



Multiple-component Decomposition from Millimeter Single-channel Data

Iván Rodríguez-Montoya^{1,2}, David Sánchez-Argüelles², Itziar Aretxaga², Emanuele Bertone², Miguel Chávez-Dagostino²,
David H. Hughes², Alfredo Montaña^{1,2}, Grant W. Wilson³, and Milagros Zeballos²

¹ Consejo Nacional de Ciencia y Tecnología. Av. Insurgentes Sur 1582, 03940, Ciudad de México, México; irodriguez@inaoep.mx

² Instituto Nacional de Astrofísica, Óptica y Electrónica. Apartado Postal 51 y 216, 72000. Puebla Pue., México; domars@inaoep.mx, itziar@inaoep.mx

³ Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

Received 2017 June 9; revised 2017 November 6; accepted 2017 November 10; published 2018 March 2

Abstract

We present an implementation of a blind source separation algorithm to remove foregrounds off millimeter surveys made by single-channel instruments. In order to make possible such a decomposition over single-wavelength data, we generate levels of artificial redundancy, then perform a blind decomposition, calibrate the resulting maps, and lastly measure physical information. We simulate the reduction pipeline using mock data: atmospheric fluctuations, extended astrophysical foregrounds, and point-like sources, but we apply the same methodology to the Aztronomical Thermal Emission Camera/ASTE survey of the Great Observatories Origins Deep Survey–South (GOODS-S). In both applications, our technique robustly decomposes redundant maps into their underlying components, reducing flux bias, improving signal-to-noise ratio, and minimizing information loss. In particular, GOODS-S is decomposed into four independent physical components: one of them is the already-known map of point sources, two are atmospheric and systematic foregrounds, and the fourth component is an extended emission that can be interpreted as the confusion background of faint sources.

Key words: atmospheric effects – methods: statistical – submillimeter: diffuse background – submillimeter: galaxies – techniques: image processing

Supporting material: tar.gz file

1. Introduction

Astronomical observations in millimeter wavelengths provide crucial information to comprehend the formation and evolution of structures in the universe at all scales, from galaxy clustering (Carlstrom et al. 2002) to circumstellar debris disks (Chavez-Dagostino et al. 2016). This observational window also led to the discovery of a whole new population of bright dust-obscured submillimeter galaxies (SMGs), mostly unresolved by single-dish telescopes, but whose detection is available within a wide range of high redshifts (Casey et al. 2014). Moreover, cold dusty sources are brighter in millimeter wavelength, allowing a relatively easy detectability from ground-based observatories.

Even though millimeter astronomy already spans a few decades, there are still some relevant challenges for ground-based observations, concerning especially foreground removal and calibration of data. First, water vapor and oxygen emit lines at different microwave lengths, making Earth’s atmosphere partially opaque to millimeter emissions, alongside the difficulty that atmospheric fluctuations are nonstationary and often abrupt. Second, even astrophysical foregrounds may hinder the inference of some physical quantities. For example, bright patches of an extended emission could be confused with SMGs or other compact sources. Conversely, point-like objects stand as a contamination for an extended source. Third, for single-channel instruments multiwavelength separation is impeded, making foreground removals quite challenging. But even with multichannel instruments, in order to maximize their profit, the challenge dwells in developing advanced decomposition algorithms.

Thus, for any ground-based millimeter-wavelength experiment, it is crucial to explore new strategies to improve data cleaning, ameliorate astrophysical component separation, and enhance sensitivity. In this spirit, here we present a new

implementation of two well-known methodologies: principal component analysis (PCA) and independent component analysis (ICA). Our main goal is to propose and test a new technique able to perform multicomponent separation in defiance of the single-channel limitation. Second, we want to propose and test strategies to calibrate the separated components. Finally, we want to apply our ideas to real data in order to probe our ability to recover previous measurements and gain insight on the potential benefits of using our PCA–ICA technique.

In previous studies, atmospheric cleaning has been attempted by removing common modes along the detector array (Sayers et al. 2010). On the other hand, ICA was used in space-based multichannel experiments to clean the cosmic microwave background from its astrophysical foregrounds (for a review see Ichiki 2014). Similar algorithms have been successfully applied to a variety of astrophysical observations, from exoplanetary light curves (Morello 2015) to forecasts of interferometric 21 cm cosmological signals (Zhang et al. 2016). In context with the literature, we are reporting the first multicomponent analysis of single-wavelength millimeter data and for a ground-based telescope. The core of our proposal relies on a technique to increase data redundancy, whose closest discussion was made by Waldmann (2014). Although we focus on the Aztronomical Thermal Emission Camera (AzTEC), a 144-bolometer camera currently operating in a single (1.1 mm) channel (Wilson et al. 2008) and coupled to the Large Millimeter Telescope (Hughes et al. 2010), our approach could be extended to other single- or even multichannel experiments.

This paper is organized as follows. In Section 2, we motivate our methodology, introducing the theoretical basis of the PCA and ICA algorithms. In Section 3, we describe the AzTEC instrument, the observational data, and the numerical code used to process the time-domain data into an astrophysical map. In

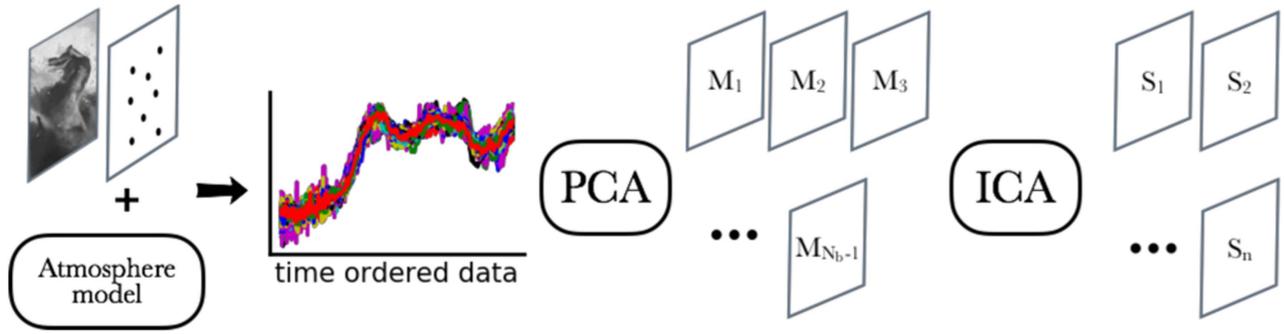


Figure 1. Schematic description of the data processing in this paper. For simulations, the time streams are built from an atmospheric model and mock (extended and point-like) astrophysical data. PCA is used in time domain to generate a series of redundant maps $\{M_i\}$, and ICA is applied in map domain for multicomponent decomposition. Later on, the independent signals $\{S_j\}$ are calibrated and astrophysical information is extracted.

Section 3.3, we introduce our proposal of using PCA in time domain followed by ICA in map domain, as an extension to the standard AzTEC pipeline. Section 4 is devoted to implementing and testing our techniques with simulations. We describe the mock data employed, the simulated reduction process, the decomposition parameters, calibration strategies, and steps to extract astrophysical information. In Section 5 we apply the same tools to the Great Observatories Origins Deep Survey–South (GOODS-S), observed with AzTEC when it was installed on the 10 m Atacama Submillimeter Telescope Experiment (ASTE; Scott et al. 2010, hereafter *KS10*), in order to recover previous measurements and discuss them in connection with our simulation results. Our conclusions are summarized in Section 6.

2. PCA and ICA Algorithms

Why PCA and ICA? Our concern is that atmospheric and astrophysical emissions are mixed along some range of scales. It is appealing that both PCA and ICA are blind (nonparametric) separation algorithms, so we do not need to rely on physical models, but just on the statistical properties of data. PCA computes uncorrelated projections of data, while ICA demands the stronger condition of statistical independence. Before going into formal details of each algorithm (for an introductory tutorial see Stone 2004), let us intuitively discuss their respective roles in our implementation (see also Figure 1).

PCA is used in time domain as follows: due to their intense brightness and angular scale, atmospheric fluctuations induce large correlations along the bolometer array. PCA computes a vector basis where the fluctuation modes are uncorrelated; the first few are attributed to atmospheric contamination and removed, leaving modes dominated by astrophysical signal and noise. PCA is very efficient, especially for point-source recovery, but some motivations to explore more advanced algorithms include that PCA uses only second-order statistical moments, i.e., non-Gaussian information is unexploited. More importantly, at least part of the astrophysical information is lost when the subset of (bad) principal components is discarded.

Formally, ICA is an extension of PCA, but the cases of interest and the criteria to apply each of these algorithms may be significantly different. Provided that the underlying emissions are non-Gaussian, ICA employs high-statistical moments to find a basis of (not only uncorrelated but) statistically independent components. Using ICA, it should be possible to reduce information loss because every component is in principle isolated, containing no information about others. In

general practice, ICA is fed with $m \geq 2$ mixed signals to be decomposed into $n \leq m$ independent components. Unfortunately, ICA bears an inherent incompatibility with single-channel instruments: a single-channel instrument yields a single map at a given wavelength band ($m = 1$), but ICA needs at least two input maps in order to decompose them into at least two independent components (see Sections 2.2 and 3.3 for details). Still, we can turn this limitation around by producing hierarchical levels of (artificial) redundancy, mimicking the maps from a multichannel survey, which can be used as input for ICA. Specifically, in this paper we propose to apply ICA in map domain over a series of redundant maps produced with PCA in time domain.

In the rest of this section we briefly overview the basics of both algorithms, so that the reader familiar with the theoretical aspects may jump directly to Section 3.

2.1. Principal Component Analysis

Let us assume that we are working with an astronomical survey whose raw data $\{x_i\}_{i=1}^{N_b}$ are a collection of N_b time streams (e.g., the number of bolometers), each of them with sampling size T . Thus, our time-ordered data can be represented by a vector \mathbf{x} , such that $\dim(\mathbf{x}) = N_b \times T$. Assume also that the time streams were centered in a pre-processing step, $x_i \leftarrow x_i - \langle x_i \rangle$, where $\langle \cdot \rangle$ denotes the expectation value.

The covariance matrix is defined by $\mathbf{C}_x = \langle \mathbf{x} \cdot \mathbf{x}^T \rangle$, $\dim(\mathbf{C}_x) = N_b \times N_b$. If it would be the case that every time stream x_i were Gaussianly distributed, then \mathbf{C}_x would contain all the information available, and higher-order statistical moments would be either zero or trivially rewritten in terms of \mathbf{C}_x .

PCA is formally an eigenvalue problem for the covariance matrix \mathbf{C}_x ; the goal is to find a vector basis $\{\mathbf{e}_i\}_{i=1}^{N_b}$ such that they project the raw data \mathbf{x} into a new set of uncorrelated components $\{y_i\}_{i=1}^{N_b}$,

$$\begin{aligned} \mathbf{y} &= \mathbf{E}^T \cdot \mathbf{x}, \\ \langle y_i \cdot y_j \rangle &= \lambda_i \delta_{ij}, \end{aligned} \quad (1)$$

where the columns of \mathbf{E}^T are the unit eigenvectors \mathbf{e}_i , and $\{\lambda_i\}_{i=1}^{N_b}$ are the eigenvalues of \mathbf{C}_x . It turns out that the new covariance matrix is diagonal, $\mathbf{C}_y = \text{diag}(\{\lambda_i\})$, whose elements are the variances of the projected time streams, $\lambda_i = \text{var}(y_i)$. The PCA distinctive step is to order the projected time streams according to an eigenvalue hierarchy $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_b}$; then, y_1 is called

the first principal component, y_2 the second principal component, and so on. When projected back to the raw data \mathbf{x} , the principal component y_1 is evidently the major source of correlation, contrary to y_{N_b} , which is the source of least correlation.

