

# Cosmology with wide-field SZ cluster surveys: selection and systematic effects

J. B. Juin<sup>1,2</sup>, D. Yvon<sup>1</sup>, A. Réfrégier<sup>1</sup>, and C. Yèche<sup>1</sup>

<sup>1</sup> CEA/DSM/DAPNIA, CE Saclay, Bât 141, 91191 Gif-sur-Yvette Cedex, France  
e-mail: yvon@hep.saclay cea.fr

<sup>2</sup> Universidad de Concepción, Astronomy group, Casilla 160-c, Concepción, Chile

Received 12 December 2005 / Accepted 15 January 2007

## ABSTRACT

The cosmological potential of large-scale structure observations for cosmology have been extensively discussed in the literature. In particular, it has recently been shown how Sunyaev-Zel'dovich (SZ) cluster surveys can be used to constrain dark energy parameters. In this paper, we study whether selection and systematic effects will limit future wide-field SZ surveys from achieving their cosmological potential. For this purpose, we use a sky simulation and an SZ-cluster detection software presented in Pires et al. (2006, A&A, 455, 741), using the future Olimpo survey as a concrete example. We show that the SZ-cluster selection function and contamination of SZ-cluster catalogues are more complex than is usually assumed. In particular, the simulated field-to-field detected cluster counts variance can be a factor 3 larger than the expected Poisson fluctuations. We also study the impact of missing redshift information and of the uncertainty of the scaling relations for low mass clusters. We quantify, through hypothesis tests, how near-future SZ experiments can be used to discriminate between different structure formation models. Using a maximum likelihood approach, we then study the impact of these systematics on the joint measurement of cosmological models and of cluster scaling relations.

**Key words.** cosmology: cosmological parameters – large-scale structure of Universe – cosmic microwave background – galaxies: clusters: general

## 1. Introduction

In the next few years, a new generation of dedicated instruments based on large-array bolometer cameras (APEX<sup>1</sup>, ACT<sup>2</sup>, SPT<sup>3</sup>, BOLOCAM<sup>4</sup>, OLIMPO (Masi et al. 2003), and improved interferometers (AMI<sup>5</sup>, AMiBA<sup>6</sup>, SZA<sup>7</sup>) will provide large amount of information on cosmic structure formation and evolution, and thus on cosmological models. The Planck satellite (Lamarre et al. 2004), to be launched in 2007, will provide a full-sky catalogue of galaxy clusters detected by their Sunyaev-Zel'dovich (SZ) signal (Sunyaev & Zel'dovich 1970, 1972). The potential of SZ observations results from the properties of the SZ effect: the lack of surface dimming and the “clean” measurement of thermal energy of the cluster gas, should afford a measure of the cluster mass function up to high redshift, with reduced systematics when combined with X-ray observations and weak lensing surveys. The distribution of cluster abundance with redshift is sensitive to the cosmological parameters  $\sigma_8$  and  $\Omega_M$ , and also to a lower extend to  $\Omega_\Lambda$  and the dark energy equation-of-state (Barbosa et al. 1996; Oukbir & Blanchard 1997; Haiman et al. 2001). Battye & Weller (2003) studied the dependence of these cosmological constraints on large scale structure formation and gas cluster physics models. Melin et al. (2005) presented a first study of the selection function of large SZ-cluster survey.

In this paper, we use simulations of the sky and of an SZ experiment, along with a recent cluster detection pipeline presented in Pires et al. (2006), to simulate future large-array bolometer observations and cluster detections. We first discuss photometric issues, and then present a detected cluster catalogue, with its selection function, and purity curves. Those are found to be complex. In particular, the contamination of the cluster catalogue is quantified as a function of cluster brightness. We also compute the count variance in the observed catalogues. We then assume that the observations and cluster detection can be statistically described by an observation model. This observation model allows us to transform a semi-analytic cosmology-motivated cluster count functions,  $\frac{dn}{dz dY(z,Y)}$ , into a set of probability density functions (pdf) of the detected cluster observed parameters  $N_{\text{Obs}}$ ,  $Y_{\text{Obs}}$ ,  $z_{\text{Obs}}$  and of the contaminants  $N_{\text{Cont}}$  and  $Y_{\text{Cont}}$ , where  $Y$  is the Compton parameter integrated over the cluster angular size, and  $z$  the redshift. This observation model is found to be accurate enough given the statistics of the upcoming SZ cluster surveys and to be computationally very effective. Based on this model, we then discuss our ability to constraint cosmology assumptions and parameters. We show, using an hypothesis test method, how future SZ surveys will make it possible to distinguish between several mass functions. We quantify the constraint that such an experiment would place on the effective “heating” parameter  $T_*$ , using the cosmological parameters values measured by WMAP, to break degeneracies. We conclude by showing how any conclusions on cosmological parameters are sensitive to inaccuracies in the observation model.

Future large-array bolometer surveys share some common features. They observe the sky in several frequency bands to facilitate astronomical source separation. They use large

<sup>1</sup> <http://bolo.berkeley.edu/apexsz>

<sup>2</sup> <http://www.hep.upenn.edu/~angelica/act/act.html>

<sup>3</sup> <http://spt.uchicago.edu/>

<sup>4</sup> [http://astro.caltech.edu/~lgg/bolocam\\_front.htm](http://astro.caltech.edu/~lgg/bolocam_front.htm)

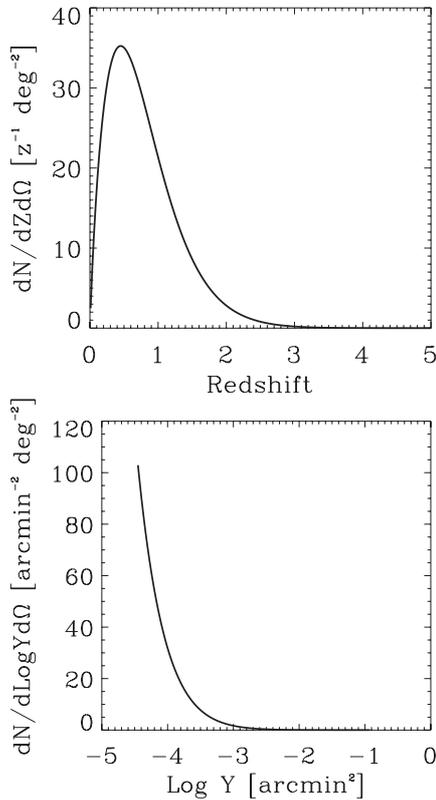
<sup>5</sup> <http://www.mrao.cam.ac.uk/telescopes/ami/index.html>

<sup>6</sup> <http://www.asiaa.sinica.edu.tw/amiba>

<sup>7</sup> <http://astro.uchicago.edu/sze>

**Table 1.** Cosmological and gas physics parameters used in the simulations. Densities relative to critical density are labelled  $\Omega$ .  $\Omega_{\text{tot}}$  is the density of universe, all components included,  $\Omega_b$  is the baryon density,  $\Omega_\Lambda$  the vacuum energy density,  $\Omega_{\text{DM}}$  the dark matter density.  $h$  is the reduced Hubble constant,  $n_s$  the spectral index of primordial density power spectrum,  $\sigma_8$  the rms density fluctuations in spheres of 8 Mpc diameter,  $f_{\text{mass}}$  the mass function used in cluster abundance computations,  $T_\star$  the cluster mass-temperature normalisation factor, and  $f_g$  the cluster gas mass fraction. The double vertical line distinguishes the primordial cosmological parameters, from the ingredients of the structure formation semi-analytic model.

