Collaboration, P. A. R. Ade, N. Aghanim, E. Angelakis, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, and et al. Planck early results. XIV. ERCSC validation and extreme radio sources . *A&A*, 536:A14, December 2011b. doi: 10.1051/0004-6361/201116475. 90

Planck Collaboration, P. A. R. Ade, N. Aghanim, F. Argüeso, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, and et al. Planck early results. XIII. Statistical properties of extragalactic radio sources in the Planck Early Release Compact Source Catalogue. *A&A*, 536:A13, December 2011c. doi: 10.1051/0004-6361/201116471. 26, 41, 90

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, M. Baker, A. Balbi, A. J. Banday, and et al. Planck early results. I. The Planck mission. *A&A*, 536:A1, December 2011d. doi: 10.1051/0004-6361/201116464. vii, 3, 64, 65

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, and et al. Planck early results. VII. The Early Release Compact Source Catalogue. *A&A*, 536:A7, December 2011e. doi: 10.1051/0004-6361/201116474. vii, xiii, 10, 23, 62, 71, 76, 77, 87, 90, 139

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, and et al. Planck early results. VIII. The all-sky early Sunyaev-Zeldovich cluster sample. *A&A*, 536:A8, December 2011f. doi: 10.1051/0004-6361/201116459. vii, xiv, 3, 51, 57, 60, 71, 81, 84, 87, 88, 89, 90, 91, 120, 121, 126, 128, 139

Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, J. G. Bartlett,

E. Battaner, K. Benabed, A. Benoît, and J.-P. et al. Bernard. Planck early results. XXVI. Detection with Planck and confirmation by XMM-Newton of PLCK G266.6-27.3, an exceptionally X-ray luminous and massive galaxy cluster at z ˜ 1. *A&A*, 536: A26, December 2011g. doi: 10.1051/0004-6361/201117430. vii, 71, 90, 139

Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, M. Bartelmann, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, J.-P. Bernard, and et al. Planck early results. X. Statistical analysis of Sunyaev-Zeldovich scaling relations for X-ray galaxy clusters. *A&A*, 536:A10, December 2011h. doi: 10.1051/0004-6361/201116457. xiv, 83

Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, M. Bartelmann, and et al. Planck early results. IX. XMM-Newton follow-up for validation of Planck cluster candidates. *A&A*, 536:A9, December 2011i. doi: 10.1051/0004-6361/201116460. vii, 71, 90, 139

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, and et al. Planck Intermediate Results. V. Pressure profiles of galaxy clusters from the Sunyaev-Zeldovich effect. *ArXiv e-prints*, July 2012a. vii, 71, 85, 86, 139

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, J.-P. Bernard, and et al. Planck Intermediate Results. IV. The XMM-Newton validation programme for new Planck galaxy clusters. *ArXiv e-prints*, May 2012b. vii, 71, 90, 139

Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, and et al. Planck intermediate results. I. Further validation of new Planck clusters with XMM-Newton. *A&A*, 543:A102, July 2012c. doi: 10.1051/0004-6361/201118731. vii, 71, 86, 90, 139

Planck HFI Core Team, P. A. R. Ade, N. Aghanim, R. Ansari, M. Arnaud, M. Ashdown, J. Aumont, and et al. Planck early results. VI. The High Frequency Instru-

ment data processing. *A&A*, 536:A6, December 2011. doi: 10.1051/0004-6361/201116462. xii, xiii, 65, 67, 68, 69, 70, 72, 80, 121

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992. ISBN 0-521-43108-5. 48, 54, 85, 136

C. L. Reichardt, B. Stalder, L. E. Bleem, T. E. Montroy, K. A. Aird, K. Andersson, R. Armstrong, M. L. N. Ashby, M. Bautz, M. Bayliss, G. Bazin, B. A. Benson, M. Brodwin, J. E. Carlstrom, and et al. Galaxy clusters discovered via the Sunyaev-Zel'dovich effect in the first 720 square degrees of the South Pole Telescope survey. *ArXiv e-prints*, March 2012. 60, 124, 126

M. Reinecke, K. Dolag, R. Hell, M. Bartelmann, and T. A. Enßlin. A simulation pipeline for the Planck mission. *A&A*, 445:373–373, January 2006. doi: 10.1051/0004-6361:20053413. 4, 92, 94