The cleaning step is simply to discard N_{atm} principal components that are attributed to large-scale foregrounds (typically the atmosphere). Then, the remaining components $\{y_i\}_{i=N_{\text{atm}}+1}^{N_b}$ are projected back into the original space. Notice that the cleaned data vector \mathbf{x}' also has $\dim(\mathbf{x}') = N_b \times T$.

2.2. Independent Component Analysis

In signal processing analysis, the *cocktail party* problem is referred to as the linear mixing of n true signals into the recordings of m sensors, under the condition $n \leq m$ (for a comprehensive review see Hyvärinen et al. 2002; Comon & Jutten 2010). In this subsection the notion of a sensor is meant quite generically. For instance, a sensor may be an astronomical survey measured at a given wavelength along with the process to make an observation map; in this scheme, the mixed and true signals live in pixel domain. Keeping this broader notion in mind, let us denote x_i as the mixed signal of the i th sensor, for $i = 1, 2, \dots, m$. The j th true signal is then denoted as s_j , for $j = 1, 2, \dots, n$. The raw data are then modeled by ICA as an instantaneous mixing of the true signals in terms of a linear combination,

$$x_i = a_{i1} s_1 + a_{i2} s_2 + \dots + a_{in} s_n. \quad (2)$$

The mixing coefficients $\{a_{ij}\}$ are real numbers that may be interpreted as the transmission/extinction information of the true signals through the sensing process. The model can also be written in matrix notation,

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s}, \quad (3)$$

$$\mathbf{s} = \mathbf{W} \cdot \mathbf{x}, \quad (4)$$

where \mathbf{x} is an array whose column vectors are the m mixed signals, and \mathbf{A} is called the mixing matrix ($m \times n$). The goal is to estimate the unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, retrieving the true signals \mathbf{s} . Since both \mathbf{W} and \mathbf{s} are simultaneously estimated, the unmixing problem becomes too complicated for classical methods.

ICA relies on the assumption that the true signals are statistically independent and non-Gaussian. Two random variables y_1 and y_2 are statistically independent if and only if their joint probability distribution is equal to the product of their marginal probability distributions,

$$p(y_1, y_2) = p(y_1) \cdot p(y_2). \quad (5)$$

Statistical independence implies uncorrelatedness, though uncorrelatedness does not necessarily imply independence. ICA then appeals to the central limit theorem, which says that the sum of two non-Gaussian distributions is more Gaussian than the initial distributions; conversely, independent signals are maximally non-Gaussian. Thus, to solve the unmixing problem in Equation (4), ICA estimates those coefficients w_{ji} that maximize the non-Gaussianities of $\{s_j\}_{j=1}^n$.

Non-Gaussianities are measured by high-order statistical moments. For example, the skewness measures the symmetry of the distribution, whereas the kurtosis measures how spiky (or flat) the distribution is. For Gaussian distributions both the

skewness and kurtosis are zero because only the first two moments are relevant, namely, the mean and variance. Although the simplest approach would seem to maximize skewness or kurtosis, they are easily biased by outliers; hence, more robust non-Gaussianity estimators should be used instead (for a concise review see Choi 2011). Here, we focus on the concept of *negentropy*.

Entropy is the basic concept in information theory, and it is particularly interesting for measuring non-Gaussianities (Hyvärinen & Oja 2000). Defined as

$$H(y) = -\int dy p(y) \log p(y), \quad (6)$$

entropy is related to the degree of information contained in the random variable y . As the variable is more unstructured and unpredictable, its entropy is larger as well. Indeed, the Gaussian distribution possesses the maximum entropy, i.e., it is the most random, least structured, and least informative distribution. In the same vein, negentropy is defined as the deviation from the maximum entropy,

$$J(y) = H(y_G) - H(y). \quad (7)$$

Here y represents the signal of interest, and y_G is a Gaussian distribution with the same mean and variance as y . Negentropy is always nonnegative, $J(y) \geq 0$, and equals zero if the signal y is Gaussianly distributed, $J(y_G) = 0$. Clearly, the larger the negentropy, the more informative is the distribution, and the more independent is the signal. Hence, negentropy is the optimal estimator of non-Gaussianity and statistical independence.

As defined in Equation (7), measuring negentropy is difficult, and some approximated estimators, like higher-order cumulants, are frequently used:

$$J(y) = [\langle G(y) \rangle - \langle G(y_G) \rangle]^2, \quad (8)$$

where G is a nonquadratic function that may be conveniently chosen. In principle, any power function higher than quadratic is a valid choice for G , but in practice, a wise choice may boost the speed of the algorithm. For instance, $G(y) = y^4$ is the kurtosis-based approximation, useful when the independent signals are flat-like distributed, but otherwise nonrobust against outliers. Some commonly used functions are

$$G_1(y) = 1/c_1 \log(\cosh(c_1 y)), \quad (9)$$

$$G_2(y) = -\exp(-y^2/2), \quad (10)$$

where $1 \leq c_1 \leq 2$ is a constant often equal to 1. With this negentropy approximation, an optimizing algorithm can find the numerical values of the unmixing coefficients such that they maximize the negentropy of the independent components expressed in Equation (4). To this end, in this paper we use FastICA (Hyvärinen & Oja 1997, 2000), which is a very efficient fixed-point algorithm that has been widely used and tested, as it is the most standard ICA algorithm.

There remain, though, two ambiguities inherent to ICA. These are natural consequences of the fact that we are dealing with a system with fewer equations than unknown variables. Henceforth, as a post-decomposition step, the independent components ought to be calibrated, before any physical information may be inferred.

The *permutation ambiguity* means that the order of the independent components is basically random; this is because Equation (2) is invariant under permutations of the mixing matrix. The seriousness of this ambiguity depends on the number of independent components and how distinctive they are. If a given problem required to control the order of many independent components, the permutation ambiguity could become too prohibitive for an ICA application.

The *scaling ambiguity* is the inability to determine the variance of each independent component. Notice that Equation (2) remains invariant under the transformation $s_j \leftarrow a_j s_j$ and $a_{ij} \leftarrow a_{ij}/a_j$, where a_j is a real scale factor. Then, one may choose arbitrary scales right after the decomposition. As a convention, we will set the scale of every s_j to have standard deviation (std) equal to 1, but scale calibrations ought to be pursued afterward.

3. The AzTEC Instrument and Pipeline

3.1. The Instrument

AzTEC (Wilson et al. 2008) is a continuum millimeter-wavelength receiver containing 144 Si₃Ni₄ spiderweb mesh bolometers. The receiver is configured to operate in the 1.1 mm atmospheric window. The bolometers are arranged in a hexagonal array divided into six slices or hexants, distributed in a closed packed configuration. The footprint of the bolometer array covers a roughly circular area of ~ 8 arcmin diameter on the sky. These time-ordered data, or *time streams*, are later processed along with the telescope pointing information to construct an image of the sky surface brightness. For a ground-based (sub)millimeter camera, a single detector time stream d can be described as

$$d = \mathcal{P}s + \mathcal{A} + N, \quad (11)$$

where s is the surface brightness distribution of astronomical objects, \mathcal{P} is the pointing matrix, \mathcal{A} is the atmosphere emission, and N is the instrumental noise. It is important to note that the atmosphere fluctuations are between 1 and 4 orders of magnitude larger than the astronomical emission. Therefore, it is necessary to calculate and remove an estimation of the atmospheric component, in order to retrieve an image of the brightness distribution of faint sources. This process is critical for ground-based observations, where both the telescope scanning pattern and the map projection code are designed to decouple the astronomical emission from the atmospheric foregrounds. In particular, for the data described in the sections below, AzTEC observations were carried out with a modified Lissajous pattern; this is a parametric curve constructed from two sinusoidal waves in orthogonal directions. The projection of the scan track over the sky, relative to the map center, can be described by

$$\delta A(t) = 5.5 \sin 9t + 2.0 \sin 9t/30, \quad (12)$$

$$\delta E(t) = 5.5 \sin 8t + 2.0 \sin 8t/30, \quad (13)$$

where t is the observation elapsed time in seconds, and δA and δE are the track offsets from the map center in azimuth and elevation, respectively. In the following subsection we briefly overview the standard map projection code, namely, the reduction pipeline.

3.2. The Reduction Pipeline

A deep millimeter survey is a number of observations of a sky patch, stored in raw data files containing the recorded time streams, along with telescope parameters for calibrations. For surveys made with the AzTEC camera, each observation contains N_b time streams, corresponding to the effective number of bolometers (i.e., the bolometers with the best electronic responsivity). The AzTEC reduction pipeline uses PCA to remove the atmospheric signal (hereafter the *cleaning process*) and projects the cleaned time streams into a bi-dimensional grid, delivering an astronomical image as a result. (For a detailed description of the AzTEC pipeline, we refer the reader to Scott et al. 2008; Wilson et al. 2008.)

Time sampling. Typically, a single observation lasts 20 minutes, but the time streams are sampled in *time chunks* of 10 or 29 s. We denote with \mathbf{x} the chunk made of N_b time streams with length T . Every time chunk is worked out sequentially for each observation, but the reduction code runs in parallel for multiple observations.

Signal conditioning. All time streams are corrected for instrumental glitches and large spikes induced by cosmic rays. A low-pass filter is applied in order to minimize the contamination of high frequencies.

Atmospheric removal. As explained in Section 2.1, the principal component is the major source of correlation and is blamed for atmospheric contamination. But also the second and third principal components are often contaminated. The question is how many principal components shall be discarded, regarding a compromise between contamination removal and information loss. Because PCA is applied to every time chunk, the amount of information contained in each principal component depends on the time-chunk length T : the larger T , the more components shall be discarded.

The PCA2.5 σ procedure. The AzTEC pipeline has a semiautomated process optimized for point sources (Wilson et al. 2008). Choosing a small time-chunk length ($T = 10$ s customarily), the number of discarded components is estimated from the eigenvalue distribution: the 2.5 std outliers are iteratively rejected, and their corresponding eigenvectors are discarded. Using this procedure, typically 12 principal components are discarded per time chunk.

Mapmaking. The sky positions per bolometer are continuously recorded, according to the telescope pointing calibration, the bolometer-array geometry, and the scanning strategy. This information is contained in an object called the pointing matrix \mathcal{P} . Using \mathcal{P} , the cleaned time streams are projected into a single grid called the co-added map.

Noise estimation. Around 100 jackknife simulations of the cleaned time streams are projected into a single *weight map* $W(p)$ that stores the inverse noise variance per pixel p . The noise of every co-added map depends on (1) the effective sensitivity σ_{eff} or the rms noise, which is nearly uniform along the map, and (2) the sample number per pixel, namely, the *hitmap* $H(p)$. Hence, the weight map $W(p)$ can be approximated by $W^{1/2}(p) \approx w_{\text{eff}} H(p)$. Both the effective weight $w_{\text{eff}} = 1/\sigma_{\text{eff}}$ and $H(p)$ can be normalized such that $0 < H(p) < 1$. This approximation is accurate to better than 0.1% for AzTEC maps. Then, the signal-to-noise ratio (S/N) map can be directly obtained from the signal and weight maps.

Filtering. The AzTEC pipeline is equipped with low- and high-pass filters. The low-pass filter is a Gaussian one with an FWHM of the size of the telescope beam, and it is necessary for

removing spurious high-frequency fluctuations. A Wiener filter with a point-like kernel can be used to boost the detection of point sources (Downes et al. 2012; Perera et al. 2013); it enhances compact sources in detriment of extended signals.

The *detection of bright sources* is performed inside the uniformly covered area. The first step is to locate the highest-S/N pixel and enclose the bright source within a beam-radius area. Then, pixels closer than twice the beam size are discarded as candidates for the next brightest source. The search continues down to a specified limit e.g., $S/N > 4$.

3.3. The ICA Extension to the Pipeline

Despite the fact that ICA is formally an extension of PCA, ICA cannot be directly applied to clean the bolometer time streams. The main obstacle is the permutation ambiguity, explained in Section 2.2. As mentioned in the preceding subsection, keeping control of the time-stream order is crucial because it contains the bolometer positions on the sky, without which the mapmaking step could not be accomplished.