$\Omega_{\text{tot}}$	$\Omega_b$	$\Omega_\Lambda$	$\Omega_{\text{DM}}$	$h$	$n_s$	$\sigma_8$	$f_{\text{mass}}$	$T_\star$	$f_g$
1	0.05	0.7	0.25	0.7	1	0.9	S. & T.	1.9	$0.9 \Omega_b / \Omega_m$



**Fig. 1.** Generated cluster distribution as a function of redshift and integrated Compton flux.

bolometer matrices to maximise redundancy of observation on the sky and speed up field coverage. They are ambitious in terms of mission length, given the technology. In this paper, we mostly use the Olimpo project as a concrete example of an upcoming bolometer-based large SZ survey. The same methods should be easily adapted for the other large-array bolometers surveys.

## 2. Montecarlo simulations

In the following, we use the sky simulation software, the instrument model and the cluster detection pipeline described in (Pires et al. 2006). We briefly summarise it here for convenience and point out the minor differences when necessary. We simulated four astrophysical contributions to the sky map: primordial CMB anisotropies (excluding the dipole), bright infrared galaxies as observed by SCUBA (Borys et al. 2003), the infrared emission of the Galaxy and SZ-clusters. All simulations use a cosmological model with parameters shown in Table 1.

Figure 1 shows the distribution of the generated cluster as a function of redshift and integrated Compton flux.

The frequency dependence of the bright infrared sources and of galactic dust are described by a grey-body spectrum. The spectral index of bright infrared sources is generated randomly

**Table 2.** Foreseen experimental features of the Olimpo balloon bolometer project.

Observation bands $\nu$ [GHz]	143	217	385	600
Bandwidth $\Delta\nu$ [GHz]	30	30	30	30
Bolometer Number	19	37	37	37
Beam Width, $FWHM$ [arcmin]	3	2	2	2
Noise level [ $\mu K_{\text{CMB}} \text{s}^{1/2}$ ]	150	200	500	5000

for each sources between 1.5 and 2. Table 2 provide the noise level and the  $FWHM$  of antenna lobe at each frequencies, which is assumed to be Gaussian. Bandwidth filters are assumed to have a top hat response.

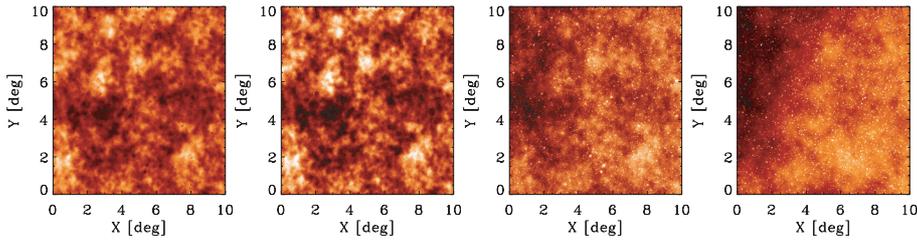
Figure 2 shows the “observed” maps, simulated using the Olimpo parameters. We then apply an Independent Component Analysis method named JADE (Cardoso 1999) on our map, after a wavelet transform. JADE separates the SZ signal from the other astrophysical sources effectively, as long as the noise level is kept low enough, and provides a noisy SZ map. Unlike in Pires et al. (2006), we convolve the noisy SZ map with a Gaussian of user chosen width: this turned out to minimise photometry biases at low brightness. Then we use the SExtractor software (Bertin & Arnouts 1996) to detect bright sources. Detected sources can be reliably associated with simulated clusters and are labelled as “true” clusters. This allows us to compute the selection function and the photometric accuracy of our simulated observations. Detections that are not associated with simulated clusters are identified as contaminants. We can then calculate the purity of our recovered sample and the brightness distribution of the contaminants.

### 2.1. Photometry

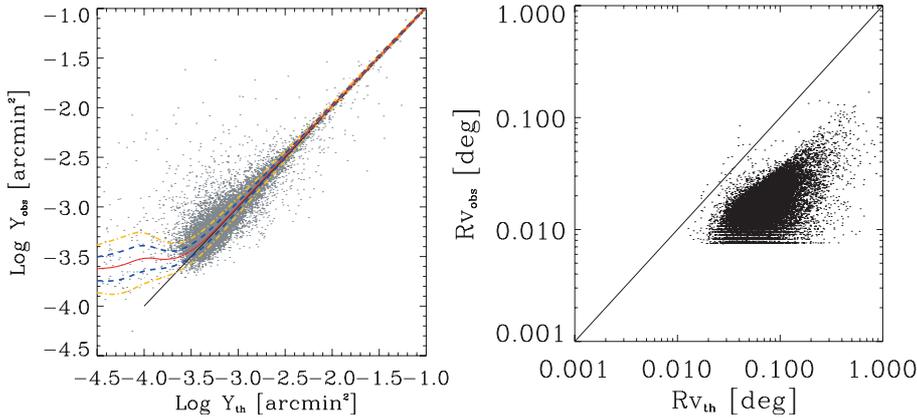
Selecting the sources associated with simulated clusters, following Melin et al. (2005), we plot in Fig. 3 the observed cluster flux  $Y_{\text{Obs}}$  versus the true simulated flux  $Y_{\text{th}}$ , and derive our photometry model, i.e. the probability density function  $\text{pdf}(Y_{\text{Obs}}|Y_{\text{th}})$ . The observed flux is overestimated at low brightness due to the Malmquist-Eddington bias (Malmquist 1920). Our first attempt for a statistical model reproduces very well the simulated photometric behaviour, except for the (small) non-Gaussian tails.

### 2.2. Cluster size reconstruction

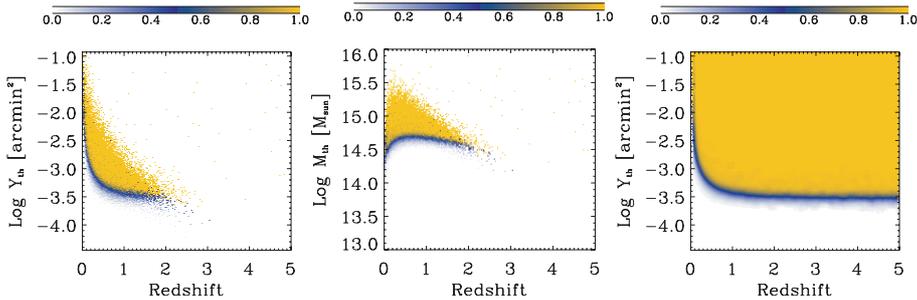
One way to infer the redshift of a cluster would be to measure its virial radius. Figure 3 plots the reconstructed clusters virial radius versus their true (simulated) virial radius. One can hardly see a significant correlation between the simulated virial radius and the observed virial radius. This is the reason why we decided to neglect this information in the following work.