Rocha, G. et al. A comparison of algorithms for detecting compact sources in Planck data - I. [in preparation], 2013. xiv, xv, 91, 94, 96, 102, 104, 105, 106, 139

J. L. Sanz, D. Herranz, and E. Martínez-Gónzalez. Optimal Detection of Sources on a Homogeneous and Isotropic Background. *ApJ*, 552:484–492, May 2001. doi: 10.1086/320550. 3

J. Sayers, N. G. Czakon, C. Bridge, S. R. Golwala, P. M. Koch, K.-Y. Lin, S. M. Molnar, and K. Umetsu. Bolocam Observations of Two Unconfirmed Galaxy Cluster Candidates from the Planck Early Sunyaev-Zel'dovich Sample. *ApJL*, 749:L15, April 2012. doi: 10.1088/2041-8205/749/1/L15. 86

B. Schäfer and M. Bartelmann. Detecting Sunyaev-Zel'dovich clusters with Planck - III. Properties of the expected SZ cluster sample. *MNRAS*, 377:253–268, May 2007. doi: 10.1111/j.1365-2966.2007.11596.x. 3

B. M. Schäfer, C. Pfrommer, R. M. Hell, and M. Bartelmann. Detecting Sunyaev-Zel'dovich clusters with Planck - II. Foreground components and optimized filtering schemes. *MNRAS*, 370:1713–1736, August 2006. doi: 10.1111/j.1365-2966.2006.10622.x. 80

S. Serjeant and D. Harrison. The local submillimetre luminosity functions and predictions from Spitzer to Herschel. *MNRAS*, 356:192–204, January 2005. doi: 10.1111/j.1365-2966.2004.08440.x. 27

D. S. Sivia and J. Skilling. *Data Analysis–A Bayesian Tutorial*. Oxford Science Publications, 2nd edition, 2006. 10, 21, 47, 53

R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173, November 1972a. 2

R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173, November 1972b. 77

L. Svensson and M. Lundberg. Land mine detection in rotationally invariant noise fields. In *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, pages 170 –173, 2001. doi: 10.1109/SSP.2001.955249. 5

J. A. Tauber, N. Mandolesi, J.-L. Puget, T. Banos, M. Bersanelli, F. R. Bouchet, R. C. Butler, J. Charra, G. Crone, J. Dodsworth, and et al. Planck pre-launch status: The Planck mission. *A&A*, 520:A1, September 2010. doi: 10.1051/0004-6361/200912983. 65

M. Tegmark and A. de Oliveira-Costa. Removing Point Sources from Cosmic Microwave Background Maps. *ApJL*, 500:L83, June 1998. doi: 10.1086/311410. 2

Harry L. Van Trees. *Detection, Estimation, and Modulation theory. Part I*. John Wiley & Sons, inc, 2001. 2, 10, 12, 14, 21, 33, 112, 113

R. Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49:71–104, March 2008. doi: 10.1080/00107510802066753. 52, 54, 55

E. M. Waldram, R. C. Bolton, G. G. Pooley, and J. M. Riley. Some estimates of the source counts at Planck Surveyor frequencies, using the 9C survey at 15 GHz. *MNRAS*, 379:1442–1452, August 2007. doi: 10.1111/j.1365-2966.2007.12025.x. 26

N. Yoshida, J. Colberg, S. D. M. White, A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, A. Jenkins, C. S. Frenk, F. R. Pearce, G. Efstathiou, J. A. Peacock, and P. A. Thomas. Simulations of deep pencil-beam redshift surveys. *MNRAS*, 325: 803–816, August 2001. doi: 10.1046/j.1365-8711.2001.04499.x. 39, 44

A. Zacchei, D. Maino, C. Baccigalupi, M. Bersanelli, A. Bonaldi, L. Bonavera, C. Burigana, R. C. Butler, F. Cuttaia, G. de Zotti, J. Dick, and et al. Planck early results. V. The Low Frequency Instrument data processing. *A&A*, 536:A5, December 2011. doi: 10.1051/0004-6361/201116484. 65, 121