For surveys made with multichannel instruments, it is possible to generate a map per channel, containing redundant information at different wavelengths. ICA is often used to gain leverage from these multiwavelength signals in map domain (e.g., Ichiki 2014). Since typically one would have only a few channels (e.g., $m=3$), the permutation ambiguity is not an issue in map domain. Unfortunately, the number of intrinsic signals within the maps could be typically larger than the number of available channels, that is, $n > m$, then preventing a proper decomposition. In single-channel instruments like AzTEC the limitation is obviously worse, lacking any leverage of multiwavelength redundancy.

With the aim to overcome the single-channel limitation, we are proposing the generation of *artificial* redundancy. Specifically, we generate redundant maps by applying different thresholds in the PCA technique: map M_1 is made by discarding the first principal component, M_2 by discarding the first and second principal components, and so on, up to M_{N_b-1} , where N_b is the effective number of bolometers. Our main assumption is that the components present in the map are coupled by different mixing coefficients in the redundant maps. We choose a relatively large time chunk (120 s) in order to generate a smoother transition in the degree of redundant information. Notice that PCA in time domain is basically employed as a filter to generate hierarchical levels of redundancy. Consequently, the redundant maps are strongly correlated at several angular scales, and now ICA can be applied on them.

Following Section 2.2, we model the set of redundant maps $\{M_i\}_{i=1}^{N_b-1}$ as a mixture of n independent components $\{S_j\}_{j=1}^n$. The i th redundant map is explicitly

$$M_i = a_{i1} S_1 + a_{i2} S_2 + \dots + a_{in} S_n, \quad (14)$$

or in matrix notation,

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{S}, \quad (15)$$

$$\mathbf{S} = \mathbf{W} \cdot \mathbf{M}. \quad (16)$$

Here \mathbf{M} is an array containing the $N_b - 1$ redundant maps, \mathbf{A} is the $(N_b - 1) \times n$ mixing matrix, $\mathbf{W} \approx \mathbf{A}^T$ is the unmixing matrix, and \mathbf{S} is an array containing the n independent maps. In this paper we use FastICA (Hyvärinen & Oja 2000) to estimate

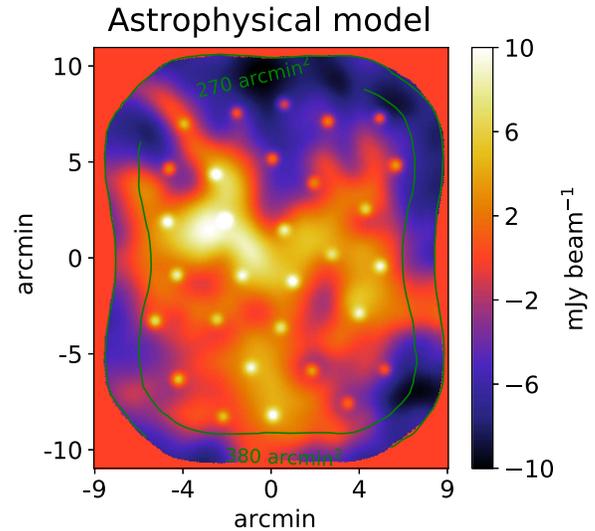


Figure 2. Astrophysical mock data used in our simulations: (1) a set of 30 beam-sized sources compose the point model P , and (2) a smoothly ($3 \times$ beam) varying source constitutes the extended model E . The inner 270 arcmin^2 contour represents the area where the telescope coverage is at least 50% of the maximum.

\mathbf{W} and \mathbf{S} by maximizing the negentropy of the independent maps.

In FastICA we simply use the `logcosh` function as in Equation (9). For visualization purposes, we also use the following expression for negentropy:

$$J(y) = 20 \frac{|\langle G(y_N) \rangle - \langle G(y) \rangle|}{\langle G(y_N) \rangle}, \quad (17)$$

which is equivalent to Equation (8) but differs by a couple of normalization factors (the constant 20 and $\langle G(y_N) \rangle$) included to increase the contrast between negentropy values. Here, y and y_N are unit-std distributions centered at zero, y represents the pixel values of a map, y_N is the normal standard distribution, and $G(y) = \log(\cosh y)$. Notice that this expression satisfies the negentropy properties $J(y) \geq 0$ and $J(y_N) = 0$.

It is wise to restrict the pixel data to the better-sampled region on the sky. We adopt the convention to feed ICA with the map area where the (outermost) coverage is at least $\sim 30\%$ of the maximum, but to extract physical information only from the 50% uniformly covered region. After a successful decomposition, in order to tackle the permutation and scaling ambiguities, we must calibrate the independent maps. To this end, we implement some calibration alternatives in Section 4.3.

To end this section, a few alternatives to generate redundancy can be mentioned from the literature. An interesting approach could be a decomposition in a convenient wavelet space, with the ancillary beneficial ability to calibrate the independent components (Waldmann 2014). Another arguably possible alternative would be to use observations taken at different time intervals as the input for ICA (see, e.g., Funaro et al. 2003; Waldmann 2012). We do not follow this approach because the amount of Gaussian noise in every individual observation is much larger than the co-added observation, thus making quite difficult the decomposition for ICA. Besides, given that this set of time-ordered maps were not observed simultaneously, they would break the basic assumption of instantaneous mixing.

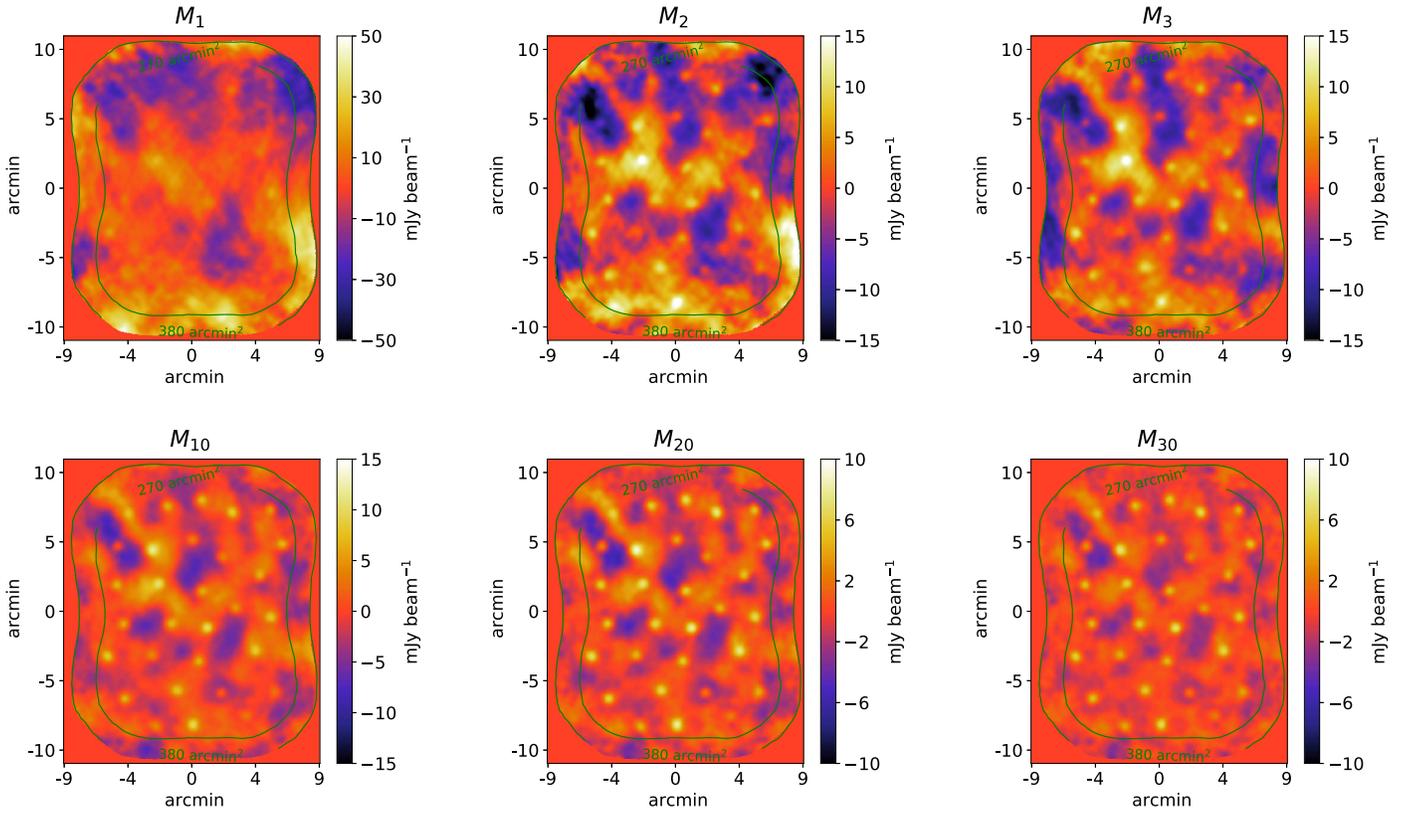


Figure 3. Redundant maps M_i made with the AzTEC pipeline, removing i principal components, respectively. Notice that atmospheric and extended foregrounds are progressively removed. Point sources persist in all redundant maps with decreasing brightness; the last (not shown) redundant maps contain nearly Gaussian noise. Contours as in Figure 2.

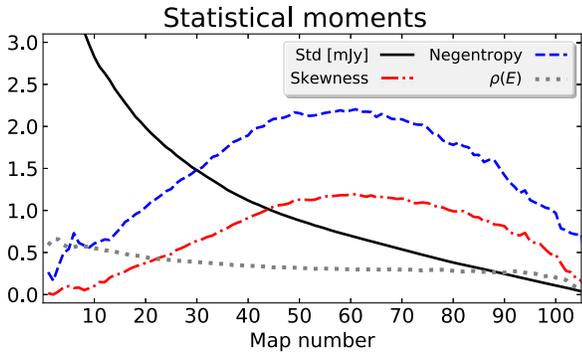


Figure 4. Statistical moments of (inner 270 arcmin² area) redundant maps. Negentropy is computed with Equation (17), and $\rho(E)$ stands for pixel correlation with the extended model E .

4. Decomposition of Mock Data

In this section, we fabricate a set of atmospheric and astrophysical signals mixed in time streams, simulate the pipeline process to perform a decomposition, propose strategies for calibration, and finally measure (mock) astrophysical information, which is useful to probe our technique.

As a benchmark, we use the AzTEC GOODS-S survey, in which a Lissajous scan was performed, the telescope beam FWHM was approximately 30 arcsec, and the pixel size was chosen to be 3 arcsec.