**Fig. 2.** Simulations of sky maps, as observed by a large-array bolometer experiment. For these simulations, we used the Olimpo experiment model. *From left to right*, 143, 217, 385, and 600 GHz bands are shown. In the two lower frequencies band, CMB primordial anisotropies are the dominant features. At higher frequencies bands, bright Infrared galaxies and Galactic dust become dominant. The SZ cluster signal is sub-dominant at all frequencies.



**Fig. 3.** *Left*: cluster reconstructed flux versus the true simulated flux, and our photometry model contours. Dashed lines are the one sigma error and dash-dotted are the 2 sigmas errors; the continuous line is the mean. 20 cumulative Monte-Carlo simulations were used for this plot. *Right*: SZ cluster reconstructed virial size versus true simulated virial size. Although the normalization is not correct, a small correlation is visible for large clusters.



**Fig. 4.** Completeness as a function of redshift for flux (*left*) and mass (*middle*), as simulated from a semi-analytic large scale structure and cosmology model. We used design parameters of the Olimpo project to model observation performance. *Right*: modelled selection function after extended simulations.

### 2.3. Completeness

From the true cluster catalogue, we computed the cluster detection probability as a function of cluster integrated fluxes, redshifts and masses. Figure 4 left and middle, show the results. We see that a selection function can not be taken as a simple cut in total flux, nor in mass. We also notice that clusters at large redshift are detected, even though very few are predicted by the cosmological model. To quantify the selection function at high redshift, we therefore introduced “by hand” in our simulated map, 10% additional high- $z$  clusters, randomly generated in the guessed  $Y$ -threshold area ( $3.5 \times 10^{-5} < Y < 10^{-3}$  and  $1.2 < z < 5$ ). We averaged 100 Monte-Carlos and we obtained the completeness map plotted at Fig. 4 right. The selection function reduces to a simple  $Y$  sensitivity curve at large redshift, when cluster sizes become smaller than angular resolution. But at redshift below 1, where we expect to detect most of the clusters, the completeness curve is strongly distorted toward high Compton flux. For convenience, we provide in Annex 6 the tabulated values of Olimpo selection function versus cluster mass and redshift.

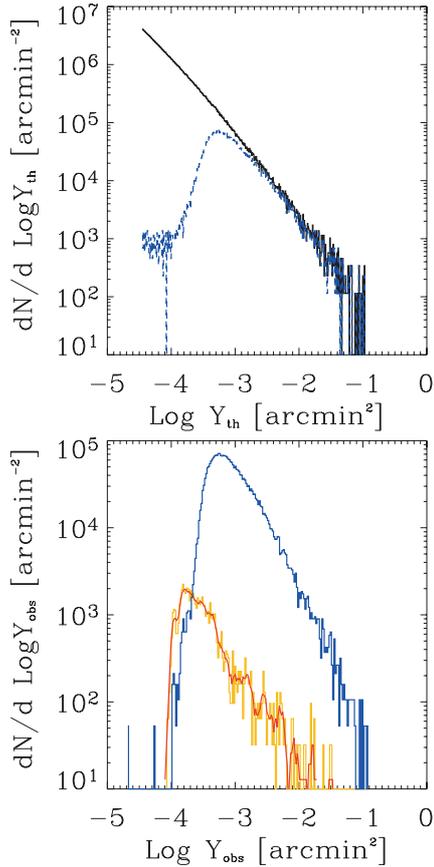
### 2.4. Purity and contamination

Future SZ-cluster experiments won’t be able to easily sort the contamination from the true clusters. Our evaluation of the

observed flux distribution of contaminants is done by selecting sources that are not associated with simulated clusters, and is shown at Fig. 5. The contamination histogram provides the red curve which we use as our modelled flux distribution of contamination. An integration over histograms shown in Fig. 5 lead to a sample purity value of 95%, tuned by choosing the detection threshold.

### 2.5. Sources counts

Cluster counts provides powerful information for large scale structure physics and cosmology. If the counts are dominated by field-to-field variations, one would expect them to follow a Poisson distribution. We found that this might not be true. Figure 6 shows the histograms of cluster, contamination and source counts for 100 simulated fields. Overprinted is the fit to simulation data. By computing the peak  $FWHM$ , we note a factor of 3 excess widths, relative to Poisson’s distribution expectation. The issue is related to the flux calibration of the observations. Astronomical observations are usually calibrated on reference objects on the sky. This is what we have done with these simulations: we selected the 100 brightest true clusters in the recovered map, and fitted a single scaling coefficient so that their recovered flux best match the simulated flux. It turns out that doing so, we introduce a random error (5%) on our qualification



**Fig. 5.** We ran 100 monte-carlo on 400 deg<sup>2</sup>. *Left:* the black curve is the histogram of generated cluster flux, compared to the blue (dashed line) histogram of true cluster detection. *Right:* the blue (dashed line) histogram is the true cluster observed flux. The flux distribution of the contamination is plotted in (light line) orange. The (thin line) red curve is our modelled flux distribution of contamination.

and thus on the  $Y$  threshold applied when selecting clusters, that lead to the observed enlarged count variance. Assuming now the (thermal) SZ frequency spectrum known, (a rigorous approximation for non relativistic cluster gas) and *perfect knowledge of the instrumentation*, we can on simulated data, compute the true calibration factor including the cluster extraction software chain. The count variance recovered using this calibration now closely match the Poisson distribution. Abnormal number of detections for some data set (i.e., the tails of the distribution) are also significantly reduced. Thus one of the challenge of future experiments will be to optimize their  $Y$  calibration, since as we will show at paragraph 4.4.1, enlarged count variance dramatically deteriorates constraints on cosmological parameters. In the following we will use both the nominal Poisson distribution and fits to the degraded simulated cluster and contamination counts (red curves in Fig. 6) to construct our observation models and support discussion.

### 3. Observation model

Our first goal in building an observation model is to identify and understand systematic effects in large-array bolometer surveys, relevant to cluster detection and cosmology. The second is to avoid to run a full Monte-Carlo chain to generate source catalogues observed by SZ-survey, a time consuming step in an analysis software that limits the number of possible iterations

in partice. This is a strong assumption, that we checked up to the precision of the statistical uncertainties of upcoming surveys.