4.1. Building Simulations

The atmospheric signals are created in time domain, where an inverse $f^{-\alpha}$ filter is used to generate atmosphere realizations

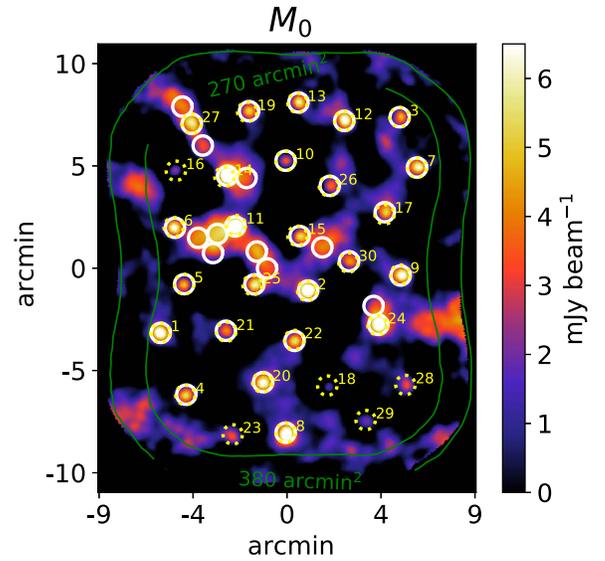


Figure 5. Reference M_0 map, computed with the PCA 2.5σ procedure, described in Section 3.2. Sources with $S/N > 4$ are counted as detections and circled with solid white lines. The 30 initial mock point sources are numbered and circled with dotted yellow lines. Notice that not all the initial/detected point sources are coincident; some detections correspond to bright patches of the extended model E .

statistically similar to observations. The i th detector signal is

$$\mathcal{S}_i = \mathcal{F}^{-1}\{(\mathcal{P}_i)^{1/2}\mathcal{G}\}, \quad (18)$$

where \mathcal{P}_i is the i th power spectrum, \mathcal{G} is the Fourier transform of a random Gaussian sequence with the same length of time

Table 1
Statistical Properties per Simulation Map

Map	Std (mJy)	Skewness	Negentropy	$\rho(P)$	$\rho(E)$
M_0	2.18	0.27	0.54	0.47	0.48
S_1	1.51	1.26	2.15	0.70	0.12
S_2	2.61	0.05	0.10	0.08	0.71

Note. Statistical moments quantified within a 270 arcmin² area. Negentropy is computed with Equation (17), $\rho(P)$ stands for the pixel correlation with the point model, and $\rho(E)$ stands for the correlation with the extended model.

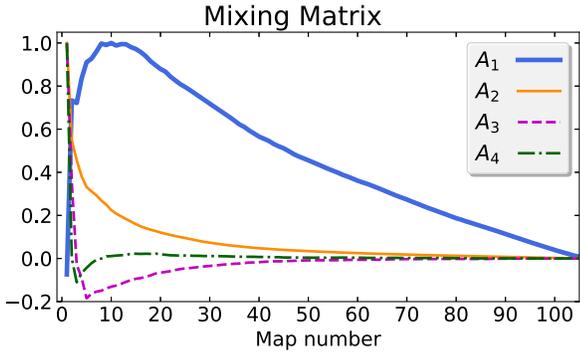


Figure 6. Mixing matrix coefficients from Equation (14). The j th component S_j in Figure 7 is coupled to the redundant maps $\{M_i\}$ by the mixing coefficients $A_j = \{a_{ij}\}$.

streams, and \mathcal{F}^{-1} is the inverse Fourier transform operator. To preserve the statistics of the correlation matrix, we use a single \mathcal{G} -realization for all the detectors. We include the effect of an elevation gradient and a differential airmass change in the line of sight,

$$F_i = F_0 e^{\tau/\tau_0} \sec(\pi/2 - \varepsilon_i), \quad (19)$$

where ε_i is the elevation track, τ is the opacity, and $F_0 \simeq 80$ mJy beam⁻¹ and $\tau_0 \simeq 0.06$ are typical flux and opacity normalization factors at 1.1 mm, respectively.⁴ The noise levels resulting from our simulations are usually smaller than real data by a factor $\delta n_S \simeq 1.25$ – 1.75 (possibly because real atmospheric data may contain additional patterns, though we expect them to be less dominant). Thus, we propagate this factor as $W_S \leftarrow \delta n_S^{-2} W_S$, allowing us to make a proper comparison between simulation and real S/N.

The mock astrophysical data, as shown in Figure 2, are a group of 30 point sources labeled as P , embedded in an extended source labeled as E . P resembles, for example, a population of SMGs, while E represents an extragalactic extended emission. The point sources are randomly located Gaussian distributions, spaced out at least 5 times the beam size, with fluxes between 4.5 and 8 mJy beam⁻¹. For E we use a facsimile of 30 Doradus in the Large Magellanic Cloud (Seale et al. 2014), adapted for our interests: smoothed with a 90 arcsec Gaussian kernel, centered at the mean, and with a maximum flux of 10 mJy beam⁻¹.

Finally, we map the mock astrophysical data back to time domain, using the GOODS-S scanning strategy and pointing information. To reduce computation time, we take advantage of the Lissajous scan continuity, applying a custom bi-linear

interpolation algorithm to efficiently convert pixel information into mock time streams.

4.2. Reduction and Decomposition of Mock Data

A total of $N_b = 106$ redundant maps are produced with the AzTEC pipeline, and a few of them are shown in Figure 3; it is interesting to see their statistical moments because they reflect their mixing degree. We adopt the 270 arcmin² (50% of uniform coverage) map area to measure information and to perform our analyses. In Figure 4 we plot the std, skewness, and our approximation to negentropy. Due to the dominant brightness of atmospheric emission, the std is high for the first map and falls off to zero at the last map. The most mixed map is M_1 ; correspondingly, its negentropy and skewness are close to zero. We also computed a set of maps with the atmospheric model only, which helped us to confirm $M_{1 \leq i \leq 5}$ as the most atmospheric-contaminated maps. However, atmospheric and extended emissions are progressively removed from redundant maps, increasing negentropy until a maximum around M_{60} ; we assert that M_{60} is the least mixed map. Afterward, only point-like sources are left, but they become gradually fainter, until the last maps are dominated by nearly Gaussian noise; likewise, negentropy and skewness fall off close to zero at M_{106} .

We also perform the customary PCA 2.5σ procedure for point sources, as explained in Section 3.2. The resulting map M_0 can be used as a reference to appreciate the leverage of our PCA-ICA technique, compared to the simplest PCA approach in time domain. M_0 is shown in Figure 5, and its statistical moments are listed in Table 1. Indeed, we do not see bright atmospheric residuals in M_0 , like those evident in $\{M_{1 \leq i \leq 5}\}$. Still, we see important residuals from the astrophysical extended model, all mixed with the point sources. As we mentioned, these astrophysical residuals can be harmful because they bias measurements intended for compact sources.

We use the following set of FastICA parameters. We choose to work with $n = 4$ independent components; this number is data dependent, but a good choice of n must yield physically meaningful and robust solutions. We test a large number of random initialization matrices and check that the solution preserves small negentropy dispersion. We use the FastICA parallel algorithm and the tolerance parameter $\text{tol} = 10^{-12}$ (Hyvärinen & Oja 2000).

The decomposition results are both the mixing matrix shown in Figure 6 and the independent components shown in Figure 7. By visual inspection, S_1 can be easily identified with the point model P and S_2 with the extended model E . S_3 is made of smooth bright fluctuations, so we identify it as an atmospheric foreground. S_4 looks less familiar because it contains symmetric stripes; actually, this pattern can be identified as an effect due to the Lissajous scanning strategy.

In order to gain an intuitive insight about the usefulness of redundancy, we also perform ICA over a highly redundant set of maps. We produce another set of redundant maps denoted by $\{N_i\}$; they contain the same mock atmosphere and E , but not P . We then perform an ICA decomposition of the $2(N_b - 1)$ redundant maps $\{M_i\}$ and $\{N_i\}$, and the results are remarkable: as seen in Figure 8, the point sources are almost perfectly isolated. The flux distribution is delta-like with a heavy positive tail due to the point-source signal; hence, detections would have extremely high S/N. Of course, this degree of redundancy is unrealistic, but at least from a qualitative point of view, this

⁴ 1 Jy = 10^{-26} W m⁻² Hz⁻¹.

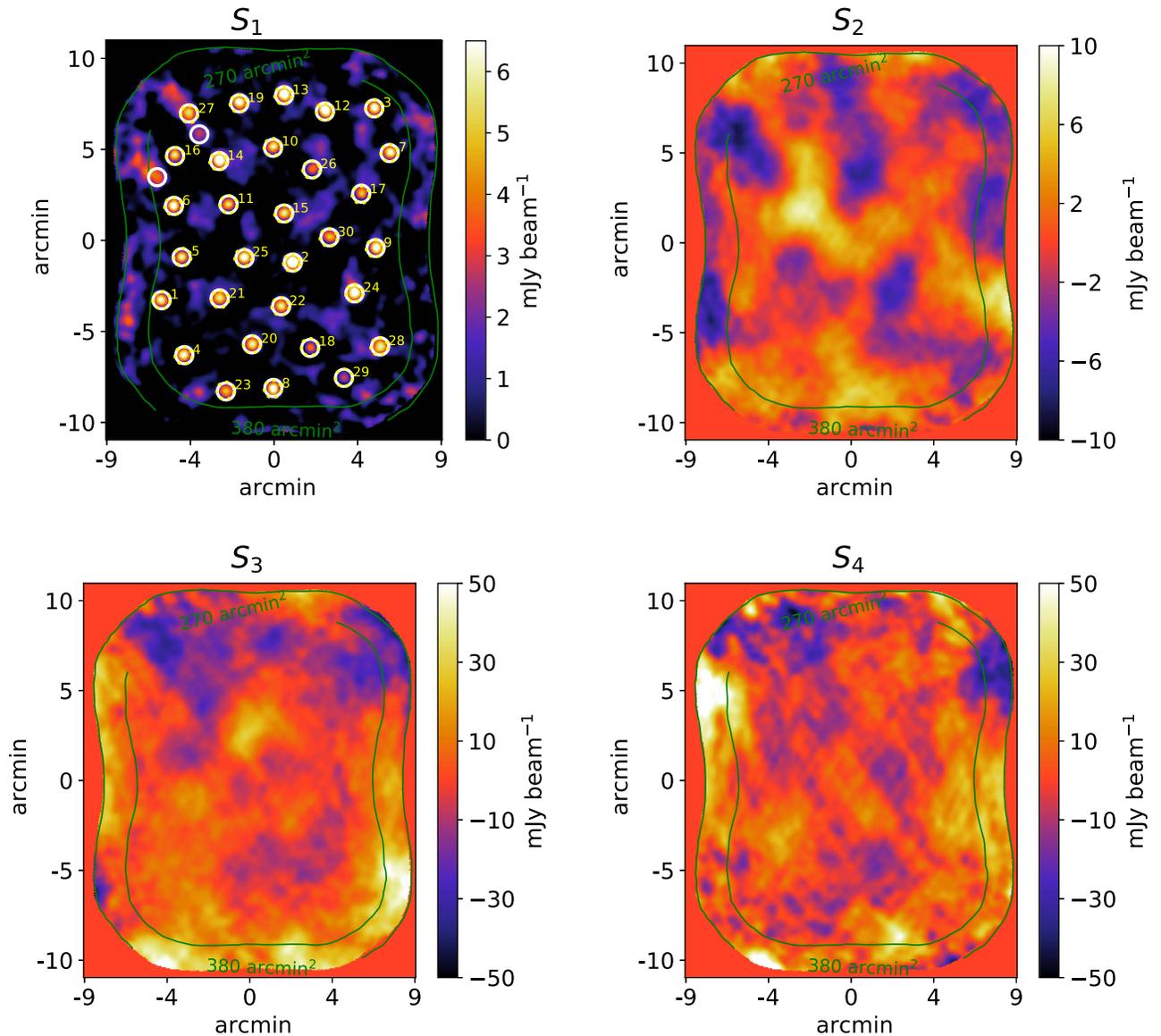


Figure 7. Independent components decomposed from redundant maps as in Equation (14) and calibrated as explained in Section 4.3. S_1 is interpreted as the point-like component of the astrophysical model. The white solid circles enclose $S/N > 4$ detections, and the yellow dotted circles locate the original mock point sources; all but two detections are coincident with the point model. S_2 is interpreted as the extended component of the astrophysical model. The component S_3 is interpreted as an atmospheric foreground. The stripes featured in map S_4 are interpreted as effects of the Lissajous scan. Contours as in Figure 2.

idealization helps us to gain intuition about what we may expect from a decomposition of redundant maps.

4.3. Calibration of Independent Components

Because of the ICA ambiguities, the independent components need some calibrations before being ready to extract physical information. Below we discuss some calibration strategies.