#### 3.1. Observation model ingredients

Semi-analytic LSS model provides the expected number  $N_{\text{clusTh}}$  of cluster of flux above a chosen threshold  $Y_{\text{Thres}}$  and the cluster probability density function  $\text{pdf}(Y_{\text{th}}, z_{\text{th}})$ . The observation model includes:

- a 2D selection function  $\text{Sel}(Y_{\text{th}}, z_{\text{th}})$  giving the probability of a cluster of flux  $Y_{\text{th}}$  and redshift  $z_{\text{th}}$  to be detected. This 2D function is calibrated on simulation as described at paragraph 2.3. The expected number of true detected cluster can then be computed by integration of

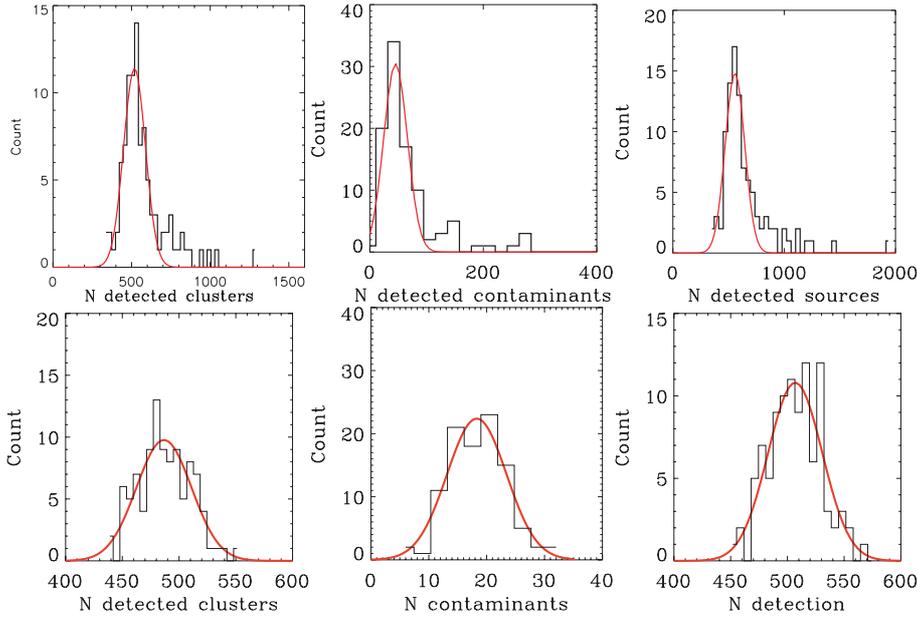
$$N_{\text{obsTh}} = \int N_{\text{clusTh}} \cdot \text{pdf}(Y_{\text{th}}, z_{\text{th}}) \cdot \text{Sel}(Y_{\text{th}}, z_{\text{th}}) dY_{\text{th}} dz_{\text{th}}; \quad (1)$$

- detection count variance model. In order be able to take into account the excess on true cluster counts described above, we introduce the pdf of observing  $N_{\text{obs}}$  clusters, given  $N_{\text{obsTh}}$ ,  $\text{pdf}(N_{\text{obs}} | N_{\text{obsTh}})$ . We simply assume that it follows a Gaussian law, with a width  $\sigma = A^{\text{clus}} \sqrt{N_{\text{obsTh}}}$ , with  $A^{\text{clus}}$  to either the value 3.0 or 1, depending of calibration methods as explained at paragraph 2.5;
- a photometric model  $\text{pdf}(Y_{\text{obs}}, z_{\text{obs}} | Y_{\text{th}}, z_{\text{th}})$  that provides the pdf of observing  $Y_{\text{obs}}$  and  $z_{\text{obs}}$  given  $Y_{\text{th}}$  and  $z_{\text{th}}$ ;
- a contamination model: we assume that contamination are driven by the foreground models and that these do not depend on cosmological parameters. The pdf of the observed flux of contaminants  $\text{pdf}(Y_{\text{Cont}})$  is deduced from the Monte-Carlo simulations of paragraph 2.4 after normalisation. The expectation of the contaminant counts and its variance are taken to be constant and fitted from the distribution shown in Fig. 6;
- an error model on the redshift determinations, when such a complementary measurement is available:  $\text{pdf}(z_{\text{obs}} | z_{\text{th}})$ .

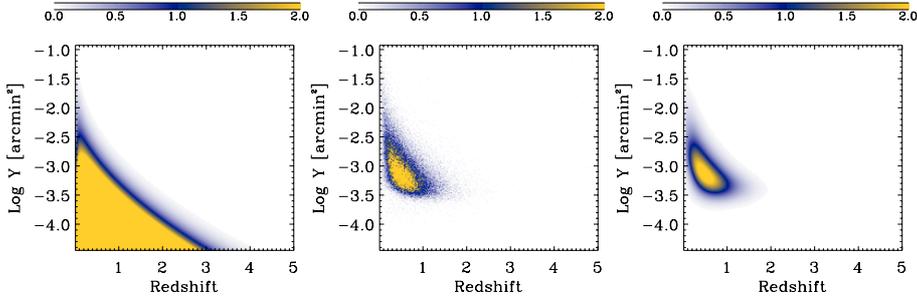
#### 3.2. Simplifying assumptions

The above components of the observation model have been derived from the simulations, given instruments parameters and for the concordance cosmological model. They have been shown to be very sensitive to experiment properties such as noise level, number of observation bands, etc. When constraining cosmological parameters, these experimental effects are expected to be under control. On the other hand, we assume that the observation model is *not sensitive* to the cosmological parameters, when these are *reasonably* close to our reference model. This assumption is strong and not obvious, since in large-array bolometers survey, the contribution of source confusion to the photometric noise may not be negligible. We checked the validity of this assumption, all other parameters being kept constant, by changing the cluster map density by a factor 1.5 and 0.75. Both recovered observation models were compatible with the above model, except for a small increase in the width of the photometry curve (paragraph 2.1) in the large density option. We assumed this to be acceptable since, would such a dramatic discrepancy of cluster density be observed, we would recalibrate our observation model on representative simulations.

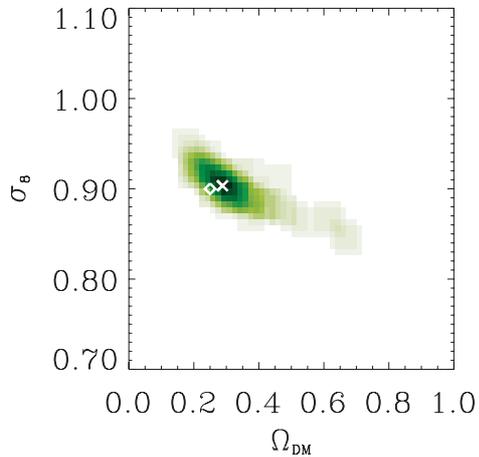
Thus given a cosmological model and observations' model, we derive a set probability density function describing our observations:  $\text{pdf}(N_{\text{Obs}})$ ,  $\text{pdf}(Y_{\text{obs}}, z_{\text{obs}})$ ,  $\text{pdf}(Y_{\text{Cont}})$ ,  $\text{pdf}(N_{\text{Cont}})$ . Figure 7 shows the distributions of flux and redshifts of detected



**Fig. 6.** From left to right, first line: true clusters, contamination and sources counts histograms for 100 simulations. Red curve fits are used in the following observations' model. Second line show same results when assuming a perfect calibration of SZ cluster observations.



**Fig. 7.** Left: the cluster distribution generated by simulations,  $\frac{dN^{\text{clus}}}{dz_{\text{th}} dY_{\text{th}}}(z_{\text{th}}, Y_{\text{th}})$ , and the observed cluster distribution,  $\frac{dN^{\text{clus}}_{\text{obs}}}{dz_{\text{th}} dY_{\text{th}}}(z_{\text{th}}, Y_{\text{th}})$ , for 100 averaged Monte-Carlo simulations (middle), and from the observation model (right). The axes are the integrated flux  $Y$  in arcmin<sup>2</sup> versus redshift.



**Fig. 8.** Probability density of the recovered cosmological parameters  $\sigma_8$  and  $\Omega_{\text{DM}}$ , for 100 full Monte-Carlo simulations and using the observation model in the Extended Likelihood computations. Diamond is the model used at the input of simulations. Cross, is the maximum of occurrence of reconstructed parameters.

clusters, generated by full Monte-Carlo and from our observation model. Those are remarkably similar, thus confirming the validity of our observation model.

We also tested whether the use of the observation model would bias the cosmological parameter estimation. For this purpose, we computed 100 full Monte-Carlo source catalogues. For each catalogue, we computed the cosmological parameters using our observations' model. Figure 8 show the surface density of

two of the cosmological parameters  $\sigma_8$  and  $\Omega_{\text{M}}$  mostly relevant for this study, fitted by the Extended Likelihood method (EL), presented at paragraph 4.1. We observe that the input cosmological parameters are well within the 68% CL contour of the distribution. Thus, the bias induced by the observation model is small compared to the statistical error of the observations.

We conclude that the use of an observation model is legitimate given the accuracy of upcoming experiments. This observation model will be improved: taking into account non Gaussian tails in our photometry model is the main improvement foreseen. In the following, all source catalogues have been generated using the observation model.

#### 4. Cosmological implications

The main physics goals of large SZ-cluster surveys are to learn more about cluster gas physics, large scale structure, and cosmology. These physical models are parametrised, and involve assumptions that can be tested by future SZ cluster surveys. In the following, we first present statistical tools and results testing the compatibility of our data, with a parametrised model family in paragraph *hypothesis tests*. Then we show how SZ cluster data can constraint the mass temperature normalisation factor  $T_*$ , using a classical *parameter estimation* method. Assuming then  $T_*$  known, we explore the potential of SZ cluster survey, for constraining cosmological parameters  $\Omega_{\text{M}}$  and  $\sigma_8$ , assuming all other parameters known. We conclude by showing the effect on cosmological parameters of oversimplifying features of the observations' model. In the following we assumed we have available a catalogue of observed sources corresponding to a nominal

Olimpo scientific flight: 490 sources observed over 300 square degrees.

#### 4.1. Extended likelihood

The tool for all the following statistical tests is the so called *extended likelihood* of the cosmological parameters  $\mathbf{C}$ , given the experimental source catalog  $\mathbf{S}$ :  $L(\mathbf{C}|\mathbf{S})$ .

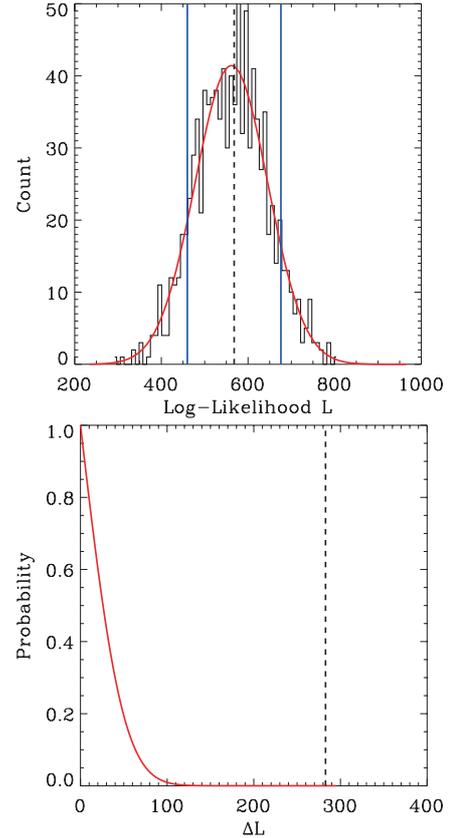
$$L(\mathbf{C}|\mathbf{S}) = \frac{dP}{dN^{\text{Sour}}} (N^{\text{Sour}}; \mathbf{C}) \prod_{i=1}^{N^{\text{Clus}}} \text{pdf}(z_i^{\text{Clus}}, Y_i^{\text{Clus}}; \mathbf{C}) \prod_{i=1}^{N_1^{\text{Sour}}} \text{pdf}(Y_i^{\text{Sour}}; \mathbf{C})$$

with  $N^{\text{Sour}} = N^{\text{Clus}} + N_1^{\text{Sour}}$ .

The likelihood incorporates three kinds of information available in the data. The first factor is the probability of observing  $N^{\text{Sour}}$  sources given the cosmological parameters  $\mathbf{C}$ . The second factor is the probability of observing a cluster with a flux  $Y$  and at redshift  $z$  (using follow-up observations). We assume that the follow-up observation established whether the source is a cluster of galaxies, or a false detection. In the latter case, this source is excluded from the likelihood, except from the first factor. The third factor is the probability of observing a source of flux  $Y$ , when no follow-up observations were available. In this case, we do not know whether this source is a SZ-cluster or a false detection. Our observation model (paragraph 3) provides these three factors in the likelihood, either directly or after integration and normalisation of the distributions.

#### 4.2. Hypothesis tests

The question we wish to answer *before* constraining parameters models is whether there exist a parametrised model which is compatible with our data. To settle this issue, we use an hypothesis test method. For a cosmological model, we generate by Monte-Carlo a large number of observed source catalogues, compute their likelihood for the given cosmological model and build an histogram of the likelihood (see Fig. 9). The normalisation of the integral of the histogram provides the probability curve of an observed catalogue to be compatible with the cosmological model. The use of the statistical observation model speeds up this work dramatically. Figure 9 *left*, shows the histogram of the likelihoods computed assuming the concordance cosmological model and a Press and Schechter mass function (Press & Schechter 1974) for the clusters. The black line shows the likelihood value computed for a source catalogue computed with the same cosmology, but with the mass function of Sheth and Tormen (Sheth et al. 2001). The probability of compatibility is lower than  $10^{-5}$ . The Press and Schechter hypothesis is thus rejected by the data. In practise, cosmological parameters are free parameters and we often obtain a compatibility valley for our parameters with our “observed” data. Other sources of constraints on cosmological parameters (such as CMB anisotropies) will allow us to select the relevant cosmological models and mass function.