The *permutation ambiguity* could be even trivially solved by eye as we just did in the previous subsection, but when the ICA decomposition is part of a pipeline, we need an autonomous algorithm to identify physical components *on the fly*. As can be seen in Figure 6, we find heuristically that the sum of the P -mixing coefficients is always the largest, followed by the E -mixing coefficients, the atmospheric foregrounds, and the scanning pattern. This effect is related to the typical angular scales of the objects contained in each independent component.

We can use this hierarchy for blind identification of components. Our main target is the point-source component because of the scale calibration described below. (An alternative criterion to handle the permutation ambiguity can be found in Waldmann 2012.)

The scaling ambiguity may be split into *sign* and *absolute-scale* ambiguities. To solve the sign ambiguity, we demand that $\sum_{i=1}^{N_b-1} a_{ij} > 0$, for each $j = 1, \dots, 4$. The intuitive reasoning is that the mixing coefficients a_{ij} represent the degree of S_j mixing into every individual M_i , and the average mixing should be positive. The absolute-scale ambiguity could be approached with the std criterion: every independent component is scaled with the std of the most *akin* redundant map. With “most akin,” we mean the M_i whose pixel correlation with S_j is maximum. This inaccurate criterion might be useful only for visualization purposes, with the atmospheric foregrounds, for instance.

We also propose a scale calibration using external artificial information. We refer to a *witness* as a mock source similar to

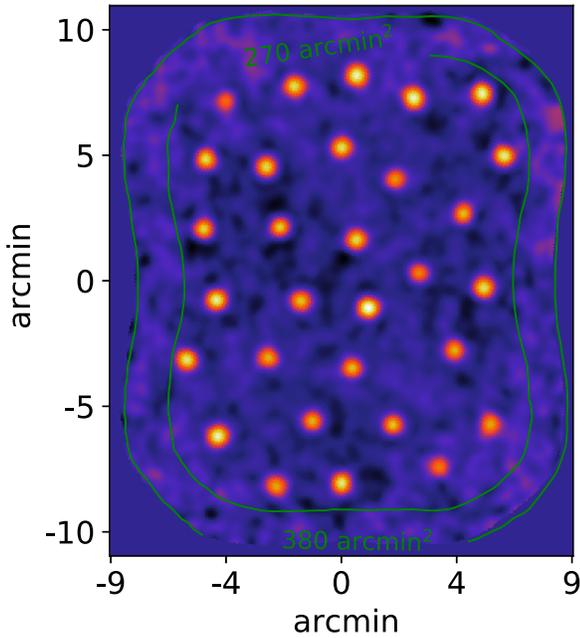


Figure 8. Point (uncalibrated) component decomposed from arbitrarily highly redundant maps (the sets M_i and N_i explained in the text).

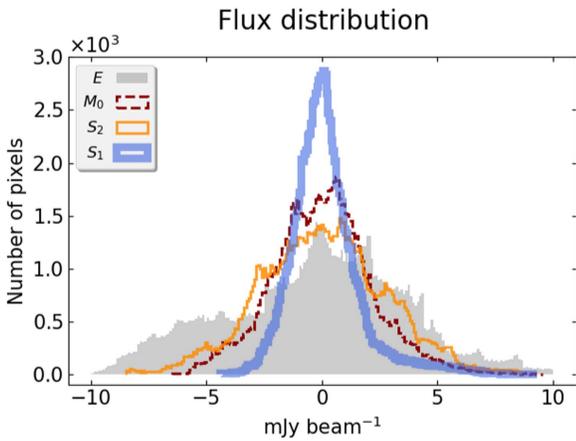


Figure 9. Flux distribution within the inner 270 arcmin² region of the following maps: the extended model E in Figure 2, the reference map M_0 of Figure 5, and the independent components S_1 and S_2 in Figure 7.

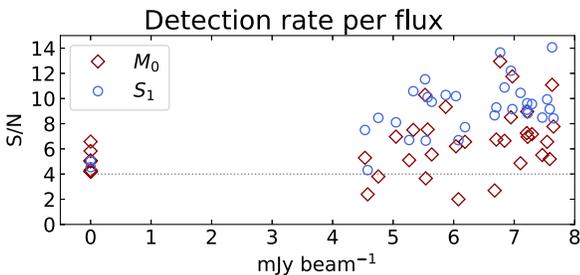


Figure 10. Detection rate (S/N) of point sources as a function of flux. Sources with $S/N > 4$ are counted as detections: 10 (2) detections on M_0 (S_1) do not match the point model.

the physical signal of interest. It is inserted into the redundant maps and retrieved after an ICA decomposition. By definition, a witness must fulfill two conditions: (i) it does not alter the statistical properties of actual data, and (ii) it is always recovered in the ICA map under calibration. These conditions

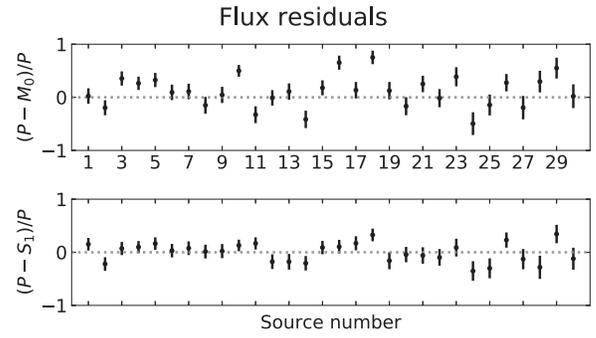


Figure 11. Point-source flux residuals: the model flux minus the measured flux in the M_0 and S_1 maps.

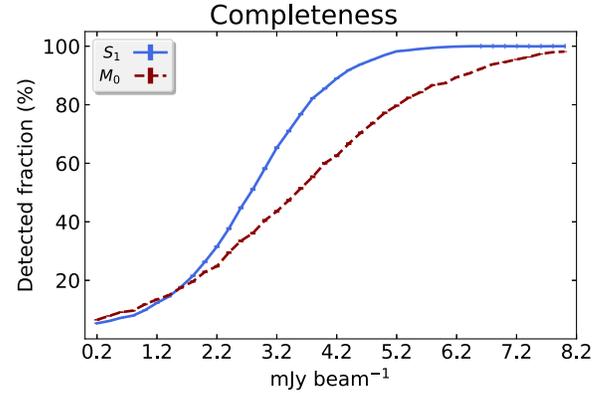


Figure 12. Completeness of M_0 and S_1 maps.

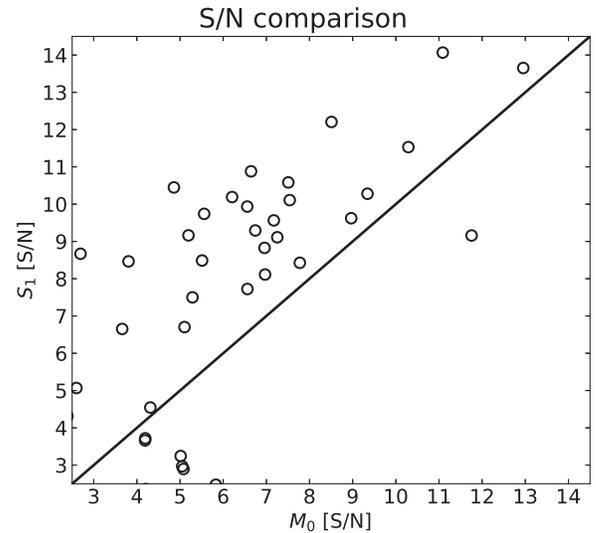


Figure 13. S/N comparison between point sources detected on M_0 or S_1 .

warrant that except for the scale, the decomposition preserves the witness information.

For the point-source component, a witness can be a two-dimensional Gaussian with beam-sized symmetrical widths. First, we insert one witness into the redundant maps, creating a new set of slightly perturbed $\{M'_i\}$ maps. The witness is wisely located in regions where the flux distribution is locally uniform and at least 3.5 beams away from any (mock) astrophysical point source. We check that the general statistical properties of the redundant maps are not modified by the insertion of a witness. After $\{M'_i\}$ are ICA-decomposed, the witness is always found in S_1 . We set to unit the scale of S_1 inside the

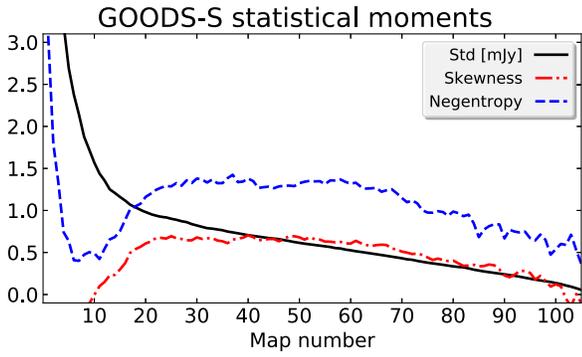


Figure 14. Statistical moments of GOODS-S redundant maps, analogous to Figure 4.

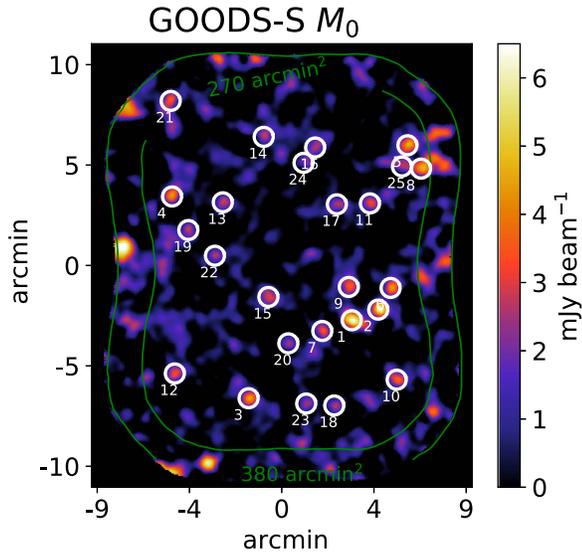


Figure 15. GOODS-S reference map M_0 computed with the PCA 2.5σ procedure, described in Section 3.2. Sources with $S/N > 4$ are counted as detections and circled with solid white lines. External contours as in Figure 2.

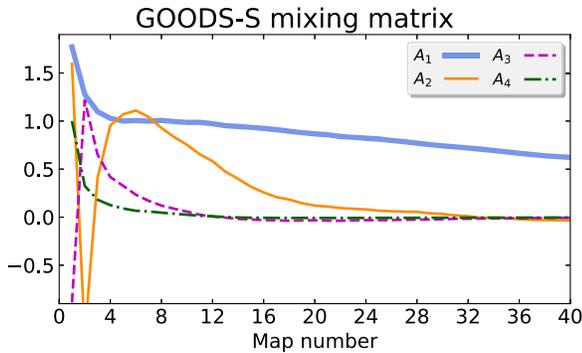


Figure 16. Mixing matrix coefficients of the GOODS-S redundant maps, decomposed with Equation (14).

Table 2
Statistical Properties of GOODS-S Maps

Map	Std (mJy)	Skewness	Negentropy
M_0	1.03	0.84	1.07
S_1	1.07	0.77	1.18
S_2	1.04	-0.67	0.64

Note. Statistical moments quantified within the 270 arcmin 2 GOODS-S field. Negentropy is computed with Equation (17).