**Fig. 9.** *Left:* histogram of log-likelihood  $L$  (black) for  $N$  Monte-Carlo catalogue of a Press-Schechter cosmological model. The peak is fitted by a Gaussian law (red line), with mean  $L_{\text{mean}}$ . *Right:* red line is the probability versus  $\Delta L = L - L_{\text{mean}}$  of observing a Press-Schechter based catalogue with  $\Delta L$ . Vertical dashed line is the  $\Delta L$  computed for a catalogue generated from a Sheth and Tormen model. The probability of compatibility is lower than  $10^{-5}$ .

#### 4.3. Parameter estimation

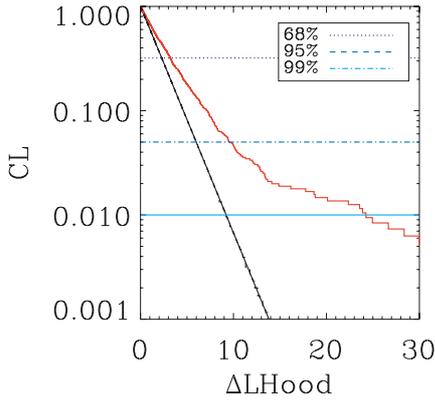
Once we have selected a parametrised model compatible with our data, the next question is to estimate a set of best cosmological parameters, in agreement with data and to compute the associated errors (or confidence levels). For this purpose, we minimise the function  $-\ln L(\mathbf{C}|\mathbf{S})$  over  $\mathbf{C}$ , vector of the model parameters, to find the best model  $\mathbf{C}_{\text{best}}$  according to our data and his likelihood  $L_{\text{best}}$ . Then we generate, according to  $\mathbf{C}_{\text{best}}$ , many source catalogues  $\mathbf{S}_i$ . We minimise the likelihood to find the best model  $\mathbf{C}_i$  matching each  $\mathbf{S}_i$ , and build the histogram of

$$\Delta_i = -2 [\ln L(\mathbf{C}_i|\mathbf{S}) - \ln(L_{\text{best}})] \quad (2)$$

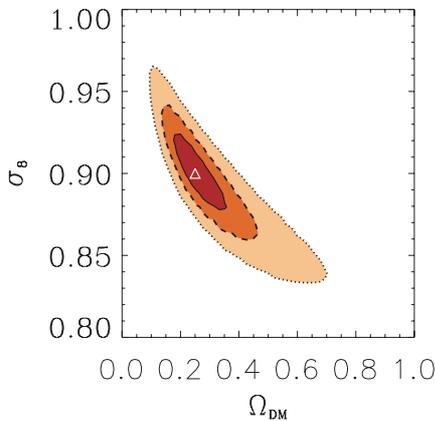
and the map of  $\Delta_i$  at position  $\mathbf{C}_i$ . Computing the normalised cumulative distribution of the variable  $\Delta$  (Fig. 10) allows us to compute  $\Delta_i$  values of the 68%, 95% and 99% confidence level limits to be used to draw contours on the model map. Would the pdf be Gaussian distributed,  $\Delta_i$  distribution would follow a  $\chi^2$  law. We notice that cluster density versus redshift and flux are *not Gaussian*.

#### 4.4. Constraints on cosmological parameters for the Olimpo SZ-cluster surveys

We now use the above tools to constraint the cosmological parameters. In the following, we will assume a  $\Lambda$ CDM cosmological model with parameter list of Table 1. The most important



**Fig. 10.** *Left:* histogram (red) of difference of log-likelihood  $\Delta_i$  for  $N = 1000$  Monte-Carlo catalogues of the best cosmological model  $C_{\min}$  according to our data set. Black line is the  $\chi^2$  law, expected for Gaussian distributions with 2 degrees of freedom ( $\Omega_M$  and  $\sigma_8$ ). 68%, 95% and 99% confidence levels are shown as dotted, dashed and dot-dashed horizontal blue lines. The  $\chi^2$  approximation is very optimistic.



**Fig. 11.** Expected constraints on  $\sigma_8$  and  $\Omega_{DM}$  from an Olimpo scientific flight, with full spectroscopic follow-up of the sources and a field to field Poisson-like count variance. All other cosmological parameters, have been set to the values in Table 1. Diamond is the initial cosmological model used to simulate data.

parameters for large scale structure formation and SZ-clusters are  $\sigma_8$  and  $\Omega_M$  as well as  $T_*$ , the normalisation of the mass to temperature scale relations (Pierpaoli et al. 2001) in cluster formation models. We plot in Fig. 11 the expected constraints on  $\sigma_8$  and  $\Omega_{DM}$  from SZ-cluster observations, assuming all other cosmological parameters known, at their simulated values. We assumed that follow-up observations provided redshifts for *all* the sources. This is the best constraint achievable, according to our observation model.

#### 4.4.1. Cluster count variance

As shown at paragraph 2.5 the simulated source count variance can be significantly larger than a Poisson's distribution of same expectation that we would assume from field to field variance. In this paragraph we quantify the degradation on confidence level due to such an enlarged count variance. Figure 12 shows the results.

Therefore calibration of Flux measurement must be a priority in the design of Large SZ cluster surveys. Count variance might be due to observations (instruments), but also to cluster

detection algorithms (Herranz et al. 2002; Pierpaoli et al. 2005; Melin 2004; Pires et al. 2006). Those should be evaluated, on their efficiency, on the purity of the recovered source catalogue, but definitely *on the source count variance* at the output of the chain. In the following, we use the poisson-like counting variance obtained with our detection chain assuming a perfect calibration.

#### 4.4.2. Degeneracy with late cluster physics

Galactic physics event heats the extragalactic cluster gas, and thus contribute to the SZ-cluster signal in addition to the gravitational potential and virialisation. Late cluster gas heating mechanisms are not well known yet. Their contribution to gas heating is commonly parametrised in the mass-temperature relation as a normalisation factor,  $T_*$ . The left panel of Fig. 13 shows the CL contours placed on  $T_*$  and  $\sigma_8$  marginalised on  $\Omega_M$ . We observe that with the input of WMAP and CFHT-LS weak-shear forecast, the residual correlation between  $T_*$  and  $\sigma_8$  is small. In addition on going X-Rays surveys from XMM satellite should provide a lot of information on cluster gas physics and allow precise determination of  $T_*$ . Thus in the following we set  $T_*$  to the value 1.9.