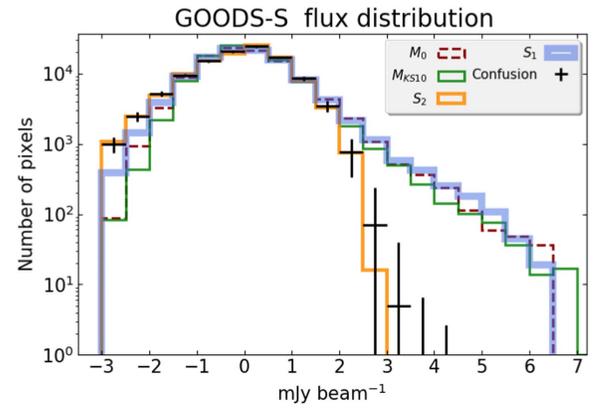


Figure 17. Flux distribution within the inner 270 arcmin 2 region of the GOODS-S field, for the following maps: the M_0 reference map computed with the PCA 2.5σ procedure, the M_{KS10} map computed with PCA 2.5σ and Wiener filtering (Scott et al. 2010), and the independent components S_1 and S_2 decomposed with our PCA-ICA technique. Black error bars represent the expectation from our simulations of the confusion background (details in text).

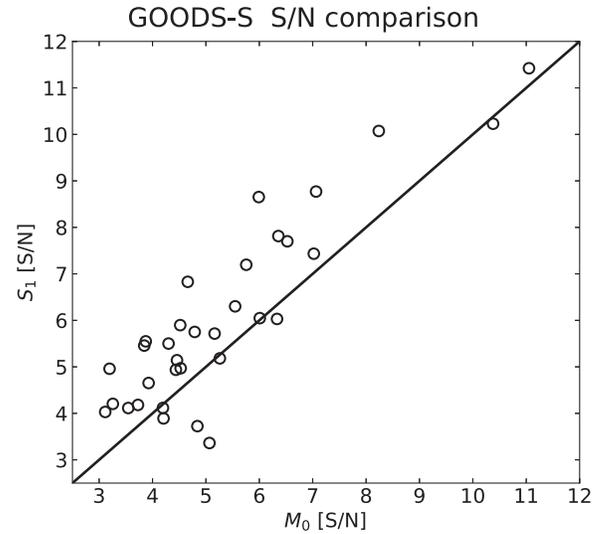


Figure 18. S/N comparison between point sources detected in the GOODS-S field maps S_1 or M_0 .

270 arcmin 2 area. We fix the sign ambiguity and make a bivariate fit to the witness in S_1 . The scale factor is found by the ratio of the actual witness flux to the resulting fitted flux. The process is randomly repeated to generate a statistical distribution, whose mean is the calibration scale a_1 , and its std is added in quadrature as a systematic error. Using about 10,000 witnesses, we find $a_1 = 1.44 \pm 0.06$, and the calibrated std is listed in Table 1. Then, the companion weight map W_1 can be computed according to our discussion in Section 3.2.

For the extended component S_2 , it is not obvious what kind of artificial signal meets the conditions to be used as a witness. Hence, a pixel-by-pixel fit to a subset of redundant maps M_i seems a better approach for calibration. We should consider the pixel correlation with S_2 and discard the most atmospheric-contaminated maps to choose that subset. For this simulation, we actually know the underlying (mock) astrophysical component, and hence we can fit S_2 directly to E in order to help us choose the subset of maps for the fit, according to their degree of pixel correlation with S_2 . Following this approach, we choose $M_{4 < i < 7}$, whose correlations with S_2 are large ($\gtrsim 0.75$). The fit is performed simultaneously for the four

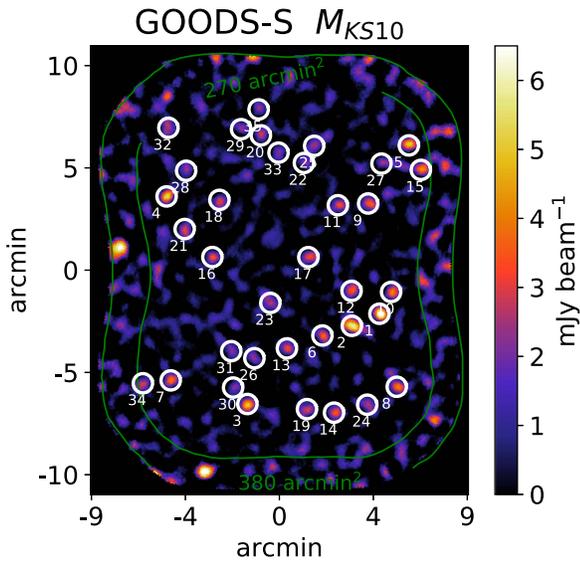


Figure 19. GOODS-S M_{KS10} map (Scott et al. 2010). Besides the PCA2.5 σ procedure, an optimized Wiener filter was used to enhance point sources. Sources with $S/N > 4$ are counted as detections and circled with solid white lines. Contours as in Figure 2.

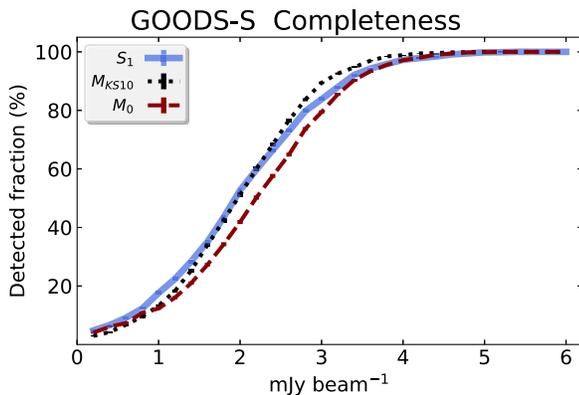


Figure 20. GOODS-S completeness of M_0 , M_{KS10} , and S_1 maps.

independent components. We use the outer 380 arcmin² area for calibration (the same as for the ICA decomposition). The resulting scale factor from this calibration is $a_2 = 2.78 \pm 0.46$, and the corresponding std is reported in Table 1.

For comparison, from the same fit, the scale factor for S_1 is found to be 1.48 ± 0.42 . Hence, the calibration scales obtained from the witness and fitting approaches are consistent within the error bars.

4.4. Inference of Astrophysical Information

After the redundant maps are decomposed and the corresponding calibrations are performed, we can analyze the M_0 , S_1 , and S_2 maps in more detail. The statistical properties of these maps can be read from Table 1 and seen on their flux distribution in Figure 9.

The point-source measurements are improved after our decomposition. In Table 1, all the statistical properties of S_1 are improved compared to M_0 . Negentropy, in particular, indicates that S_1 contains more information than M_0 ; as a consequence, we should expect an S/N boost in S_1 compared to M_0 . Furthermore, M_0 evidently contains extended emission residuals, so we could expect more biased flux measurements in M_0 . We also read in

Table 1 that M_0 is equally correlated to the punctual P and extended E models, whereas S_1 is mostly correlated with P . From the histograms in Figure 9, we see that the flux distribution of M_0 is relatively flat, resembling the distribution of the extended model E . Conversely, the S_1 is sharply peaked at its mean, with a hard positive tail corresponding to the signal of point sources, as expected from our discussion in Section 4.2.

The extended source is fairly isolated after decomposition. The S_2 flux distribution in Figure 9 is much more spread out than S_1 ; actually, it resembles the flux distribution of E more closely than M_0 . Because S_2 is mostly correlated with the extended model E but negligibly correlated to the point model P (Table 1), we can assert that the astrophysical information contained in S_2 is not contaminated by the point sources.

Now we can detect point sources as described in Section 3.2. In M_0 , we detect 35 bright sources with $S/N > 4$, from which 10 do not match any mock point sources, and 5 of the 30 mock point sources are not detected in M_0 . On the other hand, in S_1 we detect 32 bright sources with $S/N > 4$, from which only 2 sources (with $S/N \lesssim 5$) do not match P . Inside 270 arcmin², the level of noise is not abruptly varying, so that one would expect that the S/N depends primarily on the flux of the point source. In Figure 10 we plot the detection rate as a function of point-source flux. The sources detected in S_1 seem to follow the expected trend, while in M_0 some bright sources (>5 mJy) are found with very low S/N . These anomalies in M_0 can be attributed to the evident extended emission residuals in Figure 5.

We also check the point-source flux recovery. Indeed, flux biases are not unexpected after the reduction process; for example, the Gaussian filter could smear fluxes in the map. Usually, these biases are negligible for blank fields (see next section), but in the presence of an extended source like in our mock data, the bias can be a problem to deal with. In Figure 11, we show the residuals between initial and measured fluxes for M_0 and S_1 respectively. The error bars are computed as usual (e.g., for the residual $R = (F_1 - F_2)/F_1$, the error is $\sigma_R = F_2/F_1 \sqrt{(\sigma_1/F_1)^2 + (\sigma_2/F_2)^2}$). For the P known point fluxes, we use the effective sensitivity $\sigma_{\text{eff}} \approx 0.58$ mJy, taken from the weight map W_1 (see Section 3.2). In M_0 we measure a significant rms of 0.31 mJy, again attributed to the remains of the extended emission. Contrarily, S_1 residuals are much less scattered, with an rms deviation of 0.19 mJy, a behavior closer to the expectation from a blank field. (Notice that these values account for statistical errors only.) Moreover, our S_1 -scale calibration gets reassured; any (positive or negative) tendency would hint at a fail in calibration. Satisfyingly, the points are roughly symmetrically distributed around zero, hence indicating an accurate S_1 -scale calibration.

The detection rate in Figure 10 is representative only of the particular set of point sources in our mock data. A deeper characterization of S_1 and M_0 requires a larger sample of point sources and is known as the *completeness* of the map. To compute it, we insert one additional point source, excluding the locations of the 30 initial point sources and surrounding (beam-radius) areas. The artificial source is said to be recovered if it is found with $S/N > 4$ around a circle of half-beam radius. For each equidistant flux step, we insert 10,000 point sources (one at a time) at random locations. We estimate the error bars assuming a binomial distribution. The results for M_0 and S_1 are shown in Figure 12.

Finally, as expected from Table 1, the point sources in S_1 are detected with higher S/N than in M_0 (see Figure 13). Only false point-source detections show higher S/N in M_0 .