#### 4.4.3. Restriction to high-mass clusters

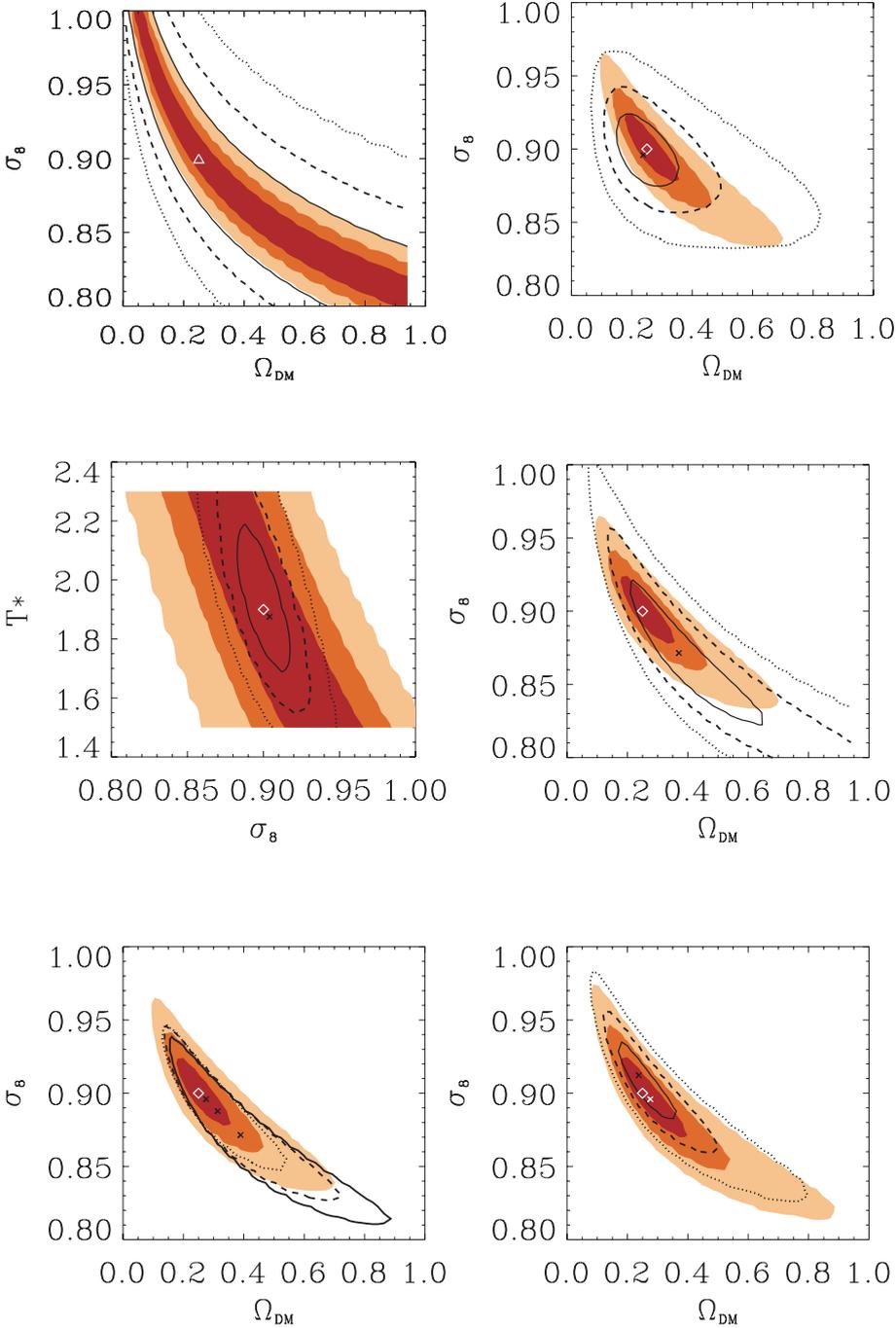
Large clusters involve hundred of galaxies. Their gravitational potential is stronger than in smaller clusters and non-gravitational physics is less important than in low mass systems. As a result, their mass to temperature scaling law is expected to show a smaller dispersion. Thus one can foresee that heavy clusters will be statistically better modelled, and that constraints based on massive cluster observations will be reliable. It is thus instructive to study the cosmological constraints which can be derived from a sample restricted to high-mass clusters. The right panel of Fig. 13 shows the confidence level map computed from a catalogue, when we select clusters of flux larger than  $7.4 \times 10^{-4}$  arcmin<sup>2</sup>. The CL contour are significantly enlarged compared to the reference contour drawn in black, because of the much smaller statistic. This is a strong motivation for theorists to understand and build a model of low-mass clusters.

#### 4.4.4. Incomplete spectroscopic follow-up

Large-array bolometer surveys will provide large cluster catalogues, including flux and positions and shapes for resolved clusters, but have to rely on follow-up experiments for redshift measurements. The left panel of Fig. 14 shows the impact of partial redshift coverage, assuming 10%, 20% or 50% random coverage of clusters candidates. Remaining catalogue contaminations have been properly taken into account. We note that no significant bias on the CL contours is seen, but that the statistics are degraded. This shows that follow-up observations will be very important for the accuracy of the physics output of large SZ-Cluster surveys.

#### 4.4.5. Neglecting contamination

Assuming now that the redshift followup of the observations will be incomplete (as is very likely to be the case in practice), we quantify now the effect of neglecting contamination in the recovered catalogue. We assume in the paragraph that 50% of the sources have a redshift. Technically, this means using



**Fig. 12.** Degradations of constraint due to the cluster count variance. White marks give the cosmological models used to simulate the data. *Left:* black lines are the  $\Omega_{\text{DM}}$  vs.  $\sigma_8$  Confidence Level (1, 2 and 3 sigmas) contours computed using the degraded count variance, all other cosmological parameters set to their simulated values. Only the information on SZ-cluster count has been used in this figure, *no redshift*. Colored are the same CL constraints, with the Poisson-like field to field cluster count variation. *Right:*  $\Omega_{\text{DM}}$   $\sigma_8$  Confidence Level (CL) contours assuming 100% follow-up for cluster redshifts, and a degraded count variance as quantified at paragraph 2.5. All other cosmological parameters set to simulated values. Lines draw the CL contours with degraded field to field cluster count variation (black cross is the reconstructed model). Colored contours are the reference constraints as in Fig. 11.

**Fig. 13.** *Left:* confidence level map, on  $\sigma_8$  and  $T_*$ , marginalised on  $\Omega_{\text{DM}}$ . Colors are computed from SZ-cluster data only, dashed lines uses WMAP and CFHTLS weak-shear Fisher matrix constraints, no systematic effect on WMAP or CFHT-LS are included. *Right:* lines are the constraints on cosmological parameters if we keep only the largest flux clusters. All other cosmological parameters, have been set. Diamond is the generated concordance model. Black cross is the reconstructed model. Colors delimit the references CL contour. Lower statistic induce heavy loss in the constraint accuracy.