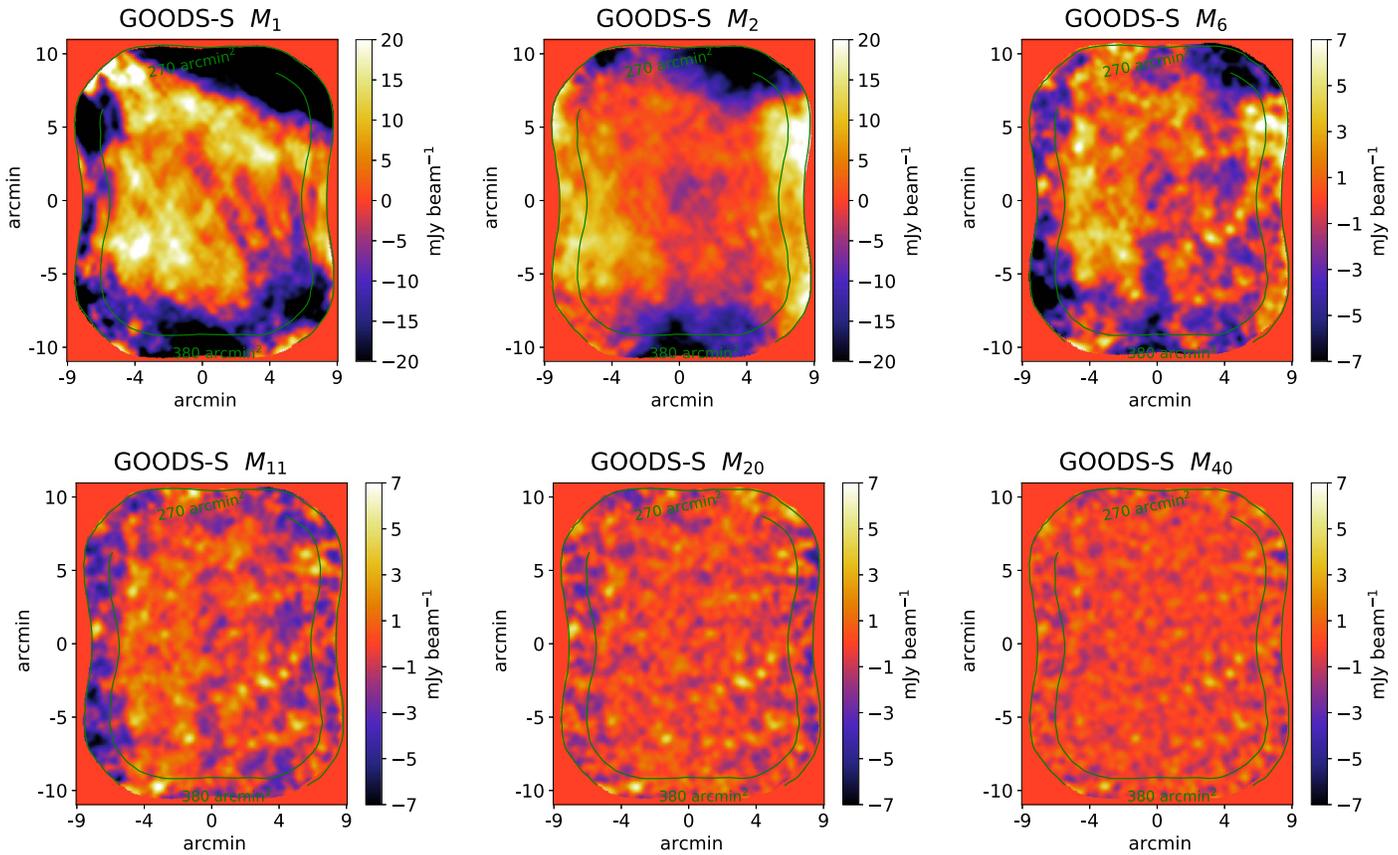


Figure 21. Redundant maps of the GOODS-S field made with the AzTEC pipeline, removing i principal components, respectively.

5. Decomposition of Real Data

Our next step is to apply our techniques to a set of real data with the aim of checking the recovering of previous results. For that, we revisit the AzTEC/ASTE GOODS-S survey, which is considered to be a blank field. For that reason, this survey has been extensively studied and used as a trial data set for extensions to the AzTEC pipeline (Scott et al. 2010, 2012; Downes et al. 2012; Yun et al. 2012). Although astrophysical foregrounds are not expected, these observations were certainly contaminated by bright atmospheric foregrounds. Hence, the AzTEC/ASTE GOODS-S survey is ideal for our purposes, mainly to test our calibration strategies on point-like sources. Besides, our previous simulations were purposely designed to be very similar to the AzTEC/ASTE GOODS-S survey, which consists of 74 observations, each one containing $N_b = 106$ effective bolometer time streams. Then, we can apply the same methodology as in Section 4.

The AzTEC/ASTE GOODS-S observations reported a 1σ depth of about $0.48\text{--}0.73\text{ mJy beam}^{-1}$, which is below the estimated confusion background limit of 2 mJy beam^{-1} (Scott et al. 2010). The confusion background is the sea of faint unresolved sources in the sky, creating an extended emission that can potentially bias detections below the confusion background limit (Hogg 2001). Unfortunately, this uncertainty cannot be reduced by increasing the observation time; however, as the confusion background should be non-Gaussianly distributed, then, an ICA decomposition is not precluded a priori.

Following the same methodology, we compute the redundant maps $\{M_i\}$ for GOODS-S. We show their statistical moments in Figure 14, as well as a few representative redundant maps in Figure 21. Notice the mixture of small

and large structures in every map. Next, we compute the reference M_0 map as shown in Figure 15, using the PCA2.5 σ procedure (see Section 3.2).⁵

We proceed with the same decomposition parameters used previously. We show the mixing matrix in Figure 16 and the independent components in Figure 22. As we discussed with simulations, the independent components can be identified either by eye or from the behavior of the mixing coefficients. S_1 is a point-source component, whose mixing coefficients extend along all the redundant maps. S_2 is an extended emission suspected for an astrophysical origin, whose mixing coefficients survive to half the redundant maps. The negative peak in A_2 is an effect due to the positive/negative borders between the M_2 and M_{14} redundant maps. S_3 is made of smooth bright fluctuations, while A_3 survives only for the first dozen maps, so it can be identified as an atmospheric foreground. Finally, because of its symmetric stripes, S_4 can be identified at least partially with a systematic effect due to the Lissajous scan. Given that A_4 mixes only $M_{1 < i < 3}$, we can assert that the Lissajous systematic harms mostly the largest angular scales.

We continue with calibration (see Section 4.3 for details). S_1 is scaled as $a_1 S_1$, with $a_1 = 1.22 \pm 0.05$ found employing the witness-based calibration; its companion weight map W_1 is computed as before, with the scale uncertainty added in quadrature as a systematic error. For S_2 , we proceed with a pixel-by-pixel fit to $M_{5 < i < 9}$, which are the most correlated with S_2 , finding $a_2 = 1.57 \pm 0.41$. (For comparison, from the same fit, the S_1 scale is found to be $a_1 = 1.22 \pm 0.45$.) We finally

⁵ The redundant and M_0 maps of the AzTEC/ASTE GOODS-S observations are available for visualization or further image processing in the .tar.gz package.

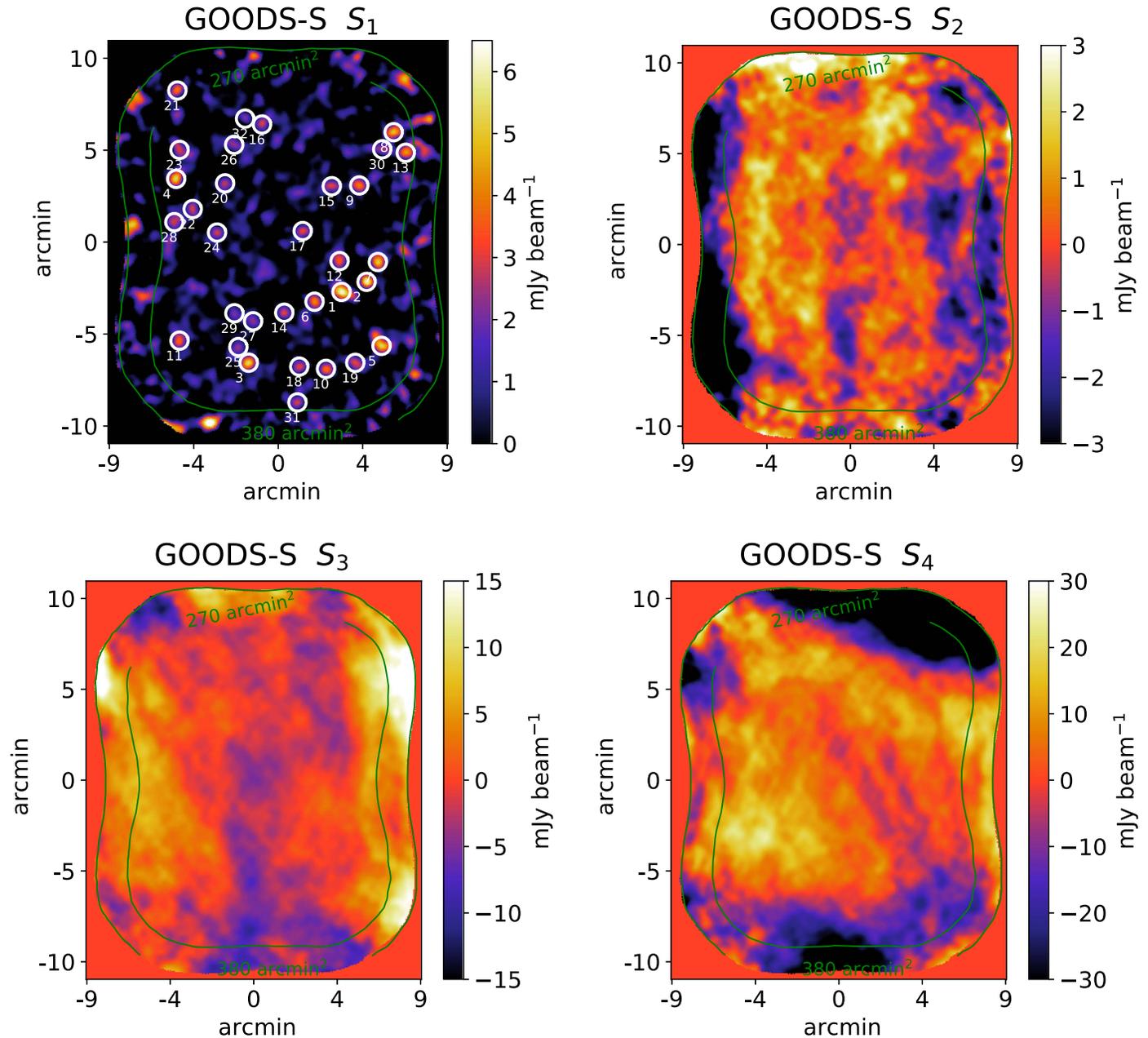


Figure 22. Independent components decomposed from GOODS-S redundant maps and calibrated as explained in Section 4.3. S_1 is a point-like component of the GOODS-S field. White circles enclose 32 bright sources found with $S/N > 4$. S_2 is an extended component, suspected of astrophysical nature, possibly the confusion background. S_3 and S_4 are interpreted as atmospheric foregrounds, and the effect of the Lissajous scanning pattern.

scale S_3 and S_4 also with the pixel-by-pixel fit to their most akin redundant maps, $M_{2,3}$ and M_1 , respectively.

The statistical properties of M_0 , S_1 , and S_2 (within 270 arcmin^2) are listed in Table 2, and their flux distribution can be observed in Figure 17. In this case, the improvement of the statistical properties of S_1 compared to M_0 can only be noticed in the degree of negentropy. Yet, the improvement is significant enough to boost the S/N of point sources, as shown in Figure 18. In M_0 , we count 25 bright sources with $S/N > 4$. In S_1 , we count 32 bright sources with $S/N > 4$. We cannot check for bias as we did with mock data, but we can compare the fluxes measured in M_0 and S_1 (see Figure 23).

To enrich our discussion, we test the ability of our technique to reproduce previously reported results from the same GOODS-S survey. In *KS10*, the authors also applied the

PCA 2.5σ procedure (with a different code), but the main difference is that they applied a Wiener filter with a specialized point-like kernel as a prior. The assumptions taken in that approach and ours are conceptually different and not mutually exclusive; thus, a direct comparison must be moderately assessed. The *KS10* map is shown in Figure 19, in which there are 35 bright sources with $S/N > 4$.