**Fig. 14.** *Left:* impact on cosmological constraints, due to an incomplete redshift follow-up of cluster candidates. Black line is the 95% CL contour assuming a 10% coverage follow-up. Dashed line assumes 20% coverage follow-up and dotted line 50% coverage. Colored contours are a copy of Fig. 11. *Right:* lines show the systematic shift in the CL contour induced by neglecting contaminants in the recovered source catalogue. This plot was generated assuming that 50% of the sources have been observed in follow-up for redshift. Colors stand for contours computed with the same dataset, but taking into account contaminations. White cross is the best model taking into account contaminants, and black cross is the biased best model. The diamond is still the simulated cosmological model.

in the first factor of the likelihood the pdf count of true detected cluster and in the third factor, the integral over  $z_{\text{obs}}$  of  $\frac{dP}{dz_{\text{obs}}dY_{\text{obs}}}(z_i^{\text{Clus}}, Y_i^{\text{Clus}}; \mathbf{C})$ . The right panel of Fig. 14 shows the results. We see that since we assumed that false detections are clusters, the reconstructed parameters are biased toward large  $\sigma_8$ , since contamination produce a spurious enhancement of the cluster distribution at low surface brightnesses. This effect is rather small, since we chose to use a 95% pure SZ Catalog.

The reduced size of the CL contour is a secondary effect of the low  $\Omega_{\text{M}}$  fitted value.

## 5. Conclusion

In this paper, we have explored, in details, the potential and limitations of upcoming wide-field SZ surveys. We used a full simulation pipeline, going from the cosmological models to

recovered cluster catalogues and constraints on cosmological parameters. We showed that the selection function and purity of the recovered catalogues are more complex than is usually assumed. We quantified the foreseen selection function, photometry, contamination and field to field count variance of the upcoming Olimpo project. We presented methods to statistically model the observations, select parametrised models compatible with data, and then constraint models parameters.

We showed that field to field count variance is likely to be enlarged if great care is not taken in the calibration of cluster flux extraction. We showed that any enlargement in cluster count variance must be taken into account in the cosmological parameter estimation, and makes the constraints worse. We showed that, using SZ Cluster data, combined with WMAP and expected CFHTLS weak-shear forecast data, little correlation is seen between the mass-temperature normalisation factor  $T_*$  and  $\sigma_8$ . Complementary X-Rays observations will be necessary to put tighter constraints on  $T_*$ , but on the other hand, we only need moderate precision on  $T_*$ , to achieve good knowledge on  $\sigma_8$ . We finally exposed the impact on cosmological parameters of systematics in observations or interpretation of our data.

This paper does not use the 2-point correlations of SZ Cluster to constraint cosmology (Mei & Bartlett 2003), nor cluster angular size. This incorporation of these informations in our simulation pipeline and the computation of their impacts on cosmological parameter constraints is left to future work.

*Acknowledgements.* We hereby acknowledge many scientific and algorithmic discussions with J. Rich, J.-P. Pansart, C. Magneville, R. Teyssier (CEA Saclay/DAPNIA) and J.-B. Melin and J. G. Bartlett (Univ. Paris 7, APC). Special thanks to P. Lutz and J. Bouchez (CEA Saclay/DAPNIA) for their help with statistic methods. We would like to acknowledge as well, the Dapnia-SAp and the Astronomy group of the University of Concepción, Chile, for computing facilities that helped to achieve this study.

## 6. Annex: selection function data

The following Table 3 samples values of the selection function as a function of mass and Compton flux.

**Table 3.** Selection function versus redshift. Value entered correspond to 90% efficiency.

Redshift	90% efficiency flux [arcmin <sup>2</sup> ]	90% efficiency Mass [ $M_\odot$ ]
0.1	$3.1 \times 10^{-3}$	$3.98 \times 10^{14}$
0.2	$1.6 \times 10^{-3}$	$4.89 \times 10^{14}$
0.3	$1.0 \times 10^{-3}$	$5.13 \times 10^{14}$
0.4	$7.9 \times 10^{-4}$	$5.19 \times 10^{14}$
0.5	$6.9 \times 10^{-4}$	$5.25 \times 10^{14}$
0.7	$5.1 \times 10^{-4}$	$5.25 \times 10^{14}$
1.0	$4.9 \times 10^{-4}$	$5.13 \times 10^{14}$
2.0	$4.0 \times 10^{-4}$	$4.07 \times 10^{14}$
3.0	$3.89 \times 10^{-4}$	$3.16 \times 10^{14}$
4.0	$3.89 \times 10^{-4}$	$2.51 \times 10^{14}$
5.0	$3.89 \times 10^{-4}$	$1.99 \times 10^{14}$

## References

- Barbosa, D., Bartlett, J. G., Blanchard, A., & Oukbir, J. 1996, *A&A*, 314, 13  
 Battye, R. A., & Weller, J. 2003, *Phys. Rev. D*, 68, 083506  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393  
 Borys, C., Chapman, S., Halpern, M., & Scott, D. 2003, *MNRAS*, 344, 385  
 Cardoso, J. F. 1999, *Neural Computation*, 11, 157  
 Haiman, Z., Mohr, J. J., & Holder, G. 2001, *ApJ*, 553, 545  
 Herranz, D., Sanz, J. L., Barreiro, R. B., & Martínez-González, E. 2002, *ApJ*, 580, 610  
 Lamarre, J., Puget, J., Bouchet, F., et al. 2004, *New Astron. Rev.*, Proc. of the workshop on The Cosmic Microwave Background and its Polarization, ed. S. Hanany, & R. A. Olive  
 Malmquist, K. G. 1920, *Medd. Lund. Astron. Obs. Ser. 2*, 22  
 Masi, S., Ade, P., de Bernardis, P., et al. 2003, *Mem. Soc. Astron. It.*, 74, 96  
 Mei, S., & Bartlett, J. G. 2003, *A&A*, 410, 767  
 Melin, J. B. 2004, Ph.D. Thesis, Univ. Denis Diderot – Paris VII  
 Melin, J. B., Bartlett, J. G., & Delabrouille, J. 2005, *A&A*, submitted  
 Oukbir, J., & Blanchard, A. 1997, *A&A*, 317, 1  
 Pierpaoli, E., Anthoine, S., Huffenberger, K., & Daubechies, I. 2005, *MNRAS*, 359, 261  
 Pierpaoli, E., Scott, D., & White, M. 2001, *MNRAS*, 325, 77  
 Pires, S., Juin, J.-B., Yvon, D., et al. 2006, *A&A*, 455, 741  
 Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425  
 Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1  
 Sunyaev, R. A., & Zel'dovich, Y. B. 1970, *Comments Astrophys. Space Phys.*, 2, 66  
 Sunyaev, R. A., & Zel'dovich, Y. B. 1972, *Comments Astrophys. Space Phys.*, 4, 173