In Figure 23 we also compare the flux measured on S_1 and M_{KS10} . For $S/N > 5$, most of the fluxes are consistent; only below $S/N \sim 5$ are there some noncoincident detections. These differences arise most likely as a result of the combined effect of persistent foreground residuals after the PCA 2.5σ procedure and their subsequent enhancement by the Wiener filter at unresolved scales. This flux comparison is interesting for the purpose of probing the recovery of high- S/N detections

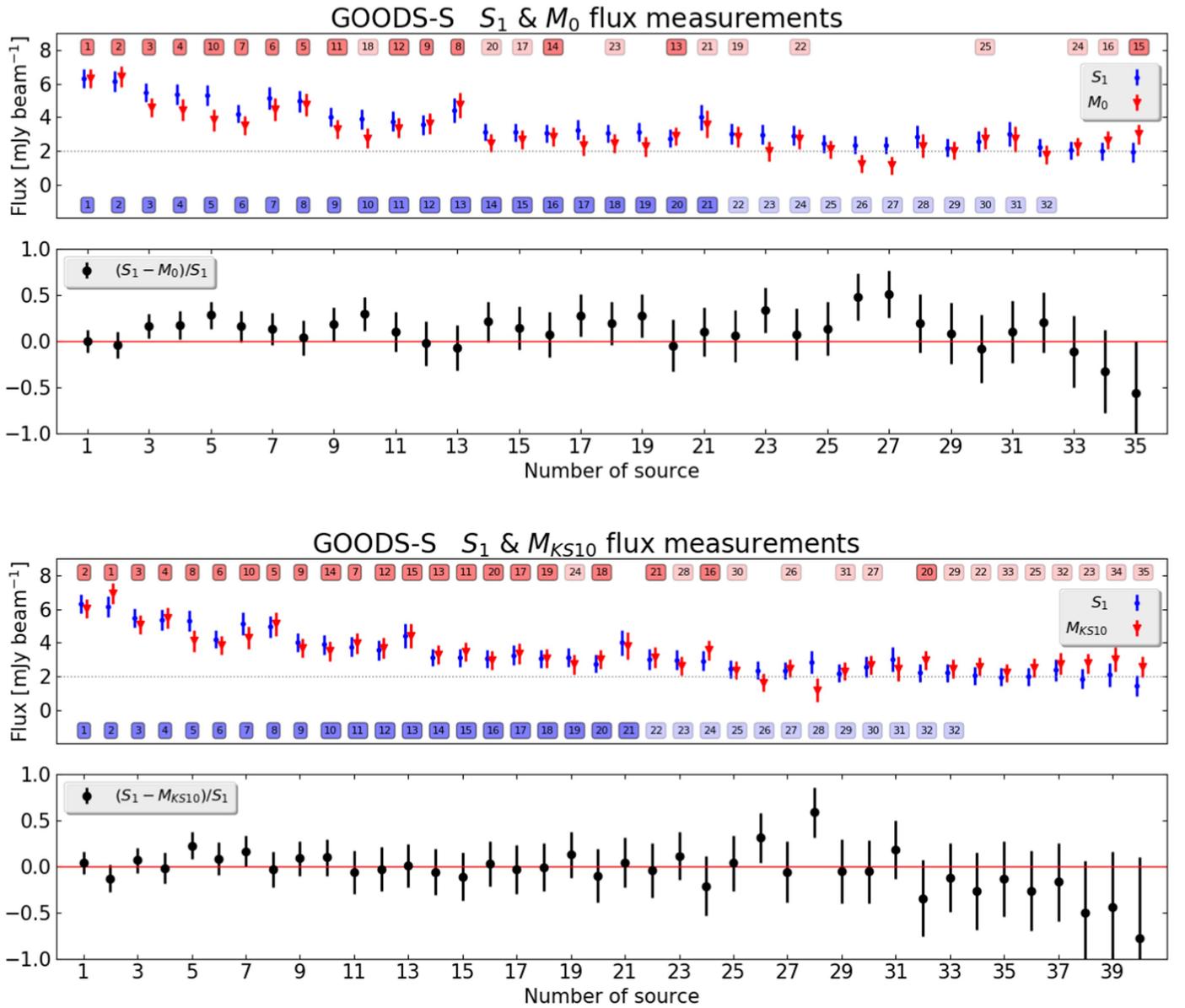


Figure 23. GOODS-S flux measurements. Top panel: bright sources detected in S_1 and M_0 . Bottom panel: bright sources detected in S_1 and M_{KS10} . Square labels indicate the detection number on each map. Dark-color labels indicate detections with $S/N > 5$; light-color labels indicate detections between $4 < S/N < 5$. The confusion limit for this survey is 2 mJy (Scott et al. 2010). The flux residuals between each pair of maps are also shown.

using our PCA-ICA approach, but lack of coincidences at low S/N should not be overstated.

As we mentioned, S_2 is suspected to have an astrophysical origin, and an intuitive prospect is the confusion background. To explore this possibility, we simulate the fainter dusty star-forming galaxy population, following the number counts measured by Fujimoto et al. (2016), which includes the deep ALMA census of faint sources ($\gtrsim 0.02$ mJy) at 1.2 mm and the bright-end AzTEC 1.1 mm counts of Scott et al. (2012). Sources are randomly distributed in space within the simulated maps (i.e., no clustering). As every realistic observation contains some degree of instrumental noise, we also add and convolve a 1.1 mJy white noise to our simulations, in order to approximately match the negative flux tail of S_2 . We generate 200 random realizations of these confusion maps. Using the weight map W_1 , we subtract bright sources ($S/N > 3.5$) and quantify the average flux distribution. As seen in Figure 17, S_2 is consistent with the flux distribution expected from

our simulated confusion background and reflects the 2 mJy confusion limit. This interpretation of S_2 could help to explain the detection differences of S_1 compared to M_0 or M_{KS10} , especially for point sources below the confusion limit (see Figure 23).

We also compute in Figure 20 the completeness for M_0 , M_{KS10} , and S_1 . As expected from simulation results, the completeness of S_1 is better compared to M_0 . Given that M_{KS10} comes from a process that suppresses extended objects in favor of point ones, we would have expected a much larger completeness for M_{KS10} compared to S_1 ; however, the results are comparable.

6. Conclusions

In this paper we are presenting a PCA-ICA algorithm capable of separating atmospheric fluctuations, extended astrophysical foregrounds, and point-like sources from single-wavelength millimeter

surveys. In order to probe the consistency of our results, we have tested our methodologies on both mock and real data.

We confirm that our PCA generation of redundant maps allows a successful application of an ICA decomposition, in defiance of the single-channel limitation. We find a good agreement between simulation inputs and the resulting independent components, along with a good degree of isolation (see Table 1).

We have proposed and tested different strategies to calibrate independent components, getting rid of the permutation and scale ambiguities, inherent to ICA. We find that our approach can be useful to remove both atmospheric and astrophysical foregrounds, minimizing information loss. Consequently, our decomposition can help prevent bias in flux measurements and boost the S/N.

We also applied our techniques to the AzTEC/ASTE survey of the GOODS-S field. We find that a PCA–ICA decomposition S_1 is preferred over the simplest PCA 2.5σ procedure M_0 , as expected from the analogous result in simulations. We confirm agreement with $S/N > 5$ detected sources in KS10, showing consistency to recover previously reported measurements. An unexpected finding of this work was the measurement of a feeble extended emission on the GOODS-S field (S_2 in Figure 22), which, according to simulations, is consistent with the flux distribution of the faintest SMGs' confusion background. We conclude that our PCA–ICA implementation is a viable and promising approach to separate atmospheric and astrophysical (extended and compact) sources.

One route to extend our work is to improve redundancy with algorithms other than PCA in time domain. One can also try different decomposition algorithms, possibly more powerful than the simplest version of FastICA. Besides, it should be possible to adapt our technique to multiwavelength data, increasing redundancy and decomposing signals in each channel, before a multiwavelength analysis. The complementary maps decomposed by ICA are interesting by themselves. Certainly, further investigations of this kind of technique not only can improve atmospheric and instrumental models but also make new astrophysical emissions available. The GOODS-S S_2 component is just an interesting example: our astrophysical simulations indicate consistency with the confusion background, yet in future analyses we shall confirm this result through exhaustive simulations of every systematic effect possibly sourcing S_2 . If confirmed, the characterization of the confusion background is of utmost importance in millimeter astronomy and cosmology, allowing us to study the clustering properties of SMGs and the distribution of matter in the universe.

We finally stress that this kind of analysis is particularly interesting at the advent of the next generation of continuum cameras, for example, MUSCAT,⁶ TolTEC,⁷ SCUBA-2 (Holland et al. 2013), NIKA2 (Calvo et al. 2016). Both MUSCAT and TolTEC are currently in construction and appointed to work on the Large Millimeter Telescope (Hughes et al. 2010). Indeed, this paper is our first step toward the ultimate goal of developing an efficient multicomponent decomposition pipeline for both instruments. MUSCAT is a single-channel

large-format camera, comprising $\simeq 1200$ detectors at the 1.1 mm wavelength band. MUSCAT will record 10 times more time streams than AzTEC. We will be able to generate many more levels of redundancy, along broader ranges of angular scales; thus, the application of our PCA–ICA algorithms will be very similar to this paper, but much more promising in quality of the decomposition. TolTEC will include 6300 detectors distributed in three channels, at 1.1, 1.4, and 2.1 mm, each of them sensitive to linear polarization. Thus, each single TolTEC observation will result in nine maps, which can be further decomposed by our PCA (or alternative) technique to generate higher levels of redundancy. Altogether, TolTEC surveys and our PCA–ICA technique bring out exciting expectations about the final data quality and possible new astrophysical fields to be uncovered.

This project was possible owing to partial support from CONACyT research grants: CB-2011/167291, CB-2015/256961, *Fronteras de la Ciencia* 2016/1848, and FONCICYT 2016/69. We thank Emmaly Aguilar and Ivnió Puerari for useful discussions and an anonymous referee for a critical review that has improved the paper.

References

- Calvo, M., Benoît, A., Catalano, A., et al. 2016, *JLTP*, 184, 816
 Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, *ARA&A*, 40, 643
 Casey, C., Narayanan, D., & Cooray, A. 2014, *PhR*, 541, 45
 Chavez-Dagostino, M., Bertone, E., Cruz-Saenz de Miera, F., et al. 2016, *MNRAS*, 462, 2285
 Choi, S. 2011, in *Handbook of Natural Computing*, ed. G. Rozenberg, T. Bäck, & J. N. Kok (Berlin: Springer), 435
 Comon, P., & Jutten, C. 2010, *Handbook of Blind Source Separation: Independent Component Analysis and Applications* (New York: Academic)
 Downes, T., Welch, D., Scott, K., et al. 2012, *MNRAS*, 423, 529
 Fujimoto, S., Ouchi, M., Ono, Y., et al. 2016, *ApJS*, 222, 1
 Funaro, M., Oja, E., & Valpola, H. 2003, *NN*, 16, 469
 Hogg, D. W. 2001, *AJ*, 121, 1207
 Holland, W., Bintley, D., Chapin, E., et al. 2013, *MNRAS*, 430, 2513
 Hughes, D., Correa, J. J., Schloerb, F., et al. 2010, *Proc. SPIE*, 7733, 773312
 Hyvärinen, A., Karhunen, J., & Oja, E. 2002, *Independent Component Analysis* (New York: Wiley)
 Hyvärinen, A., & Oja, E. 1997, *Neural Computation*, 9, 1483
 Hyvärinen, A., & Oja, E. 2000, *NN*, 13, 411
 Ichiki, K. 2014, *PTEP*, 2014, 06B109
 Morello, G. 2015, *ApJ*, 808, 56
 Perera, T. A., Wilson, G. W., Scott, K. S., et al. 2013, *PASP*, 125, 838
 Sayers, J., Golwala, S. R., Ade, P. A. R., et al. 2010, *ApJ*, 708, 1674
 Scott, K., Yun, M., Wilson, G., et al. 2010, *MNRAS*, 405, 2260
 Scott, K. S., Austermann, J. E., Perera, T. A., et al. 2008, *MNRAS*, 385, 2225
 Scott, K. S., Wilson, G. W., Aretxaga, I., et al. 2012, *MNRAS*, 423, 575
 Seale, J. P., Meixner, M., Sewilo, M., et al. 2014, *AJ*, 148, 124
 Stone, J. V. 2004, *Independent Component Analysis: A Tutorial Introduction* (Cambridge, MA: MIT Press)
 Waldmann, I. P. 2012, *ApJ*, 747, 12
 Waldmann, I. P. 2014, *ApJ*, 780, 23
 Wilson, G. W., Austermann, J. E., Perera, T. A., et al. 2008, *MNRAS*, 386, 807
 Yun, M. S., Scott, K., Guo, Y., et al. 2012, *MNRAS*, 420, 957
 Zhang, L., Bunn, E. F., Karacki, A., et al. 2016, *ApJS*, 222, 3

⁶ <http://gtr.rcuk.ac.uk/projects?ref=ST%2FP002803%2F1>

⁷ <http://toltec.astro.umass.edu/>